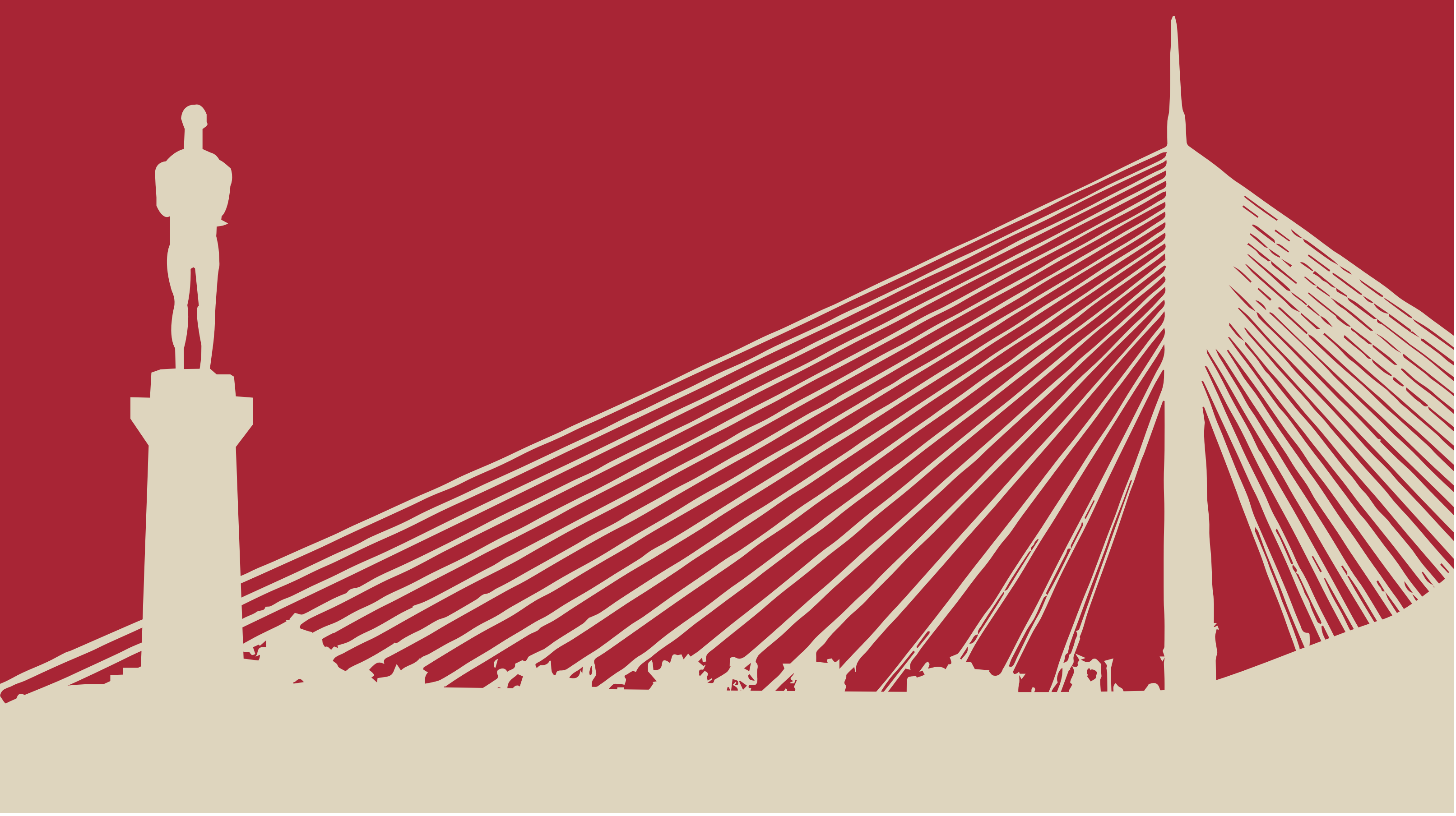
BALCOR
2018

CONFERENCE PROCEEDINGS

XIII Balkan Conference on Operational Research



BALCOR 2018

Belgrade, 25-28 May, 2018

XIII Balkan Conference on Operational Research

Proceedings

Editors

Urošević Dragan

Vidović Milorad

Dragana Makajić-Nikolić

Belgrade, 2018

Publisher

The Mathematical Institute of the Serbian Academy of Sciences and Arts (SANU)
Kneza Mihaila 36, 11001 Belgrade, Serbia

Conference Organizers

The Mathematical Institute of the Serbian Academy of Sciences and Arts (SANU), Belgrade, Serbia
University of Belgrade, Faculty of Organizational Sciences, Belgrade, Serbia
University of Belgrade, Faculty of Transport and Traffic Engineering. Belgrade, Serbia
The Yugoslav Society for Applied and Industrial Mathematics (JUPIM), Belgrade, Serbia
Society of Operations Researchers of Serbia (DOPIS), Belgrade, Serbia

Supported by

Ministry of Education, Science and Technological Development of the Republic of Serbia
EURO – The Association of European Operational Research Societies

Year

2018

ISBN: 978-86-80593-64-7

Welcome to Serbia and the XIII Balkan Conference on Operational research

Dear conference participant,

On behalf of the Scientific and Organizing Committee we are pleased to welcome you to BALCOR 2018 (The XIII Balkan Conference on Operational Research). Balkan conferences on operational research have being held since 1988 with the general aim to facilitate the exchange of scientific and technical information related to Operational research and to promote international cooperation, especially among the Balkan countries.

This year, invited speakers at Balcor 2018 are the following well-known scientists: Oleg Khamisov, Russian Academy of Sciences, Russia, Panos Pardalos, University of Florida, USA and Leonidas Pitsoulis, Aristotle University of Thessaloniki, Greece. Furthermore, two tutorials on heuristics and graph theory and combinatorial optimization will be held by Saïd Hanafi, University of Valenciennes, France and Damir Gainanov, Ural Federal University/ Moscow Aviation Institute, Russia.

The conference program contains 166 contributions by 249 authors from 37 countries. The submitted papers passed the refereeing process and all accepted papers are published in the Conference Proceedings or Springer Proceedings in Business and Economics.

Extended versions of all presented papers could be submitted by authors after the conference, to the three special issues, of IMA Journal of Management Mathematics (SJR: 0.857, IF: 1.488), Operational Research - An International Journal (SJR: 0.406, IF: 1.065) or Yugoslav Journal of Operations Research – YUJOR (SJR: 0.32). Such papers will be peer reviewed according to journals standards.

We hope that you, participants, will find at BALCOR 2018 excellent opportunity to follow presentations in your field of interest, do participate in discussions and debates and also take your time to see Belgrade with it rivers, fortress and other touristic attractions.
Welcome and enjoy your stay in Serbia.

Nenad Mladenović,
Conference Chair

Gordana Savić,
Organizing Committee chair

Program Scientific Committee

- **Mladenović Nenad, Mathematical Institute (Serbia), Conference chair**
- **Martić Milan, University of Belgrade (Serbia), SC co-chair**
- **Urošević Dragan, Mathematical Institute (Serbia), SC co-chair**
- **Vidović Milorad, University of Belgrade (Serbia), SC co-chair**
- Babić Zoran, University of Split (Croatia)
- Bilgiç Taner, Boğaziçi University (Turkey)
- Brailsford Sally, University of Southampton (United Kingdom)
- Čangalović Mirjana, University of Belgrade (Serbia)
- Davidović Tatjana, Mathematical Institute (Serbia)
- Demange Marc, ESSEC Business School in Paris (France)
- Duarte Abraham, Universidad Rey Juan Carlos (Spain)
- Dzițac Ioan, Aurel Vlaicu University of Arad (Romania)
- Figueira Jose Rui, Technical University of Lisbon (Portugal)
- Gavranović Haris, International University of Sarajevo (Bosnia and Hercegovina)
- Georgiou Andreas, University of Macedonia (Greece)
- Hertz Alan, GERAD - Ecole des HEC (Canada)
- Iosifescu Marius, Vice-President of the Romanian Academy (Romania)
- Iyigun Cem, Middle East Technical University (Turkey)
- Jacimović Milojica, University of Montenegro (Montenegro)
- Kalimoldaev Maksat, Institute of Problems of Information and Control (Kazakhstan)
- Kovačević-Vujčić Vera, University of Belgrade (Serbia)
- Kratica Jozef, Mathematical Institute (Serbia)
- Labbé Martine, Université Libre de Bruxelles (Belgium)
- Laslo Zohar, SCE - Shamoon College of Engineering (Israel)
- Lesaja Goran, Georgia Southern University (USA)
- Matsatsinis Nikolaos, Technical University of Crete (Greece)
- Mierlus Mazilu Ion, Technical University of Civil Engineering (Romania)
- Migdalas Athanasios, Aristotle University of Thessaloniki (Greece)
- Mihaljević Miodrag, Serbian Academy of Sciences and Arts (Serbia)
- Moreno Perez Jose, University of La Laguna (Spain)
- Nace Dritan, Universite de Technologie de Compiegne (France)
- Ognjanović Zoran, University of Belgrade (Serbia)
- Oğuz Ceyda, Koç University (Turkey)
- Pardalos Panos, University of Florida (USA)
- Paschos Vangelis, Universite Paris Dauphine (France)
- Preda Vasile, University of Bucharest (Romania)
- Radenković Božidar, University of Belgrade (Serbia)
- Radojević Dragan, Mihajlo Pupin Institute (Serbia, O.R. Society chairman)
- Sahin Guvenc, Sabanci University (Turkey)
- Samaras Nikolaos, University of Macedonia (Greece)
- Sburlan Silviu, Naval Academy Mircea cel Batran Constanta (Romania)
- Sifaleras Angelo, University of Macedonia (Greece)
- Siskos Yannis, University of Pireaus (Greece)
- Slowinski Roman, Poznan University of Technology (Poland)
- Speranza Grazia, University of Brescia (Italy)

- Stancu-Minasian, Ioan Romanian Academy (Romania)
- Stanić Stanko, University of Banja Luka (Republic of Srpska, Bosnia and Herzegovina)
- Stanojević Milan, University of Belgrade (Serbia)
- Stanojević Milorad, University of Belgrade (Serbia)
- Teodorović Dušan, University of Belgrade (Serbia)
- Todosioska Blagorodna, University "Ss. Cyril and Methodius"- Skopje (Macedonia)
- Trandafir Romica, Technical University of Civil Engineering (Romania)
- Tsoukias Alexis, Universite Paris Dauphine (France)
- Vujošević Mirko, University of Belgrade (Serbia)
- Yannacopoulos Dionysius, Technological Educational Institution of Piraeus (Greece)
- Yedilkhan Amirgaliyev, Institute of Information and Computational Technologies (Kazakhstan)
- Zadnik-Stirn Lidija, University of Ljubljana (Slovenia)
- Zopounidis Constantin, Technical University of Crete (Greece)

Honorary Committee Chair

- Cvetković Dragoš, Serbian Academy of Sciences and Arts (Serbia)
- Milovanović Gradimir, Serbian Academy of Sciences and Arts (Serbia)

Honorary Scientific Committee

- Guberinić Slobodan, University of Belgrade (Serbia)
- Manos Basil, Aristotle University of Thessaloniki (Greece)
- Krčevinac Slobodan, University of Belgrade (Serbia)
- Papathanasiou Byron, Aristotle University of Thessaloniki (Greece)
- Papachristos Sotirios, University of Ioannina (Greece)
- Petrović Radivoj, University of Belgrade (Serbia)
- Tsouros Constantin, Aristotle University of Thessaloniki (Greece)

Organizing Committee

- **Gordana Savić, University of Belgrade, Faculty of Organizational Sciences (Serbia)**
- Bogdana Stanojević, Mathematical Institute, SANU (Serbia)
- Nebojša Nikolić, University of Belgrade, Faculty of Organizational Sciences (Serbia)
- Marija Kuzmanović, University of Belgrade, Faculty of Organizational Sciences (Serbia)
- Dragana Makajić-Nikolić, University of Belgrade, Faculty of Organizational Sciences (Serbia)
- Biljana Panić, University of Belgrade, Faculty of Organizational Sciences (Serbia)
- Bisera Andrić-Gušavac, University of Belgrade, Faculty of Organizational Sciences (Serbia)
- Milena Popović, University of Belgrade, Faculty of Organizational Sciences (Serbia)
- Dušan Džamić, University of Belgrade, Faculty of Organizational Sciences (Serbia)
- Minja Marinović, University of Belgrade, Faculty of Organizational Sciences (Serbia)
- Milos Nikolić, University of Belgrade, Faculty of Transport and Traffic Engineering (Serbia)
- Gordana Nastić, Mathematical Institute, SANU (Serbia)

Contents

B15. O.R. in Agriculture and Forest Management

B16. Operations Management

B17. Optimization Software

B19. Risk Analysis & Management

B20. Statistics, Stochastics & Simulation

B21. Structural Optimization

B22. Traffic & Transportation

B1
Banking & Financial Decision Making

BALCOR 2018
XIII Balkan Conference on Operational Research

VARIABLE NEIGHBORHOOD SEARCH FOR CARDINALITY CONSTRAINED PORTFOLIO OPTIMIZATION

ANDRIJANA BAČEVIĆ[1]

[1] University of Belgrade, Faculty of Organizational Sciences, andrijana.bacevic@fon.bg.ac.rs

---

Abstract: *In this paper we consider portfolio selection problems in order to provide the investor a suitable balance between risk and return. We consider unconstrained portfolio optimization model of Markowitz and one more realistic including cardinality constraint. We have proposed an Variable Neighborhood Search (VNS) based heuristic to solve the portfolio optimization problem with cardinality constraint. The proposed algorithm is tested on random generated and on the well-known instance problems from OR-Library. Experimental results show that the proposed heuristic provides quality solutions.*

Keywords: *portfolio optimization, cardinality constrained, variable neighborhood search*

1. INTRODUCTION

Managing financial portfolios is primarily concerned with choosing the proportions of various assets to be held in a portfolio, in such a way to make the portfolio better than any other according to given criterion. Expected returns and risks are the most important criteria in portfolio optimization problems. Investors generally prefer to minimize risk while maximizing returns.

The mean–variance (MV) portfolio optimization theory of Harry Markowitz is an investment theory based on the idea that risk-averse investors can construct portfolios to maximize expected return subject to achieving a specified level of calculated risk or equivalently, minimize risk subject to obtaining a predetermined level of expected return . By varying the level of expected return, the Markowitz model determines the so-called efficient frontier, as the set of non-dominated portfolios regarding the two goals (variance and mean of the return). For every level of desired mean return, this efficiency frontier indicates the best investment strategy. The portfolios on the efficient frontier can be found by quadratic programming (QP) using widely open-source or commercial QP solvers. Integer constraints that limit a portfolio to have a specified number of assets change the classical QP into a MIQP (mixed-integer quadratic programming), which can no longer be solved in polynomial time. These requirements come from real-world practice and allow the investors to invest partially in smaller portfolios.

1.1. Literature review

Markowitz mean-variance optimization model is introduced by Markowitz (1952). Since then, many refinements have been proposed to make the Markowitz model more realistic. Brito and Vicente (2013) propose a novel approach to handle cardinality in portfolio selection, by means of a bi-objective cardinality/mean-variance problem, allowing the investor to analyze the efficient trade-off between return and risk and number of active positions.

Markowitz basic model has been widely studied and efficient algorithms have been proposed. Bienstock (1996) proposed a branch-and-cut algorithm. Li et al. (2006) proposed a convergent Lagrangian method as an exact solution scheme. As we mentioned, basic Markowitz portfolio optimization problem can be solved using widely open-source or commercial quadratic programming solvers.

Since exact methods are able to solve only a fraction of practically useful Markowitz models, numerous heuristics have also been proposed for portfolio optimization with real-world constraints. Chang et al. (2000) presented three heuristic algorithms based upon genetic algorithms, tabu search and simulated annealing for finding the cardinality constrained efficient frontier. Cesarone et al. (2009) proposed heuristic approach that starts from a pair of assets and tries to add one asset at the time in an optimal manner by exploiting some recent theoretical results on Quadratic Programming. Deng et al. (2012) proposed improved PSO which increases exploration in the initial search steps and improves convergence speed in the final search steps.

Markowitz theory assumes that the means and covariances of the underlying asset returns are known. In practice, they are unknown and have to be estimated from historical data. Lai et al. (2011) proposed a novel approach and extended Markowitz mean–variance portfolio optimization theory to the case where the means and covariances of the asset returns are unknown.

1.2. Outline

The objective of the research is to develop heuristic method to deliver high-quality solutions for the mean-variance model when enriched by additional practical constraints. The paper is organized as follows: unconstrained and cardinality constrained model of the portfolio optimization problem is stated in section 2; Variable Neighborhood Search algorithm and its main components in detail is described in section 3. Computational results are discussed in section 4; conclusions and final comments are given in section 5.

2. MARKOWITZ MEAN - VARIANCE APPROACH

This section presents the unconstrained portfolio optimization problem and one more realistic including cardinality constraint. A survey on "traditional" portfolio selection according to Markowitz as well as subsequent models, constrained and unconstrained, can be found in (Elton et al., 2009).

2.1. Unconstrained model

The Markowitz model describes a market with n assets characterized by:

- r_i – the expected return of asset i, $(i = 1,2,\dots n)$,
- σ_{ij} – the covariance between assets i and j, $(i = 1,2,\dots n,\ j = 1,2,\dots n)$.

The decision variables are:

- w_i – the proportion held of asset i , $(i = 1,2,\dots n)$.

According to individual preferences, an investor defines weighting parameter λ $(0 \leq \lambda \leq 1)$. This parameter, called risk tolerance, represent the trade-off between risk and return. If $\lambda = 0$ optimal portfolio includes only one asset with the highest return disregarding risk. If $\lambda = 1$ optimal portfolio includes numerous assets with minimal total risk disregarding return.

Using the above notation and Markowitz mean-variance approach, we have that portfolio optimization problem is:

$$\min \quad \lambda \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \cdot \sigma_{ij} \cdot w_j - (1-\lambda) \cdot \sum_{i=1}^{n} w_i \cdot r_i \tag{1}$$

subject to:

$$\sum_{i=1}^{n} w_i = 1, \tag{2}$$

$$w_i \geq 0 \quad \text{for } i = 1,2\dots n, \tag{3}$$

The objective function (1) minimizes the trade-off between risk, and expected return. Constraints (2) and (3) are used to define that portfolio must be fully invested and consist of long positions (no negative weights), respectively.

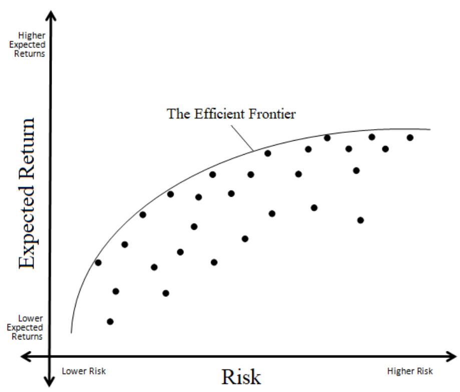


Figure 1 Efficient Frontier example

By resolving this QP for varying values of λ, we can trace out the efficient frontier, a smooth non-decreasing curve that gives the best possible trade-off risk against return. One example is shown in Figure 1. Each dot represents a portfolio. The dots that are closest to the efficient frontier curve are the portfolios that are expected to show the best performance with smallest risk. The curve represents the set of Pareto-optimal (non-dominated) portfolios.

2.2. Cardinality constrained model

Previous model can be extended to the cardinality constrained case (Chang et al., 2000) by introducing:

- k as the desired number of different assets in the portfolio,
- m_i as the minimum proportion of the portfolio allocated to asset i,
- M_i as the maximum proportion of the portfolio allocated to asset i,

New binary variables are:

- $x_i = \begin{cases} 1 & \text{if asset } i \text{ is included in portfolio,} \\ 0 & \text{otherwise} \end{cases}$,

The cardinality constrained portfolio optimization problem is:

$$\min \quad \lambda \cdot \sum_{i=1}^{n} \sum_{j=1}^{n} w_i \cdot \sigma_{ij} \cdot w_j - (1-\lambda) \cdot \sum_{i=1}^{n} w_i \cdot r_i \tag{4}$$

subject to:

$$\sum_{i=1}^{n} w_i = 1, \tag{5}$$

$$\sum_{i=1}^{n} x_i = k, \tag{6}$$

$$x_i \cdot m_i \leq w_i \leq x_i \cdot M_i, \quad i = 1, 2 \ldots n, \tag{7}$$

$$x_i \in \{0, 1\}, \quad i = 1, 2 \ldots n, \tag{8}$$

Constraint (6) is used to define that exactly k assets are held. Constraint (7) ensures that if any of asset i is included in portfolio its proportion w_i must lie between m_i and M_i otherwise its proportion w_i is zero.

3. VARIABLE NEIGHBORHOOD SEARCH

Variable Neighborhood Search (VNS), proposed by Mladenović and Hansen (1997), is a metaheuristic method for solving a set of combinatorial optimization and global optimization problems. The main idea of VNS algorithm is to explore the search space with a systematic change of neighborhood. The success of VNS comes from the fact that a local minimum for one neighborhood structure may not be a local minimum for a different neighborhood structure enabling improvements towards the global minimum.

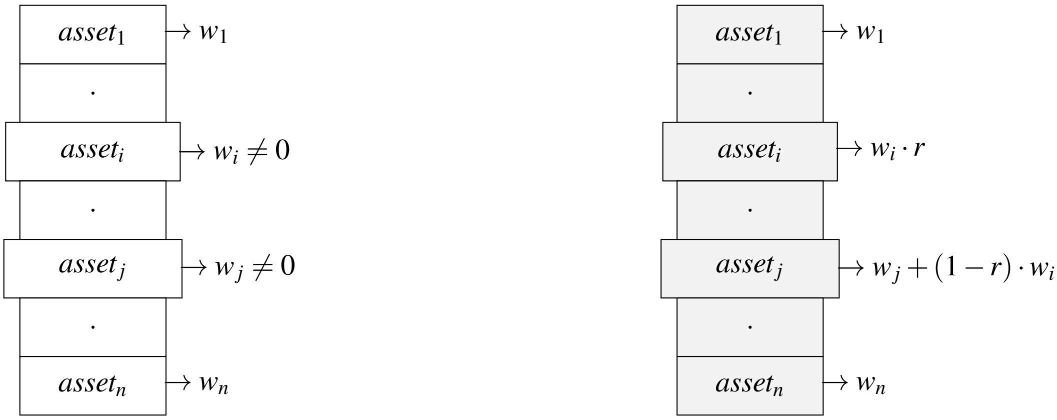
In this paper, we proposed a heuristic based on VNS to solve the constrained portfolio optimization problem. A solution represents an array w (length n) of associated weights w_i for each asset i. The main steps of the basic VNS are given as follows:

1. Initialization: Select a set of neighborhood structures $N_p (p = 1, \ldots, p_{max})$, find an initial solution w, set $p = 1$, choose a stopping condition.
2. Shaking: Generate a solution $w' \in N_p(x)$ at random.
3. Local search: Apply a local search method starting with w' to find local minimum w''.
4. Move or not: If w'' is better than the incumbent, then set $w = w''$ and $p = 1$, otherwise set $p = p + 1$ (or if $p = p_{max}$ set $p = 1$).
5. Test stop condition: If stop condition is not satisfied then go to step 2. Otherwise return the best solution w.

A initial solution is generated by random selection of k assets with weight $\frac{1}{k}$. In the shaking step algorithm explores different types of p-neighborhoods, focusing on different aspects and properties of the solutions:

- type 1 – remove p randomly chosen assets from portfolio and then rescale w such that $\sum_{i=1}^{n} w_i = 1$,
- type 2 – add p randomly chosen assets which is not included in portfolio, and then rescale w,
- type 3 – p times change weights for two randomly chosen assets.
- type 4 – combines type 1 (remove) and type 2 (add) in one step.

For local search we use best-improvement strategy and neighborhood which contains solutions obtained by changing weights (Figure 2) for each pair of assets in portfolio.



Original portfolio   One portfolio from neighborhood

Figure 2 Solution neighborhood in local search

Local search is the most time-consuming component of VNS algorithm. In order to speed up the local search, we also implemented fast update (complexity $O(n)$) of objective function after changing weights for a pair of assets. If value w_j change with value w'_j then:

$$f_{new} = f_{current} + 2\lambda \cdot \Delta w_j \cdot \sum_{i=1, i\neq j}^{n} w_i \cdot \sigma_{ij} + (\Delta w_j)^2 \cdot \sigma_{jj} + (1-\lambda) \cdot \Delta w_j \cdot r_j \tag{9}$$

where $\Delta w_j = w'_j - w_j$. This speedup makes exploring higher number of feasible solutions possible. The pseudo-code of the local search is given in Algorithm 1.

input : $w_{initial}$
output : solution w_{best}

$w_{best} \leftarrow w_{initial}$;
$w_{temp} \leftarrow w_{initial}$;
ActiveAssets $\leftarrow \{i \mid w_{temp(i)} > 0\}$;
for i in ActiveAssets **do**
    for j in ActiveAssets **do**
        $r \leftarrow$ random$(0,1)$;
        $w_{temp(i)} \leftarrow r \cdot w_{temp(i)}$;
        $w_{temp(j)} \leftarrow w_{temp(j)} + (1-r) \cdot w_{temp(i)}$;
        if $f(w_{temp}) < f(w_{initial})$ **then**
            $w_{best} \leftarrow w_{temp}$;
        else
            $w_{temp} \leftarrow w_{initial}$;
        end
    end
end

Algorithm 1: Local search procedure

All constraints are actively checked (before move or not step) using one procedure, which makes adding, deleting or changing one or many of constraints simple enough. For this study there is only cardinality constraint but any other constraints, i.e minimum transaction cost or sector capitalization can be added.

4. COMPUTATIONAL RESULTS

In this section, we evaluate efficiency of proposed VNS heuristic through a set of computational experiments. Computational experiments were first performed on problem instances that were generated randomly.

In order to illustrate the effect of cardinality constraints we consider the small five asset example. Figure 3 show comparisons of unconstrained efficient frontier and cardinality constrained efficient frontier obtained by VNS. The proposed VNS searches for efficient frontiers by testing different values for the risk tolerance parameter λ in the cardinality constrained Markowitz portfolio model.

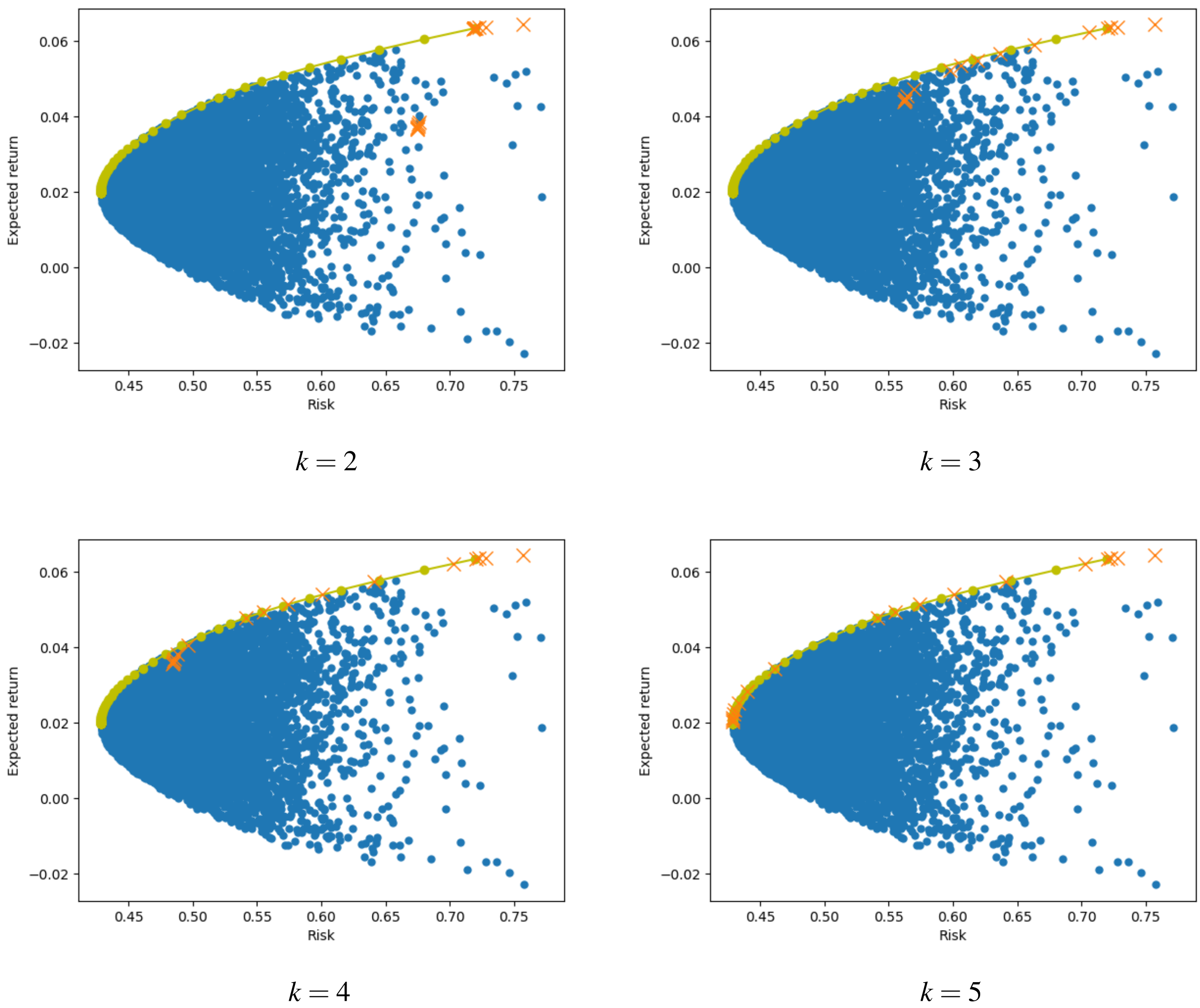


Figure 3 Comparison of solutions obtained by VNS heuristic for different cardinality value k and risk tolerance parameter λ ("$\bullet$" represents a portfolio, "$\bullet - \bullet - \bullet$" curve represents unconstrained efficient frontier and "$\times$" represent portfolio obtained by VNS)

We have also performed experiments on real-world instances which are available at OR-Library. These five benchmark datasets consists of 31, 85, 89, 98 and 225 assets and represent Hang Seng in Hong Kong, DAX in Germany, FTSE in UK, S&P in USA and Nikkei in Japan indices, respectively. The cardinality constraint used the values $k = 10$ and $k = 20$ and objective function used risk tolerance $\lambda = 1/2$. Table 1 shows the risk and expected return for obtained portfolio, for each instance. For portfolios with cardinality constraint, obtained by VNS, we calculated Euclidean distance from the unconstrained efficient portfolio. Experimental results show that the proposed heuristic provides quality solutions for cardinality constrained problem in both case ($k = 10$ and $k = 20$). Obtained solutions are very close to unconstrained efficient frontier curve.

Table 1: Comparison results

			unconstrained efficient frontiers		VNS - cardinality constrained					
					$k = 10$			$k = 20$		
	n	name	risk	return	risk	return	distance	risk	return	distance
1	31	Hang Seng	0.002492	0.009213	0.002493	0.009213	1.79E-07	0.002492	0.009213	1.98E-07
2	85	DAX	0.001028	0.009248	0.001038	0.009259	1.46E-05	0.001034	0.009254	7.67E-06
3	89	FTSE	0.001268	0.007994	0.001042	0.006186	1.82E-03	0.001364	0.008079	1.28E-04
4	98	S&P	0.001290	0.008562	0.001487	0.008347	2.92E-04	0.001001	0.006184	2.40E-03
5	225	Nikkei	0.000728	0.003631	0.001088	0.001905	1.76E-03	0.001007	0.002847	8.32E-04

5. CONCLUSION

In this paper, we have proposed a VNS based heuristic to solve the portfolio optimization problem with cardinality constraint. The proposed algorithm is tested on random generated and on the well-known instance problems from OR-Library. Experimental results show that the proposed heuristic provides quality solutions but unfortunately no comparison with other results is possible at this stage. In a forthcoming paper we will extend this algorithm to different version of the portfolio optimization, possibly considering additional complex constraints. Possible hybridization of VNS with other metaheuristics will also be considered.

Acknowledgment

I thank members from Laboratory for Operations Research - Dr Jovan Petrić and Laboratory for Mathematics, Faculty of Organizational Sciences for support and to Dragana Makajić-Nikolić for comments on earlier draft of this paper.

REFERENCES

Bienstock, D. (1996). Computational study of a family of mixed-integer quadratic programming problems. *Mathematical programming*, 74(2):121–140.

Brito, R. P. and Vicente, L. N. (2013). Efficient cardinality/mean-variance portfolios. In *IFIP Conference on System Modeling and Optimization*, pages 52–73. Springer.

Cesarone, F., Scozzari, A., and Tardella, F. (2009). Efficient algorithms for mean-variance portfolio optimization with hard real-world constraints. *Giornale dell'Istituto Italiano degli Attuari*, 72:37–56.

Chang, T.-J., Meade, N., Beasley, J. E., and Sharaiha, Y. M. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers & Operations Research*, 27(13):1271–1302.

Deng, G.-F., Lin, W.-T., and Lo, C.-C. (2012). Markowitz-based portfolio selection with cardinality constraints using improved particle swarm optimization. *Expert Systems with Applications*, 39(4):4558–4566.

Elton, E. J., Gruber, M. J., Brown, S. J., and Goetzmann, W. N. (2009). *Modern portfolio theory and investment analysis*. John Wiley & Sons.

Lai, T. L., Xing, H., and Chen, Z. (2011). Mean-variance portfolio optimization when means and covariances are unknown. *The Annals of Applied Statistics*, pages 798–823.

Li, D., Sun, X., and Wang, J. (2006). Optimal lot solution to cardinality constrained mean–variance formulation for portfolio selection. *Mathematical Finance*, 16(1):83–101.

Markowitz, H. (1952). Portfolio selection. *The journal of finance*, 7(1):77–91.

Mladenović, N. and Hansen, P. (1997). Variable neighborhood search. *Computers & operations research*, 24(11):1097–1100.

B2
Combinatorial Optimization & Integer Programming

XIII Balkan Conference on Operational Research

SOME STATIC ROMAN DOMINATION NUMBERS FOR FLOWER SNARKS

Zoran Maksimović[1], Jozef Kratica[2], Aleksandar Savić[3], Milena Bogdanović[4]

[1] University of Defence, Military Academy, zoran.maksimovic@gmail.com
[2] Mathematical Institute, Serbian Academy of Sciences and Arts, jkratica@mi.sanu.ac.rs
[3] University of Belgrade, Faculty of Mathematics, aleks3rd@gmail.com
[4] University of Niš, Pedagogical Faculty, mb2001969@beotel.net

Abstract: *This paper is devoted to the problem of finding the Roman, restrained Roman and signed total Roman domination number for flower snark graphs. The exact values of Roman and restrained Roman domination number are determined and proved. For signed total Roman domination number tight upper bound is presented.*

Keywords: *Roman domination, restrained Roman domination, signed total Roman domination, flower snarks graphs*

1. INTRODUCTION

The problem of Roman domination on graphs has arisen and was motivated by article of Ian Stewart in (Stewart (1999)), though concept was even earlier suggested by ReVelle (1997). The original problem was formulated in military history.

At the beginning of the 4th century AD, Roman Empire was consolidated and reformed by Emperor Constantine I (306 – 337). The previous century brought much destruction both through barbarian invasions and civil wars and is known as The Crisis of the Third Century. The Empire was economically exhausted and in one period (258 – 274) split in three part all of them boasting the Roman inheritance: central government (in Italy, Africa, Balkans and Central Europe), Gallic Empire (Hispania, Gallia, Germania and Britannia) and Palmyrene Empire (Middle East, Asia Minor and Egypt). After military victories of the central government under Aurelian (270 – 275) and administrative reorganization under Diocletian (284 – 305), Empire was stabilized both militarily and economically. Barbarians were driven out, Empire was united, inflation was put under control by government intervention. Another round of civil war after the death of Constantius I (father of Constantine I) has put on the throne as a sole ruler in 324, Emperor Constantine.

He did some remarkable things: reorganized the economy by introducing golden coin solidus (or denarius), allowing religious tolerance (Milan Edict in 312), built new capital (Constantinople) and finally reorganizing army. In the third century, borders were easily punctuated by barbarians because Roman strategy was based on preemptive strikes on large groups of barbarians on their territory, outside the Emipre. Civil wars and quick succession of short-lived emperors, prevented the strategy of such attacks and empire switched to defensive, at first only on the borders and afterwards defensive-in-depth. Constantine I organized army in order to implement this defensive-in-depth, which placed armed forces not only on the borders but throughout all of the Empire. To achieve this, army was organized in stationary units, stationed mostly in border regions (limitanei) and garrisons on key points and mobile troops (comitatensis) stationed on few key points in the Empire which will act quickly in support to endangered regions. Mobile units were better equipped, trained and payed than stationary units. To accomplish this strategy it was necessary to decide were to establish units for quick response and were to put stationary units. Major problem was payment of the army (which was up to 3/4 of the tax revenues), so optimal disposition was of utmost importance. Constantine I decided to differentiate communities of the Empire in 3 categories: those with mobile and stationary troops, those with only stationary troops and those without troops at all. The condition was that communities without troops must be in neighborhood of communities with mobile troops, so that in case of attack they can be defended. Detailed explanations of strategies applied by Roman Empire through history can be found in (Luttwak (2016)).

This disposition represents very interesting optimization problem. The territory of the Empire is represented by the graph $G = (V, E)$ in which communities are vertices and edges exist between neighboring communities. Formally, problem can be formulated as finding function $f : V \to \{0, 1, 2\}$ such that value $\sum_{v \in V} f(v)$ is minimal, while any vertex v with $f(v) = 0$ must be adjacent to some vertex u with $f(u) = 2$. This function is called *Roman dominating function* (RDF) and was introduced by Cockayne et al. (2004). If we define $w(f) =$

$f(V) = \sum_{v\in V} f(v)$ as a weight of function f, than *Roman domination number*, denoted as $\gamma_R(G)$, is defined as $\gamma_R(G) = \min\{w(f)|f \,\text{is a RDF on}\, G\}$.

Let us present the notion of Roman domination with the following example.

Example. On Figure 1 a simplified territory of the Roman Empire is presented. Find an optimal RDF and γ_R.

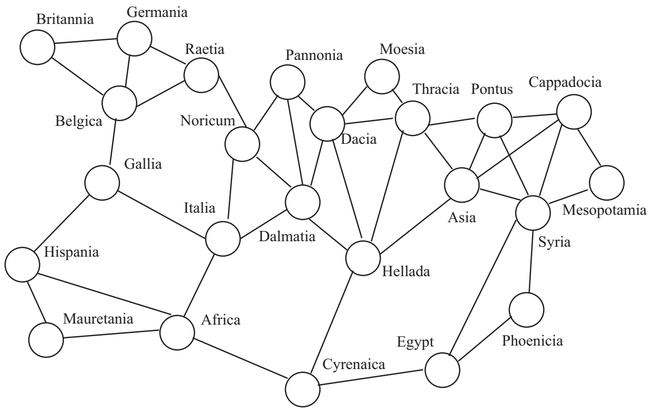


Figure 1 Graph of Roman Empire

The Roman domination number of graph on Figure 1 has value 9. Provinces with 2 units are Belgica, Africa, Dacia and Syria. Province with 1 unit is Noricum. As it can be seen Belgica protects Britannia, Germania, Raetia and Gallia. Africa protects Hispania, Mauretania, Italia and Cyrenaica. Dacia protects Pannonia, Dalmatia, Hellada, Moesia and Thracia. Syria protects Asia, Pontus, Cappadocia, Mesopotamia, Phoenicia and Egypt. The only province that was not protected is Noricum so it must have 1 unit to garrison it and protect it.

The Roman dominating function f partitions the set V in three disjunct sets V_0, V_1 and V_2, where $V_i = \{v \in V|f(v) = i\}$. Numbers $n_i = |V_i|$, $i = 0,1,2$ will represent cardinality of sets V_i. Now, the weight of a Roman dominating function f is equal to $f(V) = \sum_{v\in V} f(v) = 2\cdot n_2 + 1\cdot n_1 + 0\cdot n_0 = 2\cdot n_2 + n_1$.

From the basic problem of Roman domination, a multitude of similar problems were formulated. In this paper we will consider two of them: Restrained Roman Domination Problem (RRDP) and Signed Total Roman Domination Problem (STRDP).

The Restrained Roman domination problem was introduced by Pushpam and Padmapriea (2015) and the problem of finding the minimal number of troops such that entire Empire would be defended but with changed conditions was compared to RDP. A community is considered to be secured if at least one troop is stationed within. The condition for a community without troops within is now that it is secure if it is adjacent to at least one community with two troops and to at least one community without troops. The appropriate function is denoted as *Restrained Roman Domination Function* (RRDF). Mathematically speaking, condition that a community is secure is that for any vertex $v \in V$ value $f(v)$ is either $f(v) \geq 1$ or $f(v) = 0$ and there exist two vertices u and w adjacent to v such that $f(u) = 0$ and $f(w) = 2$. Appropriate domination number, denoted as $\gamma_{rR}(G) = min\{w(f)|f \,\text{is a RRDF on}\, G\}$.

Similarly, the Signed Total Roman Domination Problem (STRDP), introduced by Volkmann (2016), can be defined as finding function $f : V \to \{-1,1,2\}$ such that (i) for every vertex $v \in V$ such that $f(v) = -1$ there is adjacent vertex u such that $f(u) = 2$ and (ii) if we denote $N(v) = \{u \in V|\{u,v\} \in E\}$ i.e. *open neighborhood* of v, for every $v \in V$ holds $f(N(v)) = \sum_{u\in N(v)} f(u) \geq 1$. Interpretation for this variant of RDP is that vertices with value $f(v) = -1$ represent weak spots in the defense. The appropriate function is denoted as *Signed Total Roman Domination Function* (STRDF) and appropriate domination number, denoted as $\gamma_{stR}(G) = min\{w(f)|f \,\text{is a STRDF on}\, G\}$. The STRDP was introduced by Ahangar et al. (2014).

Example 2. Find optimal RRDF and SRDF and respective domination numbers γ_{rR} and γ_{stR} for graph on Figure 1.

RRD number is 9. Number of troops in the provinces is the same as for RD. Any province has one neighbor without unit which can be checked directly, for example province of Hellada is a neighbor to Cyrenaica.

STRD number is 9. Provinces with 2 units are Germania, Hispania, Dalmatia, Dacia, Hellada, Egypt and Syria. Provinces with 1 unit are Belgica, Gallia, Mauretania, Raetia, Asia and Phoenicia. All other provinces are weak spots. Number of troops in the neighborhood of any province is greater or equal to 1 which can be ascertained by direct probe. For example, number of troops in the neighborhood of Thracia is 1, since Dacia and Hellada have 2 troops, while Moesia, Asia and Pontus are weak spots (value is -1). The fact that every province which is weak spot has a neighbor with 2 units can be also easily checked.

Except in the military history, domination in graph have found various applications in different kinds of networks, like social networks, biological networks, distributed networks etc. as in (Behtoei et al. (2014)). Specifically, some facility location problems can be interpreted as Roman domination (Chambers et al. (2009)). Instead of interpreting $f(v)$ as the number of units placed at v, it can be viewed as a cost function. Units with cost 2 may be able to serve neighboring locations, while units with cost 1 can serve only their own location. For example, in a communication network, wireless hubs are more expensive but can serve neighboring locations, while wired hubs are low-range but are cheaper.

The Flower Snarks are specific class of regular graphs and are shown on Figure 2. The Flower snark graphs were introduced by Isaacs (1975) as an example of a cubic bridgeless graph family that is not 3-edge-colorable. The degree of vertices in Flower snarks are 3. The set of vertices is $V = \{a_i, b_i, c_i, d_i, |i = 0, \ldots, r-1\}$. The set of edges can be generated through neighborhood of characteristic vertices. So, noting that indices are by modulo r, we have

$$N(a_i) = \{b_i, c_i, d_i\}, N(b_i) = \{b_{i-1}, a_i, b_{i+1}\}, i = 0, \ldots, r-1$$

and

$$N(c_i) = \{c_{i-1}, a_i, c_{i+1}\}, N(d_i) = \{d_{i-1}, a_i, d_{i+1}\}, i = 1, \ldots, r-2$$

while

$$N(c_0) = \{d_{r-1}, a_0, c_1\}, N(c_{r-1}) = \{c_{r-2}, a_{r-1}, d_0\}, N(d_0) = \{c_{r-1}, a_0, d_1\}, N(d_{r-1}) = \{d_{r-2}, a_{r-1}, c_0\},$$

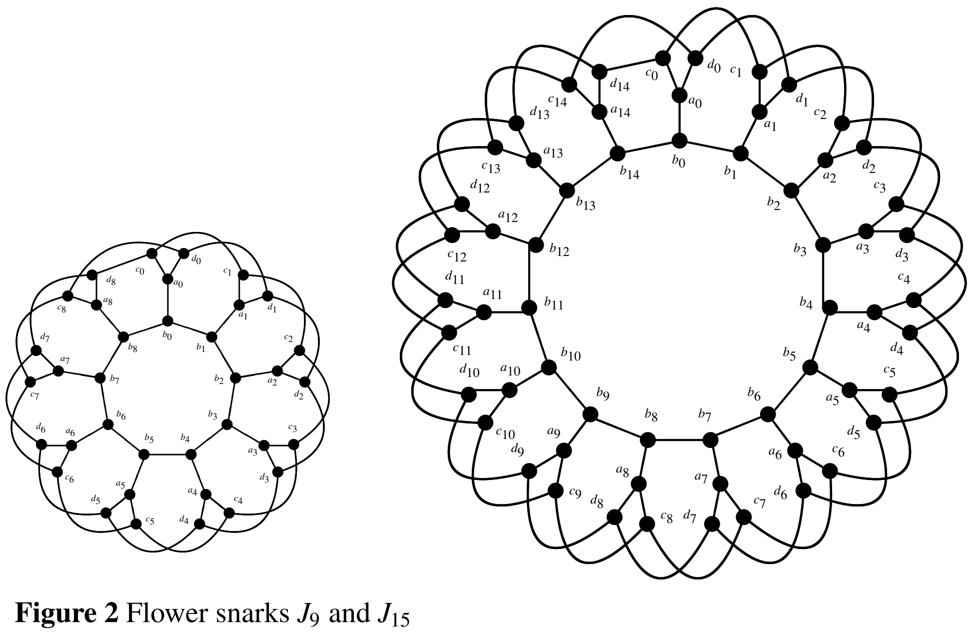


Figure 2 Flower snarks J_9 and J_{15}

Let $G = (V, E)$ be a graph, where order of G is $n = |V|$ and size of G is $m = |E(G)|$.

Let δ and Δ be the minimum and the maximum degree of vertices in G, respectively.

A set $S \subseteq V$ is the *dominating set* if every vertex in $V \setminus S$ is adjacent to at least one vertex in S. The domination number $\gamma(G)$ is the minimum cardinality of the dominating set in G.

2. PREVIOUS WORK

2.1. Roman domination problem

There is extensive literature on the domination set problems in previous decades, and RDP as one of these problems was also intensively studied.

As mentioned above, the definition of RDF is given in (Cockayne et al. (2004)) and some basic properties of these functions were studied. The authors also found $\gamma_R(G)$ for some classes of graphs. In their article they established following relationship between domination and Roman domination numbers for arbitrary graph G:

$$\gamma(G) \leq \gamma_R(G) \leq 2 \cdot \gamma(G).$$

They characterized the graphs for which $\gamma_R(G) = \gamma(G) + k$ for $k \leq 2$. Xing et al. (2006) extended this result to arbitrary k. Similar result for trees is given by Song and Wang (2006) with $\gamma_R(T) = \gamma(T) + 3$.

Of special interest for this paper are graphs with $\gamma_R(G) = 2\gamma(G)$ which are called Roman graphs, because we will prove that flower snark graphs are Roman graphs. Some well-known classes of graphs such as $P_{3k}, P_{3k+2}, C_{3k}, C_{3k+2}$ for $k \geq 1$ and $K_{m,n}$ for $\min(n,m) \neq 2$ were proved to be Roman graphs by Xueliang et al. (2009). They also proved that some regular graphs are Roman graphs: some subclasses of circulant and generalized Petersen graphs, Cartesian product of graphs $C_{5m} \square C_{5n}$, where $m \geq 1, n \geq 1$. Characterization of Roman trees was presented by Henning (2002).

The proof that RDP is NP-hard in a general case was given by Dreyer (2000). In (Shang and Hu (2007)) it was proven that Roman domination problem in unit disk graphs is also NP-hard. Nascimento and Sampaio (2015) proved that the RDP is NP-hard even for the subgraphs of grid graphs and bipartite planar graphs with $\Delta = 4$. Nevertheless, for some classes of graphs solution could be find in polynomial time. Roman domination number for interval graphs could be calculated by linear-time algorithms (Liedloff et al. (2005)). The polynomial-time algorithm for AT-free graphs was presented in the same article.

It is of interest to find lower and upper bounds, both in general case and for some specific classes of graphs.

The following lower bound, established by Cockayne et al. (2004), is very useful thanks to its simplicity:

▶ Proposition 1. For any graph G, $\gamma_R(G) \geq \frac{2|V(G)|}{1+\Delta(G)}$.

In (Chambers et al. (2009)) it was proved that $\gamma_R(G) \leq 4n/5$ if $\delta(G) \geq 1$ and $\gamma_R(G) \leq 8n/11$ if $\delta(G) \geq 2$. Similarly, if $\delta \geq 3$ than $\gamma_R(G) \leq 2n/3$ (Liu and Chang (2012b)).

Some lower and upper bounds using the diameter and the girth were proposed by Mobaraky and Sheikholeslami (2008). These bounds were improved by Bermudo et al. (2014) for $\delta(G) \geq 2$

$$\gamma_R(G) \leq n - \left(\left\lfloor \frac{Diam(G)}{3} \right\rfloor + 1 \right) (\delta(G) - 1)$$

In the same paper two useful upper bounds were proposed:

▶ Proposition 2. Let G be a graph of order n. Then $\gamma_R(G) \leq \left\lfloor \frac{2n\delta(G)}{3\delta(G)-1} \right\rfloor$.

▶ Proposition 3. If G is a graph of order n and size m, then $\gamma_R(G) \leq \min \left\{ \left\lfloor \frac{3\Delta(G)n-2m}{3\Delta(G)-1} \right\rfloor, \left\lfloor \frac{(3\Delta(G)+4)n-2m}{3\Delta(G)+4} \right\rfloor \right\}$.

Moreover, if G is a C_5-free graph, then $\gamma_R(G) \leq \left\lfloor \frac{(3\Delta(G)+2)n-2m}{3\Delta(G)+2} \right\rfloor$.

Another upper bound $\gamma_R(G) \leq 2\left(1 - \frac{2^{1/\delta(G)}\delta(G)}{(1+\delta(G))^{1+1/\delta(G)}}\right)n$, including the proof that it is asymptotically best possible is given by Zverovich and Poghosyan (2011).

One of the most used estimates of the RD number is given by Cockayne et al. (2004) and concerns relationship between RD number and domination number of the same graph, namely, $\gamma(G) \leq \gamma_R(G) \leq 2\gamma(G)$. Combining works by Favaron et al. (2009) and Bermudo and Fernau (2012), Bermudo et al. (2014) derived:

▶ Proposition 4. If G is a graph of order $n \geq 3$, then $n - \gamma(G)(\Delta(G) - 1) \leq \gamma_R(G) \leq n - \frac{\gamma(G)}{2}$.

Another lower bound using total domination number, $\gamma_t(G) \leq \gamma_R(G)$ is presented by Hedetniemi et al. (2013).

There are some specific bounds for connected graphs given in (Liu and Chang (2012a); Favaron et al. (2009)) respectively, $\gamma_R(G) \leq \max\{\lceil 2n/3 \rceil, \lceil 23n/34 \rceil\}$ and $\gamma_R(G) + \gamma(G)/2 \leq n$ where $n \geq 3$ is order of G. Finally, let G be a nontrivial, connected graph with maximum degree Δ. Then $\gamma_R(G) \geq \frac{\Delta+1}{\Delta}\gamma(G)$. The proof is given by Chellali et al. (2016).

The differential of a vertex set S is defined as $\partial(S) = |B(S)| - |S|$, where $B(S)$ is the set of vertices in $V \setminus S$ that have a neighbor in the vertex set S, and the differential of a graph is defined as $\partial(G) = \max\{\partial(S) | S \subseteq V\}$. A relation between the Roman domination number and the differential of a graph is studied by Bermudo et al. (2014).

For the Roman domination problem several integer linear programming (ILP) formulations was proposed. The first formulation was introduced by ReVelle and Rosing (2000). Another ILP formulations for Roman domination were proposed by Burger et al. (2013). These formulations were improved using a fewer number of constraints by Ivanović (2016).

Two approximation schemes, one 5-approximation algorithm of linear time and a one of polynomial-time were discussed by Shang and Hu (2007).

Since RDP could not be easily solved in the general case, of interest was to study it for different classes of graphs:

- interval graphs, cographs, asteroidal triple-free graphs and graphs with a d-octopus by Liedloff et al. (2005);
- corona graphs by Yero et al. (2013);
- grid graphs by Currò (2014);
- Generalized Sierpiński graphs by Ramezani et al. (2016);
- Generalized Petersen Graphs $GP(n;2)$ by Wang et al. (2011) and $GP(n;3)$ and $GP(n;4)$ by Zhiqiang Zhang and Xu (2014);
- cardinal product of paths and cycles in Klobučar and Puljić (2014, 2015);
- strongly chordal graphs by Liu and Chang (2013);
- digraphs by Sheikholeslami and Volkmann (2011);
- complementary prisms by Al Hashim (2017).

and others.

2.2. Restrained Roman domination

The problem of Restrained Roman domination is NP-hard which was proved in general case by Rad and Krzywkowski (2015). The Restrained Roman domination on graph $G = (V,E)$ is closely related to the Restrained domination problem which is to find a set of restrained domination of minimal cardinality, where the set of restrained domination $S \subseteq V$ is a set of vertices with neighbors both in S and in $V \setminus S$. The minimal cardinality of restrained domination set is denoted as γ_{str}.

Some of the properties of RRDF are given by Pushpam and Padmapriea (2015) and can be summarized as

► Proposition 5.
- For any graph G $\gamma_{str}(G) \leq \gamma_{rR} \leq 2\gamma_{str}(G)$;
- if G is a graph of order n and has a vertex od degree $n-1$ and $\delta(G) > 1$, then $\gamma_{rR} = 2(\gamma_{str}(G) = 1)$;
- if a graph G has C_3, then $\gamma_{rR} < n$;

Some bounds for the RRD number for connected graphs are given by Rad and Krzywkowski (2015).

► Proposition 6.
- For every connected graph G of order n, $\gamma_{rR}(G) < n+1-\lfloor (diam(G)-2)/3 \rfloor$;
- for every connected graph G of order n and circumference $g(G)$ holds $\gamma_{rR}(G) < n+1-\lfloor (g(G)-2)/3 \rfloor$;
- for every connected graph G of order n and size m the inequality $\gamma_{rR}(G) \leq 2m-n+2$ holds, and inequality is strict if and only if G is a tree with $\gamma_{rR} = n$.
- for every graph G of order n and if $\delta > 0$ and $n < \delta(\delta-1)/(\ln\delta - \ln 2 + 1)$ than holds

$$\gamma_{rR} \leq n \left(\frac{2\ln(1+\delta) - \ln 4 + 2}{\delta + 1} \right)$$

In (Pushpam and Padmapriea (2015)) are given values of γ_{rR} for some classes of graphs:

► Proposition 7.
- For $n \geq 4$, $\gamma_{rR}(P_n) = \frac{2n+3+r}{3}, \quad n \equiv r (\bmod 3)$;
- $\gamma_{rR}(C_n) = \begin{cases} \frac{2n+3+r}{3}, & n \equiv r(\bmod 3),\ r \in \{1,2\} \\ \frac{2n}{3}, & n \equiv 0(\bmod 3) \end{cases}$;
- $\gamma_{rR}(K_n) = 2$;
- $\gamma_{rR}(K_{m,n}) = 4$.

Several relationships between γ_{rR} and γ_{rst} were proved in (Pushpam and Padmapriea (2015)). Especially, they characterized graphs where $\gamma rR = \gamma_{rst} + k$, $k \in \{1,2\}$ and some other bound on γ_{rR} for trees and bipartite graphs.

2.3. Signed Total Roman domination

Some useful bounds on STRD number are given by Volkmann (2016).

► Proposition 8. Let $f = (V_{-1}, V_1, V_2)$ be a STRDF in a graph G of order n. Let $\delta = \delta(G) \geq 1$ and $\Delta = \Delta(G)$. Then the following holds

- $(2\Delta-1)|V_2| + (\Delta-1)|V_1| \geq (\delta+1)|V_{-1}|$;
- $(2\Delta+\delta)|V_2| + (\Delta+\delta)|V_1| \geq (\delta+1)n$;
- $(\Delta+\delta)w(f) \geq (\delta+2-\Delta)n + (\delta-\Delta)|V_2|$;

- $w(f) \geq (\delta + 2 - 2\Delta)n/(2\Delta + \delta) + |V_2|$.

In the same paper exact values of γ_{stR} for some classes of graphs are proved:

- if $n \geq 3$ then $\gamma_{stR}(K_{1,n-1}) = \gamma_{stR}(K_n) = 3$;
- if $n \geq 1, \gamma_{stR}(K_{n,n}) = 2$, unless $n = 3$ in which case $\gamma_{stR}(K_{3,3}) = 4$;
- if C_n be a cycle of order $n \geq 3$, then $\gamma_{stR}(C_n) = n/2$ if $n \equiv 0(\mod 4)$, $\gamma_{stR}(C_n) = (n+3)/2$ if $n \equiv 1, 3(\mod 4)$ and $\gamma_{stR}(C_n) = (n+6)/2$ if $n \equiv 2(\mod 4)$;
- if P_n be a path of order $n \geq 3$, then $\gamma_{stR}(P_n) = n/2$ if $n \equiv 0(\mod 4)$, $\gamma_{stR}(P_n) = \lceil (n+3)/2 \rceil$. otherwise.

Also, there are presented and proved some bounds

- let G be a graph of order n, $\delta \geq 1$. If $\delta < \Delta$, then $\gamma_{stR}(G) \geq \left\lceil \frac{(2\delta+3-2\Delta)n}{2\Delta+\delta} \right\rceil$;
- let G be a graph of order n, $\delta \geq 1$. Then a) $\gamma_{stR}(G) \leq n$, b) if $\delta \geq 3$ then $\gamma_{stR}(G) \leq n-1$;
- let G be a graph of order $n \geq 3$ and $\delta \geq 1$, then $\gamma_{stR}(G) \geq \frac{3}{2}(1+\sqrt{2n+1}) - n$;
- if G is connected graph of order $n \geq 3$ and size m, then $\gamma_{stR}(G) \geq \frac{11n-12m}{4}$;
- if G is bipartite graph of order $n \geq 3$ with $\delta(G) \geq 1$, then $\gamma_{stR}(G) \geq 3\sqrt{n} - n$;
- if G is graph of order $n \geq 3$ with $\delta(G) \geq 1$, then $\gamma_{stR}(G) \geq \max\{\Delta(G)+1-n, \delta(G)+4-n\}$;
- if T is a tree of order n and $\Delta(G) \geq 2$, then $\gamma_{stR}(T) \geq \Delta(G)+4-n$;
- if G is a graph of order n with $\delta(G) \geq 1$, then $\gamma_{stR}(G) \geq \left(1 + \left\lfloor \frac{diam(G)}{3} \right\rfloor\right)(\delta(G)+1) - n$;
- if G is an r-regular graph of order n such $r \geq 1$ and $n-r-1 \geq 1$ then $\gamma_{stR}(G) + \gamma_{stR}(\bar{G}) \geq \frac{4n}{n-1}$, and if n is even then $\gamma_{stR}(G) + \gamma_{stR}(\bar{G}) \geq 4(n-1)/(n-2)$ where $\bar{G}$ is complement of G.

3. THE NEW RESULTS

▶ Theorem 1. $\gamma_R(J_r) = 2r$

Proof. Step 1. $\gamma_R(J_r) \leq 2r$
Let us define the function $f : V_{J_r} \to \{0, 1, 2\}$ as partition (V_0, V_1, V_2), such that $V_2 = \{a_i | i = 0, ..., r-1\}$, $V_0 = V_{J_r} \setminus V_2$ and $V_1 = \emptyset$. Since $\bigcup\limits_{i=0}^{r-1} \mathbb{N}[a_i] = \bigcup\limits_{i=0}^{r-1} \{a_i, b_i, c_i, d_i\} = V_{J_r}$ implying that each vertex from V_0 has at least one a-vertex as its neighbor. Since all a-vertices are in V_2, then each vertex from V_0 has at least one neighbor from V_2, so f is Roman domination function with value $f(V_{J_r}) = 2r$ so $\gamma_R(J_r) \leq 2r$.
Step 2. $\gamma_R(J_r) \geq 2r$
It is easy to see that J_r is a regular graph of degree 3, with $4r$ vertices. Then, by Proposition 1 it holds $\gamma_R(J_r) \geq \lceil \frac{2 \cdot 4 \cdot r}{1+3} \rceil = 2r$. ◀

▶ Proposition 9. $\gamma_{rR}(J_r) = 2r$

Proof. The function f defined in the proof of Theorem 1 has additional property that each vertex from V_0 has at least neighbor from V_0:

- $b_i, b_{i+1} \in V_0$ and $b_{i+1} \in \mathbb{N}[b_i]$;
- If $i \leq r-2$ then $c_i, c_{i+1} \in V_0$ and $c_{i+1} \in \mathbb{N}[c_i]$;
- If $i = r-1$ then $c_{r-1}, d_0 \in V_0$ and $c_{r-1} \in \mathbb{N}[d_0]$;
- If $i \leq r-2$ then $d_i, d_{i+1} \in V_0$ and $d_{i+1} \in \mathbb{N}[d_i]$;
- If $i = r-1$ then $d_{r-1}, c_0 \in V_0$ and $d_{r-1} \in \mathbb{N}[c_0]$.

Having in mind the proof of Theorem 1, the function f is restrained Roman dominating function of J_r, so $\gamma_{rR}(J_r) \leq 2r$. On the other hand, for any graph G holds $\gamma_{rR}(G) \geq \gamma_R(G)$, implying $\gamma_{rR}(J_r) \geq \gamma_R(J_r) = 2r$. Therefore, $\gamma_{rR}(J_r) = 2r$. ◀

▶ Theorem 2. $\gamma_{stR}(J_r) \leq 3r$

Proof. Let us define the function $f : V_G \to \{-1, 1, 2\}$ as in (1):

$$f(v) = \begin{cases} -1, \ v = a_i \\ 1, \ v = c_i \vee v = d_i \\ 2, \ v = b_i \end{cases} \qquad i = 0, \ldots, r-1 \tag{1}$$

Firstly, each vertex with value -1 has (at least) one neighbor with value 2, since $f(a_i) = -1$, $f(b_i) = 2$ and $b_i \in N(a_i)$.

Secondly, the sum of function values in the open neighborhood of each vertex is at least 1, since:

- $f(N(a_i)) = f(b_i) + f(c_i) + f(d_i) = 2+1+1 = 4 \geq 1$;
- $f(N(b_i)) = f(a_i) + f(b_{i-1}) + f(b_{i+1}) = -1+2+2 = 3 \geq 1$;
- If $1 \leq i \leq r-2$, $f(N(c_i)) = f(a_i) + f(c_{i-1}) + f(c_{i+1}) = -1+1+1 = 1 \geq 1$;
- If $i = 0$, $f(N(c_0)) = f(a_0) + f(d_{r-1}) + f(c_1) = -1+1+1 = 1 \geq 1$;
- If $i = r-1$, $f(N(c_{r-1})) = f(a_{r-1}) + f(c_{r-2}) + f(d_0) = -1+1+1 = 1 \geq 1$;
- If $1 \leq i \leq r-2$, $f(N(d_i)) = f(a_i) + f(d_{i-1}) + f(d_{i+1}) = -1+1+1 = 1 \geq 1$;
- If $i = 0$, $f(N(d_0)) = f(a_0) + f(c_{r-1}) + f(d_1) = -1+1+1 = 1 \geq 1$;
- If $i = r-1$, $f(N(d_{r-1})) = f(a_{r-1}) + f(d_{r-2}) + f(c_0) = -1+1+1 = 1 \geq 1$.

Finally, $f(V_{J_r}) = f\left(\bigcup_{i=0}^{r-1} \{a_i, b_i, c_i, d_i\}\right) = \sum_{i=0}^{r-1} (f(a_i) + f(b_i) + f(c_i) + f(d_i)) = r \cdot (-1+2+1+1) = 3r$. Therefore f is the signed total Roman domination function of flower snarks graphs with value $3r$, so $\gamma_{stR}(J_r) \leq 3r$. ◀

4. CONCLUSION

In this paper we have found and proved the exact value for the Roman and restrained Roman domination number of flower snark graphs. The upper bound of the signed total Roman domination number is given, along with the appropriate signed total Roman domination sets.

In the future work the problem of finding these Roman domination numbers for other challenging classes of graphs could be considered. Another direction of future research would be to determine other Roman domination numbers for flower snark graphs.

Acknowledgement

This research was partially supported by Serbian Ministry of Education, Science and Technological Development under the grants no. 174010 and 174033.

REFERENCES

Ahangar, H. A., Henning, M. A., Löwenstein, C., Zhao, Y., and Samodivkin, V. (2014). Signed Roman domination in graphs. *Journal of Combinatorial Optimization*, 27(2):241–255.

Al Hashim, A. (2017). *Roman domination in complementary prisms*. PhD thesis, East Tennessee State University.

Behtoei, A., Vatandoost, E., and Abad, F. A. R. (2014). Signed Roman domination number and join of graphs. *arXiv preprint arXiv:1411.7208*.

Bermudo, S. and Fernau, H. (2012). Lower bounds on the differential of a graph. *Discrete Mathematics*, 312(22):3236–3250.

Bermudo, S., Fernau, H., and Sigarreta, J. M. (2014). The differential and the Roman domination number of a graph. *Applicable Analysis and Discrete Mathematics*, pages 155–171.

Burger, A., De Villiers, A., and Van Vuuren, J. (2013). A binary programming approach towards achieving effective graph protection. In *Proc. 2013 ORSSA Annual Conf., ORSSA*, pages 19–30.

Chambers, E. W., Kinnersley, B., Prince, N., and West, D. B. (2009). Extremal problems for Roman domination. *SIAM Journal on Discrete Mathematics*, 23(3):1575–1586.

Chellali, M., Haynes, T. W., and Hedetniemi, S. T. (2016). Lower bounds on the Roman and independent Roman domination numbers. *Appl. Anal. Discrete Math*, 10:65–72.

Cockayne, E. J., Dreyer Jr, P. A., Hedetniemi, S. M., and Hedetniemi, S. T. (2004). Roman domination in graphs. *Discrete Mathematics*, 278(1-3):11–22.

Currò, V. (2014). *The Roman domination problem on grid graphs*. PhD thesis, Università di Catania.

Dreyer, P. A. (2000). *Applications and variations of domination in graphs*. PhD thesis, Rutgers University.

Favaron, O., Karami, H., Khoeilar, R., and Sheikholeslami, S. M. (2009). On the Roman domination number of a graph. *Discrete Mathematics*, 309(10):3447–3451.

Hedetniemi, S., Rubalcaba, R., Slater, P., and Walsh, M. (2013). Few compare to the great Roman empire. *Congr. Numer*, 217:129–136.

Henning, M. A. (2002). A characterization of Roman trees. *Discussiones Mathematicae Graph Theory*, 22(2):325–334.

Isaacs, R. (1975). Infinite families of nontrivial trivalent graphs which are not Tait colorable. *The American Mathematical Monthly*, 82(3):221–239.

Ivanović, M. (2016). Improved mixed integer linear programing formulations for roman domination problem. *Publications de l'Institut Mathematique*, 99(113):51–58.

Klobučar, A. and Puljić, I. (2014). Some results for Roman domination number on cardinal product of paths and cycles. *Kragujevac Journal of Mathematics*, 38(1):83–94.

Klobučar, A. and Puljić, I. (2015). Roman domination number on cardinal product of paths and cycles. *Croatian Operational Research Review*, 6(1):71–78.

Liedloff, M., Kloks, T., Liu, J., and Peng, S.-L. (2005). Roman domination over some graph classes. In *International Workshop on Graph-Theoretic Concepts in Computer Science*, pages 103–114. Springer.

Liu, C.-H. and Chang, G. J. (2012a). Roman domination on 2-connected graphs. *SIAM Journal on Discrete Mathematics*, 26(1):193–205.

Liu, C.-H. and Chang, G. J. (2012b). Upper bounds on Roman domination numbers of graphs. *Discrete Mathematics*, 312(7):1386–1391.

Liu, C.-H. and Chang, G. J. (2013). Roman domination on strongly chordal graphs. *Journal of Combinatorial Optimization*, 26(3):608–619.

Luttwak, E. (2016). *The Grand Strategy of the Roman Empire: From the First Century CE to the Third*. JHU Press.

Mobaraky, B. and Sheikholeslami, S. (2008). Bounds on Roman domination numbers of graphs. *Matematički vesnik*, 60(4):247–253.

Nascimento, S. and Sampaio, R. (2015). Roman domination in subgraphs of grids. *Electronic Notes in Discrete Mathematics*, 50:77–82.

Pushpam, R. L. and Padmapriea, S. (2015). Restrained Roman domination in graphs. *Transactions on combinatorics*, 4(1):1–17.

Rad, N. J. and Krzywkowski, M. (2015). On the restrained Roman domination in graphs. In *Colourings, Independence and Domination - CID 2015*. University of Zielona Gora. `www.krzywkowski.pl/sub1.pdf`.

Ramezani, F., Rodriguez-Bazan, E. D., and Rodriguez-Velazquez, J. A. (2016). On the Roman domination number of generalized Sierpinski graphs. *arXiv preprint arXiv:1605.06918*.

ReVelle, C. S. (1997). Can you protect the Roman Empire. *Johns Hopkins Magazine*, 49(2):40.

ReVelle, C. S. and Rosing, K. E. (2000). Defendens imperium romanum: a classical problem in military strategy. *The American Mathematical Monthly*, 107(7):585–594.

Shang, W. and Hu, X. (2007). The Roman domination problem in unit disk graphs. In *International Conference on Computational Science*, pages 305–312. Springer.

Sheikholeslami, S. and Volkmann, L. (2011). The Roman domination number of a digraph. *Acta Universitatis Apulensis*, 27:77–86.

Song, X.-x. and Wang, X.-f. (2006). Roman domination number and domination number of a tree. *Chinese Quarterly Journal of Mathematics*, 21(3):358.

Stewart, I. (1999). Defend the Roman empire! *Scientific American*, 281(6):136–138.

Volkmann, L. (2016). Signed total Roman domination in graphs. *Journal of Combinatorial Optimization*, 32(3):855–871.

Wang, H., Xu, X., Yang, Y., and Ji, C. (2011). Roman domination number of Generalized Petersen Graphs P (n, 2). *arXiv preprint arXiv:1103.2419*.

Xing, H.-M., Chen, X., and Chen, X.-G. (2006). A note on Roman domination in graphs. *Discrete mathematics*, 306(24):3338–3340.

Xueliang, F., Yuansheng, Y., and Baoqi, J. (2009). Roman domination in regular graphs. *Discrete Mathematics*, 309(6):1528–1537.

Yero, I. G., Kuziak, D., and Aguilar, A. R. (2013). Coloring, location and domination of corona graphs. *Aequationes mathematicae*, 86(1-2):1–21.

Zhiqiang Zhang, Z. S. and Xu, X. (2014). On the Roman domination numbers of generalized Petersen graphs. *Journal of Combinatorial Mathematics and Combinatorial Computing*, 89:311–320.

Zverovich, V. and Poghosyan, A. (2011). On Roman, global and restrained domination in graphs. *Graphs and Combinatorics*, 27(5):755–768.

k-METRIC ANTIDIMENSION OF WHEELS AND GRID GRAPHS

Mirjana Čangalović[1], Vera Kovačević-Vujčić[2], Jozef Kratica[3]

[1] University of Belgrade, Faculty of Organizational Sciences, canga@fon.bg.ac.rs

[2] University of Belgrade, Faculty of Organizational Sciences, verakov@fon.bg.ac.rs

[3] Mathematical Institute, Belgrade, jkratica@mi.sanu.ac.rs

Abstract: *In this paper we study the k-metric antidimension problem on two special classes of graphs: wheels W_n and grid graphs $G_{m,n}$. We prove that W_n is n-metric antidimensional and find the k-metric antidimension for each k where it exists. For $G_{m,n}$ we find the k-metric antidimension for $k = 1,2$. Additionally, we determine 4-metric antidimension in the case when m and n are both odd.*

Keywords: *k-metric antidimension, Wheel graphs, Grid graphs*

1. INTRODUCTION

The concept of metric dimension of a graph G was introduced independently by Slater (1975) and Harary and Melter (1976). It is based on the notion of resolving set R of vertices which has the property that each vertex is uniquely identified by its metric coordinates with respect to R. The minimal cardinality of resolving sets is called the metric dimension of graph G.

Some interesting applications of the metric dimension include navigation of robots in networks (Khuller et al., 1996), applications to chemistry (Johnson, 1993, 1998) and application in computer graphics (Melter and Tomescu, 1984).

There are several variations of the metric dimension concept:

- Weighted metric dimension (Epstein et al., 2015);
- Resolving dominating sets (Brigham et al., 2003);
- Local metric dimension (Okamoto et al., 2010);
- Independent resolving sets (Chartrand et al., 2003);
- Strong metric dimension (Sebő and Tannier, 2004);
- Minimal doubly resolving sets (Cáceres et al., 2007);
- k-metric dimension (Estrada-Moreno et al., 2015);
- Simultaneous metric dimension (Ramírez-Cruz et al., 2014).

Recently, Trujillo-Rasua and Yero (2016a) introduced the concepts of k-antiresolving set S and k-metric antidimension. Different vertices of $V \setminus S$ now can have the same metric coordinates with respect to S, but no vertex can be identified with probability higher than $1/k$. For a given k, the minimal cardinality of k-antiresolving sets represents the k-metric antidimension of graph G. Zhang and Gao (2017) have proved that the problem of finding the k-metric antidimension is NP-hard in general case.

The concept of k-metric antidimension has been used to define privacy measures aimed at evaluating the resistance of social graphs to active attacks. Trujillo-Rasua and Yero (2016a) define a novel privacy measure, so called (k,l)-anonymity. Mauw et al. (2016) propose the first privacy-preserving transformation method for social graphs that counteracts active attacks.

The k-metric antidimension of special classes of graphs has been studied by several authors. Trujillo-Rasua and Yero (2016a,b) consider the k-metric antidimension of paths, cycles, trees and complete bipartite graphs. They also provided efficient algorithms which can be used to decide whether a tree or a unicyclic graph is 1-metric antidimensional.

Zhang and Gao (2017) analyze the size of k-antiresolving sets in random graphs and in the case of Erdos-Renyi random graphs establish three bounds on the size.

Kratica et al. (2018) study mathematical properties of the k-antiresolving sets and the k-metric antidimension of some generalized Petersen graphs. In this paper the analysis is extended to wheels and grid graphs.

2. PROBLEM DEFINITION

Let $G = (V, E)$ be a simple connected undirected graph. The degree of a vertex of a graph is the number of edges incident to that vertex. The maximum degree of a graph G, denoted by Δ_G, is the maximum degree of its vertices.

Let us denote by $d(u, v)$ the length of the shortest path between vertices u and v. The metric representation $r(v|S)$ of vertex v with respect to an ordered set of vertices $S = \{u_1, ..., u_t\}$ is defined as $r(v|S) = (d(v, u_1), ..., d(v, u_t))$. Value $d(v, u_i)$ represents the metric coordinate of v with respect to vertex u_i, $i = 1, ..., t$. The following definitions introduce the concepts of k-antiresolving set, k-metric antidimension of graph G and the notion of k-metric antidimensional graph.

▶ Definition 1. (Trujillo-Rasua and Yero (2016a)) Set S is called a k-antiresolving set for G if k is the largest positive integer such that for every vertex $v \in V \setminus S$ there exist at least $k-1$ different vertices $v_1, ..., v_{k-1} \in V \setminus S$ with $r(v|S) = r(v_1|S) = ... = r(v_{k-1}|S)$, i.e. v and $v_1, ..., v_{k-1}$ have the same metric representation with respect to S.

▶ Definition 2. (Trujillo-Rasua and Yero (2016a)) For fixed k, the k-metric antidimension of graph G, denoted by $adim_k(G)$, is the minimum cardinality amongst all k-antiresolving sets in G. A k-antiresolving set of cardinality $adim_k(G)$ is called a k-antiresolving basis of G.

▶ Definition 3. (Trujillo-Rasua and Yero (2016a)) Graph G is k-metric antidimensional if k is the largest integer such that G contains a k-antiresolving set.

Now the k-metric antidimension problem can be formulated as follows: for a given k find the k-metric antidimension of graph G if it exists. The following two properties will be used in the proofs of theorems in Section 3 and Section 4.

▶ Property 1. (Trujillo-Rasua and Yero (2016a)) If G is a connected k-metric antidimensional graph of maximum degree Δ_G, then $1 \le k \le \Delta_G$.

Property 2 presents a simple necessary and sufficient condition for a set of vertices to be k-antiresolving. Let $S \subset V$ be a subset of vertices of G and let ρ_S be equivalence relation on G defined by

$$(\forall a, b \in V)\ (a \rho_S b \ \Leftrightarrow\ r(a|S) = r(b|S))$$

and let $S_1, ..., S_m$ be the equivalence classes of ρ_S. It is easy to see that the following property is satisfied.

▶ Property 2. (Kratica et al. (2018)) Let k be a fixed integer, $k \ge 1$. Then S is a k-antiresolving set for G if and only if $\min\limits_{1\le i\le m} |S_i| = k$.

3. WHEELS

Wheel $W_n = (V, E)$ of dimension n is a graph with $n+1$ vertices and $2n$ edges, where central vertex v_0 is connected to all vertices, while other vertices v_i, $i = 1, ...n$, are connected as in a cycle. Hence central vertex v_0 has n neighbours, while all other vertices $v_i, i = 1, ..., n$, have three neighbours.

▶ Theorem 1. For $n \ge 6$ graph W_n is n-metric antidimensional and

$$adim_k(W_n) = \begin{cases} 2,\ k = 1 \vee k = 2 \\ 1,\ k = 3 \vee k = n \end{cases}$$

Proof. Step 1. $adim_1(W_n) \le 2$

Let us consider set $S = \{v_1, v_2\}$. The equivalence classes of ρ_S are given in Table 1. More precisely, the first column of Table 1 contains set S, while in the second and the third column the equivalence classes of relation ρ_S and their cardinalities are given. In the fourth column the corresponding metric representations with respect to S are shown. Since the minimal cardinality of equivalence classes is one, according to Property 2, it follows that $S = \{v_1, v_2\}$ is 1-antiresolving set.

Step 2. $adim_1(W_n) = 2$

Suppose the contrary, that $adim_1(W_n) = 1$. Then there exists an 1-antiresolving set S of cardinality one. We have two cases:

- Case 1. Suppose that $S = \{v_0\}$. From Table 1 it is evident that there exists only one equivalence class $\{v_i | i = 1, ..., n\}$ of cardinality n.

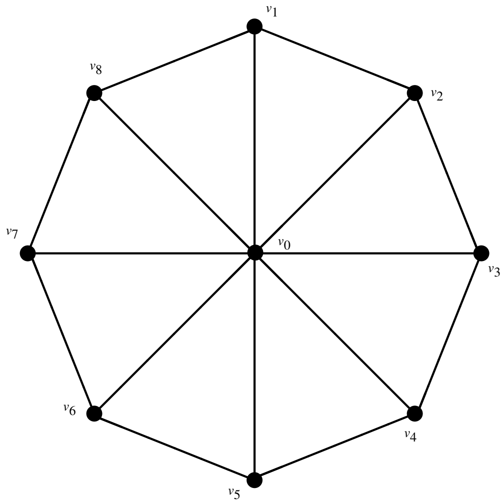


Figure 1 Graph W_8

- Case 2. Suppose that $S = \{v_i\}$, $1 \le i \le n$. Without loss of generality we can assume $i = 1$. From Table 1 it can be seen that for $n \ge 6$ the equivalence class with minimal cardinality is $\{v_0, v_2, v_n\}$.

In both cases we have a contradiction with the assumption that there exists an 1-antiresolving set S of cardinality one, so $adim_1(W_n) \ge 2$. According to Step 1, $adim_1(W_n) = 2$.

Step 3. $adim_2(W_n) = 2$

Let $S = \{v_0, v_i\}$, $1 \le i \le n$. Without loss of generality we can assume $i = 1$. Since $\{v_2, v_n\}$ is the equivalence class of minimal cardinality (see Table 1), according to Property 2 it follows that $S = \{v_0, v_1\}$ is a 2-antiresolving set. As it can be seen from Step 2, singleton sets S have equivalence classes with cardinality of at least three, so such sets can not be 2-antiresolving sets. Therefore, $S = \{v_0, v_1\}$ is a 2-antiresolving basis of W_n and $adim_2(W_n) = 2$.

Step 4. $adim_3(W_n) = 1$

Let us consider set $S = \{v_1\}$. Since the equivalence class with the minimal cardinality is $\{v_0, v_2, v_n\}$, according to Property 2 it follows that $S = \{v_1\}$ is a 3-antiresolving set. As $|S| = 1$, $S = \{v_1\}$ is a 3-antiresolving basis of W_n, so $adim_3(W_n) = 1$.

Step 5. $adim_n(W_n) = 1$

Let $S = \{v_0\}$. There is only one equivalence class $\{v_i | i = 1, ..., n\}$ of cardinality n, so it is obvious that $S = \{v_0\}$ is an n-antiresolving set of W_n. Since $|S| = 1$, it is also an n-antiresolving basis of W_n. ◀

Table 1: Equivalence classes of ρ_S on W_n

S	*Equivalence class*	*Card.*	*M. rep.*
$\{v_1, v_2\}$	$\{v_0\}$	1	(1,1)
	$\{v_n\}$	1	(1,2)
	$\{v_3\}$	1	(2,1)
	$\{v_i \| i = 4, ..., n-1\}$	$n-4$	(2,2)
$\{v_0, v_1\}$	$\{v_2, v_n\}$	2	(1,1)
	$\{v_i \| i = 3, ..., n-1\}$	$n-3$	(1,2)
$\{v_1\}$	$\{v_0, v_2, v_n\}$	3	(1)
	$\{v_i \| i = 3, ..., n-1\}$	$n-3$	(2)
$\{v_0\}$	$\{v_i \| i = 1, ..., n\}$	n	(1)

Values $adim_k(W_n)$ for $n \in \{3, 4, 5\}$ and the corresponding k-antiresolving bases are obtained by total enumeration and presented in Table 2.

▶ Theorem 2. For $l \in \{4, ..., n-1\}$ there does not exist $S \subset V$ such that S is an l-antiresolving set.

Proof. From the proof of Theorem 1 it follows that for $n \ge 6$, $S = \{v_0\}$ is an n-antiresolving set of W_n and $S = \{v_i\}$, $1 \le i \le n$, is a 3-antiresolving set of W_n. Also, $S = \{v_0, v_i\}$, $1 \le i \le n$, is a 2-antiresolving set. Let us consider all other possibilities for S.

Table 2: $adim_k(W_n)$ for $n \in \{3,4,5\}$

n	k	*Basis*	$adim_k(W_n)$
3	1	$\{v_1, v_2, v_3\}$	3
	2	$\{v_1, v_2\}$	2
	3	$\{v_1\}$	1
4	1	$\{v_1\}$	1
	2	$\{v_0, v_2, v_4\}$	3
	3	$\{v_1, v_3\}$	2
	4	$\{v_0\}$	1
5	1	$\{v_1, v_2\}$	2
	2	$\{v_1\}$	1
	3, 4	*does not exist*	*undefined*
	5	$\{v_0\}$	1

Case 1. $|S| \geq 2$, $v_0 \notin S$.

- If $S = \{v_1, ..., v_n\}$ then there is only one equivalence class $\{v_0\}$ with respect to S and hence S is an 1-antiresolving set.
- If $S \subset \{v_1, ..., v_n\}$ then there exists $v_i \in S$ such that $v_{i'} \notin S$ or $v_{i''} \notin S$, where $v_{i'}$ and $v_{i''}$ represent the previous and the next vertex in the cycle, respectively. Formally, $i' = \begin{cases} i-1, \; 2 \leq i \leq n \\ n, \; i = 1 \end{cases}$ and $i'' = \begin{cases} i+1, \; 1 \leq i \leq n-1 \\ 1, \; i = n \end{cases}$. Then v_0 and either $v_{i'}$ or $v_{i''}$, or both, are the only vertices from $V \setminus S$ which have 1 as a coordinate with respect to v_i. It follows that the equivalence class of ρ_S with minimal cardinality, has cardinality less or equal to three, i.e. S is an l-antiresolving set, where $l \leq 3$.

Case 2. $|S| \geq 3$, $v_0 \in S$.

As $v_0 \in S$ and $S \subset V$, it follows that there exists $v_i \in S$ such that $v_{i'} \notin S$ or $v_{i''} \notin S$, where $v_{i'}$ and $v_{i''}$ again represent the previous and the next vertex in the cycle, respectively. Now $v_{i'}$ or $v_{i''}$, or both, are the only vertices from $V \setminus S$ which have 1 as a coordinate with respect to v_i. Similarly as in Case 1 it follows that S is an l-antiresolving set where $l \leq 2$. ◀

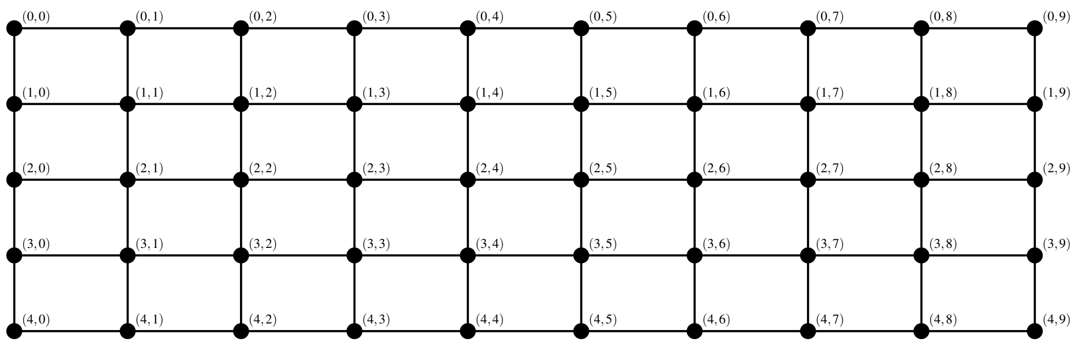


Figure 2 Graph $G_{5,10}$

4. GRID GRAPHS

Grid graph $G_{m,n} = (V,E)$ of dimension $m \cdot n,\ m,n > 1$, can be considered as a graph with set of vertices $V = \{(i,j) | 0 \le i \le m-1, 0 \le j \le n-1\}$, where vertices are organized in m rows with n vertices in each row. Two vertices are adjacent if they belong to the same row and to adjacent columns, or to the same column and to adjacent rows. Formally, vertices (i,j) and (i',j') are adjacent if $i = i'$ and $|j-j'| = 1$ or $|i-i'| = 1$ and $j = j'$, for $0 \le i,i' \le m-1$ and $0 \le j,j' \le n-1$. The interior vertices of the grid graph have 4 neighbors, the vertices along the sides have 3 neighbors and only four of them (in the corners of the rectangle) have 2 neighbors.

It is obvious that a grid graph $G_{m,n}$ can be viewed as a Cartesian product of paths P_m and P_n, i.e. $G_{m,n} \cong P_m \Box P_n$. Moreover, grid graphs of the dimensions $m \cdot n$ and $n \cdot m$ are isomorphic ($G_{m,n} \cong G_{n,m}$), so without loss of generality we can suppose that $m \le n$.

Since $G_{1,n} \cong P_n$ and the k-metric antidimension of paths is considered in (Trujillo-Rasua and Yero, 2016a), we consider work only cases when $m,n \ge 2$.

► Theorem 3. For $m,n \ge 2$ it follows:

- $adim_1(G_{m,n}) = 1$;
- $adim_2(G_{m,n}) = \begin{cases} 2, & m,n \ both \ even \\ 1, & otherwise \end{cases}$;
- $adim_4(G_{m,n}) = 1$ for m, n both odd.

Proof. Step 1. $adim_1(G_{m,n}) = 1$
Let us consider set $S = \{(0,0)\}$. The equivalence classes of ρ_S are given in Table 3 (item 1). Since the equivalence class with minimal cardinality is $\{(m-1,n-1)\}$, according to Property 2, it follows that $S = \{(0,0)\}$ is an 1-antiresolving set. As $|S| = 1$, $S = \{(0,0)\}$ is an 1-antiresolving basis of $G_{m,n}$, so $adim_1(G_{m,n}) = 1$.

Step 2. If both m and n are even, $adim_2(G_{m,n}) = 2$
Let us define set $S = \{(0,0),(m-1,n-1)\}$. The equivalence classes of ρ_S are given in Table 3 (item 2). Since the equivalence classes with minimal cardinality are $\{(0,1),(1,0)\}$ and $\{(m-2,n-1),(m-1,n-2)\}$, according to Property 2, it follows that $S = \{(0,0),(m-1,n-1)\}$ is a 2-antiresolving set. Next we will prove that there does not exists a 2-antiresolving set S of $G_{m,n}$ of cardinality one. Suppose the contrary, that there exists a 2-antiresolving set of cardinality one, $S = \{(i',j')\}$. Let us define $i'' = \begin{cases} 0, & i' \ge m/2 \\ m-1, & i' < m/2 \end{cases}$ and $j'' = \begin{cases} 0, & j' \ge n/2 \\ n-1, & j' < n/2 \end{cases}$. Vertex (i'',j'') is the unique most distant vertex from vertex (i',j'), so $\{(i'',j'')\}$ is an equivalence class of ρ_S of cardinality one, which is in contradiction with the assumption that $S = \{(i',j')\}$ is a 2-antiresolving set of $G_{m,n}$. Hence $adim_2(G_{m,n}) = 2$.

Step 3. If either m or n is odd, $adim_2(G_{m,n}) = 1$
Case 1. Both m and n are odd.
Let $p = \lfloor m/2 \rfloor$ and $q = \lfloor n/2 \rfloor$, i.e. $m = 2p+1$ and $n = 2q+1$. Let us define set $S = \{(p-1,q)\}$. Since both m and n are odd and $m,n \ge 2$ it follows that $m,n \ge 3$. As $m = 2p+1 \ge 3$ and $n = 2q+1 \ge 3$ then $p,q \ge 1$. Equivalence classes of ρ_S are given in Table 3 (item 3). For each r, $2 \le r \le p+q$ there exists vertex (i',j') such that $j' < q$ and $d((p-1,q),(i',j')) = r$. Due to symmetry of $G_{m,n}$, $d((p-1,q),(i',2q-j')) = r$, and hence each equivalence class ρ_S has cardinality at least two. Since equivalence class $\{(m-1,0),(m-1,n-1)\}$ has cardinality two it follows that $S = \{(p-1,q)\}$ is a 2-antiresolving set. As $|S| = 1$, $S = \{(p-1,q)\}$ is a 2-antiresolving basis of $G_{m,n}$, so $adim_2(G_{m,n}) = 1$.
Case 2. m is even and n is odd.
Let $p = \lfloor m/2 \rfloor$ and $q = \lfloor n/2 \rfloor$, i.e. $m = 2p$ and $n = 2q+1$. Let us define set $S = \{(p,q)\}$. Since n is odd and $n \ge 2$ it follows that $n \ge 3$. As $n = 2q+1 \ge 3$ then $q \ge 1$. The equivalence classes of ρ_S are given in Table 3 (item 4). Similarly as in Case 1, for each r, $2 \le r \le p+q-1$ there exists vertex (i',j') such that $j' < q$ and $d((p,q),(i',j')) = r$. Due to symmetry of $G_{m,n}$, $d((p,q),(i',2q-j')) = r$, and hence each equivalence class ρ_S has cardinality at least two. Since equivalence class $\{(0,0),(0,n-1)\}$ has cardinality two it follows that $S = \{(p,q)\}$ is a 2-antiresolving set. As $|S| = 1$, $S = \{(p,q)\}$ is a 2-antiresolving basis of $G_{m,n}$, so $adim_2(G_{m,n}) = 1$.
Case 3. m is odd and n is even.
Having in mind the fact that $G_{m,n} \cong G_{n,m}$, this case is reduced to Case 2.

Step 4. If both m and n are odd, $adim_4(G_{m,n}) = 1$
Let $p = \lfloor m/2 \rfloor$ and $q = \lfloor n/2 \rfloor$, i.e. $m = 2p+1$ and $n = 2q+1$. Let us define set $S = \{(p,q)\}$. Since both m and n are odd and $m,n \geq 2$ it follows that $m,n \geq 3$. As $m = 2p+1 \geq 3$ and $n = 2q+1 \geq 3$ then $p,q \geq 1$. The equivalence classes of ρ_S are given in Table 3 (item 5). For each r, $2 \leq r \leq p+q-1$ there exists vertex (i', j') such that $i' < p$, $j' < q$ and $d((p,q),(i',j')) = r$. Due to symmetries of $G_{m,n}$ we have $d((p,q),(i',2q-j')) = r$, $d((p,q),(2p-i',j')) = r$ and $d((p,q),(2p-i',2q-j')) = r$ and hence each equivalence class of ρ_S has cardinality at least four. Since equivalence class $\{(0,0),(0,n-1),(m-1,0),(m-1,n-1)\}$ has cardinality four, it follows that $S = \{(p,q)\}$ is a 4-antiresolving set. As $|S| = 1$, $S = \{(p,q)\}$ is a 4-antiresolving basis of $G_{m,n}$, so $adim_4(G_{m,n}) = 1$.

◀

5. CONCLUSION

In this article the k-metric antidimension problem is considered on wheels and grid graphs. Exact values of the k-metric antidimension of wheels W_n are obtained for $k \in \{1,2,3,n\}$ and it is proved that the k-metric antidimension does not exists for $4 \leq k \leq n-1$. In the case of grid graphs $G_{m,n}$ the exact values of the k-metric antidimension are obtained for $k \in \{1,2\}$ for arbitrary m, n and for $k = 4$ when m and n are both odd.

In future reseach it would be interesting to identify the cases when k-metric antidimension of $G_{m,n}$ does not exist. Also the k-metric antidimension problem could be considered on some other challenging classes of graphs.

Acknowledgement

This research was partially supported by Serbian Ministry of Education, Science and Technological Development under the grants no. 174010 and 174033.

REFERENCES

Brigham, R. C., Chartrand, G., Dutton, R. D., and Zhang, P. (2003). Resolving domination in graphs. *Mathematica Bohemica*, 128(1):25–36.

Cáceres, J., Hernando, C., Mora, M., Pelayo, I. M., Puertas, M. L., Seara, C., and Wood, D. R. (2007). On the metric dimension of cartesian products of graphs. *SIAM Journal on Discrete Mathematics*, 21(2):423–441.

Chartrand, G., Saenpholphat, V., and Zhang, P. (2003). The independent resolving number of a graph. *Mathematica Bohemica*, 128(4):379–393.

Epstein, L., Levin, A., and Woeginger, G. J. (2015). The (weighted) metric dimension of graphs: hard and easy cases. *Algorithmica*, 72(4):1130–1171.

Estrada-Moreno, A., Rodrıguez-Velázquez, J. A., and Yero, I. G. (2015). The k-metric dimension of a graph. *Appl. Math*, 9(6):2829–2840.

Harary, F. and Melter, R. (1976). On the metric dimension of a graph. *Ars Combin*, 2(191-195):1.

Johnson, M. (1993). Structure-activity maps for visualizing the graph variables arising in drug design. *Journal of Biopharmaceutical Statistics*, 3(2):203–236.

Johnson, M. (1998). Browsable structure-activity datasets. *Advances in Molecular Similarity*, pages 153–170.

Khuller, S., Raghavachari, B., and Rosenfeld, A. (1996). Landmarks in graphs. *Discrete Applied Mathematics*, 70(3):217–229.

Kratica, J., Kovačević-Vujčić, V., and Čangalović, M. (2018). k-metric antidimension of some generalized Petersen graphs. *submitted*.

Mauw, S., Trujillo-Rasua, R., and Xuan, B. (2016). Counteracting active attacks in social network graphs. In *IFIP Annual Conference on Data and Applications Security and Privacy*, pages 233–248. Springer.

Melter, R. A. and Tomescu, I. (1984). Metric bases in digital geometry. *Computer Vision, Graphics, and Image Processing*, 25(1):113–121.

Okamoto, F., Phinezy, B., and Zhang, P. (2010). The local metric dimension of a graph. *Mathematica Bohemica*, 135(3):239–255.

Ramírez-Cruz, Y., Oellermann, O. R., and Rodríguez-Velázquez, J. A. (2014). Simultaneous resolvability in graph families. *Electronic Notes in Discrete Mathematics*, 46:241–248.

Sebő, A. and Tannier, E. (2004). On metric generators of graphs. *Mathematics of Operations Research*, 29(2):383–393.

Slater, P. J. (1975). Leaves of trees. *Congr. Numer*, 14(549-559):37.

Trujillo-Rasua, R. and Yero, I. (2016a). k-metric antidimension: A privacy measure for social graphs. *Information Sciences*, 328:403–417.

Trujillo-Rasua, R. and Yero, I. G. (2016b). Characterizing 1-metric antidimensional trees and unicyclic graphs. *The Computer Journal*, 59(8):1264–1273.

Zhang, C. and Gao, Y. (2017). On the complexity of k-metric antidimension problem and the size of k-antiresolving sets in random graphs. In *International Computing and Combinatorics Conference*, pages 555–567. Springer.

Table 3: Equivalence classes of ρ_S on $G_{m,n}$

S	*Equivalence class*	*Card.*	*M. rep.*	*Condidtion*
$\{(0,0)\}$	$\{(0,1),(1,0)\}$	2	(1)	
	$\{(i,j)\|i+j=r, i=0,...,r\}$	$r+1$	(r)	$r=2,...,m-1$
	$\{(i,j)\|i+j=r, i=0,...,m-1\}$	m	(r)	$r=m,...,n-1$
	$\{(i,j)\|i+j=r, i=r-n+1,...,m-1\}$	$m+n-r-1$	(r)	$r=n,...,m+n-3$
	$\{(m-1,n-1)\}$	1	$(m+n-2)$	
$\{(0,0),(m-1,n-1)\}$	$\{(0,1),(1,0)\}$	2	$(1,m+n-3)$	
	$\{(i,j)\|i+j=r, i=0,...,r\}$	$r+1$	$(r,m+n-2-r)$	$r=2,...,m-1$
	$\{(i,j)\|i+j=r, i=0,...,m-1\}$	m	$(r,m+n-2-r)$	$r=m,...,n-1$
	$\{(i,j)\|i+j=r, i=r-n+1,...,m-1\}$	$m+n-r-1$	$(r,m+n-2-r)$	$r=n,...,m+n-3$
	$\{(m-2,n-1),(m-1,n-2)\}$	2	$(m+n-3,1)$	
$\{(p-1,q)\}$	$\{(p-2,q),(p,q),(p-1,q-1),(p-1,q+1)\}$	4	(1)	$m=2p+1$ and $n=2q+1$
	$\{(p-i-1,q-j),(p-i-1,q+j),(p+i-1,q-j),(p+i-1,q+j)\|i+j=r, 0\leq i\leq m-1, 0\leq j\leq n-1\}$	≥ 2	(r)	$r=2,...,p+q$
	$\{(m-1,0),(m-1,n-1)\}$	2	(p+q+1)	
$\{(p,q)\}$	$\{(p-1,q),(p+1,q),(p,q-1),(p,q+1)\}$	4	(1)	$m=2p$ and $n=2q+1$
	$\{(p-i,q-j),(p-i,q+j),(p+i,q-j),(p+i,q+j)\|i+j=r, 0\leq i\leq m-1, 0\leq j\leq n-1\}$	≥ 2	(r)	$r=2,...,p+q-1$
	$\{(0,0),(0,n-1)\}$	2	(p+q)	
$\{(p,q)\}$	$\{(p-1,q),(p+1,q),(p,q-1),(p,q+1)\}$	4	(1)	$m=2p+1$ and $n=2q+1$
	$\{(p-i,q-j),(p-i,q+j),(p+i,q-j),(p+i,q+j)\|i+j=r, 0\leq i\leq m-1, 0\leq j\leq n-1\}$	≥ 4	(r)	$r=2,...,p+q-1$
	$\{(0,0),(0,n-1),(m-1,0),(m-1,n-1)\}$	4	(p+q)	

BALCOR 2018
XIII Balkan Conference on Operational Research

A NEW LINEAR-TIME ALGORITHM FOR COMPUTING THE WEAK ROMAN DOMINATION NUMBER OF A BLOCK GRAPH

MARIJA IVANOVIĆ[1], DRAGAN UROŠEVIĆ[2]

[1] University of Belgrade, Faculty of Mathemaics, marijai@math.rs
[2] Mathematical Institute, Belgrade, draganu@mi.sanu.ac.rs

Abstract: *In this paper we show that the known linear-time algorithm for solving the weak Roman domination problem on a block graph, from the literature, does not always find a weak Roman Domination function (WRDF) of minimal total weight. Furthermore, we present our newly developed linear-time algorithm that finds a WRDF of minimal total weight for the block graph.*

Keywords: *weak Roman domination number, block graph, linear time algorithm*

1. INTRODUCTION

The Roman domination problem starts with the assumption that the Roman Empire can be represented as a graph such that every vertex represents a province. Two vertices are adjacent if the corresponding provinces are neighbors, or there is direct connection between them, allowing fast traveling from one province to another. Assuming that a province is safe from the attack if at least one legion is stationed in it and that the unsafe province can be defended if it has a neighbor with two stationed legions, Stewart (1999) initiated a new variant of the domination problem named *Roman domination problem*. Since the Roman domination problem was defined by Stewart and ReVelle and Rosing (Stewart (1999); ReVelle and Rosing (2000)), many articles have been published (i.e. Dreyer (2000); Henning (2002); Cockayne et al. (2004); Chambers et al. (2009); Liu and Chang (2013); Ahangar et al. (2014); Beeler et al. (2016)). In this paper we are focused on one of its variant, named *weak Roman domination problem*. The weak Roman domination problem, introduced by Henning and Hedetniemi (2003), can be described as follows.

Let $G = (V,E)$ be a graph, $f : V \to \{0,1,2\}$ a function, and let $f(u)$ denote the weight of a vertex $u \in V$. A vertex u with $f(u) = 0$ is undefended if it is not adjacent to the neighbor with positive weight. The function f is called a *weak Roman dominating function* (WRDF) if every vertex u with $f(u) = 0$ is adjacent to a vertex v with $f(v) > 0$ such that the function $f' : V \to \{0,1,2\}$, defined as $f'(u) = 1$, $f'(v) = f(v) - 1$ and $f'(w) = f(w)$ when $w \in V \setminus \{u, v\}$, has no undefended vertices. Assuming that $f(u)$ is equal to a number of legions assigned to the province represented by the vertex u, the number of legions assigned to graph G, with respect to the function f, is called *total function weight* and is calculated as $w(f) = \sum_{u \in V} f(u)$. The *weak Roman domination number*, denoted by $\gamma_r(G)$, is equal to the minimal total weight of all possible WRDF that can be defined for that graph, G ($\gamma_r(G) = \min\{w(f) | f \text{ is a WRDF for } G\}$). A WRDF for graph G of weight γ_r is called γ_r-function. The problem of finding γ_r-function for the graph G is called the *weak Roman domination problem* (WRDP).

Although several structural results on the WRDP are known (see Henning and Hedetniemi (2003); Cockayne et al. (2004); Mai and Pushpam (2011); Chellali et al. (2014)), only few algorithmic results exist. For instance, integer linear programming (ILP) formulation, which corresponds to the Henning and Hedetniemi (2003) definition of the WRDP, was given in Ivanović (2017). Proof that the WRDP is NP-hard even when restricted to bipartite or chordal graphs is given in Henning and Hedetniemi (2003). Since the trivial enumeration algorithm for solving this problem runs in $O^*(3^n)$ in polynomial space, Chapelle et al. (2017) proposed algorithm for solving WRDP in $O^*(2^n)$ time needed exponential space, and $O^*(2.2279^n)$ algorithm using polynomial space (the notation $O^*(f(n))$ suppresses polynomial factors). In the same paper, it was shown that the WRDP can be solved in a linear-time on interval graphs. Linear-time algorithm for finding a WRDF for a block graph is proposed by Liu et al. (2010) but, as it will be shown in this paper, the proposed algorithm finds WRDF which is not always γ_r-function. Inspired by their algorithm, we developed an algorithm for solving the WRDP on a block graph which also runs in a linear-time.

Our paper is organized as follows. In Section 2 we give definitions and the algorithm for solving the WRDP on a block graph (known from the literature). Also, we give an example to illustrate that this algorithm does not find any γ_r-function for some block graphs. In Section 3, we present our newly developed algorithm for solving the WRDP on a block graph. Conclusion and References are given in the final two sections.

2. PRELIMINARIES AND NOTATIONS

Let $G = (V,E)$ be a graph with a vertex set V and a set of edges E. The following graph terms are taken from Harray (1969).

- A *subgraph* of G is a graph having all its vertices and edges in G. For any set S of vertices of G, a *induced subgraph* is the maximal subgraph of G with vertex set S (i.e. $H = (S,E')$, $S \subset V$ and for all vertices $u,v \in S$, $e = (u,v) \in E'$ iff $e \in E$).
- A *clique* is a subset of vertices of an undirected graph such that every two distinct vertices in that subset are adjacent, i.e. clique is an induced subgraph, which is complete.
- A graph G is *connected* if there is a path between every pair of its vertices.
  A graph G is *disconnected* if in G there are two vertices such that they are not endpoints of any path in G.
- A maximal connected subset $H = (V_H,E_H)$ of a graph G is called a *component of a graph G* if:
  1) for every two vertices $v_1,v_2 \in V_H$ $(v_1,v_2) \in E_H$ iff $(v_1,v_2) \in E$ and
  2) there is no path whose endpoints are in V_H and $V \setminus V_H$.
- A *cut-vertex* (articulation point) of a graph is the one whose removal increases the number of components.
- A *biconnected graph* is a connected and *non-separable graph*, meaning that the graph will remain connected after the removal of any vertex.
- A *biconnected component* (also known as a *block* or 2-connected component) is a maximal non-separable subgraph.
- If G is non-separable, then G itself is often called a *block*.
- A *bipartite graph* (or bi-graph) is a graph whose vertices can be divided into two disjoint and independent sets, U and V, such that every edge connects one vertex in U with one vertex in V.
- A tree T is a connected graph with no cycles.
- An undirected graph G is called an *intersection graph* if it is formed from a family of sets S_i, $i = 0,1,2,\ldots$ by creating one vertex v_i for each set S_i, and connecting two vertices v_i and v_j by an edge whenever the corresponding two sets have a nonempty intersection, that is $E(G) = \{\{v_i,v_j\}|S_i \cap S_j \neq \emptyset\}$.

For a rooted tree, we will use the same definition as it was used in Liu et al. (2010):

- The distance between two vertices u and v, denoted as $d(u,v)$, will be calculated as the minimum number of edges that we pass when walking from u to v; the distance between two adjacent vertices is equal to 1.
- A tree T is a rooted tree if it has a vertex r, called the *root* of the tree, vertices with degree 1, called a *leaves*, and non-leaf vertices, called *internal-nodes*.
- For two adjacent vertices u and v of the rooted tree T, vertex u will be called *parent* of v if $d(r,u) < d(r,v)$ with respect to the given root r. Vertex v will be called a *child of the vertex u*.
- Any vertex can have only one parent but many children.

Now, a *block graph* or a clique tree is a type of an undirected graph in which every biconnected component is a clique. Block graphs, derived from graph G, are usually denoted by $B(G)$, and can be defined as follows: blocks of G correspond to the vertices of $B(G)$ such that two vertices in $B(G)$ are adjacent whenever the corresponding blocks contain a common cut-vertex in G. Block graphs may be characterized as intersection graphs of blocks of arbitrary undirected graphs. For more details see Harary (1963).

For a connected block graph G, a *block-cut-vertex tree* (cut-tree for short) of G, denoted by $bc(G)$, is defined such that each node in the tree represents either a biconnected component or a cut-vertex of G, and the node that represents a cut-vertex is connected to all nodes representing any biconnected component which contains that cut-vertex.

A block graph and its cut-tree are illustrated in Figures 1 and 2.

In Figure 2, leaves B_1 and B_2 represent two component blocks, and vertices C_j, $j = 1,\ldots,5$ represent one component block (i.e. C_j is a block which contains only a cut-vertex of a graph G), while inner vertices B_i, $i = 3,\ldots,6$, represent empty blocks (since the corresponding cut-vertices of the graph G are adjacent and have an empty intersection, the intersection of the corresponding C_j vertices of the tree T_G will be empty sets, i.e. empty blocks).

In Liu et al. (2010), a function $f' : V \to \{0^-,0^+,1,2\}$ was designed such that the sign 0^- were used when a vertex was not dominated, and 0^+ if a vertex had been dominated at that moment. They also expected that only cut vertices could be of positive weight. Starting from a cut-tree T_G of a block graph G, they assigned 0^- to every leaf. Then, they computed weights of the inner vertices as follows. Let u be the current vertex.

If u is a cut vertex,

i. $f(u) = 2$ if there are more than two undefended vertices (vertices whose weights are equal to zero and whose examined neighbors are also with weights equal to zero) in the sets of children and grandchildren of u.

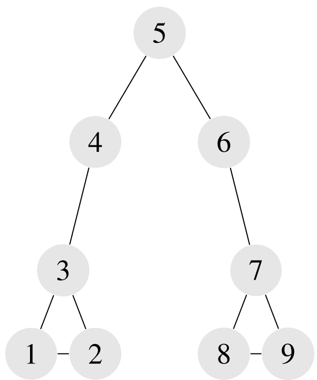
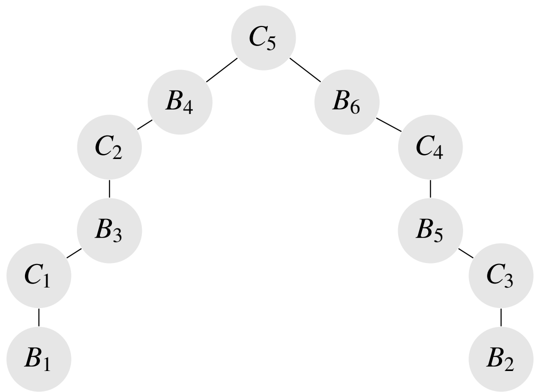


Figure 1 Block graph G



$B_1 = \{1,2\}$
$B_2 = \{8,9\}$
$B_3 = \emptyset$
$B_4 = \emptyset$
$B_5 = \emptyset$
$B_6 = \emptyset$
$C_1 = \{3\}$
$C_2 = \{4\}$
$C_3 = \{7\}$
$C_4 = \{6\}$
$C_5 = \{5\}$

Figure 2 Cut-tree of a block graph G, T_G

ii. $f(u) = 1$ if there is only one undefended vertex in the sets of children and grandchildren of u.
iii. $f(u) = 0^-$ if all children and grandchildren of u are either 0^+ or 1.
iv. $f(u) = 0^+$ if in the sets of children and grandchildren of u, there is a vertex v with $f(v) = 2$.

else

i. $f(u) = 0^+$ if the number of nodes in u is 0.
ii. $f(u) = 0^+$ if one children of u has value 2.
iii. $f(u) = 0^-$ otherwise.

Although Liu et al. (2010) claim that their algorithm finds a weak Roman domination function f of minimal weight, the found WRDF is not always γ_r-function. According to their algorithm, the minimal total weight of the cut-tree presented in Figure 2 is equal to 4: weight 2 is assigned to C_5, weight 1 to C_1 and C_3, and 0 to all other vertices. But, the optimal solution value of the illustrated cut-tree is equal to 3: weight 1 is assigned to C_1, C_3, and C_5, and weight equal to 0 to all other vertices. Indeed, if we apply the found solution to the corresponding block graph, we will notice that with weight of vertices 3, 5 and 7, which is equal to 1, and weight of all other vertices, which is equal to 0, all vertices will be considered as the currently defended and defended in case of one attack. More precisely, with the presented legion arrangement, in case of one attack, vertices 1 and 2 will be defended by the legion stationed at vertex 3; vertices 4 and 6 will be defended by the legion stationed at vertex 5; vertices 8 and 9 will be defended by the legion stationed at vertex 7, and all other vertices will be defended by their own stationed legion. So, all listed moves will not affect safety of any vertex.

3. LINEAR-TIME ALGORITHM FOR SOLVING THE WEAK ROMAN DOMINATION PROBLEM ON A BLOCK GRAPH

Based on the above considerations, we assume that G is a connected block graph. Its cut tree $T_G = (V_{T_G}, E_{T_G})$, which can be formed by depth first search in $O(|V| + |E|)$ time, is used as the input of our algorithm.

For a WRDF f let $f(i) \in \{0,1,2\}$ represents a number of legions assigned to a vertex i (let $f(i)$ be a weight of the vertex i). The output of the algorithm is γ_r-function, both for the cut-tree T_G and the block graph G, i.e. output is a function f, which assigns a value $f(i)$ to every vertex of the formed cut-tree T_G (i.e. for every vertex $i \in V_{T_G}$ that represents some cut-vertex of G, the corresponding cut-vertex in G will be of the same weight, while to all other vertices in G, which are not cut-vertices, the assigned weight is equal to 0) and is of minimal total weight for that tree.

Weight will be assigned to vertices of the cut-tree T_G post-order traversal (complexity of post-order traversal is $O(|V_{T_G}|)$), i.e. weight will be assigned first to leaves and last to the root of the tree; the weight can be assigned to a parent vertex only if it is assigned to all of its children vertices.

Since the WRDP is a dynamical problem, we say that a legion could be sent from one vertex (whose weight is positive) to its attacked neighbor vertex (whose weight is equal to zero); we decrease weight of the first vertex by one and increase weight of the second vertex by one. So, a vertex i can be considered as defended currently (i.e. it has positive weight or a neighbor j whose weight is positive), but it could be considered as undefended if its weight is equal to zero and the only neighbor j, whose weight is equal to one, sends a legion to defend neighbor k (i,k are neighbors to j, $k \neq i$), whose weight is also equal to zero. Therefore, we mark a vertex i as "defended by the stationed legion", "defended by the neighbors' legion", or "undefended", and we consider variants of these marks in accordance with the marks given for children and the parent. In other words, the mark of every vertex i can be determined by the flowchart-a shown in Figure 3.

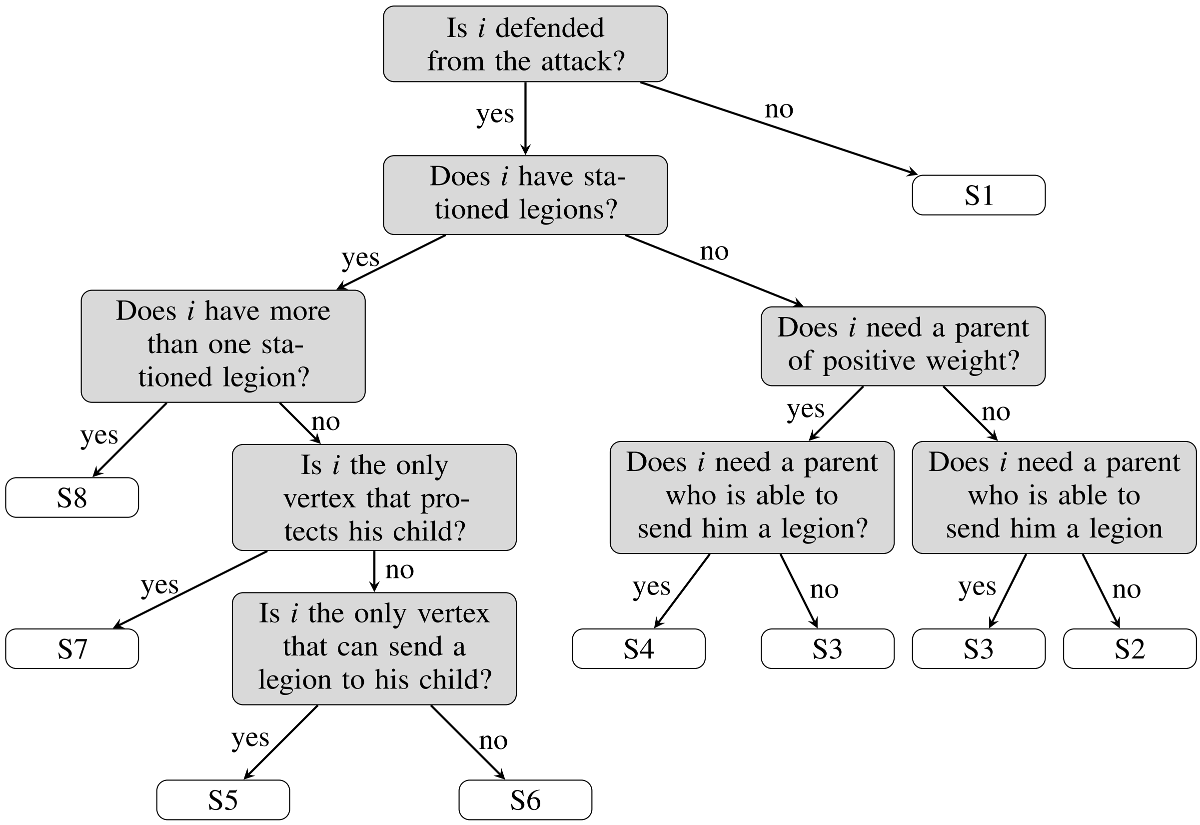


Figure 3 Flowchart of possible vertex marks

From the above, there are 8 different marks for every vertex i which can be interpreted as follows:

- S1 means that a vertex i has no child or his children don't have stationed legions, i.e. vertex i must have the parent who can protect him and send a legion in case of the attack; the parent can not send a legion to any other child or to the parent unless he has two stationed legions;
- S2 means that a vertex i has a child of positive weight j who can send a legion in case that it is attacked and, in case that a child k of j is attacked, j is not obligated to send a legion to k;
- S3 means that a vertex i has a child of positive weight j who can send a legion in case that it is attacked, and who also protects his own child k to whom it is obligated to send a legion when it is attacked. Therefore, a vertex i needs the parent with stationed legion to protect him when j sends his legion to the attacked child k. Also, the parent is not obligated to send him a legion in case when it is attacked;
- S4 means that a vertex i has a child of positive weight j and that j is the only neighbor with positive weight to his own child k, to whom he is obligated to send a legion when it is attacked. Therefore, a vertex i needs the parent who will protect him and is obligated to send a legion in case the child is attacked;
- S5 means that a vertex i has a stationed legion and does not have any unprotected child and is not obligated to send a legion to the children. Vertex i can send a legion to a parent in case that the parent is attacked;
- S6 means that a vertex i has one stationed legion and an unprotected child to whom it is not obligated to send a legion in case of the attack. Vertex i can send a legion to parent in case that parent is attacked;
- S7 means that a vertex i has one stationed legion and an unprotected child to whom it is obligated to send the legion in case that the child is attacked. Vertex i can not send the legion to its parent in case that the parent is attacked;
- S8 means that a vertex i has two stationed legions and, if it has an unprotected child, it can send one legion to attacked child, also it can send a legion to the parent in case that the parent is without station legions and needs a protection.

Our algorithm finds γ_r-function for the formed cut-tree T_G and the found solution is applied to the original block graph. Let $F(i)$ be equal to a number of legions stationed at vertex i, i.e. let $F(i)$ represents a weight of the vertex i. Since the weights of all vertices of the cut-tree T_G will be assigned post-order transversal, let i_1 represents the first and i_n, $n = |V_{T_G}|$ the last vertex to whom we assign their weights. Weight of a vertex i_m, $m = 1, \dots, n$, will be calculated as follows:

- If i_m represents a leaf, i_m will be marked with S1 and $F(i_m) = 0$.

- If i_m represents an empty block, it will be marked with S2 and $F(i_m) = 0$.
- Let i_m represents a nonempty block or a cut-vertex. If i_m represents a nonempty block, its mark will be determined in accordance with the marks of its children; in case that i_m represents a cut-vertex, its mark will be in accordance with the marks of its children and, whenever a child represents an empty-blocks, in accordance with the marks of the children of that child, too. For a vertex i_m, a number of the considered vertices with mark S1 will be denoted as num_1, a number of the considered vertices with mark S2 will be denoted as num_2, and so on (num_k represents a number of vertices with mark Sk of the considered vertices for a vertex i_m, $k = 1,\dots,8$).
  Depending on values num_k, $k = 1,\dots,8$, mark of the vertex i_m will be determined in the following way:

- If i_m is a non-empty block then, whenever conditions $num_2 + num_3 + num_4 = 0$ and $num_5 + num_6 + num_7 + num_8 > 0$ hold, it will be marked as
  1. S3 when $num_7 > 0$ and $num_5 + num_6 + num_8 = 0$,
  2. S2 when $num_5 + num_6 + num_8 > 0$,

  otherwise, it will be marked as S1. For all cases $F(i_m) = 0$.
- If i_m is a cut-vertex and
  1. one of the following two conditions hold:
     1) $num_1 > 0$ and $num_4 > 0$,
     2) $num_1 \geq 2$,

     it will be marked as S8, $F(i_m) = 2$.
  2. otherwise, if
     1) conditions $num_1 = 1$ and $num_4 = 0$ hold, it will be marked as S7.
     2) conditions $num_1 = 0$ and $num_4 > 0$ hold, it will be marked as S6.
     3) conditions $num_1 = 0$, $num_4 = 0$ and $num_3 > 0$ hold, it will be marked as S5.
     4) condition $num_1 + num_3 + num_4 = 0$ together with one of the following two conditions hold:
        i) $num_5 + num_8 > 0$,
        ii) $num_6 + num7 > 1$ i $num_6 > 0$,

     it will be marked as S2.
     5) conditions $num_1 + num_3 + num_4 + num_5 + num_6 + num_8 = 0$ and $num_7 > 0$ hold, it will be marked as S4.
     6) conditions $num_1 + num_3 + num_4 + num_5 + num_7 + num_8 = 0$ and $num_6 = 1$ hold, it will be marked as S3.
     7) conditions $num_1 + num_3 + num_4 + num_5 + num_6 + num_7 + num_8 = 0$ hold, it will be marked as S1.

     In cases 2.1)-2.3) $F(i_m) = 1$, while in cases 2.4)-2.7) $F(i_m) = 0$.

Since $F(i_m) \in \{0,1,2\}$ for every vertex i_m of the cut-tree T_G, and weights are computed so that every vertex has the positive weight, or has a neighbor whose weight is positive, with addition that, in case of an attack, it is possible to send a legion from one vertex with positive weight to the attacked neighbor without violating the safety of any vertex, constructed function F is the WRDF for the cut-tree T_G. Now, we will show that F is also the WRDF of minimal total weight.

Suppose the opposite, let F' be the WRDF for the cut-tree T_G of smaller total weight than the total weight of F. It is enough to show that total weight of F' is by one smaller than total weight of F. Therefore, let $F'(i) = F(i) - 1$ for some vertex i and $F'(j) = F(j)$ for all other vertices of the cut-tree T_G. Since $F'(i) \in \{0,1,2\}$, only two cases are possible for the vertex i: 1) $F(i) = 2$, and 2) $F(i) = 1$.

Let us first consider case 1). By the construction, $F(i) = 2$ holds if vertex i has at least two undefended children to whom it is the only adjacent vertex with positive weight and the only neighbor who can send a legion in case that someone attacks them. By the assumption, $F'(i) = F(i) - 1 = 1$ and $F'(j) = F(j)$ for all other vertices of the cut-tree T_G. Now, if a child of i, to whom i is obligated to send a legion in case of an attack, is attacked, i must send a legion in order to defend it. At this movement i will be left out of any stationed legion (will have weight equal to 0), and therefore other children, which were protected by the legion stationed at i, will become undefended, meaning that function F', constructed by the given assumption, is not a WRDF for the cut-tree T_G. So, this case is not possible.

Let us now consider case 2). By the construction of the function F, it follows that $F(i) = 1$ when

i) i has only one undefended child to whom it is the only neighbor with positive weight and to whom it is obligated to send a legion in case of the attack, and does not have any other undefended child or a grandchild (if child is an empty set) and is not obligated to send a legion to any of them.

ii) i has only one child (a grandchild if child is an empty set) to whom it is obligated to send a legion in case that the child is attacked and has no undefended children or grandchildren (if child is an empty set),

iii) i has no undefended children (nor grandchildren if child is an empty set), but has at least one child (grandchild if child is empty set) who needs the parent of positive weight.

Now, by the assumption, $F'(i) = F(i) - 1 = 0$ and $F'(j) = F(j)$ for all other vertices. In case i) i must be able to protect its unprotected child and must be able to send him a legion. Since $F'(i) = 0$, it follows that F', constructed by the assumption, is not a WRDF for the cut-tree T_G, the given child of i is left to be unprotected. In case ii) i must be able to send a legion to its child and with $F'(i) = 0$ we can conclude again that F', constructed by the given assumption, is not a WRDF for the cut-tree T_G. In case iii) weight of i is equal to zero, which is contrary to a child' needs, meaning again that F' can not be a WRDF for the cut-tree T_G.

So, since it not possible to lower the weight by one of any vertex whose weight is positive, it follows that total weight can not be lowered as well. Therefore, constructed WRDF F is with minimal total weight, implying that γ_r for the given T_G can be calculated as $\sum_{i \in V_{T_G}} F(i)$.

Now, since the weights of all vertices of the cut-tree T_G are known, constructed solution can be applied to the original block graph as follows: For every cut-vertex of the cut-tree T_G, the corresponding cut-vertex of the block graph will have the same weight. All other vertices will have weight equal to zero. Therefore, γ_r for the block graph can be also calculated as $\sum_{i \in V_{T_G}} F(i)$.

As it was earlier mentioned, the presented algorithm runs in a linear-time: transformation of the block graph into a cut-tree is in $O(|V| + |E|)$, vertices of the cut-tree are sorted in accordance with postfix index in $O(|V_{T_G}|)$, and the algorithm runs through every instance of the cut-tree T_G once, i.e. total minimal weight of that cut-tree is found in $O(|V_{T_G}|)$.

3.1. COMPUTATIONAL RESULTS

Results presented in Liu et al. (2010) are theoretical and without a source code or an executable file which can be used for testings. Therefore, in order to compare their algorithm with ours, we have programmed their algorithm and our newly developed algorithm in MATLAB2016b, and tested on Intel(R) Core(TM) i7-4700MQ CPU @ 2.40GHz, with 8GB RAM under Windows 10. Testings were performed on the set of instances that can be downloaded from http://pallini.di.uniroma1.it/Graphs.html. Results of testings are presented in Table 1: Name, number of vertices and number of edges for every instance are given in the first three columns. The optimal solution value, calculated using optimization solver CPLEX and ILP formulation provided by Ivanović (2017), are given in the fourth column. The following two columns (*val* and *t*) contain value and execution time obtained by the algorithm developed by Liu et al. (2010). For every tested instance, the last two columns contain value (*val*) and execution time (*t*) obtained by the algorithm presented in this paper.

Table 1: Computational results

Instance				Alg. developed by Liu et al. (2010)		Our algorithm	
Name	|V|	|E|	*opt*	*val*	*t*	*val*	*t*
rantree-10	10	9	5	5	0.1037	5	0.0103
rantree-20	20	19	10	10	0.1511	10	0.0041
rantree-50	50	49	25	25	0.3022	25	0.0062
rantree-100	100	99	48	49	0.4254	48	0.0084
rantree-200	200	199	103	105	0.5338	103	0.0132
rantree-300	300	299	145	153	0.7563	145	0.0184
rantree-400	400	399	201	208	1.0028	201	0.0255
rantree-500	500	499	249	255	1.2237	249	0.0296
rantree-600	600	599	299	313	1.461	299	0.0381
rantree-700	700	699	346	360	1.7541	346	0.0401
rantree-800	800	799	391	403	2.5813	391	0.0469
rantree-900	900	899	452	463	2.1692	452	0.0541
rantree-1000	1000	999	490	510	2.4234	490	0.0590

Once more, as it can be seen from Table 1: our algorithm finds solution value equal to the optimal value (from column *opt*) for every tested instance, while the algorithm developed by Liu et al. (2010) finds solution value equal to the optimal value (from column *opt*) only for the first three tested instances and greater solution value for all other tested instances.

4. CONCLUSION

The algorithm for solving the weak Roman domination problem on a block graph, which is known from the literature, is considered in this paper. It is shown that the considered algorithm does not always compute the optimal solution value on a block graph and, therefore, a new linear-time algorithm for solving the weak Roman domination problem on a block graph is presented. Since the WRDP is NP-hard problem, we believe that our algorithm, with some modifications, can also be used for solving the WRDP on some other graph classes. Also, we think that a similar approach can be used for solving the Roman domination problem on a block graph as well.

Acknowledgement

This research was partially supported by Serbian Ministry of Education, Science and Technological Developments under grants no. TR36006 and ON174010.

REFERENCES

Ahangar, H., Henning, M., Löwenstein, C., Zhao, Y., and Samodivkin, V. (2014). Signed roman domination in graphs. *Journal of Combinatorial Optimization*, 27(2):241–255.

Beeler, R. A., Haynes, T. W., and Hedetniemi, S. T. (2016). Double roman domination. *Discrete Applied Mathematics*, 211:23–29.

Chambers, E., Kinnersley, B., Prince, N., and West, D. (2009). Extremal problems for roman domination. *SIAM Journal on Discrete Mathematics*, 23(3):1575–1586.

Chapelle, M., Cochefert, M., Couturier, J.-F., Kratsch, D., Letourneur, R., Liedloff, M., and Perez, A. (2017). Exact algorithms for weak roman domination. *Discrete Applied Mathematics*, in press.

Chellali, M., Haynes, T., and Hedetniemi, S. (2014). Bounds on weak roman and 2-rainbow domination numbers. *Discrete Applied Mathematics*, 178:27–32.

Cockayne, E., P.A.Jr., D., Hedetniemi, S., and Hedetniemi, S. (2004). Roman domination in graphs. *Discrete Mathematics*, 278 (1-3):11–22.

Dreyer, J. (2000). *Applications and variations of domination in graphs*. PhD thesis.

Harary, F. (1963). A characterization of block-graphs. *Canadian Mathematical Bulletin*, 6(1):1–6.

Harray, F. (1969). *GRAPH THEORY*. Addison-Wesley MR0256911.

Henning, M. (2002). A characterization of roman trees. *Discussiones Mathematicae Graph Theory*, 22(2):325–334.

Henning, M. and Hedetniemi, S. (2003). Defending the roman empire - a new strategy. *Discrete Mathematics*, 266:239–251.

Ivanović, M. (2017). Improved integer linear programming formulation for weak roman domination problem. *Soft Computing*, pages 1–11.

Liu, C. H. and Chang, G. J. (2013). Roman domination on strongly chordal graphs. *Journal of Combinatorial Optimization*, 26(3):608–619.

Liu, C. S., Peng, S. L., and Tang, C. Y. (2010). Weak roman domination on block graphs. In *The 27th Workshop on Combinatorial Mathematics and Computation Theory*, pages 86–89. Providence University, Taichung, Taiwan.

Mai, T. and Pushpam, P. (2011). Weak roman domination in graphs. *Discuss.Math.Graph Theory*, 31(1):161–170.

ReVelle, C. and Rosing, K. (2000). Defendens imperium romanum: a classical problem in military strategy. *Amer.Math.*, 107 (7):585–594.

Stewart, I. (1999). Defending the roman empire! *Sci. Amer.*, 281 (6):136–139.

B3
Data Envelopment Analysis & Performance Measurement

MEASURING THE VOLUME OF TRANSPORT USING THE BOD-CIDI MODEL: THE CASE OF EUROPE

MILICA MARIČIĆ[1], MARINA DOBROTA[2], MILICA BULAJIĆ[3], VELJKO JEREMIĆ[4]

[1] University of Belgrade, Faculty of Organizational Sciences, milica.maricic@fon.bg.ac.rs
[2] University of Belgrade, Faculty of Organizational Sciences, dobrota.marina@fon.bg.ac.rs
[3] University of Belgrade, Faculty of Organizational Sciences, bulajic.milica@fon.bg.ac.rs
[4] University of Belgrade, Faculty of Organizational Sciences, jeremic.veljko@fon.bg.ac.rs

---

Abstract: *Measuring the volume of transport is important in evaluating existing services and infrastructure. Therefore, in this paper, we propose a novel composite indicator which can serve as a measure of the volume of transport in European countries. The implied composite indicator consists of five respective indicators which are mainly focused on freight transport, but also include the passenger transport. A ubiquitous issue when creating a composite indicator refers to individual weights that each indicator receives. For this reason, in the process of composite indicator development, we applied a multivariate approach for defining weights, based on the Benefit-of-the-Doubt (BoD) model and the CIDI methodology. By combining these approaches, some of the limitations of the original BoD model were mastered. Application of this combined approach enabled us to propose entity-specific weights which maximise the value of the composite indicator. The presented approach might initiate further research on the topic of composite indicators, the application of optimisation methods in composite indicator creation, and on ranking countries based on the volume of transport.*

Keywords: *Statistics and Operational Research, Benefit-of-Doubt, CIDI Methodology, Transport.*

1. INTRODUCTION

Global transport and all of its segments represent a fundamental part of any economy and society. Transport enables economic growth and job creation all over the world, while mobility is vital for the internal market and the quality of life of citizens as they enjoy their freedom to travel (European Commission, 2011). In developed economies, transport investments and improved technology over the last century have resulted in a continuous decline in transport costs, which in turn stimulated growth and economic development (Berg et al., 2017). The future prosperity of economies will depend on their ability to remain fully and competitively integrated into the world economy, and the efficient transport is vital in making this happen (European Commission, 2011).

The concept of transport can be decoupled to cover environmental issues (e.g. carbon), social issues (e.g. fatalities), as well as the economic dimensions (e.g. income growth and transport activities) and effects (Loo & Banister, 2016). While numerous papers cover the environmental issues (Hao et al., 2015; Lu, 2016; Xylia & Silveira, 2017; Andrés & Padilla, 2018;) and social issues (Moeinaddini et al., 2015; Mononen & Leviäkangas, 2016), this paper focuses on volume of transport in countries as one of the main tokens of economic dimension.

The volume of transport has gradually increased over the past years in Europe. A constant supply of goods is vital for the liveability (Mackett & Thoreau, 2015) and the economy of any country in the world (Kin et al., 2017). As a consequence, a large number of freight vehicle movements is generated (Lindholm & Browne, 2013; Kin et al., 2017). The increase of road transport is reflected on the highways and speed roads in regions which have higher economic potential. The increase of rail transport is not as massive as in road transport, but yet quite noticeable (Černá, Zitrický, & Daniš, 2017). For example, according to the research of Frey, Hartwig, and Doll, (2014), modal share in freight transport in Germany is dominated by road haulage, serving roughly 63 % of tonne-kilometres of transport. The remaining performance is carried by rail (22%), inland waterways (12%), and pipelines (3%).

Trade globalisation also influences the increase of transport volume in intermodal transport (De Borger, 2018). In March 2011, the European Commission (European Commission, 2011) adopted a comprehensive strategy (Transport 2050) for a competitive transport system: its goal is to increase mobility, to remove major barriers in key areas, especially regarding the fuel employment. As the largest emerging country and the world's most populous nation, China has experienced a boost in transport turnover in the last three decades (Zhang, Chen, & Huang, 2016). Thus, among others, the development of transport volume in freight transport in Europe largely depends on the transport and business processes between the European Union and China, and it is an opportunity for many European countries to obtain transit transport flows (Černá et al., 2017).

Transportation became an important topic in composite indicators of sustainability and competitiveness of countries (Annoni & Kosovska, 2010; Luzzati & Gucciardi, 2015). However, composite indicators on the sole topic of transportation have also been devised. For example, Stenström et al. (2015) proposed an integrity index for benchmarking and monitoring rail infrastructure. Angelopoulos (2017) created a composite indicator for the port industry and seaborne trade. More recently, a comprehensive composite indicator for managing sustainable urban freight transport was proposed (Buldeo Rai et al., 2018). Lead by their examples, in this paper we propose a composite indicator that would consist of five respective indicators, all of which singly indicate the volume of transport. As the composite indicator aims at ranking 29 European countries, we call the index European Transportation Index (ETI).

The question which attracts the attention of composite index creators and other experts is the weighting procedure of indicators which make the composite indicator (Amado, São José, & Santos, 2016; Jeremic et al., 2011; Nardo et al., 2005). Commonly the weighting schemes are based on expert opinion; however, such approach to weighting has often been criticised (Soh, 2014; Milenkovic et al., 2016; Becker et al., 2017). Thus, in this paper we employ the Benefit-of-the-Doubt-CIDI approach (BoD-CIDI), devised by Maricic et al. (2016), to obtain data-driven weights to create the ETI. Namely, the Composite I-distance Methodology (CIDI), based on the I-distance method, is used with a lot of success to obtain a data-driven weighting scheme (Dobrota et al., 2015; Dobrota et al., 2016) while the Benefit-of-the-Doubt approach is employed to obtain entity specific weights (Cherchye et al., 2008).

The following chapter sees the presentation of the proposed composite index framework while the statistical methodology and the optimisation problem have been elaborated in detail in Section 3. The obtained results are provided in Section 4, while the concluding remarks are given in the final chapter

2. EUROPEAN TRANSPORTATION INDEX FRAMEWORK

To perform the research for this paper, we collected the data on 29 countries in Europe. The chosen countries include the data on EU-28, excluding Malta due to severe lack of data and expanded by Norway and Switzerland. Regarding the Brexit issue, for the time being, the United Kingdom remains a full member of the European Union.

We considered a broad set of indicators that measure the volume of transport in these countries. The indicators are mainly focused on freight transport, but also include the passenger transport. The rail and inland waterways are partially excluded from this study due to the lack of data. For example, passenger transport by rail and waterways are not included in the study, while the freight transport has been processed, and includes transport by road, rail, and inland waterways. Rail and inland waterways transport are based on movements on national territory, regardless of the nationality of the vehicle or vessel, while road transport is based on all movements of vehicles registered in the reporting country (Eurostat, 2018a).

All the data were retrieved from Eurostat and are publicly available (Eurostat, 2018a). The data were acquired for the year 2015, the last available year on Eurostat. The final set of five indicators used in this research is given below:

- *Freight transport per GDP* – This indicator measures the volume of freight transport relative to GDP. It is defined as the ratio between tonne-kilometres (one tone of goods transported over one kilometre) and GDP, and it refers to inland transport only.
- *Motorization rate* - This indicator is defined as the number of passenger cars per 1 000 inhabitants. A passenger car can be defined as a road motor vehicle, other than a motorcycle, planned for the transport of passengers and aimed to seat no more than nine passengers (including the driver). The term "passenger car" therefore covers microcars (that do not need a permit to be driven), cars, taxis, hired passenger cars, or pick-ups provided that they have fewer than ten seats.
- *Passenger transport per GDP* – It measures the volume of passenger transport relative to GDP. It is computed as the ratio between the volume of inland passenger transport measured in passenger-kilometres and GDP. It includes transport on national territory by passenger car, bus and coach, and train.
- *Goods transport by road TKM* – The indicator measures goods transport by road given in million tonne-kilometre (TKM). It covers the carriage of goods by road using goods road transport vehicles registered in the reporting countries. They may exclude from the scope of these statistics the goods road transport

vehicles whose load capacity is lower than 3.5 tons, or the maximum permissible laden weight is lower than 6 tons.

- *Air transport of goods per GDP* – This indicator shows the volume of goods transported by air, measured in tones and related to GDP. The statistics cover the total quantity of freight and mail loaded/unloaded.

For the need of the research, all data were normalised. The normalisation was necessary to get the finest results. It was performed by a min-max normalisation method, with respect to zero as the minimum value. Thus, all the values are transformed into a range 0-1, but the value 0 has been bypassed.

After the indicators were chosen and normalized, the next step in the process of creating the ETI index was the application of the BoD-CIDI methodology. The BoD-CIDI methodology is briefly presented in the following section.

3. WEIGHTING METHODOLOGY

The Benefit-of-the-Doubt (BoD) model was originally proposed by Melyn and Moesen (1991), whereas it has its roots in the Data Envelopment Analysis (DEA) (Charnes, Cooper, & Rhodes, 1978). The main idea of the DEA method is to calculate the maximum efficiency of decision-making units (DMUs) based on the information on their inputs and outputs. Similarly, the BoD model aims at maximising the overall index value while having no prior information on the indicator weights. Clearly, there are similarities between DEA and BoD models: first, between their aims and second, in the lack of available information on weights (Cherchye et al., 2008; Maricic et al., 2016).

The original BoD model has several benefits, but there are some drawbacks. One of them is that the model has full freedom (Rogge, 2012). The full freedom is allowed by the constraint which states that the weight should be more or equal to zero. This constraint means that the model can take into account only the value of one indicator whose value is the highest compared to others and assign zero weights to other indicators. Therefore, additional weight constraints are needed and recommended (Cherchye et al., 2007).

Maricic et al. (2016) proposed to constrain the BoD model using the Composite I-distance Indicator (CIDI) methodology. The CIDI methodology is a recently developed statistical methodology for creating composite indicators using the results of the I-distance method devised by Ivanovic (1977) (Dobrota et al., 2016). The I-distance method overcomes the multiple issues in the creation of composite indicators: normalisation, weighting, and aggregation (Jeremic et al., 2011). However, the question of sensitivity and comparability of results emerged, so the CIDI methodology was created. Namely, it creates a comparable composite index, using the weights which derive from the I-distance method. The new weights are formed by dividing the Pearson's correlation coefficient of an indicator and the I-distance value by the sum of correlation coefficients (Dobrota et al., 2015; Maricic et al., 2015; Maricic et al., 2016).

The BoD-CIDI optimisation problem can be formulated as:

$$CI_c = \max_{w_{c,i}} \sum_{i=1}^{q} w_{c,i} y_{c,i} \qquad (1)$$

s.t

$$\sum_{i=1}^{q} w_{c,i} = 1 \quad i = 1,...,q \quad j = 1,...,c,...n \qquad (2)$$

$$w_{c,i} \geq 0.75 \cdot w_{CIDIi} \qquad (3)$$

$$w_{c,i} \leq 1.25 \cdot w_{CIDIi} \qquad (4)$$

In the BoD-CIDI model, equation (1) is the objective function which computes the composite indicator and it is the same as the objective function in the original BoD model. Equation (2) guarantees that the sum of weights assigned to the indicators of the observed entity will be 1. The equations (3) and (4) ensure that the new weights will be within the interval of ±25% of the weights suggested by the CIDI methodology. The chosen interval around the CIDI weights was used to ensure a wide enough interval to have robustness checks (Saisana & Saltelli, 2014). Using the suggested constraints, the proposed model guarantees that all indicators will be taken into account and that no indicator will be assigned zero weight. The BoD-CIDI model, therefore, overcomes the observed problem of the BoD model: full freedom. However, the model has its limitations. Namely, prior to solving it, the indicator values should be normalised to the range between 0 and 1. Accordingly, the question of the type of normalisation arises. Herein we limit ourselves to the application of the proposed model on the already normalised data.

4. RESULTS

The dataset needed for the analysis contained all five indicator values for 29 countries for the year 2015. As the dataset was normalised in the previous steps, the next step in the creation of the ETI was the application of the CIDI methodology to obtain the data-driven weights. The CIDI weights and the upper and lower bounds of constraints of indicator weights for the BoD-CIDI model are presented in Table 1.

Table 1: CIDI weights along with the upper and lower bounds of constraints of indicator weights

Indicator	CIDI weight	Lower bound	Upper bound
Passenger transport per GDP	0.280	0.210	0.350
Motorization rate	0.270	0.203	0.338
Goods transport by road TKM	0.200	0.150	0.250
Air transport of goods per GDP	0.170	0.128	0.213
Freight transport per GDP	0.080	0.060	0.100

According to the CIDI methodology, the most important indicator for the ranking process is *Passenger transport per GDP* followed by *Motorization rate*. The least important indicator is *Freight transport per GDP* with only 8% of weight. The obtained weights acted as input for the construction of the BoD-CIDI model constraints. Namely, the lower bound of the BoD weights constraint is calculated as 75% of the weights assigned by the CIDI methodology. Similarly, the upper bound is calculated, whereas it is 25% above the weights assigned by the CIDI methodology. We used the ±25% interval as suggested in previous studies (Saisana & Saltelli, 2014; Maricic et al., 2016).

After obtaining the model constraints, the BoD-CIDI model was solved using Excel Solver. The model was run separately on each of the 29 observed countries to obtain country-specific weights. Those weights were used in the next step to calculate the value of the ETI. The ETI values and ranks for the top and bottom ten countries are presented in Table 2.

Table 2: Values and ranks of top and bottom ten countries according to the ETI

Country	ETI value	ETI rank	Country	ETI value	ETI rank
Germany	0.861	1	Portugal	0.521	20
Italy	0.677	2	Czech Republic	0.517	21
Poland	0.677	3	Sweden	0.506	22
France	0.662	4	Croatia	0.493	23
Greece	0.626	5	Hungary	0.480	24
United Kingdom	0.623	6	Ireland	0.445	25
Spain	0.620	7	Romania	0.442	26
Cyprus	0.595	8	Latvia	0.439	27
Luxembourg	0.592	9	Lithuania	0.420	28
Finland	0.588	10	Slovakia	0.408	29

According to the results, Germany tops the list, followed by Italy and Poland. Germany is known for its highly developed transportation infrastructure. Nevertheless, their air transport is also immense. Namely, airports in Germany dealt with 4.3 million tonnes of air freight in 2015, considerably more than in any other EU Member State (Eurostat, 2018b). Transportation and logistics have been of great importance in Italy since the earliest times. Therefore it has significant road infrastructure such as motorways, freight hubs, alpine tunnels, and railway lines (Statista, 2018). In Poland, on the other hand, a Centre for EU Transport Projects (CEUPT) has been formed in 2008 to implement key transport projects financed from EU funds (CEUPT, 2018). The results of their work are visible, as Poland is on a high third position.

On the other side of the ranking lay Slovakia and Lithuania. The European Commission (European Commission, 2018b) reports that the ratings for Slovakia's road infrastructure have improved slightly, but that the air transport infrastructure is still rated poorest in the EU. In case of Lithuania, again the quality of their air transport infrastructure is below the EU average (European Commission, 2018a).

Alongside the elaboration of the ETI values and ranks, the Country-specific weights should be observed. The assigned weights of the top and bottom five countries are accordingly presented in Table 3.

Table 3: Assigned weights to top and bottom five countries using the BoD-CIDI model

Country	Passenger transport per GDP	Motorization rate	Goods transport by road TKM	Air transport of goods per GDP	Freight transport per GDP	ETI rank
Germany	0.21*	0.267	0.25**	0.213**	0.06*	1
Italy	0.324	0.338**	0.15*	0.128*	0.06*	2
Poland	0.21*	0.338**	0.25**	0.128*	0.074	3
France	0.324	0.338**	0.15*	0.128*	0.06*	4
Greece	0.35**	0.312	0.15*	0.128*	0.06*	5
…	…	…	…	…	…	…
Ireland	0.324	0.338**	0.15*	0.128*	0.06*	25
Romania	0.35**	0.272	0.15*	0.128*	0.1**	26
Latvia	0.35**	0.272	0.15*	0.128*	0.1**	27
Lithuania	0.284	0.338**	0.15*	0.128*	0.1**	28
Slovakia	0.284	0.338**	0.15*	0.128*	0.1**	29

Note: * The weight restriction attains the lower bound, ** The weight restriction attains the upper bound

Table 3 provides evidence that the proposed model can be solved without constraint violation. The sum of assigned weights is 1 and the weights are in the defined interval. Also, it gives insight on how the model assigns weights. Take the example of Germany whose results were 0.682, 0.829, 1, 1, and 0.565, respectively. The upper weight bound was assigned to indicators whose values were maximum, and not to the indicator that was the most important for the ranking process (Maricic et al., 2016). Among the top five countries, it can be observed that they have high values of indicator *Motorization rate* as the weight of this indicator attained the upper bound. Taking a closer look on the five countries on the bottom of the list it can be observed that the weights attain the lower bound for indicators *Air transport of goods per GDP* and *Goods transport by road TKM*. This indicates that these countries have low values of these indicators and that there is place for improvement in those segments of transportation.

To additionally explore the weights assigned by the BoD-CIDI model we provide the descriptive statistics of the assigned weights in Table 4. Comparing the mean values of assigned indicator weights with the CIDI weights, it is noticeable that the weights of *Passenger transport per GDP* and *Motorization rate* are higher, while the rest mean values are below the weights suggested by CIDI. When it comes to the standard deviation of weights, weights of indicator *Passenger transport per GDP* show the highest variability, 0.0399. This indicates that the weights of this indicator vary the most. Contrarily, the weights of *Air transport of goods per GDP* vary the least. Finally, when analysing the minimum and maximum values of weights, it is interesting to note there is just one indicator whose weights did not attain the upper or the lower bound. Namely, the lowest assigned weight of indicator *Motorization rate* is 0.272 compared to 0.203 which is the lower constraint.

Table 4: Descriptive statistics of the assigned weights

Indicator	Mean	Standard deviation	Minimum	Maximum
Passenger transport per GDP	0.3152	0.0399	0.210	0.350
Motorization rate	0.3206	0.0267	0.272	0.338
Goods transport by road TKM	0.1603	0.0305	0.150	0.25
Air transport of goods per GDP	0.1309	0.0155	0.128	0.213
Freight transport per GDP	0.0729	0.0184	0.06	0.100

5. CONCLUSION

Composite indexes slowly but surely became a valuable asset for government representatives and decision makers as they can facilitate comparison, benchmarking, and communication (Yigitcanlar & Dur, 2010; Singh et al., 2012). Nevertheless, their methodologies have been criticised, especially when it comes to the weighting processes based on expert opinions (Giannetti et al., 2009; Jeremic et al., 2011). Herein, we aim to expand the current literature on composite indicators by suggesting a novel composite indicator in the sphere of transport whose weighting scheme is based on an optimisation model and statistical methodology.

The outlined paper aims at creating a composite index based on the volume of air transport, road transport, and passenger transport. The indicator data is based on the official statistics collected by the Eurostat. What makes this composite indicator stand out is the proposed weighting method which assigns entity specific, data-driven weights. The BoD-CIDI model applied is an extension of the original BoD model which aims to maximise the overall index value by assigning entity specific weights (Maricic et al., 2016). The proposed model is constrained using the results of the CIDI method, surpassing the issue of full freedom of the original BoD model (Rogge, 2012). The BoD-CIDI model was employed on the chosen five indicators to create the ETI. The ETI results show that Germany and Italy lead the way, closely followed by Poland.

The additional analysis of the obtained weighting schemes can provide valuable information for policy and decision makers on the level of countries and on the EU level. The results of the created composite indicator show that the countries at the bottom of the ranking have lower values of indicators *Air transport of goods per GDP* and *Goods transport by road TKM.* This means that if the countries are to advance in the ranking, these two aspects should be improved.

The presented paper has several benefits which should be mentioned. First, it provides a novel composite indicator which ranks European countries based on their volume of transport. Secondly, it shows that the BoD-CIDI model devised by Maricic and associates (2016) can be successfully used in the process of creation of composite indicators.

Nevertheless, during research, we encountered several limitations. The first limitation of the study is the fact that the last available data are for the year 2015. It would be quite engaging to perform the same study for future years and to compare the results, to investigate what changes would occur in the ranking of countries in Europe. It would also enable the analysis of the assignment of specific weights for each indicator and determination of possible changes over time. The second limitation of this study is the inability to include rail and inland waterways. Namely, they are only partially included in the study through the measure of freight transport. This is especially important because one of the goals of European Commission's strategy (Transport 2050) for a competitive transport system is the following: "by 2030, 30% of road freight over 300 km should shift to other modes such as rail or waterborne transport" (European Commission, 2011). However, due to the lack of the data, the length or passenger transport by rail and waterways had to be excluded at the time. For future research, it would be a preference for these data to be provided and included in the study, regarding the fact that they make a large share of total transport of a country.

We hope that the proposed index based on the recently developed BoD-CIDI model can initiate further research on the index itself and on the weighing schemes of composite indices devised using optimisation and statistical methods.

Acknowledgement

The research outlined in this paper was supported by the Serbian Ministry of Education and Science grant III47003.

REFERENCES

Amado, C. A. F., São José, J. M. S., & Santos, S. P. (2016). Measuring active ageing: A Data Envelopment Analysis approach. *European Journal of Operational Research*, *255*, 207–223. https://doi.org/10.1016/j.ejor.2016.04.048

Andrés, L., & Padilla, E. (2018). Driving factors of GHG emissions in the EU transport activity. *Transport Policy*, *61*, 60–74. https://doi.org/10.1016/j.tranpol.2017.10.008

Angelopoulos, J. (2017). Creating and assessing composite indicators: Dynamic applications for the port industry and seaborne trade. *Maritime Economics & Logistics*, *19*(1), 126–159. https://doi.org/10.1057/s41278-016-0050-8

Annoni, P., & Kosovska, K. (2010). *EU regional competitiveness index 2010*.

Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators*, *80*, 12–22. https://doi.org/10.1016/j.ecolind.2017.03.056

Berg, C. N., Deichmann, U., Liu, Y., & Selod, H. (2017). Transport Policies and Development. *The Journal of Development Studies*, *53*(4), 465–480. https://doi.org/10.1080/00220388.2016.1199857

Buldeo Rai, H., van Lier, T., Meers, D., & Macharis, C. (2018). An indicator approach to sustainable urban freight transport. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, *11*(1), 81–102. https://doi.org/10.1080/17549175.2017.1363076

Černá, L., Zitrický, V., & Daniš, J. (2017). The Methodology of Selecting the Transport Mode for Companies on the Slovak Transport Market. *Open Engineering*, *7*(1). https://doi.org/10.1515/eng-2017-0002

CEUPT. (2018). „On the right track. Transport Infrastructure in Poland – Lessons and Challenges" – conference report. Retrieved February 15, 2018, from https://cupt.gov.pl/en/news/1078-conference-report-on-the-right-track-transport-infrastructure-in-poland-lessons-and-challenges

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Cherchye, L., Moesen, W., Rogge, N., & Puyenbroeck, T. Van. (2007). An introduction to "benefit of the doubt" composite indicators. *Social Indicators Research*, *82*(1), 111–145. https://doi.org/10.1007/s11205-006-9029-7

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., … Tarantola, S. (2008). Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index. *Journal of the Operational Research Society*, *59*(2), 239–251. https://doi.org/10.1057/palgrave.jors.2602445

De Borger, B. (2018). Special issue "The political economy of transport decisions." *Economics of Transportation*. https://doi.org/10.1016/j.ecotra.2018.01.002

Dobrota, M., Bulajic, M., Bornmann, L., & Jeremic, V. (2016). A new approach to the QS university ranking using the composite I-distance indicator: Uncertainty and sensitivity analyses. *Journal of the Association for Information Science and Technology*, *67*(1), 200–211. https://doi.org/10.1002/asi.23355

Dobrota, M., Martic, M., Bulajic, M., & Jeremic, V. (2015). Two-phased composite I-distance indicator approach for evaluation of countries' information development. *Telecommunications Policy*, *39*(5), 406–420. https://doi.org/10.1016/j.telpol.2015.03.003

European Commission. (2011). *White Paper on transport: Roadmap to a Single European Transport Area– Towards a competitive and resource efficient transport system*. https://doi.org/10.2832/30955

European Commission. (2018a). Mobility and transport - Lithuania. Retrieved February 10, 2018, from https://ec.europa.eu/transport/facts-fundings/scoreboard/countries/lithuania/investments-infrastructure_en

European Commission. (2018b). Mobility and transport - Slovakia. Retrieved February 10, 2018, from https://ec.europa.eu/transport/facts-fundings/scoreboard/countries/slovakia/investments-infrastructure_en

Eurostat. (2018a). European statistics.

Eurostat. (2018b). Freight transport statistics. Retrieved February 4, 2018, from http://ec.europa.eu/eurostat/statistics-explained/index.php/Freight_transport_statistics

Frey, K., Hartwig, J., & Doll, C. (2014). Accelerating a shift from road to rail freight transport in Germany – Three scenarios. *Transport Research Arena (TRA) 5th Conference: Transport Solutions from Research to Deployment*, 10p.

Giannetti, B. F., Bonilla, S. H., Silva, C. C., & Almeida, C. M. V. B. (2009). The reliability of experts' opinions in constructing a composite environmental index: The case of ESI 2005. *Journal of Environmental Management*, *90*(8), 2448–2459. https://doi.org/10.1016/j.jenvman.2008.12.018

Hao, H., Geng, Y., Li, W., & Guo, B. (2015). Energy consumption and GHG emissions from China's freight transport sector: Scenarios through 2050. *Energy Policy*, *85*, 94–101. https://doi.org/10.1016/j.enpol.2015.05.016

Ivanovic, B. (1977). *Classification theory*. Institut za ekonomiku industrije, Beograd.

Jeremic, V., Bulajic, M., Martic, M., & Radojicic, Z. (2011). A fresh approach to evaluating the academic ranking of world universities. *Scientometrics*, *87*(3), 587–596. https://doi.org/10.1007/s11192-011-0361-6

Kin, B., Verlinde, S., Mommens, K., & Macharis, C. (2017). A stakeholder-based methodology to enhance the success of urban freight transport measures in a multi-level governance context. *Research in Transportation Economics*, *65*, 10–23. https://doi.org/10.1016/j.retrec.2017.08.003

Lindholm, M., & Browne, M. (2013). Local authority cooperation with urban freight stakeholders: A comparison of partnership approaches. *European Journal of Transport and Infrastructure Research*, *13*(1), 20–38.

Loo, B. P. Y., & Banister, D. (2016). Decoupling transport from economic growth: Extending the debate to include environmental and social externalities. *Journal of Transport Geography*, *57*, 134–144. https://doi.org/10.1016/j.jtrangeo.2016.10.006

Lu, S.-M. (2016). A low-carbon transport infrastructure in Taiwan based on the implementation of energy-saving measures. *Renewable and Sustainable Energy Reviews*, *58*, 499–509.

https://doi.org/10.1016/j.rser.2015.12.242
Luzzati, T., & Gucciardi, G. (2015). A non-simplistic approach to composite indicators and rankings: an illustration by comparing the sustainability of the EU Countries. *Ecological Economics*, *113*, 25–38. https://doi.org/10.1016/j.ecolecon.2015.02.018
Mackett, R. L., & Thoreau, R. (2015). Transport, social exclusion and health. *Journal of Transport & Health*, *2*(4), 610–617. https://doi.org/10.1016/j.jth.2015.07.006
Maricic, M., Bulajic, M., Martic, M., & Dobrota, M. (2015). Measuring the ict development: the fusion of biased and objective approach. *Naval Academy Scientific Bulletin*, *18*(2), 326–334.
Maricic, M., Bulajic, M., Radojicic, Z., & Jeremic, V. (2016). Multivariate approach to imposing additional constraints on the Benefit-of-the-Doubt model: The case of QS World University Rankings by Subject. *Croatian Review of Economic, Business and Social Statistics*, *2*(1), 1–14. https://doi.org/10.1515/crebss-2016-0005
Melyn, W., & Moesen, W. (1991). *Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available*.
Milenkovic, M. J., Brajovic, B., Milenkovic, D., Vukmirovic, D., & Jeremic, V. (2016). Beyond the equal-weight framework of the Networked Readiness Index: a multilevel I-distance methodology. *Information Development*, *32*(4), 1120–1136. https://doi.org/10.1177/0266666915593136
Moeinaddini, M., Asadi-Shekari, Z., Sultan, Z., & Zaly Shah, M. (2015). Analyzing the relationships between the number of deaths in road accidents and the work travel mode choice at the city level. *Safety Science*, *72*, 249–254. https://doi.org/10.1016/j.ssci.2014.09.015
Mononen, P., & Leviäkangas, P. (2016). Transport safety agency's success indicators – How well does a performance management system perform? *Transport Policy*, *45*, 230–239. https://doi.org/10.1016/j.tranpol.2015.03.015
Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. (2005). *Handbook on constructing composite indicators*. *OECD Statistics Working Papers*. https://doi.org/10.1787/533411815016
Rogge, N. (2012). Undesirable specialization in the construction of composite policy indicators: The environmental performance index. *Ecological Indicators*, *23*, 143–154. https://doi.org/10.1016/j.ecolind.2012.03.020
Saisana, M., & Saltelli, A. (2014). JCR statistical audit of the WJP Rule of Law index 2014. In *World Justice Project: The World Justice Project Rule of Law Index 2014* (pp. 188–197). World Justice Project.
Singh, R. K., Murty, H. R., Gupta, S. K., & Dikshit, A. K. (2012). An overview of sustainability assessment methodologies. *Ecological Indicators*, *15*(1), 281–299. https://doi.org/10.1016/j.ecolind.2011.01.007
Soh, K. (2014). Nominal versus attained weights in Universitas 21 Ranking. *Studies in Higher Education*, *39*(6), 944–951. https://doi.org/10.1080/03075079.2012.754866
Statista. (2018). Transportation industry in Italy - Statistics & Facts. Retrieved February 6, 2018, from https://www.statista.com/topics/3690/transportation-industry-in-italy/
Stenström, C., Parida, A., Lundberg, J., & Kumar, U. (2015). Development of an integrity index for benchmarking and monitoring rail infrastructure: application of composite indicators. *International Journal of Rail Transportation*, *3*(2), 61–80. https://doi.org/10.1080/23248378.2015.1015220
Xylia, M., & Silveira, S. (2017). On the road to fossil-free public transport: The case of Swedish bus fleets. *Energy Policy*, *100*, 397–412. https://doi.org/10.1016/j.enpol.2016.02.024
Yigitcanlar, T., & Dur, F. (2010). Developing a Sustainability Assessment Model: The Sustainable Infrastructure, Land-Use, Environment and Transport Model. *Sustainability*, *2*(12), 321–340. https://doi.org/10.3390/su2010321
Zhang, H., Chen, W., & Huang, W. (2016). TIMES modelling of transport sector in China and USA: Comparisons from a decarbonization perspective. *Applied Energy*, *162*, 1505–1514. https://doi.org/10.1016/j.apenergy.2015.08.124

THE EFFICIENCY ASSESSMENT OF 'SMART' PERFORMANCES AND THE QUALITY OF LIFE IN CEE CITIES

JELENA STANKOVIĆ[1], IVANA MARJANOVIĆ[2]

[1]University of Niš, Faculty of Economics, jelena.stankovic@eknfak.ni.ac.rs
[2]University of Niš, Faculty of Economics, ivana.veselinovic@eknfak.ni.ac.rs

Abstract: *The current way of cities' functioning is becoming unsustainable. Traditional cities can be categorised as resource consumers. Namely, cities base their functioning on the consumption of (mostly external) resources, often creating negative externalities (pollution, garbage, etc.). Since the cities represent the bearers of social and economic development and the urbanization is a process that cannot be stopped, the focus of policymakers and researches is shifted towards finding a solution that will respond to this new challenges. One of the possible answers is the concept of smart cities. Smart city initiatives are consideredto be a way to make urban areas more sustainable. However, the issue of cities' efficiency in fulfilling citizens' basic needs is not often considered. Therefore, the subject of this paper is to explore the efficiency of "smart" performances of CEE cities in creating preconditions for higher quality life oftheir citizens, using Data Envelopment Analysis. The research is based on data from Urban Audit Perception Survey conducted by Eurostat and results show that more than 60% of observed CEE cities achieve efficiency frontier.*

Keywords: *smart cities, Data Envelopment Analysis, quality of life.*

1. INTRODUCTION

Modern trends of population migrations, which are primarily reflected in the transition from rural to urban areas point to the conclusion that cities need to become more sustainable and self-sufficient in order to provide quality life for their citizens. According to the United Nations (World urbanisation prospects 2014), more than 60% of the world's population will live in an urban area by 2050, while the same projection for Europe predicts that the share of the population living in urban areas will rise to just over 80% by 2050 (Urban Europe Statistics on cities, towns and suburbs 2016).Urbanisation creates a number of economic and environmental problems that reduce quality of life in urban settlements, such aspollution, sanitation, ageing population, transport, inequality, safety issues and others. In order to respond to these problemscities should have smart infrastructure and the sustainable use of resources available in the immediate surroundings of cities - the potential that has so far been insufficiently exploited in a large number of European cities. The concept that meets these conditions is the so-called concept of smart cities. The idea of smart cities is not new and the hasteneddevelopment of technology, globalization and the aspiration towards sustainable development in recent years has contributed to the emergence of many smart cities throughout the world. These are mainly large metropolitan cities that have used different solutions and technologies to make them more efficient.However, most of the cities in Central and Eastern Europe (abbreviated CEE) are smaller and in the development of the smart city concept they face different problems and challenges. Therefore, the aim of this paper is to identify efficientcities in CEE in terms of achieving higher quality life for their citizens and to set guidelines for inefficient cities how to use their resources to improve the level of achieved citizens' quality of life.

Bearing in mind the stated goal, the paper next to the introduction and conclusion contains the following units: a) a brief literature review on the development of the *smart city* concept; b) model development and methodology and c) presentation and discussion of main results obtained by the Data Envelopment Analysis.

2. LITERATURE REVIEW

The concept 'smart city' emerged as a response to the problems generated by the rapid urbanization. However, despite the continuingacademic research in the recent years, there is no generally accepted, comprehensive definition of the 'smart city'. Also, there are several versions of the origin of the concept itself. Gabrys (2014) claims that the 'smart city' concept dates back to the 1960s whenit was called the 'cybernetically planned cities'. That concept wasadditionally developed in proposals for networked or computable cities in urban development plans from the 1980s onward (Börjesson*et al*. 2015). According to

Dameri and Cocchia (2013) the concept was presented in 1994. On the other hand, Neirotti et al. (2014) claim that the origin of the concept can be found in the late 1990s referring smart growth movement (Bibri and Krogstie 2017).With the launch of IBM's Smarter Planet project in 2008 concept of 'smart cities' gained a lot of attention(Palmisano 2008) and since then it has continued to develop and progress. Regarding the meaning of the 'smart city' a lot of interpretation can be found in the literature. The EU defines a smart city as 'a place where traditional networks and services are made more efficient with the use of digital and telecommunication technologies, for the benefit of its inhabitants and businesses' (Urban Europe Statistics on cities, towns and suburbs 2016). Martinez-Balleste et al. (2013) describe smart cities as 'cities strongly founded on ICT (Information and Communication Technology, abbreviated hereafter as ICT) that invest in human and social capital to improve the quality of life of their citizens by fostering economic growth, participatory governance, wise management of resources, sustainability, and efficient mobility, whilst they guarantee the privacy and security of the citizens'.According to Ferrara (2016) 'smart cities are characterized by high-quality public services, better living standards, new job opportunities generated by a more innovative entrepreneurial ecosystem, greater environmental sustainability and lower public resources through the involvement of private finance and networking of large industrial groups and SMEs'. Komninos (2011) defines smart cities as territories with high capacity for learning and innovation, which is built-in the creativity of their population, their institutions of knowledge creation, and their digital infrastructure for communication and knowledge management. Lombardi et al. (2012) consider that the smart city has the characteristics of application of information and communications technology with consideration of their effects on human capital, social and relational capital, and environmental issues.Marsal-Llacuna et al. (2014) claim that smart cities initiatives try to 'improve urban performance by using data, information and information technologies to provide more efficient services to citizens, to monitor and optimize existing infrastructure, to increase collaboration among different economic actors, and to encourage innovative business models in both the private and public sectors'. Kourtit and Nijkamp (2012) consider that smart cities are the result of knowledge-intensive and creative strategies aiming at enhancing the socio-economic, ecological, logistic and competitive performance of cities. They also claim that such smart cities are based on a promising mix of human capital (e.g. skilled labor force), infrastructural capital (e.g. high-tech communication facilities), social capital (e.g. intense and open network linkages) and entrepreneurial capital (e.g. creative and risk-taking business activities). Meijer and Bolivar (2013) believe a city to be smart when 'investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance'.Caragliu et al. (2011a) believe a city to be smart when 'investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance'. Albino et al. (2015) tried to systematize the basic characteristics of smart cities. According to them smart city has:

- a city's networked infrastructure that enables political efficiency and social and cultural development;
- an emphasis on business-led urban development and creative activities for the promotion of urban growth;
- social inclusion of various urban residents and social capital in urban development;
- a natural environment as a strategic component for the future.

In essence, there are two mainstream approaches to smart city (BibriandKrogstie 2017): one is the technology and ICT–oriented approach, which focuses on the efficiency and technological advancement of the city's hard infrastructure and technology (Angelidou 2014) and the other people–oriented approach where main focus is, according to Angelidou (2014), on the soft infrastructure and people(i.e. social and human capital, knowledge, inclusion, participation, social innovation and equity).

Different methods and measurement that can be used to evaluate performance in smart cities can be found in the literature (see Albino et al. 2015). However, the assessment of smart cities' efficiency is a topic that is not often dealt with in research. Data Envelopment Analysis (abbreviated DEA) can be used for this purposes. Several researches have been conducted regarding cities' efficiency. Sueyoshi (1992) has measured the industrial performance of Chinese cities by data envelopment analysis with intention to improve the economic achievement within Chinese cities.Charnes et al. (1989) used the data envelopment analysis to evaluate efficiency in the economic performance of Chinese cities. They claim that DEA can be used to identify sources, and estimate amounts of inefficiencies in each city's performance.Deilmann et al.(2016)analyzed urban efficiency by applying DEA to an investigation of 116 cities throughout Germany.

In their research they have developed two separate models, economic and ecological, in order to allow more precise identification of the relevance of individual parameters during the evaluation process. The results allowed a ranking of cities as well as an estimation of the ratios of economic and ecological efficiencies of the investigated cities. Wang et al. (2015) used meta-frontier and DEA to investigate the environmental protection mechanisms and economic development of 211 cities in China from the perspective of environmental efficiency. Caragliu et al. (2012) measure urban performances of nine selected European cities in the North Sea along multiple dimensions and evaluate their performance and rank their relative efficiency performance using a new variant of DEA. Kourtit et al. (2013) conduct DEA analysis which is aimed to give another approach in cities' ranking. Their results offer a meaningful contribution to decisionsupport and planning for the efficiency improvement of strategic urban policy. Also, they have found prominent differences compared with standard ranking and benchmarking procedures. An and et al. (2017) use DEA: AR-CCR model to determine the relative efficiency of each of the sustainability indicators for a smart city in the context of input and output criteria. They have done the decomposition of efficiency measures in order tohighpoint which sustainability indicator the country needs to focus based on the importance of the input criteria to achieve the desired outputs.

3. MODEL DEVELOPMENT AND METHODOLOGY

In order to determine relative efficiency cities attain in achieving quality life of their citizensthe data on citizens' perceptions of different categories of urban life are used for model development. We have analyzed 26 indicators that refer to smart performance of a city across five thematic categories: (1) infrastructure, (2) liveability and housing conditions, (3) environment, (4) employment and finance and (5) governance, urban safety, trust and social cohesion. These categories represent inputs in our model. In addition, we included two indicators that refer to citizens' perceptions on quality of life in the city, answering to the following questions: (1) Are you satisfied to live in this city and (2) Are you satisfied with the live you lead? Indicators represent the subjective perception of the inhabitants of the fulfilment of certain standards of urban life in their cities and these perceptions represent outputs of model which is the subject of the analysis.

In order to analyze the efficiency of selected smart cities, the DEA method will be applied. DEA is a mathematical, non-parametric approach for calculating efficiency which is based on linear programming. It is used for measuring the performance of decision-making units (abbreviated hereafter as DMU).The method consists in reducing multiple inputs to one 'virtual' input, and multiple outputs down to one 'virtual' output, using weighting coefficients, wherecorresponding linear programming model is formed and solved for each organizational unit. Efficiency is assessed as the ratio of the weighted sum of outputs and the weighted sum of inputs. Obtained efficiency ranges from 0 to 1, and each deviation of 1 is attributed to excess of inputs or lack of outputs. In literature and in practice there are two most commonly used DEA models: CCR (Charnes, Cooper and Rhodes (Charnes 1978)) model is more restrictive and the efficiency ascertained by CCR is always less or equal to the efficiency assessed by BCC model (Banker, Charnes and Cooper, (Banker et al. 1984)). There are two basic orientations of DEA models. Input-oriented models are used to test if a DMU under evaluation can reduce its inputs while keeping the outputs at their current levels while output-oriented models are used to test if a DMU under evaluation can increase its outputs while keeping the inputs at their current levels (Zhu 2014).

Further analysis will be conducted using CCR output-oriented model which tends to maximize output at a given input level. CCR DEA output-oriented linear programming model can be mathematically represented as follows (Cooper et al, 2011):

$$\min q = \sum_{i=1}^{m} v_i x_{i0} \tag{1}$$

Subject to

$$\sum_{r=1}^{s} \mu_r y_{r0} = 1$$

$$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} \mu_r y_{rj} \geq 0, j = 1 \ldots n$$

$$\mu_r, v_i \geq \varepsilon$$

Wherein:

y_{rj} – the value of the output r (r = 1, 2,…, s) for DMU j (j = 1, 2, …, n);

x_{ij} – the value of the input i (i = 1, 2, …, m) for DMU j (j = 1, 2, …, n);

μ_r– weighting coefficient of outputy_{rj} and v_i –weighting coefficient of output x_{rj} .

Resolving is often performed using a dual problem. Output – oriented CCR DEA model can be represented mathematically (Cooper et al, 2011):

$$\max \varphi + \varepsilon(\sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+) \quad (2)$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{i0}, i = 1, \dots, m$$

$$\sum_{j=1}^{n} \lambda_j y_{rj} - s_r^+ = \varphi y_{r0}, r = 1, \dots, s$$

$$\lambda_j \geq 0, j = 1,2\dots n$$

wherein $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$, $\lambda \geq 0$ vector of weighting coefficients allocated to units to be assessed, with s^+ands^- as vectors of positive and negative adjusting variables (slacks) in the input and output constraints, ε is infinitely small constant.This model is solved in two steps. First step consists incalculating φ^* by ignoring theslacks. Then in the second step the slacks are optimize by fixingφ^*in the following linear programming problem(Cooper et al, 2011):

$$\max z = \sum_{i=1}^{m} s_i^- + \sum_{r=1}^{s} s_r^+ \quad (3)$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{i0}, i = 1, \dots, m$$

$$\sum_{j=1}^{n} \lambda_j y_{kj} - s_r^+ = \varphi^* y_{r0}, r = 1, \dots, s$$

$$\lambda_j \geq 0$$

Decision unit is efficient if the following two conditions are met: the optimal value of the variable φ^* is equal to 1 and the optimal values of allslacks s_i^+ and s_i^- are equal to zero. Decision unit is considered to be weakly efficient if the optimum value of the variable φ^* is 1 and the optimal values of all slacks s_i^+ and s_i^- differ from zero. If the resulting value is higher than 1 decision unit is considered inefficient and it is possible to achieve higher level of output with same resources. However, in a standard DEA model all efficient DMUs get the score 1 leading to the problem of adequate ranking of efficient DMUs. In order to make a distinction between efficient units Andersen and Petersen (1993) suggested the use of the Super-Efficiency model for ranking efficient units. They have developed a modelwhich aims to identify the relative importance of each individual efficient DMU by designing and measuring a score for its 'degree of influence' if this efficient firm is omittedfrom the efficiency frontier (Kourtit el al. 2013). If omitted DMU has high degree of influence and outstrips the other DMUs, that DMU is considered to be super-efficient. The super efficiency estimate assumes the modification of the DEA model so that efficient units can be assigned an index greater than 1 and thus allows discrimination between them. The modification of the primary model consists in the fact that the restriction corresponding to DMU_k is removed from the set of constraints imposed by the second relation in the primary CCR DEA model:

$$\min q = \sum_{i=1}^{m} v_i x_{i0} \quad (4)$$

Subject to

$$\sum_{r=1}^{s} \mu_r y_{r0} = 1$$

$$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} \mu_r y_{rj} \geq 0, j = 1 \dots n, j \neq k$$

$$\mu_r, v_i \geq \varepsilon$$

In a dual CCR model, the DMU_k whose efficiency is being evaluated is not considered when defining an input-output mix of a composite unit. In this way, an efficient unit is compared withthe new efficiency frontier that is formed without taking this unit into account.

4. EMPIRICAL DATA AND ANALYSIS

In order to analyze quality of life in Central and Eastern European cities, *Urban Audit Perception Survey* was used in this paper (Statistical Office of the European Union - EUROSTAT 2015).

Urban Audit represents the Eurostat's statistics on cities that provides a range of indicators covering most aspects relating to quality of life in cities. The coverage of the survey includes several aspects of quality of life such as demography, housing, health, economic activity, labor market, income disparity, educational qualifications, environment, climate, travel patterns, tourism and cultural infrastructure (Stanković et al. 2017). The data for the analysis are gathered from the most recent perceptions survey (2015) and they include data for 23 Central and Eastern European Cities (Sofia, Burgas, Prague, Ostrava, Tallinn, Zagreb, Riga, Vilnius, Budapest, Miskolc, Warsaw, Kraków, Gdansk, Bialystok, Bucharest, Cluj-Napoca, Piatra Neamt, Ljubljana, Bratislava, Kosice, Ankara, Antalya and Istanbul).

The survey examined citizens' opinion about their overall satisfaction with regard to the cities they live in and also on their satisfaction with different aspects of urban life. The respondents have expressed their opinion through some of the alternative responses: very satisfied, rather satisfied, rather unsatisfied, not at all satisfied and don't know/no answer. Each response has been assigned a value of the Likert scale (4 for 'very satisfied', 1 for 'not at all satisfied', while the 'don't know/no answer' responses are not taken into account). The survey results have been calculated as mean values of the responses and these means represent 26 indicators that refer to smart performance of a city.After performed calculation results are grouped into categories and the value for each of the category is calculated using weights obtained by the AHP method as shown in Stanković et al. (2017). The relative importance of all indicators is shown in Table 1.

Table 1: Relative importance of all indicators in the model

Category	Indicator	Weights
Infrastructure	Public transport in the city, for example bus, tram or metro	0.064756
	Availability of retail shops	0.024884
	Public spaces in this city such as markets, squares, pedestrian areas	0.014265
Liveability and housing conditions	Health care services offered by doctors and hospitals in this city	0.026218
	Schools in the city	0.077630
	Sports facilities such as sport fields and indoor sport halls in the city	0.008927
	Cultural facilities such as concert halls, theatres, museums and libraries in the city	0.006135
	In this city, it is easy to find good housing at a reasonable price	0.046432
	State of streets and buildings in my neighbourhood	0.018495
Environment	Green spaces such as public parks or gardens	0.002465
	This city is committed to the fight against climate change	0.001534
	The quality of the air in the city	0.018183
	The noise level in the city	0.009303
	The cleanliness in the city	0.005905
Employment and finance	In this city it is easy to find a good job	0.172543
	You have difficulty paying your bills at the end of the month	0.086272
	Your personal job situation	0.172543
	The financial situation of your household	0.086272
Governance, urban safety, trust and social cohesion	When you contact administrative services of this city, they help you efficiently	0.044763
	The public administration of the city can be trusted	0.044763
	Foreigners who live in this city are well integrated	0.005186
	The presence of foreigners is good for this city	0.005186
	Generally speaking, most people in this city can be trusted	0.020019
	You feel safe in this city	0.008651
	You feel safe in the neighborhood you live in	0.008651
	Most people in my neighbourhood can be trusted	0.020019
Σ	Σ	1.00000

Source: Stanković et al. (2017)

The process of obtaining the values of categories is following: the value of an indicator is multiplied by the corresponding weight of that indicator, determined by the AHP method. Then, indicators from the same category are summed and the resulting value represents the value of that category. This is done for each of the category.

Beside stated categories the respondents expressed their satisfaction with the city they live in, as well as with the life they lead, the data that will be also included in the model.

Table 2: Descriptive Statistics of included categories and indicators of citizen satisfaction

	N	Minimum	Maximum	Mean	Std. Deviation
Infrastructure	23	0.28	0.36	0.3151	0.02267
Liveability and housing conditions	23	0.41	0.57	0.4920	0.03583
Environment	23	0.08	0.12	0.0982	0.01212
Employment and finance	23	0.51	0.79	0.6589	0.07191
Governance, urban safety, trust and social cohesion	23	0.17	0.23	0.2026	0.01467
You are satisfied to life in this city	23	1.70	2.37	2.0609	0.19363
The life you lead	23	2.76	3.74	3.4257	0.24041

Source: Authors' calculation

The analysis of smart cities' efficiency level was performed using Efficiency Measurement System Software (abbreviated EMS) developed by Holger Scheel fromthe University of Dortmund in Germany (Scheel, 2000). EMS represents a software specifically designed to measure the efficiency applying the DEA method. The model used for analysis is CCR output-oriented model with five input and two output variables (Table 3). Also Super-Efficiency DEA is performed on the same model in order to gain ranking of units on the efficiency frontier.

Table 3: Input and output variables for DEA model

Input variables					Output variables	
Infrastructure	Liveability and housing conditions	Environment	Employment and finance	Governance, urban safety, trust and social cohesion	You are satisfied to life in this city	The life you lead

Source: Author's preview

Obtained results show that 14 out of 23 cities are efficient, which means that they are all located on the efficiency frontier, while for others there is still some space for an improvement of their performance. Further analysis was conducted using Super-Efficiency DEA. Namely, the results obtained by CCR DEA do not provide clear picture regarding efficient cities. This means that for some of the efficient cities it isequally easy to maintain their position or to improve theirfuture position (Caragliu et al. 2011a).Super-Efficiency results show that with the inputs of super-efficient unit the other DMUs or a convex combination of them would be able to produce lesser output than what actually super-efficient unit produces (Chandrababu and Hariprasad 2015). In our analysis, Super-Efficiency scores show that Ostrava isthe city with the highest efficiency, meaning Ostrava is the best ranked city in terms of providing the highest level of the quality of lifeto its citizens' with the given amount of inputs in comparison with the other analyzed cities.

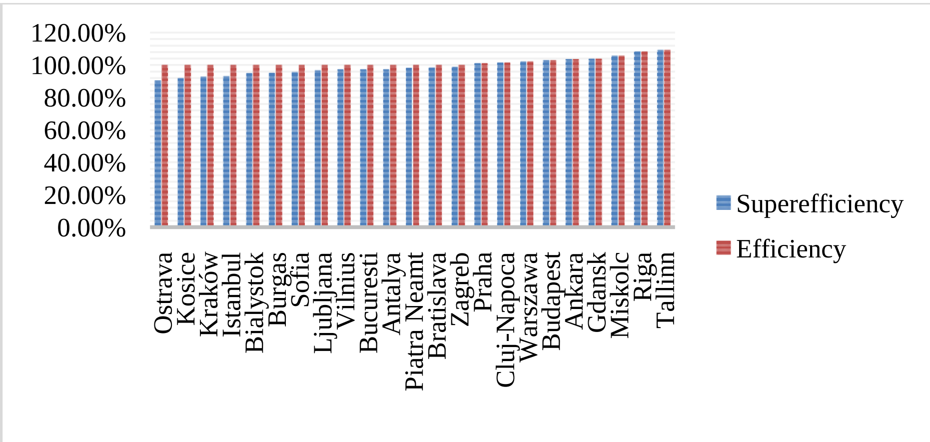


Figure 1: Efficiency and Super-Efficiency scores (Source*:* Authors' preview)

Regarding the inefficient units, they haven't managed to provide the quality of life to its citizens at the level they would do if they have been efficient. For each of the inefficient cities benchmarks are determined. This benchmarks serve as a guidelines to the inefficient cities in terms of what should they strive to. The results show that in order to be efficient Praha should focus on using its resources more like Vilnius and Bratislava. For Cluj-Napoca strongest benchmarks are Bratislava and Bucuresti. Warszawa should look up to Bratislava as its strongest benchmark. Other inefficient cities and its benchmarks are shown in Table 4.

Table 4: Benchmarks for inefficient cities in output-oriented CCR model

City	Benchmarks (λ in parenthesis)
Praha	Vilnius (0,45), Bratislava (0,40), Istanbul (0,14)
Cluj-Napoca	Burgas (0,08), Bucuresti (0,47), Bratislava (0,54), Kosice (0,03)
Warszawa	Burgas (0,07) Vilnius (0,10) Kraków (0,03), Bratislava (0,64), Kosice (0,19)
Budapest	Burgas (0,40), Vilnius (0,22), Kosice (0,33)
Ankara	Ljubljana (0,05), Kosice (0,12), Antalya (0,65), Istanbul (0,17)
Gdansk	Burgas (0,43), Zagreb (0,07) Vilnius (0,34), Kosice (0,18)
Miskolc	Burgas (0,24), Zagreb (0,09) 14 (0,52), Kosice (0,04)
Riga	Sofia (0,35), Vilnius (0,33), Kosice (0,31)
Tallinn	Burgas (0,26) Vilnius (0,75)
Istanbul	Vilnius (0,45), Bratislava (0,40), Istanbul (0,14)

Source*:* Authors preview according to the results of EMS

5. CONCLUSION

Continuing processes of urbanization highlight the importance of adequate analysis of urban development. Cities are considered to be drivers of economic development with high concentration of employment and economic activity. However, modern cities confront with many challenges, namely environmental and social. Therefore, the smart cities concept which is considered to be a solution to many problems of urban living gained a lot of attention recently.

The analysis in this paper is focused on citizens' perception on fulfilment of different demands concerning life satisfaction and quality of life in Central and Eastern European cities. The findings of the performed procedures allow useful comparisons between the cities' efficiency. The results of the research point out that more than 60% of smart cities in Central and Eastern Europe are considered to be efficient in providing quality life for its citizens. Further analysis of efficient cities resulted in their ranking and in the identification of the city with highest output gains, which is shown to be Ostrava. Also, CCR DEA analysis provided some guidelines for inefficient cities in terms of benchmarks they should follow in order to achieve efficient performance.

The scientific contribution of the research is reflected in the fact that it represents the first type of such an assessment of the efficiency of the smart cities in Central and Eastern Europe. Similar research has been done by Kortit et al. (2013) where the efficiency of world smart cities was assessed. Therefore, the goal of this research was not ranking of cities based on a comprehensive set of indicators but the assessment of the efficiency.

Our research faces the usual limitations regarding the usage of data on self-reported life satisfaction and subjective evaluations provided by the respondents. Further research could be done by expanding the data with the data from previous Urban Audits and by performing Window DEA analysis in order to observe time dimension of efficiency.

REFERENCES

Albino, V., Berardi, U., & Dangelico, R. M. (2015). Smart cities: Definitions, dimensions, performance, and initiatives. *Journal of Urban Technology*, 22(1), 3-21.

An and, A., Rufuss, D. D. W., Rajkumar, V., &Suganthi, L. (2017). Evaluation of Sustainability Indicators in Smart Cities for India Using MCDM Approach. *Energy Procedia*, 141, 211-215.

Andersen, P., & Petersen, N. C. (1993). A procedure for ranking efficient units in data envelopment analysis. *Management science*, 39(10), 1261-1264.

Angelidou, M. (2014). Smart city policies: A spatial approach. *Cities*, 41, S3-S11.

Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management science*, 30(9), 1078-1092.

Bibri, S. E., & Krogstie, J. (2017). Smart sustainable cities of the future: An extensive interdisciplinary literature review. *Sustainable Cities and Society*, 31, 183-212.

Börjesson Rivera, M., Eriksson, E., & Wangel, J. (2015). ICT practices in smart sustainable cities: In the intersection of technological solutions and practices of everyday life. In *EnviroInfo& ICT4S*, Atlantis Press, 317-324
Caragliu, A., Del Bo, C., & Nijkamp, P. (2011a). Smart cities in Europe. *Journal of urban technology*, 18(2), 65-82.
Caragliu, A., Del Bo, C., Kourtit, K., & Nijkamp, P. (2011b). Comparative performance assessment of Smart Cities around the North Sea basin. *Network Industries Quarterly*, 13(3), 15-17.
Caragliu, A., Del Bo, C., Kourtit, K., Nijkamp, P., & Suzuki, S. (2012). In search of incredible cities by means of super-efficiency data envelopment analysis. *Studies in Regional Science*, 42(1), 129-144.
Chandrababu, S., & Hariprasad, S. A. (2015). Concept of Input-Output oriented Super-efficiency in Decision Making Units. *American Journal of Engineering Research*,4(1), 76-81.
Cooper, W. W., Seiford, L. M., & Zhu, J. (2011). Data envelopment analysis: History, models, and interpretations. In *Handbook on data envelopment analysis*, Springer, Boston, MA, 1-39
Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6), 429-444.
Charnes, A., Cooper, W. W., & Li, S. (1989). Using data envelopment analysis to evaluate efficiency in the economic performance of Chinese cities. *Socio-Economic Planning Sciences*, 23(6), 325-344.
Deilmann, C., Lehmann, I., Reißmann, D., & Hennersdorf, J. (2016). Data envelopment analysis of cities–Investigation of the ecological and economic efficiency of cities using a benchmarking concept from production management. *Ecological indicators*, 67, 798-806.
Ferrara, G. (2016). SMART CITY: A GEOGRAPHICAL PERSPECTIVE. *RevistaRomână de GeografiePolitică Year XVIII*, (2), 43-48.
Gabrys, J. (2014). Programming environments: environmentality and citizen sensing in the smart city. *Environment and Planning D: Society and Space*, 32(1), 30-48.
Komninos, N. (2011). Intelligent Cities: Variable Geometries of Spatial Intelligence. Intelligent Buildings International, 3(3): 172–188
Kourtit, K., & Nijkamp, P. (2012). Smart cities in the innovation age. *Innovation: The European Journal of Social Science Research*, 25(2), 93-95.
Kourtit, K., Nijkamp, P., & Suzuki, S. (2013). The rat race between world cities: In search of Exceptional Places by means of super-efficient data development analysis. *Computers, Environment and Urban Systems*, 38, 67-77.
Lombardi, P., Giordano, S., Farouh, H., & Yousef, W. (2012). Modelling the smart city performance. *Innovation: The European Journal of Social Science Research*, 25(2), 137-149.
Marsal-Llacuna, M. L., Colomer-Llinàs, J., & Meléndez-Frigola, J. (2015). Lessons in urban monitoring taken from sustainable and livable cities to better address the Smart Cities initiative. *Technological Forecasting and Social Change*, 90, 611-622.
Martinez-Balleste, A., Pérez-Martínez, P. A., & Solanas, A. (2013). The pursuit of citizens' privacy: a privacy-aware smart city is possible. *IEEE Communications Magazine*, 51(6), 136-141.
Meijer, A., & Bolivar, M. (2013, September). Governing the smart city: Scaling-up the search for socio-techno synergy. In *2013EGPA Conference Proceedings*.
Palmisano, S. J. (2008). A smarter planet: the next leadership agenda. *IBM. November*, 6, 1-8.
Stanković, J., Džunić, M., Marinković S. & Džunić Ž. (2017). A Multi-Criteria Evaluation of the European Cities' Smart Performance: Economic, Social and Environmental Aspects. *Proceedings of Rijeka Faculty of Economics: Journal of Economics and Business*, 35(2): 519-550
Scheel, H. (2000). EMS: efficiency measurement system user's manual. *Operations Research and Witshaftinsformetik. University of Dortmund, Germany*.
Statistical Office of the European Union - EUROSTAT 2015. (2015). Available at: http://ec.europa.eu/eurostat/web/cities/data/database
Sueyoshi, T. (1992). Measuring the industrial performance of Chinese cities by data envelopment analysis. *Socio-Economic Planning Sciences*, 26(2), 75-88.
Urban Europe Statistics on cities, towns and suburbs.(2016). Available at: http://ec.europa.eu/eurostat/documents/3217494/7596823/KS-01-16-691-EN-N.pdf/0abf140c-ccc7-4a7f-b236-682effcde10f
Wang, Q., Zhao, Z., Shen, N., & Liu, T. (2015). Have Chinese cities achieved the win–win between environmental protection and economic development? From the perspective of environmental efficiency. *Ecological indicators*, 51, 151-158.
World urbanisation prospects. (2014). Available at: https://esa.un.org/unpd/wup/publications/files/wup2014-highlights.pdf
Zhu, J. (2014). *Quantitative models for performance evaluation and benchmarking: data envelopment analysis with spreadsheets* (Vol. 213). Springer.

BALCOR 2018
XIII Balkan Conference on Operational Research

TECHNOLOGY-BASED CRITICAL SUCCESS FACTORS OF SECONDARY EDUCATION EFFICIENCY: A CROSS-COUNTRY ANALYSIS

NEMANJA MILANOVIĆ[1], MILOŠ MILOSAVLJEVIĆ[2], GORDANA SAVIĆ[3]

University of Belgrade, Faculty of Organizational Sciences,
[1]nemanja.milanovic@fon.bg.ac.rs; [2]milos.milosavljevic@fon.bg.ac.rs; [3]gordana.savic@fon.bg.ac.rs

---

Abstract: *The improvement of secondary education efficiency is believed to be a strong source of social, economic and technological development of a country. It has received immense attention of scholars and practitioners in the last few decades, particularly in the field of measuring a cross-country efficiency. However, there is a scarcity of comprehensive input-output cross-country analyses on education efficiency, particularly those using Data Envelopment Analysis (DEA). The aim of this study is to analyze the efficiency of secondary education among selected European countries. Using an approach of critical success factors, the study developed a matrix of input variables for the analysis of efficiency. The results of the Program for International Student Assessment (PISA) were used as an output, whereas a set of technology based, social economic and instructor-based factors were used as inputs in DEA. The results indicate that technology integration, social and economic background, and instructor-based factors play an important role in education efficiency.*

Keywords: *Education DEA efficiency, technology integration, critical success factor, cross-country analysis.*

1. INTRODUCTION

In the last few decades, the efficiency of secondary education has been growing among policy holders, decision makers, scholars and other stakeholders. Education is believed to be a strong source of human capital formation (De la Fuente and Ciccone 2002) and creator of economic well-being (Blundell et al. 1999). Therefore, any improvement in education efficiency – greater outputs for the given amount of resources invested – would contribute not only to the education alone, but to social, economic and technological development of a country. There are many contentions that technology is one of the prime drivers of an education paradigm change (Polly et al. 2010). Nonetheless, there is a lack of cross-country analyses on this issue. The aim of this study is to analyze the efficiency of education using a set of technology variables as the main drivers of this efficiency. The motivation for this study is based on various initiatives in EU countries, such as "ICT in schools" (European Commission 2013), "1 to 1" initiative (Balanskat and Garoia 2010), and the Turkish FATIH project (Education Reform Initiative 2013).

A broad body of evidence can be found for the measures of educational efficiency (Sutherland et al. 2007; Perelman and Santin 2011; Afonso et al. 2013; Bogetoft et al. 2015). Other studies focus on technology use and technology integration in education (Horzum et al. 2014; Demir et al. 2014). Some studies even combine technology drivers for measuring education efficiency (Güngören et al.2014; Harris et al. 2009). However, there is a scarcity of comprehensive input-output cross-country analyses on education efficiency, particularly those using Data Envelopment Analysis. DEA is an approach to relative efficiency measurement using multiple incommensurate inputs and outputs. If a suitable set of measures can be defined, DEA provides an efficiency measure not relying on the application of a common weighting of the inputs and outputs. Additionally, the method identifies peer units and targets for inefficient units.

The remainder of this paper is organized as follows. Section 2 explains the influence of technology on education efficiency. Section 3 elaborates on critical success factors of technology integration. Section 4 explains the methodology for the comparative study, presents the indicators for input and output variables and deals with the application of DEA in assessing input-output efficiency of secondary education efficiency. Section 5 elaborates and discusses the results, addresses theoretical contributions and recommendations, limitation of the study, further recommendation, and concluding remarks.

2. TECHNOLOGY INTEGRATION AND EDUCATION EFFICIENCY

Technology use in education has received an immense attention nowadays. Koc (2005) encompasses current research in two opposing ideological streams. The first stream is based on modernistic optimism, where technology is seen as a paramount factor that will change the paradigm of teaching and learning. This approach has been advocated by many scholars arguing that technology can improve education (Jonassen 1995; Hastings and Tracey 2004). Alongside the empirical evidence, a wide set of theories, such as the objectivist approach, constructivist approach, cooperative model, cognitive information processing model, and sociocultural model of learning (Leidner and Jarvenpaa 1995), support the thesis that technology can boost education. This optimism fuels up the idea of complete replacement of traditional lecture delivery. Thus, they try to path the way for the expansion of the walls of the physical classroom aiming to create a "virtual" space for learning (Dickson and Segars 1999).

Postmodernists do not intend to neglect the importance of technology, but rather accept the idea that there is no one-way-fits-all approach. For instance, some studies find that electronic textbooks may facilitate learning (Maynard and Cheyne 2005), while the others find no statistical differences in students' learning when they use electronic and paper versions of textbooks (McFall 2005). Moreover, postmodernism questions the rationale of colossal investments in education technology. Education institutions have spent large funds on technology in the last decade, but that technology has not been effectively integrated in curricula (Keengwe et al. 2008). The main theory supporting postmodernism is critical theory, proponing that technology is an ambivalent process of development offering different possibilities (Feenberg 1991).

This study is based on a postmodernism idea. Namely, we assume that technology affects education efficiency, following the findings of positive effects of technology on both teachers' and students' attitudes, such as computer anxiety, perceived importance of computers, and computer enjoyment (Christensen 2002), and positive impact of technology integration on the quality of education, teaching and learning (Trotter 1997). In the pursuit for education improvement, it is not only technology that is needed in the classroom, but a whole set of other social, economic, contextual and instructor-based factors need to be considered.

3. CRITICAL SUCCESS FACTORS OF TECHNOLOGY INTEGRATION

Although technology facilitates learning and teaching, there is still an ongoing debate on the key factors of effective technology use in education. Technology can positively affect education efficiency when the right conditions are met (Honey et al. 2000). These critical success factors (CSFs) are defined in a broader literature as "those few key areas of activity in which favorable results are absolutely necessary" for stakeholders to reach their goals (Rockart 1982). From a micro-perspective, Volery and Lord (2000) conducted a survey among 47 students enrolled in Global Business, and reported that technology, instructor characteristics and student characteristics are the main drivers of efficient online education. Using a confirmatory factor modeling approach, Selim (2007) draws on eight factors that assist in efficient technology integration: instructor characteristics (attitude towards and control of the technology, and teaching style), student characteristics (computer competency, interactive collaboration, and e-learning course content and design), technology (ease of access and infrastructure), and support.

A stream of research has been investigating CSFs from an institutional perspective, as well. For instance, Papp (2000) developed a new set of CSFs: intellectual property, suitability of the course for e-learning environment, building the e-learning course, e-learning course content, e-learning course maintenance, e-learning platform, and measuring the success of an e-learning course. On the other side, Soong et al. 2001 used a multiple case-study approach and found that the most appropriate set of institution-based critical success factors are: human factors, technical competencies, mindset, level of collaboration, and IT infrastructure and technical support. Nevertheless, a little has been done in defining a set of system-level critical success factors. Accordingly, we developed a set of critical success factors on the system level. They encompass technology usage, social and economic conditions, and instructor-based factors.

The first CSF of education efficiency is general technology use by students. Numerous empirical findings confirm that the use of technology facilitates learning in schools, improved writing, mathematical and technical skills of students (Lei and Zhao 2008; Holcomb 2009). However, there are voices of critiques towards the technology education utopia epitomized in Cuban (2001). Even though the effects of technology use on learning results still ambiguous, this study assumes that availability of internet to students, availability of ICT to students, and use of school computers for learning purposes.

Economic, social and contextual factors play an important role in integrating teaching and technology (Koehler and Mishra 2009). This is to some extent more indirect influence compared to technology-learning scheme. By investing into the tech classrooms, administrators are investing into the future of education and provide students with knowledge and skills needed for their communities (Erol et al. 2016). However, the voices of scholars and practitioners are not as unanimous as one might think. Some studies find no statistical significance in the number of students per computer in schools and traditional educational outcomes (Alspaugh 1999) and claim that technology-based reforms of education are simply a well-travelled myth

(Weston and Bain 2010). In spite of this claims, the study assumes that larger capital investments and expenditures in secondary education will affect technology use and consequently learning outcomes.

Even though the policies are widespread, technology integration is argued to be based on the personal characteristics of instructors rather than on policies and strategies (Pierson 2001; Cox and Graham 2009). Thus, different teacher demographic characteristics, such as years of education and previous experiences with technologies, are an important determinant of successful implementation of integration and the efficiency of education. This has been examined in a lot of studies including the path models of Van Braak et al. (2004), and Inan and Lowther (2010). For instance, Volery and Lord (2000) states that „students are more likely to experience positive learning outcomes if their classroom instructor has a positive attitude towards distributed learning and promotes the technology". Following the aforementioned, the study assumes that instructor-based factors play an important role in technology integration and education efficiency.

Table 1. Input Variables, Indicators, Explanations, and Data Sources

Inputs	Indicator	Brief definition	Source
Technology use			
Availability of internet to students (I1)	Students per internet-connected computer	Number of students per internet connected computer	European Commission Survey of Schools: ICT in Education
Availability of ICT to students (I2)	Availability of ICT to students and teachers in classrooms	Percentage of students in schools where both teachers and students use ICT equipment in lessons	European Commission Survey of Schools: ICT in Education
Use of school computers for learning purposes (I3)	Use of school computers for learning purposes weakly	Percentage of students declaring they use a school computer/laptop for learning purposes during lessons at least weekly	European Commission Survey of Schools: ICT in Education
Social and economic factors			
Capital expenditure (I4)	Capital expenditure as percentage of total expenditure in secondary public institutions	Percentage of total expenditure in secondary public institutions	OECD database
Government expenditure (I5)	Government expenditure per lower secondary student	Government expenditure per lower secondary student as % of GDP per capita (%)	UNESCO database
Students to teaching staff ratio (I6)	Ratio of students to teaching staff in educational institutions in upper secondary education	The number of students (in full-time equivalent) to the number of teachers (in full-time equivalent)	UNESCO database
Instructor-based determinants			
ICT based activities (I7)	Teachers' ICT based activates frequency with the class	The frequency of ICT-based activities ranging from 1 to 4 (1 meaning never or almost never; 2: several times a month; 3: at least once a week; and 4: every day or almost every day)	European Commission Survey of Schools: ICT in Education
ICT usage in lectures (I8)	Teachers' use of ICT in more than 25% of lessons	Percentage of teachers using ICT in more than 25% of lessons	European Commission Survey of Schools: ICT in Education
ICT knowledge and skills (I9)	Teacher's compulsory participation in ICT training	The percentage of students taught by teachers for whom participation in ICT training is compulsory	European Commission Survey of Schools: ICT in Education

4. METHODOLOGY

4.1. Methods for comparative analysis

The study is based on generic quality criteria for cross-country comparison (Cacace et al. 2013). Firstly, we developed a set of variables based on postmodernism critical theory of technology integration in education thoroughly discussed in a previous section. Secondly, we carefully selected the comparator countries in order to tackle the aim of the study. The rationale for the selection of EU and non-EU countries is twofold. Analyzed countries have similarities in their aims regarding the technology use in education, but they also have differences. For instance, some countries included technology as a compulsory in a teaching program, but there are no minimum standards across the continent (Yildirim and Göktas 2007). On the other side, these countries share similar openness and disclosure of data on education efficiency (PISA test). The data is comparable in terms of indicators and time horizon. It should be noted that the number of countries is lower than the number of European countries participated in PISA, as some input measures were not available.

Thirdly, the study was based on the quantitative analysis, as it compared multiple countries. A case study approach would be appropriate for comparison of two countries, a mixed method for a few countries, but quantitative approach is the most valuable for the examination of differences among a broad set of countries

(Milosavljević et al. 2016). Fourthly, the study was based on thoroughly selected variables and their measures. All the variables used for comparison were complex and reflected the phenomenon of technology use, social and economic, and instructor-based determinants for the efficiency of education outputs, thus driving the attention to the complexity of comparison. Fifthly, we used accurate data from reliable sources. The most important sources for the formulation of measures were from peer-reviewed assessments. The data depicting these measures were secondary and retrieved from relevant and unambiguous sources, such as the OECD education efficiency, European Commission data-sets, and UNESCO databases. Moreover, we provided the explanation for any potential weakness of the data used in the study, as they may potentially incorporate pitfalls (Atkinson and Brandolini 2001). Finally, the cross-country comparison provided useful data for the implications and recommendations for policy holders and other researchers, thus contributing to the existing body of knowledge in the field of education efficiency.

4.2. Measures

Education efficiency was the output in this study. The metrics for the efficiency was rewarded from the Program for International Student Assessment (PISA) 2012, which measures the efficiency of 15-year olds scholastic performance in reading, mathematics, problem solving and science. This metrics has already been used in input-output efficiency analyses (Afonso and Aubyn 2006). It should be noted that the results for specific categories are highly correlated. However, we used all three measures as Benefit of Doubt DEA is not affected by any multicollinearity. The results are displayed in the last three columns of Table 1. Input measures and indicators were developed according to the previous literature review on technology integration critical success factors of education efficiency. Some of these measures, such as socio-economic ones, were used in other studies (Hanushek and Luque 2003; Hanushek 2003). The list of measures and indicators, their brief explanation and the source of data are presented in Table 1.

This study examined the education efficiency using three groups of inputs – technology use, social and economic factor and instructor-based indicators. The results for the inputs are presented in Table 2.

Table 2. Inputs for the education efficiency analysis

		Inputs									**Outputs**		
		Technology use			**Social and economic factors**			**Instructor based indicators**			**Math**	**Reading**	**Science**
No.	**Country**	I1	I2	I3	I4	I5	I6	I7	I8	I9			
1	Austria	37	84	40	2.2	29.8	8.2	1.89	22	1	506	490	506
2	Belgium	63	63	47	2.7	35.3	9.2	1.8	36	12	515	509	505
3	Bulgaria	125	47	71	4.6	22.6	13.4	1.76	26	49	439	436	446
4	Cyprus	40	14	68	6.3	38	9.7	1.79	52	26	440	449	438
5	Czech Republic	33	88	62	5.6	25.9	11.3	1.94	36	36	499	493	508
6	Denmark	4	89	70	9	23.8	14	2.11	40	47	500	496	498
7	Estonia	18	64	51	10.4	26.1	7.9	2.16	57	28	521	516	541
8	Finland	24	57	27	10.4	32.2	90	1.85	29	22	519	524	545
9	France	63	66	41	8.8	24.8	14.7	1.82	39	21	495	505	499
10	Hungary	18	49	66	5.6	19.2	9.7	1.89	31	32	477	488	494
11	Ireland	22	50	54	4.3	23.8	14	1.98	60	13	501	523	522
12	Italy	125	79	46	1.6	23.7	10.8	1.76	33	4	485	490	494
13	Latvia	45	51	45	11.8	16	7.4	2.17	39	53	491	489	502
14	Lithuania	48	55	66	8.9	66	7.8	2.27	29	74	479	477	496
15	Norway	4	87	58	13.4	20.1	8.5	2.21	31	41	489	504	495
16	Poland	63	45	65	4.8	24	9.1	1.89	19	16	518	518	526
17	Portugal	43	36	50	1.6	31.4	10	1.95	66	49	487	488	489
18	Romania	250	65	50	7.2	12.2	4	2.03	35	56	445	438	439
19	Slovakia	34	77	69	2.9	16.9	11.2	2.11	48	20	482	463	471
20	Slovenia	37	76	31	5.9	32.6	11	1.89	40	44	501	481	514
21	Spain	7	74	52	13.3	18.9	7.3	1.96	27	17	484	488	496
22	Sweden	2	91	55	7.8	26.4	11.2	1.93	40	38	478	483	485
23	Turkey	167	40	66	11.3	24.4	9.5	2.14	60	36	448	475	463

4.3. The "Benefit-of-Doubt" DEA Model

DEA is a semi-parametric linear programming technique used for the assessment of a set of Decision Making Units (Charnes et al.1978). It has already been used in the examination of education efficiency. However, current studies use other variables as inputs or outputs and put completely different spotlight to these issues (Kirjavaine and Loikkanent 1998; Abbott and Doucouliagos 2003; Afonso and Aubyn 2006; Johnes 2006). The custom-made model used for this study is the "Benefit-of-Doubts" (BoD) DEA model developed by Melyn and Moesen (1991). The model is as follows (Savić and Martić, 2016):

$$(\max) h_k = \sum_{r=1}^{s} u_r y_{rk} \quad (1)$$

s.t.

$$\sum_{r=1}^{s} u_r y_{rj} \leq 1, \quad j = 1, \ldots, n \quad (2)$$

$$u_1, \ldots, u_s \geq 0 \quad (3)$$

This model provided the most favourable aggregated performance score for each DMU_k (k=1,…,n) in terms of all used component indicators. The resulting performance score h_k identifies the closeness of observed DMU_k to its best practice. This is the best-fit model for a two-stage analysis of education efficiency as it constructs a multidimensional measure of efficiency (Verschelde and Rogge, 2012). Accordingly, we analyzed inputs and outputs in two phases. The aim of the first phase was to determine the efficiency of education relative to the three groups of measures (technology usage, social and economic conditions, instructor-based factors). These indices are based on three models in which outputs remain the same, while inputs are varied. The model is displayed in Figure 1. The efficiency scores (E1-E3) were used in the second phase for the construction of a composite index of the relative efficiency of education. Thus, the model incorporates multiple inputs and a single output is fixed for the same value for each of the examined countries.

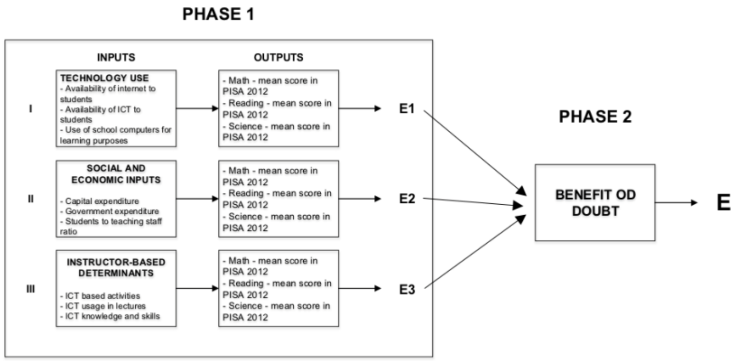


Figure 1. Benefit of Doubt DEA Model for Technology-Based Education Efficiency

4.4. Results

As the results of employed efficiency analysis indicate, the most important determinants for the education efficiency are instructor-based inputs (average efficiency=0.896), followed by technology usage, while social and economic inputs are the least influential (average efficiency=0.728). The results of this phase are displayed in Table 4. It is interesting to note that no country has the highest score for all three indices obtained in the first stage. Hypothetically, such a country would be a role model. In the second phase, efficiencies obtained are used as outputs in BoD DEA model (1-3) for each country. The results are displayed in last two columns of Table 4.

All countries are grouped in five tiers. Tier 1 countries are the ones with an overall score of 1.00 (Poland, Italy, Belgium, Ireland and Finland). These countries have the highest efficiency scores with the fewest resources invested; their secondary education systems generate the highest results. As aforementioned, there is no single country with top scores in all categories. Accordingly, tier one countries will serve as role models for the study. Tier 2 countries have average score from .95 to .99 (Slovakia, France, Portugal, Austria, Slovenia and Bulgaria). Tier 3 countries range from 0.9 to .95 (Czech Republic, Hungary, Latvia, and Estonia). Tier 4 includes countries with the score ranging from .80 to .90 (Spain, Sweden, Cyprus, Denmark and Norway). An interesting fact is that this tier includes majority of North European countries. For the given amount of investment, their secondary education has very poor outcomes. Finally, three countries are listed as tier 5 countries (Romania, Lithuania and Turkey). These countries should improve their resource base for the secondary education in order to improve outcomes.

5. DISCUSSION AND CONCLUSION

5.1. Summary of key findings

The aim of this paper was to analyze the efficiency of secondary education using two-phase DEA approach. For the outputs, we selected student performance across the selected countries (PISA results) and compared them against the inputs structured as technology use variables, socio-economic variables, and instructor-based variables as critical success factors of technology integration.

Table 4. Efficiency indicators, ranks and average weights indices

No.	Country	Technology usage	Social and economic factors	Instructor based indicators	Efficiency Score	Rank
1	Austria	0.923	0.682	0.833	0.966	9
2	Belgium	0.704	0.679	0.814	1.000	1
3	Bulgaria	1.000	0.487	0.781	0.950	11
4	Cyprus	0.959	0.701	0.851	0.842	18
5	Czech Republic	0.979	0.580	0.873	0.935	12
6	Denmark	1.000	0.737	0.951	0.839	19
7	Estonia	0.728	0.882	1.000	0.910	15
8	Finland	0.626	0.882	0.824	1.000	1
9	France	0.727	0.731	0.828	0.973	7
10	Hungary	0.932	0.587	0.844	0.932	13
11	Ireland	0.765	0.471	0.960	1.000	1
12	Italy	0.868	0.527	0.793	1.000	1
13	Latvia	0.641	0.923	0.975	0.924	14
14	Lithuania	0.933	1.000	1.000	0.760	22
15	Norway	0.963	1.000	0.978	0.812	20
16	Poland	0.916	0.592	0.833	1.000	1
17	Portugal	0.705	0.616	1.000	0.971	8
18	Romania	0.729	1.000	0.910	0.775	21
19	Slovakia	0.982	0.459	0.964	0.986	6
20	Slovenia	0.835	0.628	0.860	0.966	10
21	Spain	0.823	1.000	0.867	0.887	16
22	Sweden	1.000	0.680	0.876	0.883	17
23	Turkey	0.930	0.891	1.000	0.756	23
	Average	**0.855**	**0.727**	**0.896**		

Countries are ranked based on their overall efficiency in the use of resources, or the factors that drive secondary education efficiency. The most important factor driving secondary education efficiency are instructor-based factors, followed by technology use. This implies that any national policy should encompass not only investments in technology, but a combination of these investments and improvements in instructors' soft skills and attitudes towards the technology use. All countries are listed in five tiers. Observed through the prism of efficiency, the results significantly differ to PISA outcomes. The results indicate that no country is the best measured by all aspects. However, tier 1 countries can serve as role models.

5.2. Strengths, limitations and further recommendations

The major strength of this study is the use of a variety of critical success technology factors as inputs for the assessment of education efficiency. New variables – technology use, socio-economic and instructor-based inputs – were applied in calculus which shed a new light to a cross country education efficiency research. The variables were drawn from the existing scholarly and practical publications. This implies that the list is not comprehensive and can be filled with additional critical success factors that could potentially reshape the results. Also, the study was based on a secondary data, which tend to input add methodological omissions to the efficiency calculation. For instance, PISA "effects" have been questioned for their impact (Grek 2009), and more generally aggregated indicators were criticized for their utility and validity (Langbein and Knack 2010; Thomas 2010). Additional data-specific ambiguity is related to their level of aggregation. The data are system-level, highly aggregated and represent approximations, rather than specific measures. Accordingly, further studies should be based on more specific, education-centric and primary data. Another strength of the study is that it actually analyses the input-output efficiency of education, using a quantitative approach. Current rankings are based on outputs (such as PISA), thus exposing only a portion of this phenomenon and leaving a lot of interrogatives unfolded. However, some pitfalls are connected to this approach. Firstly, this study is cross-sectional and does not use time-series. The use of time-series would capture on evolutionary characteristics of technology factors of education efficiency. Being quantitative by nature, this study lacks of the qualitative comments, which would contribute to the thorough examinations and exploration of the secondary education efficiency phenomenon. Further studies should incorporate time-series and qualitative in-depth comments on technology use in education.

5.3. Implications

Secondary education efficiency is high on the agenda of policy holders, decisions makers, scholars and other interested parties. A vital step for the overall improvement in education efficiency should be based on further technological, social, economic, and instructor-based factors of selected countries. Although the technology integration, and economic investments, are is recognized as a high priority, other determinants should also be within the scope of policy holders. The findings of this study suggested that the analyzed countries significantly differ regarding the efficiency of secondary education. However, seen through the prism of input-output analysis, the rankings give a new perspective. Tier 1 countries (Poland, Italy, Belgium, Ireland and Finland) can serve as a role model country, as they have the highest results for the resources invested in secondary education. Other countries, particularly those in the last tier (Romania, Lithuania and Turkey) should invest heavily into the drivers of education efficiency. Although there are critiques directed to the large-scale technology integration projects in other places in the world, Turkey (and other Tier 5 countries specified in this study) should remain on the course of extensive investments in secondary education.

5.4. Conclusion

The findings of this study illustrate the efficiency of secondary education in 23 European countries. Using an input-output analysis based on DEA, we ranked the countries in terms of their secondary education efficiency. In a methodological term, we used two-staged semi-parametric procedure. After delivering results and scores for each country (DMUs), we divided them in five tiers. This classification significantly differs to any other study of secondary education efficiency. Results also indicate that technology integration, social and economic background, and instructor-based factors play an important role in education efficiency.

Secondary schooling is a vital part of economic and societal growth. Accordingly, all the resources consumed by secondary education should lead to skill-based technology change in these countries. This study adds to these policy-based comments by providing efficiency-based ranking and directing them to the most important factors that need to be "pushed" in order to improve the educational and overall performances.

REFERENCES

Abbott, M., & Doucouliagos, C. (2003). The efficiency of Australian universities: a data envelopment analysis. Economics of Education review, 22(1), 89-97.

Afonso, A., & Aubyn, M. S. (2006). Cross-country efficiency of secondary education provision: A semi-parametric analysis with non-discretionary inputs. *Economic modelling*, *23*(3), 476-491.

Afonso, A., Ayadi, M., & Ramzi, S. (2013). Assessment of Efficiency in Basic and Secondary Education in Tunisia: A Regional Analysis. SSRN Electronic Journal.

Alspaugh, J. W. (1999). The relationship between the number of students per computer and educational outcomes. *Journal of educational computing research*, *21*(2), 141-150.

Atkinson, A. B., & Brandolini, A. (2001). Promise and pitfalls in the use of" secondary" data-sets: Income inequality in OECD countries as a case study. *Journal of economic literature*, *39*(3), 771-799.

Balanskat, A., & Garoia, V. (2010). Netbooks on the rise: European overview of national laptop and netbook initiatives in schools. *European Schoolnet*.

Blundell, R., Dearden, L., Meghir, C., & Sianesi, B. (1999). Human capital investment: the returns from education and training to the individual, the firm and the economy. *Fiscal studies*, *20*(1), 1-23.

Bogetoft, P., Heinesen, E., & Tranæs, T. (2015). The efficiency of educational production: A comparison of the Nordic countries with other OECD countries. Economic Modelling, 50, 310–321.

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, *2*(6), 429-444.

Christensen, R. (2002). Effects of technology integration education on the attitudes of teachers and students. *Journal of Research on technology in Education*, *34*(4), 411-433.

Cox, S., & Graham, C. R. (2009). Using an elaborated model of the TPACK framework to analyze and depict teacher knowledge. *TechTrends*, *53*(5), 60-69. ISSN 1436-4522 (online)

Cuban, L. (2009). *Oversold and underused: Computers in the classroom*. HU Press. ISBN 9780674011090

De la Fuente, A., & Ciccone, A. (2002). *Human capital in a global and knowledge based economy*. Institute for Economic Analysis.

Demir, Ö., Yasar, S., Sert, G., & Yurdugül, H. (2014). Examination of the Relationship between Students' Attitudes towards Computer and Self-Directed Learning with Technology. *Egitim ve Bilim*, *39*(176).

Erol, M., Özcan, A., & Luft, J.A. (2016). Lisans Derslerini SCALE-UP Sınıflarında Yürütmek İçin Altı Sebep: Yükseköğretim Otoriteleri ve Öğretim Elemanlarına Öneriler. *Egitim ve Bilim 41*(184).

European Commission. (2013). ICT in schools survey – many children not getting what they need; teachers need more training and support. Brussels: European Commission, *Press release*.

Feenberg, A. (1991). *Critical theory of technology*. Oxford, UK: Oxford University Press. ISBN: 978-1-405-14601-2

Güngören, Ö.C., Bektas, M., Öztürk, E., & Horzum, M.B. (2014). Acceptance of TPC Scale-Validity and Reliability Study. *Egitim ve Bilim*, *39*(176).

Hanushek, E. A. (2003). The Failure of Input-based Schooling Policies*. *The economic journal*, *113*(485), F64-F98.

Hanushek, E. A., & Luque, J. A. (2003). Efficiency and equity in schools around the world. *Economics of education Review*, *22*(5), 481-502.

Harris, J., Mishra, P., & Koehler, M. (2009). Teachers' technological pedagogical content knowledge and learning activity types: Curriculum-based technology integration reframed. Journal of Research on Technology in Education, 41(4), 393-416.

Hastings, N. B., & Tracey, M. W. (2004). Does media affect learning: where are we now?. *TechTrends*, *49*(2), 28-30.

Holcomb, L. B. (2009). Results & lessons learned from1: 1 laptop initiatives: A collective review. *TechTrends*, *53*(6), 49-55. ISSN: ISSN-8756-3894

Honey, M., Culp, K. M., & Carrigg, F. (2000). Perspectives on technology and education research: Lessons from the past and present. *Journal of Educational Computing Research*, *23*(1), 5-14.

Horzum, M. B., Öztürk, E., Bektas, M., Güngören, Ö. C., & Çakir, Ö. (2014). Secondary School Students Tablet Computer Acceptance and Readiness: A Structural Equation Modelling. *Egitim ve Bilim*, *39*(176).

Inan, F. A., & Lowther, D. L. (2010). Factors affecting technology integration in K-12 classrooms: A path model. *Educational Technology Research and Development*, *58*(2), 137-154.

Johnes, J. (2006). Data envelopment analysis and its application to the measurement of efficiency in higher education. *Economics of Education Review*, *25*(3), 273-288.

Jonassen, D. H. (1995). Supporting communities of learners with technology: A vision for integrating technology with learning in schools. *Educational technology*, *35*(4), 60-63.

Keengwe, J., Onchwari, G., & Wachira, P. (2008). Computer technology integration and student learning: Barriers and promise. *Journal of Science Education and Technology*, *17*(6), 560-565. ISSN 1059-0145

Kirjavainen, T., & Loikkanent, H. A. (1998). Efficiency differences of Finnish senior secondary schools: an application of DEA and Tobit analysis. *Economics of Education Review*, *17*(4), 377-394.

Koc, M. (2005). Questioning technology use in educational reform: From ideological, theoretical and practical perspectives. *Malaysian online journal of instructional technology*, *2*(2), 72-81. ISSN: 1823-1144

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)?. *Contemporary issues in technology and teacher education*, *9*(1), 60-70.

Langbein, L., & Knack, S. (2010). The worldwide governance indicators: six, one, or none?. *The Journal of Development Studies*, *46*(2), 350-370.

Lei, J., & Zhao, Y. (2008). One-to-one computing: What does it bring to schools?. *Journal of Educational Computing Research*, *39*(2), 97-122.

Leidner, D. E., & Jarvenpaa, S. L. (1995). The use of information technology to enhance management school education: A theoretical view. *MIS quarterly*, 265-291.

Maynard, S., & Cheyne, E. (2005). Can electronic textbooks help children to learn?. *The Electronic Library*, *23*(1)

McFall, R. (2005). Electronic textbooks that transform how textbooks are used. *The electronic library*, *23*(1), 72-81.

Melyn, W., & Moesen, W. (1991). Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available. *Public Economics Research Papers*, 1-24.

Milosavljević, M., Milanović, N., & Benković, S. (2016). Politics, policies and public procurement efficiency: A quantitative study of 25 European countries. *Lex Localis - Journal of Local Self-Government*, 14(3).

Papp, R. (2000). Critical success factors for distance learning. *AMCIS 2000 Proceedings*, 104.

Perelman, S., & Santin, D. (2011). Measuring educational efficiency at student level with parametric stochastic distance functions: an application to Spanish PISA results. *Education Economics*, *19*(1), 29-49.

Pierson, M. E. (2001). Technology integration practice as a function of pedagogical expertise. *Journal of Research on Computing in Education*,*33*(4), 413-430.

Polly, D., Mims, C., Shepherd, C. E., & Inan, F. (2010). Evidence of impact: Transforming teacher education with preparing tomorrow's teachers to teach with technology (PT3) grants. Teaching and Teacher Education, 26(4), 863-870. ISSN 0742-051X

Pouezevara, S., Dinçer, A., Kipp, S., & Sariisik, Y. (2003). Turkey's FATIH project: A plan to conquer the digital divide or a technological leap of faith. *Turkey: RTI International & Education Reform Initiative (ERI)*.

Rockart, J. F. (1982). *The changing role of the information systems executive: a critical success factors perspective*. Boston: Massachusetts Institute of Technology.

Savić, G., & Martić, M. (2016). Composite Indicators Construction by Data Envelopment Analysis: Methodological Background (chapter 5) in Emerging Trends in the Development and Application of Composite Indicators“, eds. Jeremic V., Dobrota, M., & Radojicic, Z. IGI Global, pp. 98-126, ISBN: 978-1-53610-566-7.

Soong, M. B., Chan, H. C., Chua, B. C., & Loh, K. F. (2001). Critical success factors for on-line course resources. *Computers & Education*, *36*(2), 101-120.

Sutherland, D., Price, R., Joumard, I., & Nicq, C. (2007). Performance indicators for public spending efficiency in primary and secondary education. *OECD Economics Department Working Paper*, 546.

Thomas, M. A. (2010). What do the worldwide governance indicators measure & quest. *European Journal of Development Research*, *22*(1), 31-54.

Trotter, A. (1997). Taking technology’s measure. *Education Week*, *17*(11), 6-11.

Van Braak, J., Tondeur, J., & Valcke, M. (2004). Explaining different types of computer use among primary school teachers. *European Journal of Psychology of Education*, *19*(4), 407-422.

Verschelde, M., & Rogge, N. (2012). An environment-adjusted evaluation of citizen satisfaction with local police effectiveness: Evidence from a conditional Data Envelopment Analysis approach. *European Journal of Operational Research*, *223*(1), 214-225.

Volery, T., & Lord, D. (2000). Critical success factors in online education. *International Journal of Educational Management*, *14*(5), 216-223.

Weston, M. E., & Bain, A. (2010). The end of techno-critique: The naked truth about 1: 1 laptop initiatives and educational change. *The Journal of Technology, Learning and Assessment*, *9*(6). ISSN: 1540-2525

Yildirim, Z., & Göktas, Y. (2007). ICT Integration in primary education and teacher education programs in Turkey and in EU countries. *Egitim ve Bilim*, *32*(143), 55.

BALCOR 2018
XIII Balkan Conference on Operational Research

ASSESSMENT OF EUROPEANS NRENS SERVICE EFFICIENCY

MARA BUVIĆ[1], GORDANA SAVIĆ[2]

[1] University of Belgrade, Computer Centre, mara@rcub.bg.ac.rs
[2] University of Belgrade, Faculty of Organizational Sciences, gordana.savic@fon.bg.ac.rs

Abstract: *NREN (National Research and Education Network) is a general term for a unique national level organization and, based on its non-profit grounds, information and communication infrastructure and services being organized for all academic and research institutions in one country, as well as for its education sector, libraries, museums, etc. This paper describes the starting analysis aiming to quantify NREN's efforts to provide various internet services for its community. Data Envelope Analysis (DEA) was applied for 38 European NREN's gathered around common regional organisation GÉANT, by splitting data set in two parts. Though, the results presented are for the part consists of 20 NREN, who perform services mainly using their own resources and cover costs from their own businesses. The aim of the study is twofold: first, to gain insight in the suitability of DEA method to analyse NRENs and second, to examine, for further analyses, the availability and quality of data on Europeans NRENs activity collected yearly by GÉANT.*

Keywords: *IT Service Efficiency, DEA, Data Envelope Analysis, National Research and Education Network.*

1. INTRODUCTION

All developed and most of the growing world economies organise the information and communication infrastructure and services for their education and research institutions through a NREN (National Research and Education Network). NREN is a term which, depending on the context, determines an organisation that provides for the functioning of the national research and educational network, but also the network itself. Some of the more prominent NRENs in the world and the region are Internet2 in the USA, CERNET in China, GRNET in Greece, AMRES in Serbia, Hungarnet in Hungary, etc. National networks are being joined together in larger regional networks such as GÉANT in Europe (G4-1. 2015-2016).

By 2010, around 62 percent of countries in the world established their NRENs (ITU, 2010). Their similarities mostly lie in their unique position and non-profit character, i.e. as a rule, each country has only one NREN, and most of them are operating on a non-profit base. However, NRENs are constituted differently, sometimes as a consortium of universities and/or its member institutions, sometimes as a company, agency, non-government organisation, etc. (The Case for NREN, 2009-2017) (Foley, M. 2017) It is not possible to determine a unique organisation frame by which the existing NRENs operate. It is chosen for and adjusted to the needs of the members and/or founders and it depends on NREN's roles, i.e. what is expected from a NREN. The basic expectation is for a NREN to offer more bandwidth that costs less, so a minimum degree of authorisation is given to a NREN in order to organise the internet lease from an internet service provider (i.e commodity service), which is more cost effective. (Foley, M. (from World Bank)) A more complex authorisation degree is given to NREN which is expected to provide the infrastructure in order to connect its members (connectivity service). The most complex organisation degree is when a NREN is required to provide a whole range of other services, such as web hosting, grid or cloud computing, Authentication and Autorization Infrastructure (AAI), distance learning, data science, learning management platform (LMS), access to digital libraries and e-magazines, etc. With the widest range of authorisation, NREN also provides services to other institutions, such as schools, libraries, museums, hospitals.

NREN's tasks evolve and can be redefined, not only to facilitate scientific discoveries and expansion of knowledge, but also to help develop human capital within the economic development of a society. (ITU. (2014).) In developed countries/economies, the development of NREN's has been moving to its more advanced version for some time now, i.e. toward organising and providing a wide range of IT services. This is why the abbreviation NRENv2 (NREN version 2) was coined, in order to distinguish the difference in the degree of development.

The significance of choosing the management and organisation model grew with the introduction of NRENs version 2. According to the ERNEST study (Robin, Arak, 2007), the more advanced versions of NRENs must continue to innovate, develop and support other so called value-added services besides providing cost effective and reliable commodity&connectivity services. The issues of organisation and management regarding the planning and delivery of new services become exceptionally significant in the transition from the best-effort services to the guaranteed end-to-end services which can be chosen and customised according to the users' needs (Robin, Arak, 2007). Regarding the less developed NRENs which are yet to embark on the journey toward more advanced services which more developed NRENs already took or are taking at the moment, there is the question of which best practices examples to follow. Although everybody agrees that the management method can greatly influence NREN's operations, there were no attempts to measure which models, measures, plans, etc. give better or worse results.

In the literature there are studies regarding the productivity indicators assessment (innovation and catch-up with the trends) for the whole ICT (information communication technology) services sector of country, where Data Envelopment Analysis (DEA) is one of the methods used in combination with other methods. The total productivity factor expressed via Malmquist productivity index (MPI) was calculated in (Shao & Lin, 2016), using DEA to calculate the MPI index. In (Kumar et. al., 2015), the authors analysed preferences/needs of the users by employing a hybrid combination of fuzzy AHP and DEA models in order to measure relative efficiency of telecommunication providers in India. (Serrano-Cinca et al., 2005) applied the DEA in order to measure the efficiency of internet companies, the so called dot-com companies which were in expansion during the time.

Although the establishment of a wider range of services in a NREN is accompanied by numerous challenges which are not always of a strictly technical nature, there is just one paper dealing with the organisation of NRENs. (Staphorstet al, 2016) described a model for technology forecast (TF) within the technological domain of a NREN and confirmed several hypotheses by using the output indicator (utilisation of NREN backbone) of this model: first, there is a positive correlation between the capabilities of NREN infrastructure/advanced services and the degree of the government/ministry influence/control over the NREN; second, there is a positive correlation between the capabilities of NREN infrastructure/advanced services and their use. The only hypothesis which was not confirmed is the positive correlation between the NREN infrastructural possibilities and possibilities provided by advance services, which is explained by the position that technology leapfrogging (conquering the technology in great leaps, without going through all development stages sequentially) is a widespread practice in the global NREN community.

According to our best knowledge, no one else tried to analyse and quantify NRENs' efforts by any method. Since NRENs operate as non-profit organisations, we have selected the DEA method in order to examine service efficiency of European NRENs. We did the experiments by using the constant return to scale (CCR) DEA model.

At least one NREN operates in 39 European countries (ITU, 2014). There are differences between them in the way they are organised, how they're funded, in type, number as well as the quality of services they provide to their users, etc. At the moment, the widest NREN service portfolio consists of 84 services divided in several groups, i.e. Network services, Security services, Identity services, Collaboration services, Multimedia services, Storage and hosting services, Professional and ISP services. The statistics on NRENs services and organization date from 2001. TERENA organization was collecting and publishing the data on European (and other) NRENs from 2001 to 2014 (TERENA, 2001-2014). After 2014, the GÉANT organisation continued with these activities. The data are available in electronic form, while the Compendium is published annually in hard copy (GÉANT, 2008-2016).

The aim of the analysis/experiments given in the paper is twofold: first, to gain insight in the suitability of DEA method to analyse NRENs and to examine the availability and quality of data on Europeans NRENs activity provided yearly by GÉANT organisation in order to answer the dilemma whether it makes sense to continue to study the service efficiency of NRENs based on these data.

The remainder of the paper is organized as follows: Section 2 describes the DEA model that we proposed for assessment the efficiency. In Section 3 we describe the data that is used. Experimental results are presented in Section 4. Finally, in Section 5 we offer conclusions and ideas for further work.

2. METODOLOGY OF NREN'S SERVICE COMPOSITE AND EFFICIENCY INDEX ASSESSMENT

We have selected the DEA method (Charnes, Cooper & Rhodes, 1978) in order to examine service efficiency of European NRENs. As a method for performance evaluation and benchmarking against best-practice, DEA is often used for measuring relative efficiency of non for profit organizations or parts of organizations, called decision making units (DMUs) in DEA terminology. The analysis is based on DMUs' multiple input and output quantities, i.e. selected inputs and outputs. By considering inputs/outputs combination that allow each DMU to be represented in the best light, DEA constructs a non-parametric frontier/surface over the data to calculate efficiency measures relative to the surface.

Following the recommendations found in the literature (Savić, 2012; Cook et al., 2014) and given the purpose of the analysis/experiments presented in this paper, we had to consider several additional questions: ”What are the decision-making units (DMUs) and the outputs and inputs to be used to characterize the performance of those DMUs?” “What is an appropriate number of DMUs, given the number of inputs and outputs chosen?” “What is the appropriate model orientation (input, output)?” “Does the analysis involve the use of ratio and raw data in the same model, and is this appropriate?” (Cook et al., 2014).

In choosing the DMU, we opted for a NREN. The number of NRENs/DMUs included in the experiment was 38. Later these were split in two groups consisting of 21 and 17 NRENs/DMUs, respectively, where the total number of inputs and outputs did not exceed 6. The inputs and outputs considered was reduced due the data availability and the last experiment was with 3 inputs and 3 outputs, for which we had data.

In choosing the orientation model, we had no dilemma that we should choose the output oriented model which enables/analyses output maximising, which can be generated with given inputs since it would not be useful to decrease already limited resources oriented toward NREN operations. Basic DEA model is used for service efficiency analysis (how well NRENs use their resources to perform services) while modified “benefit of doubt” model is applied for composing service index as a measure of performing services and ranking NRENs.

2.1. Mathematical Model

Considering the aim of the analysis, we started from two basic DEA models, i.e. we performed the starting experiments with the constant return to scale model (CRS) and then with the variable return to scale model (VRS). The variable return to scale model BCC measures only technical efficiency, i.e. it gives the DMU (Decision Making Unit) efficiency measure which ignores the business scale influence by comparing each DMU only with other units of similar scale. We have noticed that the application of the variable return to scale model gives significantly more units that are efficient, and this is why we opted for the analysis by output CCR model which measures the total efficiency of DMU units. Dual BCC model is obtained by adding a convexity constraint to a dual form of CCR model.

Given: n, the number of countries/NRENs to assess,

m, the number of inputs, s, the number of outputs,

x_{ij}, the value of input i for country/NREN/ DMU_j $\left(x_{ij} > 0,\ i = 1..m,\ j = 1..n\right)$,

y_{rj}, the value of output r for country/NREN/ DMU_j $\left(y_{rj} > 0,\ r = 1..s,\ j = 1..n\right)$,

the CRS model assumed the output orientation is written in envelope form as follows:

$$\max\ \theta_k + \varepsilon\left(\sum_{r=1}^{s} s_r^+ + \sum_{i=1}^{m} s_i^-\right) \qquad \forall DMU_k\ (k = 1..n) \tag{1}$$

Subject to:

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = x_{ik}, \qquad i = 1,2,\ldots m, \tag{2}$$

$$\theta_k y_{rk} - \sum_{j=1}^{n} \lambda_j y_{rj} + s_r^+ = 0, \quad r = 1,2,\ldots s, \tag{3}$$

$$\theta \text{ unlimited}, \qquad s_r^+ \ge 0, \quad r = 1,2,\ldots s,\ s_i^- \ge 0, \quad i = 1,2,\ldots m, \tag{4}$$

$$\lambda_j \ge 0, \quad j = 1,2,\ldots n. \tag{5}$$

where we used following notation for variables:

θ_k: the efficiency score for the i-th country, this scalar indicates the value/level which the DMU should reach by proportionally increasing its outputs if it intent to be efficient ($1/\theta_k \le 1$ represent input-oriented efficiency index),

s_i^-: slack variable, indicates the necessary reduction of the input i of the DMU_k to become efficient,

s_i^+: slack variable, indicates the necessary increase of the output r of DMU_k in order to become efficient,

λ_j: dual weight, shows the importance assigned to DMU_j in the input /output mix of the hypothetical composite unit with which the DMU_k is directly compared, i.e. the maximum output values that can be achieved with the existing input level of the DMU_k.

The area of DEA deployment has been broadly widespread and this has led to the development of numerous models. Lately, DEA is also applied to create composite indices using the so called „benefit of the doubt“ model (Melyn & Moesen,1991; Cherchye et al, 2007). The modification of the basic DEA model is reflected in the fact that data only consists outputs (supposing that there is just one input with a value equal to 1 for all DMUs). More details are given in the book chapter by Savic & Martic (2017).

3. DATA

This study is based on the data that we were able to access and download from the GÉANT organization website https://www.geant.org/ (under the NREN Compendium). Since 2000, GÉANT and its predecessor organizations have conducted a survey to collect detailed information from NRENs about the services they provide and their organisation. The findings from a survey are published in a single reference publication, named Compendium of NRENs in Europe. The Compendium is published annually.

The data used in the study come from a survey for which all European NRENs were invited to contribute on the NREN's status as of January 2015. We noticed a significant number of received responses for that year's Compendium. The missing data for four counties were found in the survey submission for 2016 (for Sweden, Belgium and Denmark) and 2013 (for Macedonia).

Before the study, we examined the NRENs' answers for the 2012-2016 Compendium Survey in order to use them as a source of data for service efficiency and productivity analysis. It turned out that a number of NRENs which sent their responses for the GÉANT survey varies greatly from year to year. Also, there are significant differences in the number of questions answered by NRENs which filled out the survey. The data set availability became an issue as soon as the analysis started to include all European NREN members of the GÉANT consortium organization. Our conclusion was that one could expect to have the issue deepened by extending the time period of the analysis from one to two or more years.

In order to select potential inputs, data availability for the following categories was examined: budget, engaged employees and collaboration with institutions/users within an NREN. In addition, the review/check-up included the data on infrastructure quality expressed by the network backbone capacity in gigabits (BBCapacity) and the number of issued server certificates (SrvCrt). In order to select potential outputs, data availability for the following categories was examined: total traffic (Traffic), generated toward or within the network (expressed in terabytes), total number of services offered by a NREN (ServiceNo), and the number of new services planned for the next budget year.

The budget refers to NREN's total budget for 2015 (or 2014 and 2015) under the question no. 3 in the 2015 survey's funding and governance section. The budget dedicated to NREN activities during one year is expressed in millions of euros, and the share by categories (such as wages, equipment, transferable capacities, etc.) was roughly estimated. At the main exchange rate of 9.1, the US dollar is used to calculate the budgets for Albania and the United Kingdom which specified their budgets in USD. NRENs' budgets depend on the size and the wealth of the country, but also on the funding model as well. The data was not available from Poland and Malta. Estimation of the NREN-related income sources (question no. 10 in funding and governance) shows that some NRENs receive all their funding from the national government and public sources, while other NRENs are funded entirely by their user institutions. Although, the most common funding models are those which combine these two approaches and use more than one source of NREN-related income.

The staff (designated as StaffFTE): The data on the employees is available in two forms: as employees' share in the budget and as the equivalent of employing one person per year, i.e. FTE (full time equivalent) which is used to calculate the work force for the GÉANT project. We opted for the data on the engagement expressed by the FTE. The StaffFTE input (see Table I) shows the total work force engaged full time or part time on activities directly related to networking for research or education, regardless of whether that staff is directly employed by the NREN organization or employed by others i.e. outsourced (question no. 26 under in the 2015 survey's funding and governance section). The survey enables an NREN to provide a breakdown of the staff in six categories, i.e. to NOC (Network operation centre) staff, other technical staff, administrative and financial staff, user support and training staff, information security staff and others not included elsewhere. Data for Poland and Malta was not available.

NRENs' infrastructure quality is described by two variables, the network backbone capacity (designated as BBcapacity) managed by NREN, and the number of server certificates (designated as SrvCertif) issued by the NREN to be installed on servers within the NREN. Question no. 11 in the 2015 survey's connectivity and traffic section refers to the backbone capacity and shows the NREN's current typical core/mean usable backbone capacity expressed in gigabits per second (Gbit/s). The NRENs which have dark fibre with a very high theoretical capacity were asked to report the usable IP capacity of its current typical core backbone. The

NRENs operating its network in star topology were asked to provide information on the maximum capacity of the network's central node. The number of servers in an NREN is estimated by the number of server certificates issued by the NREN, supposing that all important servers within the NREN should be protected by installing a security certificate on it. Questions no. 26 and 27 in the 2015 survey's middleware section refer to the variable SrvCertif and show the number of valid server certificates outstanding on December 31, 2014 (including the certificates added in 2014).

In our model, the backbone capacity, i.e. the BBcapacity variable, was used as an output, whereas the SrvCertif had been taken as an input. The following was the reason behind this decision: the backbone capacity is directly related to the available budget, i.e. in the long term it depends on NREN ability to invest in the development/improvement of its physical infrastructure. On the other hand, the number of servers in a network depends on investments from both the NREN and its client institutions, thus one can expect that the number of servers increases with either the number of client institutions joining NREN or the number of services in NREN's service portfolio. To a great extent, servers influence the total traffic in a network which described as the next variable.

Traffic quantities are designated as TotalTrafficINOUT and TTrafficPCapita variables. Question no. 21 in the 2015 survey's connectivity and traffic section refers to the TotalTrafficINOUT variable and shows the total IPv4 traffic during 2014 in TByte. The survey proposed a distinction between different types of traffic based on the source or destination and by IPv4 or IPv6 address stack. We only considered IPv4 traffic whether generated from/send to NREN customers as a part of the remit of the NERN's domain (named "customer domains traffic") or generated toward outside NREN's domain /received from external domains, such as GÉANT, general/commercial Internet, Internet exchange, peering, other NERNs, etc. (called external "domains traffic"). The TotalTrafficINOUT variable was calculated as a sum of the customer domains traffic and the external domains traffic where both data were available. Otherwise, it was calculated as a double value of traffic for domains for which NRENs provide information. Based on an estimation of the total amount of Layer 3 traffic (including possible VPNs) in TByte, and knowing the estimation for total population in millions (available at the wiki address https://en.wikipedia.org/wiki/List_of_European_countries_by_population), we estimated the traffic per capita in megabits, i.e. the values of the TTrafficPCapita output.

The service variable, used as an output, refers to a total number of services offered by an NREN in its service portfolio, where ServiceWeighted is its weighted version. The questions about services had been spread over different sections of the survey. Thus, we decided to follow the matrix of services from the on line report available at https://compendiumdatabase.geant.org/reports/nrens_services to identify services implemented and planned to be implemented in each of the 43 NRENs. The complete NREN service portfolio covered by the report consists of 84 services divided in several groups, i.e. Network services, Security services, Identity services, Collaboration services, Multimedia services, Storage and hosting services, Professional and ISP services. For each of these services, their weight was calculated as the ratio of the number of NRENs not implementing the service among the 43 NRENs and the total number of NRENs in GÉANT. Thus, the lower weight was assigned to the most common services available in a larger number of NRENs, intending to positively discriminate emerging services.

4. RESULTS AND DISCISSION

The initial number of 43 counties, planned to be included in the analysis, was reduced to 38, first by excluding Poland and Malta who did not prove information about their budget and staffing in any of the reviewed surveys, then by excluding Albania, Armenia and Azerbaijan who did not report neither their backbone capacity (BBcapacity) nor any information about traffic. The certification service is not implemented in Ukraine, Iceland, Belarus, Azerbaijan, whereas Greece, Slovakia and Switzerland did not provide information about issued certificates. Anyway, we did not consider excluding them, as SrvCerif was used to estimate the number of servers in the network.

The analyses showed that the basic CRS model made some of the undeveloped countries ”falsely” efficient, which is explained by the influence of the model for GÉANT subscription and membership fee that greatly supports the developing small and emerging NRENs in providing cost effective commodity and connectivity services. (GÉANT. (2015)) To analyse the countries' GÉANT fee to total budget ratio, we split the data set in two parts – those with the ratio less than 15% were placed in group 1 and those with the ratio greater than 15% in group 2. Group 1 included 21 countries at first (finally 20 countries when Iceland was excluded). Group 2 consisted of 17 countries. For now, group 1 is more interesting for the analysis since NRENs perform services mainly using their own resources and cover costs from their own businesses.

As mentioned, DEA is a data-driven method and the results rely on data, therefore we formulated two data models presented in Table 1 together with descriptive statistics of data for 20 countries in Group 1. The only difference in the two presented models lies in the treatment of the traffic. The Data Model 1 assumes that the total traffic is given per capita, while the Data Model 2 includes the total input/output traffic. Obviously, the composite index includes only outputs in both presented models.

Table 1: Data models and descriptive statistics

	Data Model 1						Data Model 2					
	Efficiency index						Efficiency index					
				Composite index						Composite index		
	Inputs			Outputs			Inputs			Outputs		
	Budget PerCapita	StaffFTE	SrvCertif	BBcapacity	TTraffic PCapita	Service Weighted	Budget PerCapita	StaffFTE	SrvCertif	BBcapacity	TotalTraffic INOUT	Service Weighted
Mean	1.690	69.317	7796.850	44.500	13.728	41.050	1.690	69.317	7796.850	44.500	196.81054	41.050
Min	0.277	12.000	0.000	10.000	1.566	22.000	0.277	12.000	0.000	10.000	9.03975	22.000
Max	4.788	147.100	65500.000	200.000	49.413	63.000	4.788	147.100	65500.000	200.000	1026.277	63.000
StDev	1.272	41.572	15778.841	53.260	10.672	10.501	1.272	41.572	15778.841	53.260	231.34539	10.501

The efficiency indexes, composite service index and the NREN ranks obtained by applying the presented DEA model (1-5) and its modification for the Data Model 1 and Data Model 2 are given in table 2.

Table 2: Results of the analyses

NRENs	Data Model 1					Data Model 2				
	Composite service		Efficiency		Rank differ.	Composite service		Efficiency		Rank differ.
	Index	Rank	index	rank		Index	rank	Index	rank	
Belgium	0.4661	19	0.559	16	3	0.4395	19	0.554	14	5
Croatia	**0.9206**	**5**	**0.47**	**19**	**-14**	**0.9206**	**4**	**0.4695**	**18**	**-14**
Czech Republic	**1**	**1**	**0.654**	**15**	**-14**	**0.9667**	**3**	**0.5145**	**17**	**-14**
Denmark	0.5553	18	0.84	13	5	0.4878	18	0.535	15	3
Finland	0.7008	11	1	1	10	0.6393	14	0.9892	10	4
France	0.6667	13	1	1	12	0.7615	7	1	1	6
Germany	0.6407	16	1	1	15	0.6416	13	1	1	12
Greece	0.7302	10	1	1	9	0.7302	10	1	1	9
Hungary	1	1	1	1	0	1	1	0.8786	11	-10
Ireland	0.8095	7	0.526	17	-10	0.8095	6	0.5262	16	-10
Italy	0.8185	6	1	1	5	0.8185	5	1	1	4
Luxembourg	0.6418	15	1	1	14	0.5873	16	1	1	15
Netherlands	0.7702	8	0.486	18	-10	0.7593	8	0.3705	19	-11
Norway	0.7309	9	0.422	20	-11	0.6445	12	0.3625	20	-8
Portugal	0.5714	17	0.808	14	3	0.5714	17	0.8079	12	5
Slovenia	1	1	1	1	0	0.731	9	0.5709	13	-4
Sweden	0.6485	14	1	1	13	0.5963	15	1	1	14
Switzerland	0.6745	12	1	1	11	0.6451	11	1	1	10
Turkey	**0.3492**	**20**	**1**	**1**	**19**	**0.3735**	**20**	**1**	**1**	**19**
United Kingdom	1	1	0.945	12	-11	1	1	1	1	0

In comparing the observed NRENs, there are only 4 of them who perform their services at top level (the Chez Republic, Hungary, Slovenia and the United Kingdom) in the case of the Data Model 1 and two of them (Hungary and the United Kingdom) in the case of the Data Model 2. The results show that there is no big difference in the efficiency and service indices obtained by the Data Model 1 or Data Model 2. Pearson correlation is 0.88 between the efficiency indices and even higher between the composite service indices (0.92). This indicates that the TTrafficPCapita and TotalTrafficINOUT outputs have no strong or the same impact on the efficiency and the composite service index. The exceptions are Hungary and Slovenia. Hungary is ranked on the first place based on the composite service and efficiency indices in the case of the Data Model 1. But it is not an efficient NREN in the case of the Data Model 2. Thus, regarding the efficiency, Hungary uses 0.81 millions of euros (below the mean), 102 StaffFTE and 1972 SrvCertif (below the mean) to perform a service of BBcapacity of 40, 13.809 of the TTrafficPCapita (above the mean) and the maximum 63 of the ServiceWeighted (above the mean) outputs. This means that both the efficiency and composite indices depend on the TTrafficPCapita and ServiceWeighted outputs. But in the case of the Data Model 2, the value of the TotalTrafficINOUT is 136 (above the average) which makes Hungary inefficient in comparison with other efficient NRENs. The conclusion is that the necessity for the Total Traffic per Capita is forecasted correctly but the TotalTrafficINOUT is not good enough in comparison to other NRENs. Also, this parameter is not as bad as to have a negative influence on the composite service indicator. A similar situation is with Slovenia. It uses 2.26 millions of euros, 66 StaffFTE (below the mean) and 934 SrvCertif (below the mean) to produce 10 BBcapacity, 49.413 TTrafficPCapita (above the mean) and 46 ServiceWeighted (above but close to the mean). Therefore, the efficiency and composite service indicators rely mainly on TTrafficPCapita. In the case of the Data Model 2, the value of the TotalTrafficINOUT is 101.926 (below the mean). This implicates that the efficiency index is decreasing from 1 to 0.57 and the composite index is decreasing from 1 to 0.31. Once again, the Total Traffic per Capita is forecasted correctly, but the TotalTrafficINOUT is far behind efficient NRENs such as the United Kingdom with the maximum value of 1026.277.

The United Kingdom is a benchmark for all other inefficient NRENs. The United Kingdom is a good service performer and efficient in using input resources which is not case with other countries excluding Hungary. A small negative correlation (-0.099 for Data Model 1 and -0.995 for Data Model 2) is evidenced between the efficiency and the composite service indices. Specific cases are Croatia, the Chez Republic with a big negative (-14) and Turkey with a big positive (19) difference between the ranks based on the composite service and efficiency indices. Croatia has only one above the mean output (ServiceWeighted=58) but it is big enough to put this NREN on the fifth place as a performer (composite service index 0.92), but it uses available resources with the efficiency of 0.47, which means that the resource levels are too big for produced outputs and there is a possibility for output improvement by 52% from the quantitative standpoint. On the other hand, the Chez Republic is ranked on the first place based on the composite service index because it achieves a high level of all outputs (BBcapacity=100, TTrafficPCapita=23.547 and ServiceWeighted= 54). But they are not good when we take available resources into the account (BudgetPCapita=2.17, StaffFTE=138.6 and SrvCertif=4581) and the efficiency index is 0.654 according the Data Model 1. On the opposite side, Turkey in not a good performer in comparison to other countries – the composite service index is 0.34920 taking into account the below mean outputs of 10 for the BBcapacity, 1.566 for the TTrafficPCapita and 22 for the ServiceWeighted. Though these outputs are good when inputs are included into the analysis – Turkey uses inputs, 0.28 million of euros, 15 StaffFTE and 1453 ServiceWeighted, in a relatively efficient way in comparison with to other NRENs in the observed set.

5. CONCLUSION

The paper presented the starting analysis aiming to quantify NRENs’ efforts to provide various internet services for their community. The Data Envelope Analysis (DEA) was applied to European NRENs gathered around the common regional organisation GÉANT. The aim of the study was twofold: first, to gain an insight in the suitability of the DEA method to analyse NRENs and second, to examine, for further analyses, the availability and the quality of data on European NRENs’ activities collected yearly by GÉANT.

The study was based on the data set from 2015. It has been shown that the data availability collected by GÉANT for a period of one year became an issue when it came to including all operating European NRENs (i.e.) 43 in the study. One can expect that the issue will be deepened by extending the time period for the analyses from one to two or more years, because NRENs’ contribution in answering to the GÉANT survey varies from year to year. As we were unable to obtain the data for five countries, the initial number of 43 counties for the analysis was reduced. The study included 38 countries, further splitted in two parts.

We performed the starting experiments with the CRS and then with the VRS. We have noticed that the application of the VRS model gives significantly more units that are efficient and this is why we opted for the analysis based on the output CRS model. The analyses showed that the basic CRS model made some of the undeveloped countries ”falsely” efficient, which is explained by the influence of the model for GÉANT subscription and the membership fee that greatly supports the developing small and emerging NRENs in providing cost effective commodity and connectivity services. To analyse the countries’ GÉANT fee vs/to total budget ratio, we split the data set in two subsets – those with the ratio less than 15% in group 1 and those with the ratio greater than 15% in group 2. The results presented are for group 1 consisted of 20 countries (with Iceland excluded), while group 2, consisted of 17 countries, need to be additionally analyzed.

This work can be used as a basis for further analyses. The presented results showed that the countries which continually innovate, develop and support advanced services are not among the most efficient. We suspect on the model used in the study should be improved to reflect the state in the technology advancement. It can be done in two ways: by using weighs (the weighted model), or by fixing a few countries as efficient. Both approaches require expert knowledge from different NRENs to be taken into the account in order to determine either the fix efficient country or the weights.

REFERENCES

Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429-444.

Cook, W. D., Tone, K., & Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega*, 44, 1-4.

Cherchye, L., Moesen, W., Rogge, N., & van Puyenbroeck, T. V. (2007). An introduction to 'benefit of the doubt' composite indicators. *Social Indicators Research*, *82*(1), 111–145.

Foley, M. (from World Bank). National Research and Education Networks -- how universities and researchers connect to each other and to the Internet. *The Case for NRENs Repository, What is an NREN, https://www.casefornrens.org/Resources_and_Tools/Document_Library/Pages/All_documents.aspx*, (accessed 28/2/2018).

G4-1. (2015-2016). The GÉANT Project (GN4-1) Highlihts. *GÉANT project & organization, https://www.geant.org/Projects/GEANT_Project_GN4/Documents/GEANT%20Project%20Highlights%20GN4-1.pdf*, (accessed 28/2/2018).

GÉANT. (2008-2016). NREN Compendium, Time period 2008-2016,https://compendiumdatabase.geant.org.

GÉANT. (2015). The Model for GEANT Subscription and Membership Fees – 2016 onward, *GÉANT Assembly Meeting: 25th-26th November 2015*, Document Code: GA(15)035a(final).

ITU. (2010). World Telecommunication/ICT Development Report 2010: Monitoring the WSIS Targets, A mid-term review. *ITU organisation website, https://www.itu.int/dms_pub/itu-d/opb/ind/D-IND-WTDR-2010-PDF-E.pdf*, (accessed 28/2/2018).

ITU. (2014). Partnership on Measuring ICT for Development, Final WSIS Targets Review, Achievements, Challenged and the Way Foreward. *ITU organisation website, http://www.itu.int/en/ITU-D/Statistics/Documents/publications/wsisreview2014/WSIS2014_review.pdf*, (accessed 28/2/2018).

Kumar, A., Shankar, R., & Debnath, R.M. (2015). Analyzing customer preference and measuring relative efficiency in telecom sector: A hybrid fuzzy AHP/DEA study. *Telematics and Informatics*, 32(3), 447-462.

Melyn, W., & Moesen, W. (1991). Towards a synthetic indicator of macroeconomic performance: unequal weighting when limited information is available. *Public Economics Research Papers*, 17(1), 1–24.

Robin, Arak. (2007). Earnest Foresight Study into Research and Education Networking, Report on Organisations and Governanace Issues. TERENA website, https://www.terena.org/publications/files/EARNEST-Organisation.pdf, (accessed 28/2/2018).

Savić, G. (2012). A Comparative Analysis of the Efficiency in the Financial Sector, Faculty of Organiyational Sciences, Belgrade (in Serbian).

Savić, G., & Martić, M. (2016). Composite Indicators Construction by Data Envelopment Analysis: Methodological Background. Emerging Trends in the Development and Application of Composite Indicators, eds. Jeremic V., Dobrota M & Radojičić Z., 98-126

Serrano-Cinca, C., Fuertes-Callén, Y., & Mar-Molinero, C. (2005). Measuring DEA efficiency in Internet companies. *Decision Support Systems*, 38(4), 557-573.

Shao, B.B., & Lin, W.T. (2016). Assessing output performance of information technology service industries: Productivity, innovation and catch-up. *International Journal of Production Economics*, 172, 43-53.

Staphorst, L., Pretorius, L., & Pretorius, M. W. (2016). Technology forecasting in the National Research and Education Network technology domain using context sensitive Data Fusion. *Technological Forecasting and Social Change*, 111, 110-123.

TERENA. (2001-2014). TERENA NREN Compendium - ISSN 1569-4496, Time period 2001-2014, https://www.terena.org/activities/compendium/.

The Case for NRENs. (2009-2017). A Repository of Resources to Support Funding, Advocacy and the Advancement of National and Regional R&D Networks, *GÉANT project & organization, https://www.casefornrens.org/Resources_and_Tools/Document_Library/Pages/All_documents.aspx*, (accessed 28/2/2018).

TEACHERS' EFFICIENCY MEASURING: AN APPLICATION OF DEA

JOVANA VUKOSAVLJEVIĆ[1], MILENA POPOVIĆ[2], MILAN MARTIĆ[3], MARINA POPOVIĆ[4]

[1] University of Belgrade, Faculty of Organizational Sciences, vukosavljevicjovana@gmail.com;
[2] University of Belgrade, Faculty of Organizational Sciences, milena.popovic@fon.bg.ac.rs;
[3] University of Belgrade, Faculty of Organizational Sciences, milan@fon.bg.ac.rs
[4] University of Belgrade, Faculty of Economics, popovic.marina85@gmail.com

---

Abstract: *The aim of this paper is to present the results of a study about the relative efficiency of teaching performances at the University of Belgrade, the Faculty of Organizational Sciences, using the Data Envelopment Analysis (DEA). DEA is a linear programming based technique for measuring the relative performance of decision-making units (DMUs) where the presence of multiple inputs and outputs makes comparisons difficult. DEA is able to use more parameters of input and output to evaluate which of teachers under examination is the most effective, and to compare other teachers with it. In this paper, teacher's efficiency measuring was analyzed in two aspects: efficiency of teaching and efficiency of research. Based on the results, relatively efficient and inefficient teachers were identified; reasons for all inefficient teachers were discovered; teacher's ranking was done. Considering growing competition in the field of education, with pointing out to the teachers on the weakness sources in their work, bigger responsibility level and commitment to the work is expected.*

Keywords: *Data envelopment analysis, university, teacher's efficiency.*

1. INTRODUCTION

Universities all around the world realize more and more that they are part of the service industry and they are facing competition pressure from different directions. On one side, students' pleasure is connected to their employment (Deahields, Kara & Kaynak, 2005; Elliott & Healy, 2001; Helgesen & Nesset, 2007), which led university authorities to direct their attention to those factors which can help them to more efficiently attract students and create simulative environment for learning (Venesaar, Ling & Voolaid, 2011). On the other side, universities are more included into different rankings, while ranking instruments unavoidably includes some of the measures of teacher's efficiency.

Each teacher efficiency measurement system should link the evaluation for improvement and the policy and research policy guidance for strengthening the system (Looney, 2011). Although lot of researches are focused on the effectiveness of teachers, and it is precisely this teacher's efficiency identified as the main component in the teaching process, but still it is not quite clear what the 'teacher's efficiency' is. Lack of clarity in literature seems to be leading to situation where researchers prioritize different aspects of teacher's efficiency (i.e. to use more student-oriented practices, teacher aligned with professional learning and development, interpersonal skills, productivity and content creation). All of this points out that 'teacher's efficiency' has multidimensional.

Productivity in higher education has an obvious multidimensional character as it relates to both production and dissemination of knowledge through its various activities of teaching, research, and outreach activities (Dundar & Lewis, 1998). Data Envelopment Analysis (DEA) is an approach to relative efficiency measurement where there are multiple incommensurate inputs and outputs. If a suitable set of measures can be defined DEA provides an efficiency measure not relying on the application of a common weighting of the inputs and outputs (Tongzon, 2001). Additionally the method identifies peer units and targets for inefficient units. Charnes, Cooper, and Rhodes (1978, 1981) introduced the method of DEA to address the problem of efficiency measurement for decision making units (DMUs) with multiple inputs and multiple outputs in the absence of market prices. They coined the phrase decision making units in order to include non-market agencies like schools, hospitals, and courts, which produce identifiable and measurable outputs from measurable inputs but generally lack market prices of outputs (and often of some inputs as well). The field of DEA is growing steadily, attracting unabated interest from the management science and economics communities, and continuing to be applied in practice to address new problems in policy making and management (Banker & Podinovski, 2017). New models and methods developed in recent years allow the assessment of the efficiency, performance, and productivity in public institutions (especially educational).

Evaluation of teacher's efficiency measuring is challenging as the criteria for evaluation are both objective and subjective (Thanassoulis et al., 2017). The academic roles of higher education institutions comprise three major components: teaching, research, and service (Edgar & Gear, 2013). In this study, the concept of research performance is examined in a two perspective: efficiency of teaching and efficiency of research. DEA method was used to analyze efficiency of 68 teachers at the University of Belgrade, the Faculty of Organizational Sciences. For long period of time there is a dilemma about relationship between teaching and scientific-research work, especially whether teaching activates disturb scientific-research work or scientific-research work contributes to the teaching efficiency. Results of the carried out study have shown that a teacher can be completely efficient only if he is committed to the teaching ad scientific-research work, equally.

The structure of the paper is built in the following way: the second section describes the DEA basics followed by literature survey in the sector of education. The third section will contain study about the relative efficiency of teaching performances at the University of Belgrade. The main conclusions are summarized in the last section.

2. DATA ENVELOPMENT ANALYSIS

DEA is a mathematical-programming-based approach for identifying the best practice when multiple performance metrics are present. Performance metrics are commonly classified as inputs and outputs of peer DMUs. This special mathematical technique was first introduced by Charnes et al. (1978), initially to evaluate relative efficiency in the non-profit sector.

There is basically input-oriented a constant returns-to-scale (CRS) model that initially introduced by Charnes et al. (1978):

$$
\begin{aligned}
&(\max)\ h_k = \sum_{r=1}^{s} u_r y_{rk} \\
&st \\
&\sum_{i=1}^{m} v_i x_{ik} = 1 \\
&\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \le 0,\ \ j = 1,\dots,n \\
&u_r \ge \varepsilon;\ \ r = 1,2,\dots,s, \\
&v_i \ge \varepsilon;\ \ i = 1,2,\dots,m,
\end{aligned}
\tag{1}
$$

where:

$u_r \ge 0,$ are weights assigned to the *r*th outputs, $r = 1,\dots,s$, and

$v_i \ge 0,$ are weights assigned to the *i*th inputs, $i = 1,\dots,m$ in order to assess DMU_k as efficient as possible.

This basic CCR DEA model, should be solved *n* times, once for each DMU_k. The index h_k shows relative efficiency of DMU_k, obtained as maximum possible achievement in comparison with the other DMUs under the evaluation. DEA empirically identifies the efficient frontier of a set of DMUs based on the input and output variables. Assume that there are *n* DMUs, and the *j*th DMU, produces *s* outputs ($y_{ij},..,y_{sj}$) by using *m* inputs ($x_{ij},..,x_{mj}$). The efficiency score of the observed DMU_k is given as virtual outputs (sum of weighted outputs). For a given set of inputs and outputs, DEA produces a single comprehensive measure of performance (efficiency score) for each DMU.

DEA method occupies an important place in the comparative efficiency studies in the public sector worldwide (Chalos & Cherian, 1995; Odeck, 2005). It is implemented in many aspects of higher education such as evaluation of universities, evaluation of study programs (or faculties) and evaluation of academic staff, including teaching evaluation. Thus, it is possible to determine which variables contribute to improving the performance of higher education, to assess the relative effectiveness of units in higher education institutions, and determining exactly which inputs and outputs contribute to achieving optimal performance.

Examples of DEA application in the area of higher education from around the globe are described in works by Kao and Hung (2008), Berbegal-Mirabent et al. (2013), Agasisti & Perez-Esparrells, (2010), Fuentes et al. (2016), Kuah and Wong (2011).

Kao and Hung (2008) applied DEA to assess the relative efficiency of the academic departments at National Cheng Kung University in Taiwan in utilizing the scarce resources in teaching students and producing research results in measured. The investigation is focused on the efficiency of resources utilization rather than academic performances. This type of assessment alleviates the problem of comparing departments with different characteristics. Although teaching and research have been considered by most people as the two major tasks of the university, they are difficult to measure. The overall efficiency shows whether resources are effectively utilized by each department separately, while pure technical efficiency helps identify weak areas in which more efforts should be made to improve the efficiency of the department. Fourteen (out of 41) inefficient departments with an unsatisfactory result of overall technical efficiency were identified and they are in this way able to calculate the number of inputs that need to be reduced and the number of outputs that need to be increased in order to increase efficiency. Teaching has the largest contribution with 58.2% of the average total score. Publications are the second with 23.7% of the average total score. And finally, external grants contribute with only 18.1% of the average total score. Via the DEA calculations, efficiency decompositions, and cluster analysis, the top administrators of a university are able to detect the departments that are inefficient in utilizing their resources and the department heads are able to identify the area which the greatest gains can be acquired from improvements in efficiency.

Berbegal-Mirabent et al. (2013) assesses the efficiency of universities in terms of teaching, research and knowledge transfer 44 state universities in Spain. The DEA model which is used has an exit orientation based on the assumption that in the public sector labour and budget tend to be fixed and that these organizations produce the maximum possible output forasmuch as resources they have. The results show that average inefficiency at Spanish universities is 12% indicating that they can increase their output by 12%, with the same input values. Also, 21 universities out of 44 are efficient, with a maximum value of 53.7%.

The goal of another study which uses DEA model in higher education (Agasisti & Perez-Esparrells, 2010), is to provide efficiency analysis of 74 Spanish and 76 Italian universities, as well as a perspective of comparison between countries, in order to identify the main similarities and differences. The authors use DEA allowing each of these 150 universities to assign differente weights to different dimensions of their activities in order to maximize their score. As inputs, we consider number of students, number of Ph.D. students, number of professors and financial resources available. As outputs, we use number of graduates as a proxy for teaching performance, and the amount of external resources attracted to research activities as a proxy for research performances. In the first step, a DEA analysis was run separately for Italian and Spanish universities. In general terms, it seems that Spanish HE has a higher average level of efficiency. The next step was to run a DEA analysis considering Italian and Spanish universities together. Here, the average level of efficiency is about 0.7 and, above all, there are more Italian efficient universities (12 out of 14 effective in previous analysis) than Spanish ones (only 3). It means that comparing all universities together, the efficiency barrier for Spanish universities has shifted, and the number of universities that are able to reach it is smaller.

In the work of Fuentes et al. (2016), the technical efficiency of the learning process in higher education is assessed using a three-step procedure that provides progress in relation to previous studies and improves the quality of the results. First, the authors use DEA with contextual variables (socio-economic and cultural levels of pupils' families and student education before the university). Secondly, the efficiency levels have been calculated to prioritize the efficiency units. Finally, through the sensitivity analysis, the contribution of each KPI was determined to the levels of efficiency without distorting the variables. Analytical data were collected from a survey of 633 students. The results show that the level of satisfaction with the course, the diversity of materials and satisfaction with teachers are the most important factors that influence the performance of the teaching. The methodology used in this paper allows better quality control of education, along with data obtained from students that serve to detect inefficiency in individual units and to improve results.

Kuah and Wong (2011) presented the DEA model for joint evaluation of the relative teaching and research efficiencies of universities in Malaysia. The inputs and outputs for university performance measurement were identified. They comprised of 16 measures in total. Joint DEA maximisation was used to model and evaluate these measures. The application of DEA enabled academics to identify deficient activities in their universities and take appropriate actions for improvement.

3. EMPIRICAL STUDY – TEACHER'S EFFICIENCY MEASURING

The main objective of this study was to measure the efficiency of teachers at the University of Belgrade, the Faculty of Organizational Sciences. The survey was fielded in June 2017. Altogether, 68 teachers (DMUs), who gave lectures in the third year of undergraduate study, are used in the survey. Teacher's efficiency measuring is analyzed in two aspects (Hattie & Marsh, 1996):

- efficiency of teaching (E_1) and
- efficiency of research (E_2).

Therefore, a summarized assessment of teacher's efficiency is calculated as follows:

$$Ej = w_1 * Ej_1 + w_2 * Ej_2,\ j = 1,2,\ldots,n \quad (2)$$

where w_1, w_2, weight coefficients that define the importance of each of the performance estimates in each sum assessment. Weight coefficients have a value of 0.5, since it is assumed that these two aspects of efficiency estimates have the same importance ($w_1 = w_2 = 0.5$).

3.1. Study design

***Efficiency of teaching* (*E1*)**

An input-oriented DEA CRS model (1) was used to estimate the efficiency of teaching. The parameters used for the CRS model are:

Inputs:

- Teacher's workload coefficient values – *I1*
- Average number of students in the third year of study in the school year 2015/16 per teacher – *I2*

Outputs:

- Average number of students who passed the exam per teacher – *O1*
- Average grade for all students who passed the exam (per teacher) – *O2*

Teacher's workload coefficient values are calculated as the reciprocal value of the number of subjects that a teacher hands over to students in the third year of study because the more number of subjects the greater workload of the teacher. The data for the input and output values were obtained from the student service of Faculty of Organizational Sciences. Each teacher is assigned an ID.

***Efficiency of research* (E_2)**

The assessment of efficiency of research was measured with the following equation:

$$E_{j2} = O_j / I_j,\ j = 1,\ldots,n \quad (3)$$

Due to the heterogeneity of the obtained results, we used normalization technique to get results in a [0-1] scale within the equation (4):

$$\bar{E}_{j2} = E_{j2} / \max E_{j2},\ j = 1,\ldots,n \quad (4)$$

The parameters used for this assessment are:

Input:

- The appointment of teachers -*I*
  - o assistant professors got value1
  - o associate professors got value 1.5
  - o full professor got value 2

Output:

- Number of scientific papers published in SCI journal list in the school year 2015/16 by teacher - *O*

The data required for the assessment of the efficiency of research are the number of scientific papers published in SCI journal list by the teacher and engagement in the project Ministry of Education, Science and Technological Development. Since no exact engagement data is available, teachers are assigned value, which is equivalent to their engagement. Thus, the values in DEA are derived from the data instead of being fixed in advance. To assess the efficiency of teachers in scientific research, data were downloaded from http://www.scopus.com/. The date showed about 11% "outliers" (number of scientific papers published in the school year 2015/16 by teacher more than 5). We have chosen to keep the outliers in the data and give all of them maximum values 5.

3.2. Analysis and results

Descriptive statistics of these parameters values, which used for the CRS model, is given in Table 1.

Table 1. Descriptive Statistics

Parameters	I1	I2	O1	O2
Min	0.33333	1.71	1.22	6.93
Max	1	963	145.56	9.77
Mean	0.786765	147.1356	47.41824	8.144412
Std. Dev.	0.266369	138.2835	40.3438	0.686743
Correlation coefficients				
I1	1			
I2	-0.30158	1		
O1	-0.47439	0.362323	1	
O2	-0.15564	-0.0862	0.151575	1

Based on the results of the correlation, it is obvious that the input *I1* (teacher's workload coefficient values) negatively affects to the outputs *O1* and *O2* (average number of students who passed the exam per teacher and average grade for all students who passed the exam (per teacher)).

Table 2. Results of the teacher's efficiency measuring

ID	Efficiency of research $\bar{E}_{j2}$	Efficiency of teaching E_{j1}	*E*	Rank
1	0.666	1	0.833	7
2	0	0.926	0.463	49
3	0.4	0.636	0.518	35
4	0	1	0.5	38
5	0	1	0.5	38
6	**1**	**1**	**1**	**1**
7	0.2	0.798	0.499	44
8	0	0.476	0.238	66
9	0.4	0.661	0.530	30
10	0	1	0.5	38
11	0	0.636	0.318	59
12	0.4	0.653	0.526	33
13	0.133	0.498	0.315	60
14	0.4	0.532	0.466	48
15	0	0.483	0.241	64
16	0	0.476	0.238	66
17	0.133	0.675	0.404	54
18	0	1	0.5	38
19	0.4	0.423	0.411	52
20	0	1	0.5	38
21	0.2	0.854	0.527	32
22	0.1	0.414	0.257	63
23	0	0.483	0.241	64
24	0.2	0.758	0.479	47
25	**1**	**1**	**1**	**1**
26	0.666	0.393	0.530	31
27	0.5	0.975	0.737	14
28	0	0.343	0.171	68
29	1	0.540	0.770	10
30	0.3	0.540	0.420	50
31	0.1	0.423	0.261	61
32	0.4	1	0.7	17
33	0.5	0.926	0.713	16
34	1	0.540	0.770	10
35	0.3	0.975	0.637	20
36	0.5	0.975	0.737	14
37	0.2	0.625	0.412	51

38	0.533	1	0.766	12
39	1	0.926	0.963	5
40	0.5	1	0.75	13
41	0.4	1	0.7	17
42	0.6	0.975	0.787	9
43	0	0.661	0.330	58
44	0.4	1	0.7	17
45	0.5	0.758	0.629	22
46	0.2	1	0.6	26
47	0.4	0.807	0.603	25
48	0.6	0.653	0.626	23
49	1	0.975	0.987	3
50	0.4	0.718	0.559	28
51	0.3	0.393	0.346	56
52	0.6	0.653	0.626	23
53	0.6	1	0.8	8
54	0.2	0.807	0.503	36
55	0.2	0.807	0.503	36
56	1	0.975	0.987	3
57	0.1	1	0.55	29
58	0	0.515	0.257	62
59	0.2	1	0.6	26
60	0.5	0.318	0.409	53
61	0.3	0.966	0.633	21
62	0.5	0.540	0.520	34
63	0.2	0.489	0.344	57
64	0	0.975	0.487	45
65	0	0.807	0.403	55
66	0.8	1	0.9	6
67	0	0.966	0.483	46
68	0	1	0.5	38

Teacher ranking was done on the summarized assessment of teachers' efficiency (E). Rank 1 are shared only two assistant professors with ID 6 and ID 25 (Table 2). Those teachers have both excellent performance grades, i.e. he is also successful in teaching and publishing scientific papers. Rank 3 (E = 0.987) are shared an assistant professors with ID 49 and ID 56, who have excellent grades but fewer than first-ranked. Six teachers are shared rank 38. They are very efective in teaching (E_{j1} = 1), but in the same time, have a very low estimate of efficiency in research (E_{j2} = 0). More than 7.35% of teachers have an efficiency index greater than 0.9, which means that, based on this summary score, most teachers are effective.

Based on the results of the teachers' efficiency of research (E_{j2}) is distinguished by 18 teachers (26%) who did not publish any scientific papers in the school year 2015/16 (Table 2), which according to this „aspect“ is classified as clusters of ineffective teachers (E_{j2} = 0). This cluster also has the three lowest ranked teachers who have a very low average grade on the subject. One more interesting note is that the assistant professors with ID 29 and ID 34 have a excellent grade of research (E_{j2} = 1) but poor in teaching (E_{j1} = 0.54).

4. CONCLUSION

Public higher education sector is under a growing pressure worldwide to increase efficiency and improve the quality of its activities. Results of teachers' efficiency measuring are used by University and faculty representatives to improve teaching efficiency and to assess who is to be promoted, full time employed, or supported to hold a particular course. Though numerous researches show that there is no a unique measure, it seems that teachers accomplish positive results if they are monitored and adequately supported. The choice of a model, or the way to measure teachers' efficiency, is made based on the issued Regulations of a faculty.

In this paper, we used DEA method for measuring teachers' efficiency and demonstrated its advantages: the obtained results can be easily interpreted; all criteria for efficient teaching are include; clear determination of low scores and insight into their causes, based on a specific criterion.

Continuous measuring of teachers' efficiency and enabling individual and collective development allow teachers become more efficient in contribution of research to effective continuing professional development activities and teaching. It is worth noted that this way of measuring must be continuously adapted to changing market pressures to ensure students’ satisfaction is achieved.

REFERENCES

Agasisti, T., & Perez-Esparrells, C. (2010). Comparing efficiency in a cross-country perspective: the case of Italian and Spanish state universities, *Higher Education*, 59(1), 85–103.

Banker, R.D., & Podinovski, V.V. (2017). Novel theory and methodology developments in data envelopment analysis. *Annals of Operations Research*, 250 (1), 1-3.

Chalos, P., & Cherian, J. (1995). An application of Data Envelopment Analysis to public sector performance measurement and accountability. *Journal of Accounting and Public Policy*, 160, 143–160.

Charnes, A., Cooper, W.W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, 429-44.

Charnes, A., Cooper, W.W., & Rhodes, E. (1981). Evaluating program and managerial efficiency: An application of data envelopment analysis to program follow through. *Management Science*, 27, 668-697.

DeShields Jr, O. W., Kara, A., & Kaynak, E. (2005), Determinants of business student satisfaction and retention in higher education. *The International Journal of Educational Management*, 19, 2/3; pg. 128.

Dundar, H. & Lewis, D.R. (1998). Determinants of research productivity in higher education. *Research in Higher Education*. 39(6), pp. 607-631.

Edgar, F., & Geare, A. (2013). Factors influencing university research performance. *Studies in Higher Education, 38(5)*, 774-792.

Elliot, K. M., & Healy M. A. (2001) Key factors influencing student satisfaction related to recruitment and retention. *Journal of Marketing for Higher Education*, 10, 1-11.

Fuentes,R., Fuster, B., & Lillo-Banulas, A. (2016). A three-stage DEA model to evaluate learning-teaching technical efficiency. *Expert Systems with Applications*, 48, 89-99.

Hattie, J., & Marsh, H. W. (1996). The relationship between research and teaching in universities. *Review of Educational Research*, 66, 507-542.

Helgesen, O., & Nesset, E. (2007) What accounts for students loyalty? Some Field Study Research. *International Journal of Education Management*, 21(2), 126-143.

http://www.scopus.com/

Kao, C., & Hung, H.T. (2008). Efficiency analysis of university departments: An empirical study. *The International Journal of Management Science*, 36, 653-664.

Kuah, Ch. T.& Wong, K. Y. (2011). Efficiency assessment of universities through data envelopment analysis. *Procedia Computer Science*, 3, 499–506.

Looney, J. (2011). Developing High-Quality Teachers: Teacher Evaluation for Improvement. *European Journal of Education*, 46, 440-455.

Odeck, J. (2005). Evaluating target achievements in the public sector: an application of a rare non-parametric DEA and Malmquist indices. *Journal of Applied Economics*, 8, 171–190.

Thanassoulis, E., Dey, P. K., Petridis, K., Goniadis, I., & Georgiou, A. C. (2017). Evaluating higher education teaching performance using combined analytic hierarchy process and data envelopment analysis. *Journal of the Operational Research Society*, 68(4), 431-445.

Tongzon, J. (2001). Efficiency measurement of selected Australian and other international ports using data envelopment analysis. *Transportation Research Part A: Policy and Practice*, 35(2), 107-122.

Venesaar, U. Ling, H. & Voolaid, K. (2011). Evaluation of the Entrepreneurship Education Programme in University: A New Approach. *Amfiteatru Economic Journal*, 8(30), 377-391.

B4
Data Sciences & Big Data

XIII Balkan Conference on Operational Research

QUALITY MEASURES OF QUANTIFIED LINGUISTIC SUMMARIES AGGREGATED BY THE UNINORM CONNECTIVE FOR DIMENSIONAL DATA STRUCTURES

Miroslav Hudec[1]

[1] University of Economics in Bratislava, Slovakia, Faculty of Economic Informatics, miroslav.hudec@euba.sk

Abstract: *Summarization is concerned with finding a compact description of the data set. The data summarization by short quantified sentences of natural language is beneficial for variety of data users. The truth value or validity of a summarized sentence usually is not sufficient. Several quality measures were suggested in the literature. However, some of them are functionally dependent, and are usually calculated from the data. Firstly, this work introduces a new measure: relevance, which is assigned by the user before calculating other measures to avoid computations for irrelevant summaries. Secondly, the paper suggests aggregation of existing quality measures (validity, simplicity and coverage) with the suggested relevance measure by uninorms to ensure full reinforcement and compensation effect. The findings are supported by illustrative example related to the star scheme of data warehouse. Finally, directions for further improvements are outlined.*

Keywords: *lingustic summary, relevance measure, aggregating quality measures, uninorms, fuzzy sets*

1. INTRODUCTION

Aggregation and summarization are keystones in interpreting knowledge form the data. Aggregation operators reduce a set of numbers into a unique representation or meaningful number (Detyniecki (2001)). Another observations emphasize that data summarization into the linguistic interpretation underlines numeric summarization and is more suitable for variety of users (Hudec (2016b); Yager et al. (1990)). An approach based on quantifiers for interpreting relational knowledge in the data is suggested by Rasmussen and Yager (1997). Such summary is of the structure *Q R entities are S*, where *Q* is a relative quantifier, *S* and *R* explains flexible subdomains of the considered attributes. An example of such summary is: *most of young customers have high turnover*.

Nowadays, data from the transactional databases and external sources are usually integrated into the company's data warehouses, i.e. into dimensional data structures, where each measured value gets its meaning in the intersection of dimensions (Kimball and Ross (2013)). This structure is suitable for asking variety of business intelligence questions (Grossmann and Rinderle-Ma (2010); Jensen et al. (2010)).

The data summarization copes with the two main issues: computational demand caused by the high number of possible summaries and records (Niewiadomski et al. (2006)), and the quality of summarized sentences (Wu et al. (2010); Pereira-Fariña et al. (2012)). The former has been mitigated by the optimization of calculations, e.g. (Hudec (2016b); Liu (2011)), whereas the latter is addressed by the quality measures and their aggregation, e.g. (Hudec (2017); Pereira-Fariña et al. (2012); Wu et al. (2010)).

In this work we introduce and formalize the relevance measure to fill the gap in measures expressing the sentence's relevance for user's tasks. The second contribution is suggestion for aggregating atomic quality measures by uninorms to avoid issues of the downward reinforcement and lack of the compensation effect of t-norms (Beliakov et al. (2007); Yager and Rybalov (1998)).

The reminder of this paper is organized as follows. The second section briefly explains linguistic summaries created on the dimensional data structures and existing quality measures. The third section introduces the relevance measure and explains its meaning and properties. The fourth section is devoted to aggregating atomic quality measures by uninorms. The fourth section is focused on illustrative example and discussion. Finally, the fifth section presents a summary of the paper and suggestions for the future work.

2. DIMENSIONAL DATA STRUCTURES AND LINGUISTIC SUMMARIES

This section gives preliminaries related to the star scheme of a data warehouse and summarization by short quantified sentences, which are used throughout the paper.

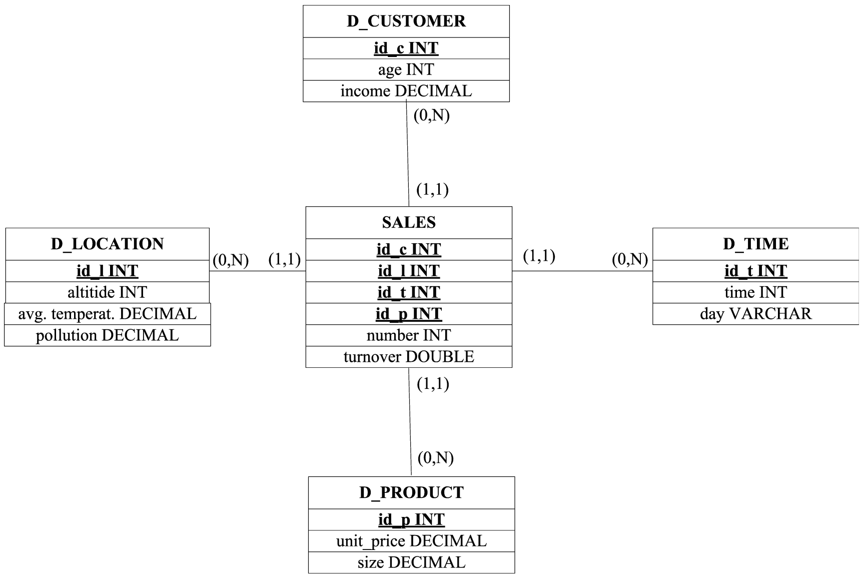


Figure 1 Structure of the STAR scheme.

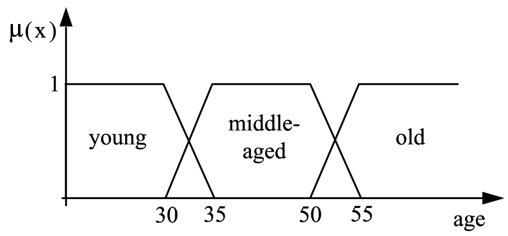


Figure 2 Attribute *age* and its linguistic terms (granules) expressed by fuzzy sets.

2.1. Dimensional data structures and business intelligence questions

The dimensional data structure usually takes the shape known as star scheme. This scheme (or variant thereof) is a structure, where a central table represents the fact on which the analysis is focused, and a number of tables, represent the dimensions required for analyses (Kimball and Ross (2013)). This structure of data warehouses is often used in business intelligence, because many questions can be asked and efficiently answered (Golfarelli and Rizzi (2009); Jensen et al. (2010)).

An example of a star scheme is shown in Figure 1. Business intelligence questions related to *who* (characteristics of customers), *when* (time stamp), *where* (parameters of location) and *what* (features of products) may be asked. Business intelligence queries related to dimensional data structures are discussed in vast literature. For instance, classical queries in (Golfarelli and Rizzi (2009); Jensen et al. (2010); Grossmann and Rinderle-Ma (2010)) fuzzy queries in (Singh et al. (2009); Fasel (2014)) and linguistic summaries over the time dimension in (Castillo-Ortega et al. (2011)).

In this paper the focus is on representative answers expressed by quantified sentences regarding all dimensions. Illustrative examples related to Figure 1 are *most of young customers have high number of ordered products* and *most of middle aged customers in early morning have high orders*. To calculate validity (truth values) of a particular summary is a trivial task, when we mathematically formalize adjectives and quantifiers by fuzzy sets. Similar holds for calculating sufficient data coverage and other quality measures. Section 2.2. provides insight into these calculations.

But, when the task is to retrieve all relevant summaries the situation is more complex. Let us have three flexible granules (small, medium, high) over attributes' domains in dimensions *d_customer*, *d_location* and *d_product*, the same granules over attribute *number* in fact table*sales*, and five granules explaining attribute *time* (early morning, late morning, around noon, early afternoon, late evening) (Figure 1). Possible overlapping granules (fuzzy sets) for customer age are shown in Figure 2. If each summary consists fo one granule from each dimension and one granule from the fact table, 2880 sentences should be evaluated. In addition, other possible combinations might be relevant, e.g. whether *most of middle-aged and low income customers during early evenings have high number of orders*. Thus, evaluation of such an amount of summaries is out of the question.

2.2. Linguistic summaries and quality issues

Linguistic summaries of the structure*Q R entities in a data set are S* developed by Rasmussen and Yager (1997) are suitable for revealing relational knowledge from the dimensional data structures. The validity (truth value) is computed as

$$v(Qx(P(x))) = \mu_Q(\frac{\sum_{i=1}^{n} t(\mu_S(x_i), \mu_R(x_i))}{\sum_{i=1}^{n} \mu_R(x_i)}) \tag{1}$$

where n is the cardinality of elements in a data set, $\frac{\sum_{i=1}^{n} t(\mu_S(x_i), \mu_R(x_i))}{\sum_{i=1}^{n} \mu_R(x_i)}$ is the proportion of the records in a data set that meets the S and belong to the R, t is a t-norm, μ_S, μ_R, and μ_Q are membership functions explaining summarizer S, restriction R and relative quantifier Q, respectively. S and R may be atomic or compound predicates. The validity v gets value from the unit interval.

Summarization is concerned with finding a compact data description. In our case, description by short quantified sentences of natural language. According to Hirota and Pedrycz (1999), the following features are essential for measuring the quality of summaries: validity, novelty, usefulness, simplicity and generality. Wu et al. (2010) have proposed equations for calculating these measures for linguistic summaries. By this way, validity corresponds with Eq. (1). The generality measure is expressed by sufficient coverage, which indicates, whether a summary is supported by sufficient subset of the data. First, the coverage ratio is calculated as (Wu et al. (2010))

$$i_c = \frac{\sum_{i=1}^{n} p_i}{n} \tag{2}$$

where n is the number of records in a considered data set and
$$p_i = \begin{cases} 1, & \text{for } \mu_S(x_i) > 0 \wedge \mu_R(x_i) > 0 \\ 0, & \text{otherwise} \end{cases}$$
Because a summary of the structure (1) covers a small subset of the whole data set, i_c is significantly smaller than 1. Thus, the following mapping converts this ratio into the degree of sufficient coverage (Wu et al. (2010))

$$C = f(i_c) = \begin{cases} 0, & \text{for } i_c \leq r_1 \\ 2(\frac{i_c - r_1}{r_2 - r_1})^2, & \text{for } r_1 < i_c \leq \frac{r_1 + r_2}{2} \\ 1 - 2(\frac{r_2 - i_c}{r_2 - r_1})^2, & \text{for } \frac{r_1 + r_2}{2} < i_c < r2 \\ 1, & \text{for } i_c \geq r_2 \end{cases} \tag{3}$$

where authors in their work suggested values 0.02 and 0.15 for r_1 and r_2, respectively. Anyway, parameters r_1 and r_2 can be adjusted according to the number of atomic predicates in the restriction part of a considered summary.

The degree of usefulness is computed as a minimum of validity and coverage (i.e. $U = \min(v, C)$). The degree of outlier O, referencing to novelty (unexpected summaries are very valuable for users if they cover the regular behaviour in the data, not in outliers), is an aggregation of validity and coverage as: "the validity degree v is very small or very high and the sufficient coverage C must be very small".

Finally, the simplicity measure expresses the length of sentence as

$$S = 2^{2-l} \tag{4}$$

where l is the total number of attributes (atomic predicates) in R and S parts of a summary. When both parts of a summary contains one attribute each, this measure gets value 1.

All aforementioned measures get values form the unit interval, which makes their aggregation easier, but some measures are functionally dependent. To keep the best value of each measure equal to 1, instead of the outlier measure, we should use its negation, more precisely calculated by the standard fuzzy negation ($\neg O = 1 - O$).

Other quality measures, which partially overlap with these are suggested in Castillo-Ortega et al. (2011). Naturally, the question which summarized sentence is better in terms of all measures appears. The basic quality ordering relation $\leq_Q$ (*having less quality than*) is defined as a binary relation (Castillo-Ortega et al. (2011))

$$\forall ls_1, ls_2 \in \mathbb{LS} \;\; ls_1 \leq_Q ls_2 \Leftrightarrow C(ls_1) \leq C(ls_2) \wedge S_{im}(ls_1) \leq S_{im}(ls_2) \wedge E_c(ls_1) \leq E_c(ls_2) \wedge A_c(ls_1) \leq A_c(ls_2) \tag{5}$$

This relation generally works when ls_i is better than ls_j by all measures. Otherwise, it is not an easy task, because some quality measures might be functionally dependent and might have different importance for users.

The quality measures in (Wu et al. (2010)), for instance are computed form the data for converting summaries of high quality into if-then rules. Quality measures suggested by Castillo-Ortega et al. (2011) are related to summarizing from the images, where users influences results by providing their opinions.

In the next section, we introduce and formalize the relevance measure to cope with the computational effort to allow filtering only relevant summaries for the users.

3. RELEVANCE MEASURE OF SUMMARIZED SENTENCES

One of quality measures should relate to the relevance of summarized sentence regarding the users goals, i.e. the impact the summarized information might have on the decision making. Even though a sentence excellently meets measures such as validity, simplicity and coverage, it might be irrelevant for the particular task in a company's department. For instance, summarized sentence *most of young employers have medium productivity* is irrelevant for the marketing department, whereas it is valuable information for the staff manager. In order to solve this problem and contribute to the reduction of over-abundant information and computational demands, we have suggested the relevance measure.

Let us have a subset of summaries $S = \{ls_1, ..., ls_m\}$, $m << M$, where M is the number of all possible summaries regarding the star scheme in Fig. 1.

Assigning relevancies (r_j) as weights to the sentences of interest by the usual property

$$\sum_{j=1}^{m} r_j = 1 \tag{6}$$

is not suitable due to two reasons:

- When user is interested in a higher number of sentences, the most relevant one gets weight significantly lower than 1.
- The ideal summarizing sentence should get value 1 as a result of aggregating all quality measures.

Thus, instead of (6), the relevance measure should meet the following property

$$\max_{j=1,...,m} r_j = 1 \quad \text{and} \quad \min_{j=1,...,m} r_j = 0 \tag{7}$$

where m is the number of sentences of interest.

Obviously, when relevance gets value of 0, then this summary is not further processed for calculating validity and other quality measures.

Let the user is interested in 5 summaries (ls_1, ..., ls_5), then he/she might assign the following values to the relevance vector $r^T = [0.75, 1, 0.2, 1, 0.85]$.

In some cases, it is more suitable for the user to assign qualitative evaluations, which are translated to the weights by the suggestive mapping shown in the Table 1:. In our case, the density of linguistic terms above value of 0.5 is higher than bellow this value. Thus, the linguistic term set is not symmetric. This is an acceptable property, because the very week importance is not a feature which attracts user's interest.

Table 1: A possible mapping from linguistic interpretation to weights

Interpretation	Weight
extremely important	1
very high importance	0.9
high importance	0.8
significantly important	0.7
medium importance	0.5
lower importance	0.3
weak importance	0.2
not at all	0

Despite the benefit of linguistic interpretation, directly assigning numerical values gives space for users to make smooth distinction among summaries, because they might assign any value form the unit interval.

To sum up, the user-in-the-loop in the first step allows him to declare, which summarized sentences should be evaluated. Consequently, in the second step user assign either numerical or linguistic weight to each summary or to the subset of summaries.

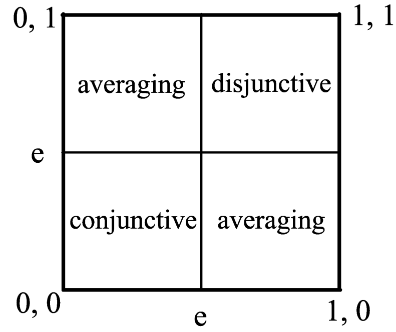


Figure 3 Graphical interpretation of two-directional reinforcement of uninorm.

A more sophisticated way is the pairwise comparison of relevancies to obtain the consistent set of preferences. This is more complex than providing weights individually. The problem is that the user might provide inconsistent preferences, for instance ls_a is slightly more relevant than ls_b, ls_b is significantly better than ls_c and ls_c is more relevant than ls_a. Thus, both types of consistency: the ordinal one—transitivity should be met (Xu et al. (2013)) and the stronger one: cardinal consistency—if ls_a is p times better than ls_b and ls_b is q times better than ls_c, then ls_a is $p\dot{q}$ times better than ls_c (Chiclana et al. (2009)) might be violated and therefore methods for checking these consistencies should be used. Anyway, this is the topic for the future research activities.

4. AGGREGATING QUALITY MEASURES BY UNINORMS

In the literature (e.g. Pereira-Fariña et al. (2012)), summaries are compared according to the preference relation *having less quality than* by Eq. (5). This corresponds to the “global” interpretation, i.e. comparison is realized among summaries, i.e. we could say that the vertical aggregation have to be realized, because summaries are usually listed like in Table 2:.

The other, less demanding option is a horizontal aggregation, which corresponds to the “local” interpretation. In this way, each summary is independently evaluated according to its respective values of quality measures. There is no need for comparison with other summaries to assess the aggregated quality measure for a particular summary.

The purpose of aggregation functions is to combine inputs, usually expressed by closed intervals [0, 1] like degrees of membership in fuzzy sets, degrees of preference, etc. to produce a real value in the unit interval. The four main classes of aggregation functions are (Calvo et al. (2002); Dubois and Prade (1985)): conjunctive, averaging, disjunctive and mixed.

A summary should meet all measures (at least partially) to be considered as a significant one. Hence, averaging and disjunctive functions are excluded. The former are limited by the property $\min(x_1, x_2, ..., x_n) \leq f(\mathbf{x}) \leq \min(x_1, x_2, ..., x_n)$, where $\mathbf{x}$ stands for vector. The latter are limited by the property $\max(x_1, x_2, ..., x_n) \leq f(\mathbf{x}) \leq 1$, i.e. they support upward reinforcement. The conjunctive aggregations (or t-norms) meet this requirement, but they lack the compensatory effect—the case of minimum t-norm (Zimmermann (2001)), or have downward reinforcement (Detyniecki (2001); Yager and Rybalov (1998)), i.e. the aggregation value is lower or equal than the value calculated by the minimum t-norm.

When all the considered quality measures in this paper (validity, coverage, simplicity and relevance) are significantly met, the summary should be emphasized by high value, i.e upward reinforcement. Analogously, when all considered measures get low values, the summary should be attenuated, i.e. downward reinforcement. It means that we need to apply full reinforcement. Such property is met by uninorm functions, which are generalization of t-norms and t-conorms by means of the same three axioms: associativity, symmetry and possession of a neutral element e (Beliakov et al. (2007)). An uninorm is graphically illustrated in Figure 3. It is clear that, when $e \longrightarrow 1$, the reinforcement emphasizes conjunction. In the opposite case ($e \longrightarrow 0$), the reinforcement emphasize disjunction.

A reasonable option is $3 - \Pi$ function suggested by (Yager and Rybalov (1996))

$$u(x_1, ..., x_n) = \frac{\Pi_n^{i=1} x_i}{\Pi_n^{i=1} x_i + \Pi_n^{i=1} (1 - x_i)} \tag{8}$$

with the convention $\frac{0}{0} = 0$ to obtain a conjunctive uninorm and $\frac{0}{0} = 1$ to acquire a disjunctive uninorm. The neutral element for this function is 0.5.

In addition, one may consider a threshold value for each quality measure (or for aggregated measure) to penalize summaries, which have very low value of considered measure(s). This is not crucial for aggregating by t-norms (especially nilpotent ones) and by uninorms due to their reinforcement effect penalizing small values. However, when one measure is very low, the uninorm pushes aggregated value form the disjunctive feature to the averaging one. This is another topic for further evaluation, which is out of the scope of this paper.

5. ILLUSTRATIVE EXAMPLE AND DISCUSSION

This section illustrates the suggested approach and provides discussion. Let us have seven summarized sentences of the user interest listed in Table 2: together with the values of four quality measures and the aggregated measure computed by Eq. (8).

Table 2: Matching degrees to quality measures and overall aggregation

Summary	Validity	Coverage	Simplicity	Relevance	Aggregation (8)
ls1	1.00	1.00	1.00	1.00	1.000
ls2	0.85	0.92	0.50	0.90	0.998
ls3	0.63	0.82	0.25	0.70	0.858
ls4	0.50	0.50	0.50	0.50	0.500
ls5	0.55	0.65	0.25	0.50	0.431
ls6	0.11	0.35	0.50	0.30	0.028
ls7	0.80	0.00	0.125	0.90	0

The summary of the ideal quality is $ls1$, because it meets all measures with the value of 1. Summary $ls2$ is very highly rated, because three of four measures are almost ideally met, whereas the simplicity measure is met with 0.5, which is the neutral element in (8). It means that the disjunctive feature is activated, i.e. this summary is reinforced to the disjunctive direction. Summary $ls6$ is reinforced to the opposite, conjunctive direction, because it meets all measures with low degrees. Summaries $ls4$ and $ls5$ are evaluated by the averaging feature of this uninorm.

Let us focus on $ls2$ and $ls4$. If we apply minimum t-norm, these summaries are indistinguishable. When we apply product t-norm, then aggregated values are low, i.e. 0.3323 and 0.0625, respectively, which might indicate to user the low quality. On the other hand, uninorm (8) reveals that $ls2$ is of higher quality than $ls4$.

We have suggested this set of measures, because they are independent and are covering main aspects of quality issues. We have also recognized possible weakness in the definition of the simplicity measure (4). For one atomic attribute in restriction and one in summarizer this measure gets value 1, which is the expected one. But, measure significantly decreases when the number of atomic predicates increases. For instance, when five attributes in summary appear, the measure get value of 0.125. Hence, in future activities we focus our attention on modifying this definition.

The other option is dividing quality measures into two sets: constraints and wishes. In this case, user might say that validity, coverage and relevance have to be satisfied (the set of constraints), and it is nice if simplicity is also satisfied (one element in the set of wishes). This way reduces the impact of simplicity to the aggregated value. Formally, we can computer the aggregated value for a summary as

$$agg(ls_i) = u(V(ls_i), C(ls_i), R(ls_i)) \text{ and if possible } S(ls_i) \quad (9)$$

The *and if possible* operator for fuzzy database queries is formalized by Bosc and Pivert (2012). The quantified version of this operator is formalized from the bipolar perspective by Kacprzyk and Zadrożny (2013) and from the asymmetric perspective by Hudec (2016a). Regarding the applicability of this operator in evaluating linguistic summaries further research is advisable.

6. CONCLUSION

This work has suggested the compact set of four quality measures for linguistically summarizing from the data warehouses. Three of them are well-known in the literature (validity, coverage and simplicity), whereas the fourth one (relevance) is suggested in this paper. The main benefit of adopting this measure is in reducing computational effort in summarizing all sentences of interest from the dimensional data structures, where large number of possible summaries can be created. Different departments and users in a company are usually interested in summarizing form different attributes, and their subdomains. The relevance measure is adjusted to meet these requirements by assigning relevance index to summarizing sentences. The other measures are calculated from the data.

The next contribution was aggregating these measures. The limitations of conjunctive functions: non-compensatory effect and downward reinforcement excludes these functions. Averaging functions as well as disjunctive cannot be applied, because we should keep value of 0 as annihilator. It led us to the conclusion that, the suitable functions are uninorms. We have aggregated quality measures by the $3-\Pi$ uninorm. Regarding the future research tasks, other uninorm functions should be examined. Threshold values of measures are mean for excluding summaries, which have low quality according to at least one measure. On the other hand, a suitable uninorm may exclude the need for the thresholds due to the downward reinforcement effect for all low values, and the property of pushing aggregated values form the disjunctive feature to the averaging one when several measures have low values.

The weak point for this aggregation is the existing definition of the simplicity measure. The value significantly decreases when the number of atomic predicates in summaries increases. For the future research, we suggest modification of the simplicity measure definition, or applying the *and if possible* operator, where validity, coverage and relevance should be satisfied, and it is beneficial if the simplicity is met as well. The simplicity measure emphasizes summaries, which in a more concise way explain data. On the other hand, in some cases, less simple summaries may excellently explain relations among specific subsets of the data.

Acknowledgement

This paper is part of a project VEGA No. 1/0373/18 entitled “Big data analytics as a tool for increasing the competitiveness of enterprises and supporting informed decisions” by the Ministry of Education, Science, Research and Sport of the Slovak Republic.

REFERENCES

Beliakov, G., Pradera, A., and Calvo, T. (2007). *Aggregation Functions: A Guide for Practitioners*. Berlin Heidelberg: Springer-Verlag.

Bosc, P. and Pivert, O. (2012). On four noncommutative fuzzy connectives and their axiomatization. *Fuzzy Sets and Systems*, 202:42–60.

Calvo, T., Kolesárová, A., Komorníková, M., and Mesiar, R. (2002). Aggregation operators: Properties, classes and construction methods. In T. Calvo, G. Mayor, R. M., editor, *Aggregation operators. New trends and applications*, page 3–104. Heidelberg: Physica-Verlag.

Castillo-Ortega, R., Marín, N., and Sánchez, D. (2011). Linguistic query answering on data cubes with time dimension. *International Journal of Intelligent Systems*, 26:1002–1021.

Chiclana, F., Herrera-Viedma, E., Alonso, S., and Herrera, F. (2009). Cardinal consistency of reciprocal preference relations: a characterization of multiplicative transitivity. *IEEE Transactions on Fuzzy Systems*, 17:14–23.

Detyniecki, M. (2001). Fundamentals on aggregation operators. In *Proceedings of the AGOP 2001*. Asturias.

Dubois, D. and Prade, H. (1985). A review of fuzzy set aggregation connectives. *Information Sciences*, 36:85–121.

Fasel, D. (2014). *Fuzzy Data Warehousing for Performance Measurement*. Switzerland: Springer International Publishing.

Golfarelli, M. and Rizzi, S. (2009). *Data Warehouse Design: Modern Principles and Methodologies*. Bologna: McGraw-Hill.

Grossmann, W. and Rinderle-Ma, S. (2010). *Fundamentals of Business Intelligence*. Springer-Verlag: Berlin Heidelberg.

Hirota, K. and Pedrycz, W. (1999). Fuzzy computing for data mining. *Proceedings of the IEEE*, 87:1575–1600.

Hudec, M. (2016a). Constraints and wishes in quantified queries merged by asymmetric conjunction. In *1st International Conference Fuzzy Management Methods (ICFMsquare 2016)*, page 25–34. Fribourg.

Hudec, M. (2016b). *Fuzziness in Information Systems - How to Deal with Crisp and Fuzzy Data in Selection, Classification, and Summarization*. Switzerland: Springer International Publishing.

Hudec, M. (2017). Merging validity and coverage for measuring quality of data summaries. In P. Kulczycki, L.T. Kóczy, R. M. and Kacprzyk, J., editors, *Information Technology and Computational Physics*, page 71–85. Switzerland: Springer International Publishing.

Jensen, C., Pedersen, T., and Thomsen, C. (2010). *Multidimensional Databases and Data Warehousing*. Morgan and Claypool Publishers.

Kacprzyk, J. and Zadrożny, S. (2013). Compound bipolar queries: combining bipolar queries and queries with fuzzy linguistic quantifiers. In *8th Conference of the European Society of Fuzzy Logic and Technology (EUSFLAT 2013)*, page 848–855. Milan.

Kimball, R. and Ross, M. (2013). *The Data Warehouse Toolkit (3rd ed.)*. New York: John Wiley & Sons.

Liu, B. (2011). Uncertain logic for modeling human language. *Journal of Uncertain Systems*, 5:3–20.

Niewiadomski, A., Ochelska, J., and Szczepaniak, P. (2006). Interval-valued linguistic summaries of databases. *Control and Cybernetics*, 35:415–443.

Pereira-Fariña, M., Eciolaza, L., and Triviño, G. (2012). Quality assessment of linguistic description of data. In *Proceedings of the ESTYLF 2012*, page 608–612. Valladolid.

Rasmussen, D. and Yager, R. (1997). Summary sql - a fuzzy tool for data mining. *Intelligent Data Analysis*, 1:49–58.

Singh, M., Tiwari, R., Mahajan, M., and Dani, D. (2009). An architecture for handling fuzzy queries in data warehouses. In et al., S. R., editor, *Contemporary computing*, page 240–249. Berlin Heidelberg: Springer-Verlag.

Wu, D., Mendel, J., and Joo, J. (2010). Linguistic summarization using if-then rules. In *Proceedings of the 2010 IEEE International Conference on Fuzzy Systems*, page 1–8. Barcelona.

Xu, Y., Patnayakuni, R., and Wang, H. (2013). The ordinal consistency of a fuzzy preference relation. *Information Sciences*, 224:152–164.

Yager, R., Ford, M., and Canas, A. (1990). An approach to the linguistic summarization of data. In *Proceedings of the 3rd International Conference of Information Processing and Management of Uncertainty in Knowledge-based Systems (IPMU 1990)*, page 456–468. Paris.

Yager, R. and Rybalov, A. (1996). Uninorm aggregation operators. *Fuzzy Sets and Systems*, 80:111–120.

Yager, R. and Rybalov, A. (1998). Full reinforcement operators in aggregation techniques. *IEEE Transactions on Systems, Man and Cybernetics*, 28:757–769.

Zimmermann, H. (2001). *Fuzzy set theory – and its applications*. London: Kluwer Academic Publishers.

B5
Decision Support Systems

DECISION SUPPORT SYSTEM FOR PRODUCT RANKING BASED ON SENTIMENT ANALYSIS AND MCDM UNDER INTUITIONISTIC FUZZY ENVIRONMENT

SEDEF ÇALI[1], ŞEBNEM YILMAZ BALAMAN[2]

[1] Dokuz Eylul University, Department of Industrial Engineering, Izmir, Turkey, sedefcali@gmail.com
[2] Dokuz Eylul University, Department of Industrial Engineering, Izmir, Turkey, s.yilmaz@deu.edu.tr

Abstract: *This study proposes a methodology integrating aspect level sentiment analysis (SA) and MCDM methods for measurement of customer satisfaction about alternative products through online customer reviews and ranking the products to recommend the best. The alternative products are evaluated according to criteria determined by frequency based aspect extraction. The sentences regarding criteria of alternatives in the customer reviews are classified as positive, negative and hesitant by lexicon-based algorithm and these classification results are transformed to intuitionistic fuzzy numbers to obtain performance scores used in decision matrix. In the ranking process, the new "IF-ELECTRE integrated with VIKOR" method is used. Entropy measure is employed to tackle the uncertainty in determination of criteria weights. The proposed methodology is applied to a case study which deals with online reviews on hotels in Turkish. Although SA algorithm is proposed to analyze Turkish reviews, it can be modified for other languages.*

Keywords: *Big Data, Customer Satisfaction, Multiple Criteria Decision Making, Entropy, Intuitionism*

1. INTRODUCTION

Online reviews are valuable sources for both companies and potential customers. Most of the potential customers tend to search for opinions of experienced users about a product before making purchase decision. On the other hand, companies can organize their production or marketing processes by analyzing negative or positive judgments about their products in the online reviews.

Sentiment analysis (SA) or opinion mining is the computational study that analyses opinionated texts, which contain people's opinions, appraisals, emotions and sentiments toward entities, individuals, topics and their attributes (Liu and Zhang 2012). SA is performed at three different levels; document level SA, sentence level SA and aspect level SA. Document level SA is regarded as identification of main sentiment about an entity from a whole document. Similarly, sentence level SA aims to determine sentiment orientation or polarity of each sentence for an entity. Both of the methods focus on extracting sentiment orientation, but they don't consider the target of sentiment. Aspect level SA studied in this paper determines the sentiment polarity of a sentence according to each aspect of an entity. Considering a review on a specific cell phone, the customer mentions not only his/her general opinion on the cell phone but also opinions on its aspects such as price and battery. The aspects in other words features of products are used as criterion in decision making procedure.

Sentiment analysis has various application area and one of them is product ranking which is also tackled in this paper. This research field has recently drowned the attention of researchers. However, there are still limited number of studies integrating SA with MCDM methods. Kang and Park (2014) proposed a study combining aspect based SA and VIKOR method for measurement of customer satisfaction on mobile service applications. They employed term presence approach to obtain weights of product features and conducted sensitivity analysis to obtain how the rankings are influenced from different customer strategies. Peng et al. (2014) extended opinion mining by integrating fuzzy PROMETHEE to evaluate Chinese reviews on mobile phones and offer to ranking list. They employed text mining techniques in determination of product features, then domain experts assessed the alternatives regarding to each key feature based on linguistic variables expressed by fuzzy numbers. They obtained the weights of criteria by calculating the document frequencies of frequent features in the study. Abirami and Askerunisa (2017) employed aspect level SA to extract customer satisfaction levels of ten hospitals in India from relevant reviews in social media. They utilized from SAW and TOPSIS methods to rank the hospitals and word count method to identify the weights of criteria for ranking process. Liu et al. (2017) is the first study employing intuitionistic fuzzy numbers (IFNs) in obtaining the sentiment scores of alternatives. They analyzed online Chinese reviews on automobiles by classifying the sentences as positive, negative and neutral. Contrary to our study, they labeled a sentence as neutral when it includes no sentiment words or both positive and negative sentiment words. The weights of criteria are identified subjectively and the ranking of products occurred with PROMETHEE II. As for our study, we bring new perspective to product ranking problems based on MCDM and SA by considering entropy weighting. In addition, we analyze hesitation expressions in classification procedure. To the best of our knowledge, this methodology is the first analysis which deals with Turkish reviews for product ranking based on MCDM and SA. The remaining of the paper involves three sections. Section 2 explains the developed methodology. Section 3 illustrates the case study and the results. Section 4 includes conclusion.

2. METHODOLOGY

This section describes the developed system coded by using JAVA computer programming language.

2.1. Data crawling and preprocessing

In this step, customer reviews are collected from the relevant web pages using web crawler. The collected reviews are parsed by ITU Turkish Natural Language Processing Web Service (Eryiğit et al. 2008, Eryiğit 2014) which presents many NLP tools such as tokenization, normalization, spelling correction, morphological analysis, POS tagging and dependency parsing.

2.2. Aspect expression extraction

This step focuses on determination of criteria from the preprocessed data set by employing frequency based aspect extraction method. The tokens with “noun” POS tag are separated since the aspects generally appear in reviews as noun or noun phrase. The separated words with high-frequency are determined. The relevant words are manually grouped together to determine the criteria. **Table 1** shows the criteria and some examples for aspect expressions pointing out the corresponding criteria.

Table 1. The determined criteria with some examples for aspect expressions describing them

Swimming	Staff & Service	Room	Location	Food	Expenditures
havuz(pool)	hizmet(service)	banyo(bath)	merkez(center)	yemek(meal)	ücret(price)
deniz (sea) sahil(beach)	çalışan(personnel garson(waiter)	mobilya(furniture) yalıtım (isolation)	ulaşım(access) …	kahvaltı (breakfast) …	ekstra(extra) …

2.3. Compiling sentiment lexicons

The frequency based method is employed to extract sentiment words and SentiTurkNet (Dehkharghani et al. 2016), the first Turkish sentiment dictionary, is utilized to extend the word lists. The obtained words are categorized according to their polarities and the criterion they describe. The system generally focuses on adjectives and adjective phrases in compiling sentiment lexicons. However, some key sentiment verbs are also considered. The examples for sentiment words and the labels showing their groups are available in **Table 2**.

Table 2. The examples for sentiment words and their labels

Sentiment word or phrase (adjectives)	Label	Sentiment verbs	Label
yardımsever (helpful), güleryüzlü (genial)	servicePos	beğenmek (to like)	generalPos
expensive (pahalı)	expenditureNeg	memnun kalmak (be satisfied with)	generalPos
fena değil (so-so), idare eder (spotty)	hesitantWords	tavsiye etmek (to recommend)	generalPos
kötü (bad), berbat (awful), zayıf (poor)	generalNeg	beğenmemek(to dislike)	generalNeg

2.4. Identifying sentiment orientation of sentences for each criterion

The developed system splits reviews into sentences using punctuations such as full stop, exclamation point, question mark, triple dot. Then, sentiment orientation of each sentence is examined for each criterion. For this purpose, the system searches the aspect expressions and related sentiment words in sentences and saved them in a list named as <AspectExpression, SentimentWords> couple. The polarities of sentiment words are given as a sentiment label to <AE, SW> couple lists. The system looks for negation words affecting each <AE, SW> couple lists. If a negation word or phrase are extracted, the sentiment label of the <AE, SW> couple list are reversed. When identifying <AE, SW> couple lists, the system obeys some Distance Rules and Linguistic Rules. These rules split the sentence into sub-sentences and simplify the classification.

Negation Handling: Negations reverse the sentiment orientations. The expression of "değil" (am/is/are not) is the common negation in Turkish. In addition, this paper considers the words like "olmalı" (must be), "olmamalı" (must not be), "olurdu" (should/would have been), "diyemem" (I cannot say), "olmadığından" (was/were not), "olmayan" (not to be) as negation expressions.

Rule L1: **If** any token "rağmen" (although), "karşın" (despite), "hariç" (except), "dışında" (without) is detected in a sentence **then** the sentence should be spitted into two from this token.

Rule L2: **If** the token whose morphological feature includes "Past", "Pres", "Neces" or "Able" is appeared in a sentence **then** the sentence should be spitted into two from this token.

Abbreviations mentioned in *Rule L2* indicate the suffix of verb and verbal in tense forms (-di/-dı etc.) in modal forms (-meli/-malı etc.). This rule is created for reviews suffering from lack of punctuations.

The sentiment orientation of each sentence regarding each criterion is determined based on following rules:

Rule S1: **If count** (positive labeled <AE, SW> couples) > **count** (negative labeled <AE, SW>) regarding criterion C_j in a sentence **then** this sentence is labeled "positive" for C_j

Rule S2: **If count** (positive labeled <AE, SW> couples) < **count** (negative labeled <AE, SW> couples) regarding criterion C_j in a sentence **then** this sentence is labeled "negative" for C_j

Rule S3: **If count** (positive labeled <AE, SW> couples) = **count** (negative labeled <AE, SW> couples) regarding criterion C_j in a sentence **or** it includes hesitant words **then** this sentence is labeled "hesitant" for C_j

The process of this step is explained with an example review and the structured forms of the review are depicted in **Table 3** and **Table 4.**

Example. "Otel çok pahalı. Hizmet idare eder. Odalar düzenli ve büyük ama mobilyalar eski. Yemekleri lezzetli değil." (The hotel is very expensive. The service is, honestly, spotty. The rooms are tidy and spacious, but the furniture are old. Foods are not delicious.)

Table 3. Structured form of the example review

Review (1)	Sen:	Aspect expression	Related criteria	Sentiment word or phrase	Related label	Negation
	1	fiyat (price)	Expenditure	pahalı(expensive)	expensiveNeg	-
	2	hizmet(service)	Service &Staff	idare eder (spotty)	hesitantWords	-
	3	oda (room)	Room	düzenli (tidy)	roomPos	-
	3	oda (room)	Room	büyük (spacious)	roomPos	-
	3	mobilya (furniture)	Room	eski (poor)	generalNeg	-
	4	yemek (food)	Food	lezzetli (delicious)	foodPos	değil(not)

Table 4. Structured form of the example review-2

Criterion	Number of sentences about the criterion	Number of sentences with positive orientation	Number of sentences with negative orientation	Number of sentences with hesitant orientation
Service &Staff	1	0	0	1
Room	1	1	0	0
Food	1	0	1	0
Expenditure	1	0	1	0

2.5. Transforming the result of aspect level SA to IFNs

IFS theory developed by Atanassov (1986) is a suitable tool to deal with vagueness and hesitancy available in real life situations. IFN which is basic component of the IFS indicates the membership degree (μ), non-membership degree (v) and hesitancy degree of an element to the set. The performance score of each alternative product with reference to each criterion can be expressed by IFNs.

Consider $X_{ij} = (\mu_{ij}, v_{ij})$ as IFN indicating performance score of ith alternative with reference to jth criterion. μ_{ij} indicates satisfaction degree of ith alternative with reference to jth criterion, v_{ij} indicates dissatisfaction degree of ith alternative with reference to jth criterion and $\pi_{ij} = 1 - \mu_{ij} - v_{ij}$ indicates hesitation degree of *i*th alternative with reference to *j*th criterion. The calculations of these indicators can be shown as follows:

$$\mu_{ij} = \frac{p_{ij}}{p_{ij}+n_{ij}+h_{ij}} \quad (1)$$

$$v_{ij} = \frac{n_{ij}}{p_{ij}+n_{ij}+h_{ij}} \quad (2)$$

$$\pi_{ij} = \frac{h_{ij}}{p_{ij}+n_{ij}+h_{ij}} \quad (3)$$

where $\mu_{ij} + v_{ij} + \pi_{ij} = 1$, i = 1, 2,…, m, j =1, 2, …, n,

p_{ij} defines the number of positive sentences regarding *j*th criterion in the reviews of *i*th alternative, n_{ij} expresses the number of negative sentences regarding *j*th criterion in the reviews of *i*th alternative, h_{ij} defines the number of hesitant sentences regarding *j*th criterion in the reviews of *i*th alternative.

2.6. Ranking the products based on IF-ELECTRE integrated with VIKOR method

A novel MCDM methodology is employed in this section. This method integrates ELECTRE I (Roy 1968) with VIKOR (Opricovic 1998) by considering the advantage of VIKOR in complete ranking and the advantage of ELECTRE I in pairwise comparison. The implementation of IF-ELECTRE integrated with VIKOR method to this problem is explained step by step.

Step 1. Establish IF group decision matrix: Let $C = \{c_1, c_2, \dots, c_n\}$ be the set of criteria and $A = \{A_1, A_2, \dots, A_m\}$ be the set of alternatives. IF-decision matrix of R brings together IFNs expressing the performance scores of each alternative with reference to each criterion calculated in previous section. The rows of decision matrix indicate each alternative, while the columns of decision matrix indicate each criterion. For instance, $r_{11} = (\mu_{11}, v_{11})$ represents the performance score of A_1 with reference to C_1.

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{bmatrix} = \begin{bmatrix} (\mu_{11}, v_{11}) & \cdots & (\mu_{1n}, v_{1n}) \\ \vdots & \ddots & \vdots \\ (\mu_{m1}, v_{m1}) & \cdots & (\mu_{mn}, v_{mn}) \end{bmatrix}$$

Step 2. Calculate the weights of criteria: Entropy measure of the criterion C_j reveals the information level transmitted by the criterion C_j. The less entropy measure has a criterion, the more relative importance it has. We employ the IF-entropy measure developed by Vlochos and Sergiadis (2007) to calculate the weights of criteria using following steps (Hung and Cheng 2009). Firstly, the entropy measures are obtained via Eq. (4).

$$E_{LT}^{IFS}(C_j) = -\frac{1}{m \ln 2} \sum_{i=1}^{m} [\mu_{ij} \ln \mu_{ij} + v_{ij} \ln v_{ij} - (1 - \pi_{ij}) \ln(1 - \pi_{ij}) - \pi_{ij} \ln 2] \quad (4)$$

After determination of divergence degrees d_j with Eq. (5), weights of criteria are computed with Eq. (6).

$$d_j = 1 - E_{LT}^{IFS}(C_j) \quad (5)$$

$$w_j = \frac{d_j}{\sum_{j=1}^{n} d_j} \quad \text{where } w_j \geq 0,\ j = 1,2, \dots n \text{ and } \sum_{j=1}^{n} w_j = 1. \quad (6)$$

Step 3. Identify the concordance and discordance sets: Three different concordance and discordance sets for each pair of alternatives are formed inspired from Wu and Chen (2011). The strong, midrange and weak concordance sets are formulated respectively as follows:

$$C'_{ab} = \{j | \mu_{aj} \geq \mu_{bj},\ v_{aj} < v_{bj} \text{ and } \pi_{aj} < \pi_{bj}\} \quad (7)$$

$$C''_{ab} = \{j | \mu_{aj} \geq \mu_{bj},\ v_{aj} < v_{bj} \text{ and } \pi_{aj} \geq \pi_{bj}\} \quad (8)$$

$$C'''_{ab} = \{j | \mu_{aj} \geq \mu_{bj},\ v_{aj} \geq v_{bj}\} \quad (9)$$

The strong, midrange and weak discordance sets are formulated respectively as follows:

$$D'_{ab} = \{j | \mu_{aj} < \mu_{bj},\ v_{aj} \geq v_{bj} \text{ and } \pi_{aj} \geq \pi_{bj}\} \quad (10)$$

$$D''_{ab} = \{j | \mu_{aj} < \mu_{bj},\ v_{aj} \geq v_{bj} \text{ and } \pi_{aj} < \pi_{bj}\} \quad (11)$$

$$D'''_{ab} = \{j | \mu_{aj} < \mu_{bj},\ v_{aj} < v_{bj}\} \quad (12)$$

Step 4. Compute the weights of concordance sets: We employed weighted distance approach adapted from (Zhang et al. 2017) for identifying the total dominance degrees of strong, midrange and weak concordance sets. The dominance degree of alternative A_a over alternative A_b in respect to criterion C_j, a member of concordance set, can be calculated with weighted distance measure $w_j * d(r_{aj}, r_{bj})$. The total dominance degrees of strong, midrange and weak concordance sets are defined as $dd' = \sum_{\substack{a,b=1 \\ a \neq b}}^{m} \sum_{j \in C'_{ab}} w_j * d(r_{aj}, r_{bj})$, $dd'' = \sum_{\substack{a,b=1 \\ a \neq b}}^{m} \sum_{j \in C''_{ab}} w_j * d(r_{aj}, r_{bj})$ and $dd'' = \sum_{\substack{a,b=1 \\ a \neq b}}^{m} \sum_{j \in C''_{ab}} w_j * d(r_{aj}, r_{bj})$,respectively.

Euclidean distance measure (Szmidt and Kacprzyk 2000) is used for calculations of distances between IFSs. The weights of strong, midrange and weak concordance sets are calculated based on total dominance degrees of these sets.

$$w_{C'} = dd'/(dd' + dd'' + dd''') \tag{13}$$
$$w_{C''} = dd''/(dd' + dd'' + dd''') \tag{14}$$
$$w_{C'''} = dd'''/(dd' + dd'' + dd''') \tag{15}$$

Let A and B be IFSs in $X = \{x_1, x_2 \dots x_n\}$, the Euclidean distance $d_E(A, B)$ is calculated with Eq. (16)

$$d_E(A, B) = \sqrt{\frac{1}{2} \sum_{j=1}^{n} \left(\mu_A(x_j) - \mu_B(x_j)\right)^2 + \left(v_A(x_j) - v_B(x_j)\right)^2 + \left(\pi_A(x_j) - \pi_B(x_j)\right)^2} \tag{16}$$

Step 5. Establish the concordance matrix: Concordance matrix is made up of concordance indices calculated on the basis of comprehensive concordance index (Figueira et al. 2010). It reflects the superiority of one alternative to another alternative. The concordance index g_{ab} between $A_a$and A_b is identified as:

$$g_{ab} = w_{C'} * \sum_{j \in C'_{ab}} w_j + w_{C''} * \sum_{j \in C''_{ab}} w_j + w_{C'''} * \sum_{j \in C'''_{ab}} w_j \tag{17}$$

Step 6. Establish the discordance matrix: Discordance matrix is composed of discordance indices which indicate the inferiority degree of alternatives according to each other. The discordance index h_{ab} between $A_a$and A_b is identified with Eq. (18) (Zhang et al. 2017).

$$h_{ab} = \frac{\max\limits_{j \in D'_{ab} \cup D''_{ab} \cup D'''_{ab}} w_j * d(r_{aj}, r_{bj})}{\max\limits_{j \in J} d(r_{aj}, r_{bj})} \tag{18}$$

Step 7. Compute the values S (group utility) and R (individual regret) for each alternative: In this step, VIKOR method can be integrated to the system. Opricovic and Tzeng (2007) presented the similarities of discordance concept of ELECTRE and R_i in VIKOR as well as the similarities of concordance concept of ELECTRE and S_i in VIKOR. They developed aggregating discordance index and aggregating concordance index. With the help of these indices, the complete ranking is realized.

The aggregating discordance index of alternative A_a is calculated as follows:

$$h_a = R_a / C \tag{19}$$
$$C = \max_j w_j \text{ and } h_a = \max_i h_{ai} \quad i \neq a \quad i = \{1,2, \dots m\} \tag{20}$$

The aggregating concordance index of alternative A_a is calculated as follows:

$$g_a = 1 - S_a \tag{21}$$
$$g_a = \sum_{i \neq a}^{m} g_{ai} / (m - 1) \tag{22}$$

where $g_{ai} = w_{C'} * \sum_{j \in C'_{ai}} w_j + w_{C''} * \sum_{j \in C''_{ai}} w_j + w_{C'''} * \sum_{j \in C'''_{ai}} w_j$

where alternatives $i = \{1,2, \dots m\}$ and criteria $j = \{1,2, \dots n\}$

Therefore, the values R and the values S of each alternative i is calculated as follows:

$$R_i = h_i * C \tag{23}$$
$$S_i = 1 - g_i \tag{24}$$

Step 8. Compute the values Q (degree of closeness) for each alternative:

$$Q_i = \gamma * \frac{S_i - S^*}{S^- - S^*} + (1 - \gamma) * \left(\frac{R_i - R^*}{R^- - R^*}\right) \qquad i = 1,2, \dots m \tag{25}$$

where $S^- = max\ S_i$; $S^* = min\ S_i$; $R^- = maxR_i$; $R^* = minR_i$; and the notation of γ indicates a weight for the strategy of maximum group utility, whereas $(1-\gamma)$ indicates a weight of the individual regret.

Step 9. Rank the alternatives, sorting by the values S, R and Q in decreasing order: Three different ranking lists are formed based on ascending order of values Q, S and R. The compromise solution is the alternative with minimum Q value only if the following conditions are satisfied:

Condition 1: Consider A_1 is the first and A_2 is the second according the ranking of Q. Alternative A_1 has a meaningful advantage over the alternative A_2 by verifying the following equation:

$$Q(A_2) - Q(A_1) \geq DQ \text{ where } DQ = 1 / (m-1) \tag{26}$$

Condition 2: The alternative A_1 has acceptable stability in decision making if it is in the first position according to both ranking of S values and ranking of R values.

If alternative A_1 satisfies the only one condition, then a set of compromise solutions is proposed. If Condition 2 is not satisfied, then A_1 and A_2 are the compromise solutions. If Condition 1 is not satisfied, then A_1, A_1,… A_m are the compromise solutions where $Q(A_m) - Q(A_1) < DQ$ and $DQ = 1 / (m-1)$.

3. CASE STUDY

In this step, the developed system is applied to hotel reviews. We have analyzed the reviews on five hotels with five stars located in Çeşme, one of the most popular touristic counties in Turkey. The alternative hotels with their notations are H1: Boyalık Beach Hotel & Spa, H2: Sheraton Cesme Hotel Resort & Spa, H3: Ilıca Hotel Spa & Wellness Resort, H4: Altın Yunus Resort & Thermal Hotel, H5: Grand Hotel Ontur. We have crawled the reviews from a hotel evaluation website otelpuan.com (https://www.otelpuan.com/tr) in July 2017. The number of reviews are 338, 317, 320, 294 and 390, respectively. We have determined the criteria at aspect expression extraction step; C1: Swimming, C2: Staff & Service, C3: Room, C4: Location, C5: Food, C6: Expenditures. The aspect level SA is conducted and the distribution of opinionated sentences can be seen in **Fig. 1**.

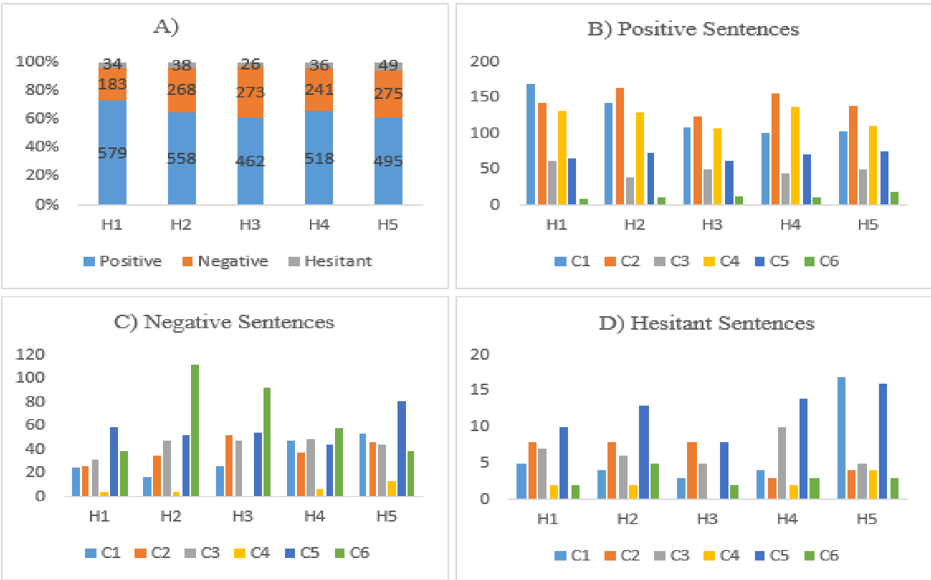


Figure 1. The distribution of opinionated sentences

Fig. 1(A) reveals that H1 has the highest rate for positive sentences, H3 has the highest rate for negative sentences and H5 has the highest rate for hesitant sentences. The total number of opinionated sentences are 4034. **Fig. 1**(B) points out that the customers of H1 have mostly satisfied with the criterion *Swimming* of the hotel and the customers of other hotels have mostly satisfied with the criterion *Staff & Service*. **Fig. 1**(C) proposes that the customers of H2, H3 and H4 have generally complained about the criterion *Expenditures* and the customers of H1 and H5 have generally complained about the criterion *Food*. Therefore, H2 should revise the prices like “extra” H5 should improve the criterion *Food* in order to increase their customer satisfaction levels. According to **Fig. 1**(D), the customers of H5 have expressed hesitation opinions on the criterion *Swimming* whereas the customers of the other hotels have expressed hesitation opinions on the criterion *Food*. The hotel H5 can enhance the facilities regarding criterion *Swimming* to attract the attention of hesitant customers. After determination of opinionated sentences, they are transformed into IFNs with Eq. (1), Eq. (2) and Eq. (3) to obtain performance scores of each hotel regarding each criterion. Hence, IF-group decision matrix is formed which can be seen at **Table 5**.

Table 5. IF-group decision matrix

R=	C1	C2	C3	C4	C5	C6
H1	(0.8492, 0.1256)	(0.8079, 0.1468)	(0.6200, 0.3100)	(0.9565, 0.0289)	(0.4850, 0.4402)	(0.1666, 0.7916)
H2	(0.8711, 0.1042)	(0.7922, 0.1690)	(0.4193, 0.5161)	(0.9558, 0.0294)	(0.5289, 0.3768)	(0.0787, 0.8818)
H3	(0.7883, 0.1897)	(0.6721, 0.2841)	(0.4903, 0.4615)	(0.9906, 0.0093)	(0.5000, 0.4354)	(0.1132, 0.8679)
H4	(0.6644, 0.3092)	(0.7948, 0.1897)	(0.4326, 0.4711)	(0.9444, 0.0416)	(0.5503, 0.3410)	(0.1408, 0.8169)
H5	(0.5930, 0.3081)	(0.7354, 0.2433)	(0.5050, 0.4444)	(0.8671, 0.1015)	(0.4360, 0.4709)	(0.3050, 0.6440)

Then, we have employed entropy concept to obtain weights of criteria. The process of determination of weights is realized with Eq. (4), Eq. (5) and Eq. (6). The weights of criteria are obtained as 0.1643, 0.1518, 0.0105, 0.4476, 0.0072 and 0.2186, respectively. It can be noticed that the most important criterion is *Location* for this dataset. We have applied the remaining steps of IF-ELECTRE integrated with VIKOR method to obtain product ranking: The strong, midrange and weak concordance/discordance sets are specified based on Eq. (7-12), respectively. The weights of concordance sets are calculated using Eq. (13), Eq. (14) and Eq. (15). The weight of strong concordance set $w_{C'}$is 0.58, the weight of midrange concordance set $w_{C''}$ is 0.40 and the weight of weak concordance set $w_{C'''}$ is 0.02. After each of concordance indices is calculated with Eq. (17), the concordance matrix G is constructed. After each of discordance indices is computed with Eq. (18), the discordance matrix H is designed.

$$G = \begin{bmatrix} - & 0.4124 & 0.2184 & 0.4684 & 0.4244 \\ 0.0982 & - & 0.1568 & 0.2747 & 0.4189 \\ 0.2640 & 0.3928 & - & 0.3613 & 0.3594 \\ 0.0029 & 0.0975 & 0.1785 & - & 0.3539 \\ 0.0876 & 0.0936 & 0.1798 & 0.0936 & - \end{bmatrix} H \begin{bmatrix} - & 0.0175 & 0.0936 & 0.0034 & 0.1370 \\ 0.0957 & - & 0.1150 & 0.0675 & 0.2036 \\ 0.1460 & 0.1517 & - & 0.1388 & 0.2186 \\ 0.1643 & 0.1643 & 0.1643 & - & 0.2186 \\ 0.1643 & 0.1643 & 0.2373 & 0.1863 & - \end{bmatrix}$$

This concordance and discordance matrices are used for obtaining the characteristics of VIKOR. The values of R_i indicating individual regret of opponent are determined based on Eq. (19, 20 and 23). The values of S_i indicating group utility of majority are determined based on Eq. (21, 22, and 25).

If decision maker assigns γ = 1, in other words, focuses on strategy of maximum group utility (min S_i), the ranking is obtained as H1>H3>H2>H4>H5. If decision maker highlights the importance of minimum individual regret (min R_i) by assigning γ = 0, the ranking is obtained as H1>H2>H3=H4>H5. If decision maker assigns γ = 0.5 by considering consensus, the ranking is obtained as H1>H3>H2>H4>H5. Consequently, H1 is recommended to decision maker by the developed system.

We have compared the result of developed system with customer ratings interval [0, 10] in the website. The average points given to hotels, respectively, 8.1, 7.8, 7.6, 7.6 and 7.4. It can be concluded that H1 is the first hotel according to customer ratings as well.

4. CONCLUSION

This paper proposes a decision support tool which can be used by potential customers and companies. The companies can measure and evaluate the customer satisfaction level in Customer Relationship Management by using the developed methodology. The potential customers can obtain the ranking lists for their alternatives according to this evaluation and make an optimal purchase decision. This paper differentiates from other studies in the literature by considering hesitation expressions indicating hesitancy of customer on products in the reviews for more reliable sentiment classification. Moreover, the methodology employs an innovative approach by employing the entropy method for obtaining the weights of criteria since it is difficult to determine realistic weights for criteria from the online reviews. For the future research, the semi-automatic SA will be transformed to complete automatic form by employing topic modeling methods in identifying the criteria.

Acknowledgement

This work was supported by Dokuz Eylül University Scientific Research Coordination Unit. Project Number: 2017.KB.FEN.026

REFERENCES

Abirami, A.M., & Askarunisa, A. (2017). Sentiment analysis model to emphasize the impact of online reviews in healthcare industry. *Online Information Review*, 41(4), 471-486.

Atanassov, K.T. (1986). Intuitionistic fuzzy sets. *Fuzzy sets and Systems*, 20(1), 87-96.

Dehkharghani, R., Saygin, Y., Yanikoglu, B., & Oflazer, K. (2016). SentiTurkNet: a Turkish polarity lexicon for sentiment analysis. *Language Resources and Evaluation*, 50(3), 667-685.

Eryiğit, G. (2014). ITU Turkish NLP Web Service, In *Proceedings of the Demonstrations at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014)*, Sweden, 1-4.

Eryiğit, G., Nivre, J., & Oflazer, K. (2008). Dependency parsing of Turkish. *Computational Linguistics*, 34(3), 357-389.

Figueira, J.R., Greco, S., Roy, B., & Słowiński, R. (2010). ELECTRE methods: main features and recent developments. In Zopounidis, C. & Pardalos, P.M. (Eds.), *Handbook of multicriteria analysis*, Berlin: Springer, 51-89.

Hung, C.C., & Chen, L.H. (2009). A multiple criteria group decision making model with entropy weight in an intuitionistic fuzzy environment. In Castillo, H. & Huang, H. (Eds.), *Intelligent automation and computer engineering.* Netherlands: Springer, 17-26.

Kang, D., & Park, Y. (2014). Review-based measurement of customer satisfaction in mobile service: sentiment analysis and VIKOR approach. *Expert Systems with Applications*, 41(4), 1041-1050.

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In Aggarwal, C. & Zhai, C. (Eds.), *Mining text data*, Boston: Springer, 415-463.

Liu, Y., Bi, J.W., & Fan, Z.P. (2017). Ranking products through online reviews: a method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion*, 36, 149-161.

Opricovic, S. (1998). Multicriteria optimization of civil engineering systems. Faculty of Civil Engineering, Belgrade, 2(1), 5-21.

Opricovic, S., & Tzeng, G.H. (2007). Extended VIKOR method in comparison with outranking methods. *European Journal of Operational Research*, 178(2), 514-529.

Peng, Y., Kou, G., & Li, J. (2014). A fuzzy promethee approach for mining customer reviews in Chinese. *Arabian Journal for Science and Engineering*, 39(6), 5245-5252.

Roy, B. (1968). Classement et choix en presence de points de vue multiples (La method ELECTRE). *Revue Francaise d'Informatique et de Recherche Operationnelle*, 8, 57–75.

Szmidt, E., & Kacprzyk, J. (2000). Distances between intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 114, 505-518.

Vlachos, I. K., & Sergiadis, G. D. (2007). Intuitionistic fuzzy information – Applications to pattern recognition. Pattern Recognition Letters, 28(2), 197–206.

Wu, M.C., & Chen, T.Y. (2011). The ELECTRE multicriteria analysis approach based on Atanassov's intuitionistic fuzzy sets. *Expert Systems with Applications*, 38(10), 12318–12327.

Zhang, H.Y., Peng, H.G., Wang, J., & Wang, J.Q. (2017). An extended outranking approach for multi-criteria decision-making problems with linguistic intuitionistic fuzzy numbers. *Applied Soft Computing*, 59, 462–474.

BALCOR 2018
XIII Balkan Conference on Operational Research

DATA STORAGE AND REPRESENTATION FOR STUDYING PROBLEMS OF PROSPECTIVE ELECTRIC POWER SYSTEMS

IVAN TROFIMOV[1], LEONID TROFIMOV[1], SERGEI PODKOVALNIKOV[1]

[1] Melentiev Energy Systems Institute Siberian Branch of the Russian Academy of Sciences, Department of Electric Power Systems. Irkutsk, Russia. t_john88@isem.irk.ru

---

Abstract: *The article describes the software tool implemented by Melentiev Energy Systems Institute SB RAS and aimed to solving problems related to study prospective electric power systems. In this paper, the Data Processing and Geo-Information System (DPGIS) means a software tool that provides collection, transfer, processing, storage and output of digital technical and economic data of different power entities. These entities are represented in DPGIS as objects of special database. Low-level database objects are power plants. Top-level database objects are countries, power systems, and interstate power interconnections. The paper discusses the example of how data storage and data representation in object-oriented database assists to improve efficiency of studied issues of prospective electric power systems expansion.*

Keywords: *geo-information system, optimization model, data processing, object-oriented database, power plant, electric power balance, electric power system.*

1. INTRODUCTION

In research of prospective electric power systems expansion, the most laborious process is the processing of large volumes of data, collected from various Internet sources & for different points and frames of time.

Similar problems of the data collection and processing concern researchers in many fields (Xuming, 2018). A lot of universal and specialized software were developed. In most cases, relational databases are using for the data storage. To process multidimensional data sets OLAP, Data Mining and others methods are used.

The issues of prospective electric power systems expansion and forecasting of electric power balances requires the data completeness, many mathematical, economic and special data transformations. To solve such problems, the using of universal methods for data storage and processing are not effective. Moreover, in many cases – it is impossible.

The most popular database management systems such as Firebird, Oracle, MS SQL and geo-information systems such as ArcGIS, MapInfo, Google Maps etc. do not assist researchers to solve specific issues. The problems dedicated to specific calculations of efficiency, forecasting and modeling of complex entities (power plants and power systems for the region and country levels), are described by a large set of various dynamic information. For this reason, a specialized decision support system[1] is necessary to study problems of prospective electric power systems.

The authors propose an original technique of data storage and processing by object-oriented database. A specialized Data Processing and Geo-Information System (DPGIS) based on object-oriented database has been developed.

DPGIS is integrated with General Algebraic Modeling System[2], and aimed to solving wide range of energy problems (Podkovalnikov, 2018).

DPGIS infrastructure includes the following program modules.

1) The first interface unit is to work with optimization model ORIRES[3], which is used to study and forecast of prospective power systems and interstate power interconnections (Beliayev, 2002).

---

[1] Decision Support System (DSS) is an information system that supports business or organizational decision-making activities for experts.
[2] The General Algebraic Modeling System (GAMS) is a high-level modeling system for mathematical optimization. GAMS is designed for modeling and solving linear, nonlinear, and mixed-integer optimization problems. The system is tailored for complex, large-scale modeling applications and allows the user to build large maintainable models that can be adapted to new situations.
[3] The ORIRES model is an optimization, linear, static and multi-node one. It optimizes the volumes and the mix of generating capacities, the transfer capacity of transmission lines, the operating modes of power plants and flows among nodes (electric power systems), for the forecasted target year.

2) The second interface unit is to study and forecast fuel and energy/power balances.

3) The cartographic unit is to construct interactive maps, with possible combination of various energy/power data diagrams, for different points and frames of time (Trofimov, 2017).

4) The graph unit is to create complex graphs and diagrams for comparison and analysis of retrospective energy statistics of different countries and regions.

5) The special Query Designer unit is to make and display various data tables, with built-in mathematical transformations (formulas for energy/power measures translating, interpolating missing values, maximum and minimum functions, data aggregation, and others).

The paper describes the example of the data storage and representation technique based on object-oriented database, which improves the efficiency of study and forecast of prospective electric power systems.

2. THE DATA STORAGE TECHNIQUE BASED ON OBJECT-ORIENTED DATABASE

The authors offer a new approach to organization of data storage and representation. We have developed an object-oriented database (OODB). Low-level database objects are power plants described by technical and economic data. Top-level database objects are countries, energy/power systems, and interstate power interconnections.

Each database object describes the properties of certain entities (power plants, power systems, countries etc.). OODB's object is a separate file, containing a set of properties and values of physical entity – energy/power parameters and values, grouped by year, day and hour. In the same directory, there is an additional file, which contains a metadata about object types and others. Any text editor can be used to view OODB files system format (Trofimov, 2015).

The content of database objects is represented by dynamic editable tables in the interface unit to work with OODB. The table contains object parameters and its values, see **Table 1.** The edited values are automatically recorded to object format.

Table 1: A fragment of data on the selected OODB's object.

Parameter type	Parameter name	Values stored by years			
		1990	1991	1992	…
type	country				
ID	Russia				
imp_coal	Coal import (ktoe)	35087 | 28011 | 23719 | …			
imp_oil	Oil import (ktoe)	18592 | 71280 | 30203 | …			
exp_coal	Coal export (ktoe)	40423| 25525 | 27867 | …			
exp_oil	Oil export (ktoe)	222907 | 175271 | 143161 | …			
el_prod	Electricity production (TWh)	1082 |1069 | 1062 | …			
...	…	…			

The technique of data storage and representation allows users to compactly store the whole data related to the certain entity. In addition, it is not necessary to create a set of auxiliary tables and indexes.

These entities are represented in OODB as objects. *Text properties* of objects are names, metadata, and links to external media (photos, videos and others). *Numerical properties* of objects are data that can be separately store by year, day and hour – it depends on objects type. In this case, all necessary data that describe the entity are collected in one special format file. This helps to visually determine the integrity of information and avoids the potential data loss/fragmentation (data distribution among many related tables). By special verification and data processing procedures, user can make various trends of numerical parameters values, even when object's name can differ for different sources. For example, on different Internet sites, the same power plant may have different name spellings.

Object-data representation allows user to compactly store large data of different object types, data collected for different points and frames of time, and from any sources. Special data processing procedures include methods for objects verification, algorithms for interpolating missing data, mathematical formulas for converting and combining various parameters. Work experience with various databases shows that such procedures are rather difficult to implement in standard SQL queries. Presented at the **Fig. 1** is OODB flowchart.

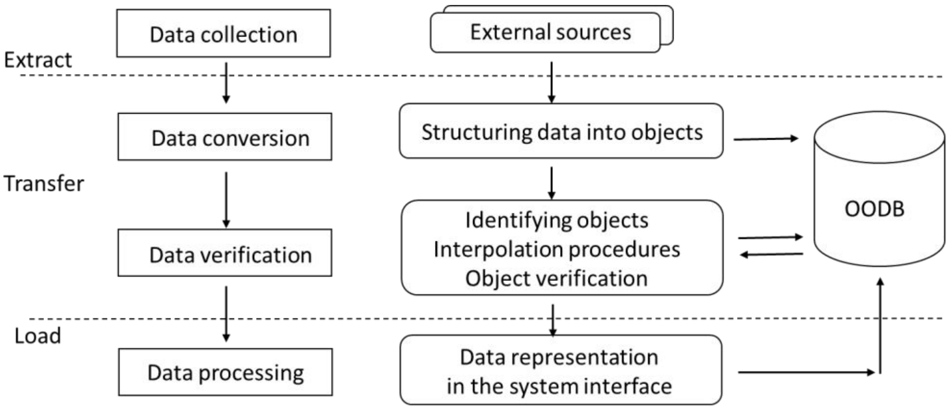


Fig.1. OODB flowchart

OODB operation consists of the following stages:

1. collection of various non-systematized data from open sources;
2. data restructuring to OODB format;
3. data extracting from OODB;
4. data processing in DPGIS interface – making interactive maps, graphs and diagrams, with the possibility of retrospective and prospective data analysis;
5. export of processed data to external Analytical Internet Service ("Energy Statistical Analytical Service," 2017) in order to present results of the research for the scientific community and attracting the interest of the world community.

3. THE DATA PROCESSING TECHNIQUE

A special Query Designer for DPGIS was developed (Trofimov, 2014). With the Query Designer unit, users do not operate with SQL query concepts, but only need to create a table with required fields. Users do not depend on programmers, and they do not need specialized knowledge to construct complex SQL queries, see **Fig. 2**. A special query technique consists of the following stages:

1) building a query by using a set of special energy field concepts;
2) reading/extracting data from OODB and performing the necessary mathematical data transformations;
3) obtaining results.

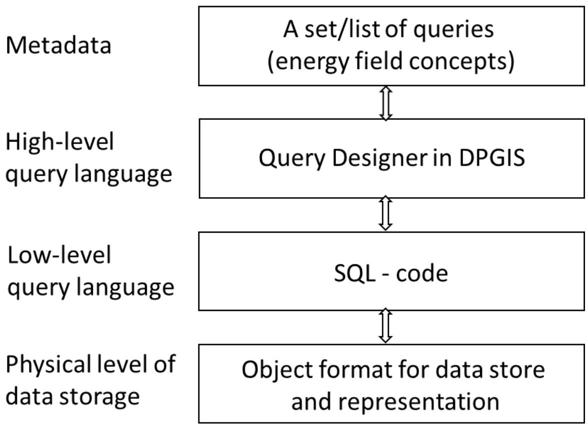


Fig. 2. Query levels to the database

This approach allows users to generate comprehensive and parametrized queries, contains special calculations, such as linear interpolation (*IL function*) of missing data of certain year, and other mathematical data transformations (addition, subtraction, multiplication, division, etc.), see **Fig. 3**.

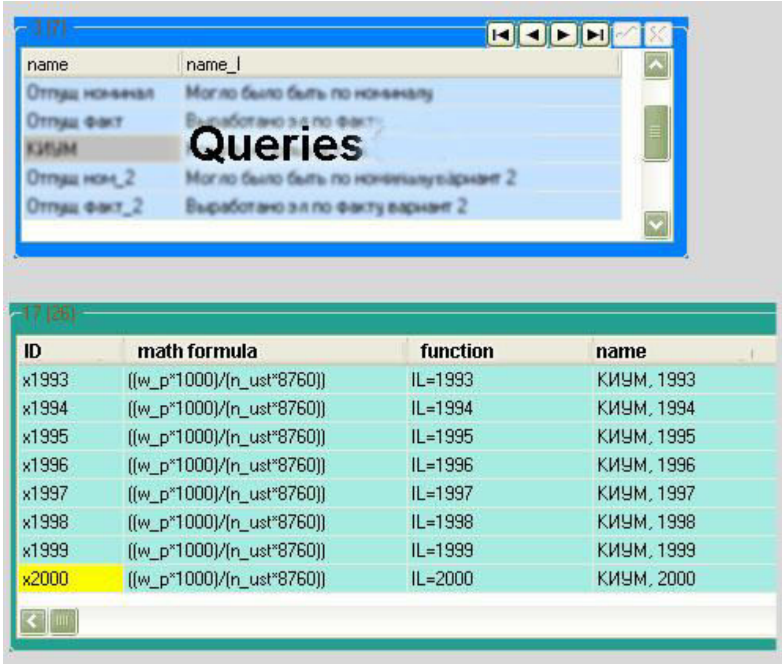


Fig. 3. Query fragment in the DPGIS interface

The Query Designer unit allows users to get effective access to data without any assistance of programmers, without special skills to work with databases and specific knowledge for building SQL queries.

4. THE GRAPH AND THE CARTOGRAPHIC UNIT INTERFACES USAGE

For comprehensive studies of prospective electric power systems expansion, it is necessary to conduct an express analysis of fuel and energy/power balances (energy resources, energy consumption, fuel mix, etc.).

We have considered an energy balance statistics distributed by the IEA ("International Energy Agency: Data Services," 2016). Information provided by the IEA contains retrospective data of energy balances of almost all countries in the world (more than 200 countries), starting from the 1970s. A multidimensional cube (OLAP-data representation) represents the data, see **Fig. 4.**

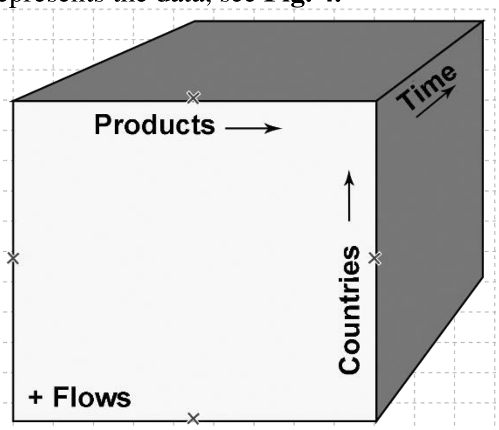


Fig. 4. Multidimensional data cube for energy balances of countries

One of cube sections is energy balance of the country, for any year. Access to this data is possible only through a special application that allows users to view any slices of this cube. Data processing is limited by the application capabilities. In addition, to study electric power systems, it is necessary to choose only the electric power component from a huge array of data.

DPGIS allows us to collect and restructure input data into the object-oriented storage. The entire volume of the above information was transferred into OODB. It was about 200 objects in OODB (by number of countries) with energy objects indicators.

We have developed special cartographic unit to make interactive maps with energy objects indicators both for retrospective data and for calculated prospective years.

The indicators display as pie charts on the map, for any region in the world. These maps are created in automatic mode in the DPGIS unit and we can view indicator's changes for different years. In addition, there is a mode of exporting constructed maps to a special Web application ("Energy Statistical Analytical Service," 2017).

For example, the map shows *electricity consumption* for all countries in the world, see **Fig. 5.** Electricity consumption is given in TWh – Terawatt hour. Sizes of pie charts are proportional to amount of electricity consumed.

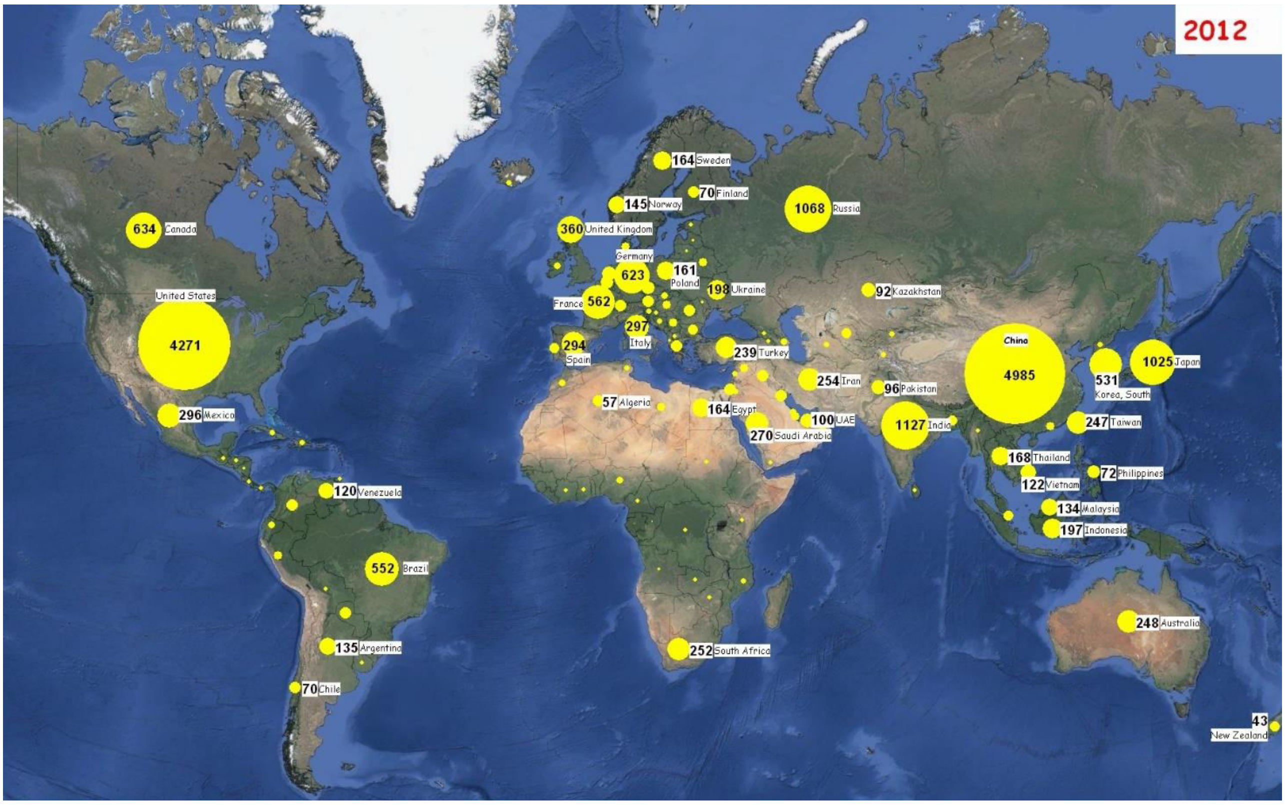


Fig. 5. World electricity consumption, 2012 (TWh)

Another example of data representation by DPGIS interface is a graphical form. The same indicator of electricity consumption can be presented on the combined graph, reflecting the growth dynamics of ten largest electricity-consumed countries, starting from 1990s, see **Fig. 6.**

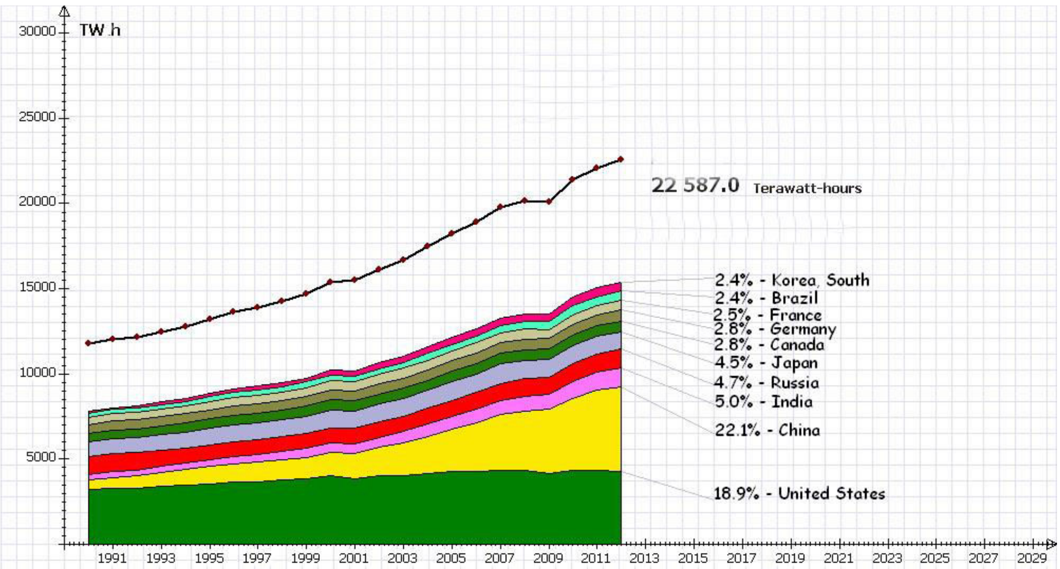


Fig. 6. Graph of 10th largest electricity consumers in the world

Visual analysis of the graph allows us to understand better regularities and changes occurring in the course of the observation. As can be seen from the **Fig.6**, in 2012 China and the US together consumed 41% of the total electricity in the world. United States electricity consumption increased quite modestly (from 3203 TWh in 1990 - to 4271 TWh in 2012), at the same time electricity consumption in China increased by more than 8 times (from 594 TWh in 1990 - to 4985 TWh in 2012).

In addition, OODB contains many other data collected from various sources. For example, data collected from open resource Enipedia.org ("Energy Pedia Portal", 2017), that contains technical and economic indicators for different countries. These indicators are capacity of power plants, geographic coordinates, volumes of electric power production and consumption, exports and imports, etc. The cartographic unit allows users to make power plant maps with coordinates, connections, volumes, capacity types and others.

The DPGIS also allows us to make the forecast of these indicators for the prospective target year.

5. EXAMPLE OF THE ORIRES MODEL USAGE

To study and forecast of prospective electric power systems expansion we use the ORIRES optimization model. The main objective of the model is to study the prospective expansion and operation of interstate power interconnections for a target prospective year. To attain this objective, many different kinds of optimization and estimation calculations are conducted (Beliayev, 2002). Detailed description of the model is beyond the area of this paper. Here it is described only in a nutshell.

The ORIRES model is *to determine the optimum installed capacity of power plants and the amount of electricity generated, for all types of power plants, for each node of the model, and for prospective target year*. As well as the optimal values of electricity transfer between nodes. The nodes are electric power systems or countries.

The objective function of the model consists of the total annual cost of the fuel component; the operational and investment cost of electricity generation; and the operational and investment cost of transmission lines. The optimal solution is determined by minimum of these costs.

A set of constraints (balance equations) is specified:

a) electricity generation with exports and imports should not exceed the given load;

b) the installed capacity for the target year must be within the specified limits;

and other operational constraints.

DPGIS is integrated with General Algebraic Modeling System GAMS, which solves equations of the model. We give an example of forecasted electric power balance of Northeast Asia (NEA) power systems for 2035 (Chudinova, 2018). **Table 2** shows electricity flows among nodes. **Table 3** shows power generation mix by type of power plants (PP).

Table 2: Electricity flows among nodes (TWh per year).

No	Power system	Electricity generation, TWh	Charge of pumped storage PP, TWh	Outflow, TWh	Inflow, TWh	Loss, TWh	Total consumption, TWh
1	East power system of Russia	**70**		-4	7		**73**
2	Siberian power system of Russia	**345**		-8	0		**338**
3	Japan	**1203**	-17				**1186**
4	North Korea	**80**					**80**
5	South Korea	**850**	-4			-1	**845**
6	Mongolia	**22**		0.4	1	-1	**22**
7	North China power system	**2912**	-8	0.1	3		**2907**
8	North-East China power system	**1171**	-14	-3	4	-17	**1141**
9	Sakhalin power system	**6**					**6**
	Total	**6661**	**-44**	**-15**	**15**	**-20**	**6597**

Table 3: Power generation mix of Northeast Asia power systems (TWh per year).

No	Power system	Coal, TWh	Gas, TWh	Nuclear TWh	Oil, TWh	Hydro, TWh	Pumped TWh	Wind, TWh	Solar, TWh	**Electricity generation, TWh**	**Total consumption, TWh**
1	East power system of Russia	22	13			35				**70**	**73**
2	Siberian power system of Russia	125	16	16		188				**345**	**338**
3	Japan	303	478	33	125	186	14	20	44	**1 203**	**1186**
4	North Korea	13	20		4	41		1	1	**80**	**80**

5	South Korea	344	156	273	4	14	3	42	13	**850**	**845**
6	Mongolia	18				3		1	0.1	**22**	**22**
7	North China power system	2 532	60	179		24	7	71	40	**2 912**	**2 907**
8	North-East China power system	819	1	238		39	10	52	12	**1 171**	**1 141**
9	Sakhalin power system	2	4							**6**	**6**
	Total	**4 179**	**748**	**741**	**132**	**495**	**33**	**187**	**110**	**6 661**	**6 597**

In this example, the ORIRES model allows one to obtain optimal electricity generation mix by type of power plants for the prospective target year.

5. CONCLUSION

Among constantly changing databases and information systems offered on the market, there are specialized decision support systems, especially in case when it is necessary to have a compact data structure and transparently store information. In this case, the procedures for extracting, processing and data visualizing must have a universal interfaces that can be customized for various issues related to study prospective electric power systems.

The authors propose a new, modern approach to analyze and process a large volume of data. The main components (units) of DPGIS related to solving problems of energy/power systems expansion have been developed.

The data storing technique based on object-oriented database, and data processing by graphical interfaces of DPGIS, considered in this paper, are universal for solving many energy issues. The proposed technique greatly simplifies calculations, reduces time for making forecasts of energy/power balances.

DPGIS is a kind of decision support system, and related to improve efficiency of studied problems of prospective electric power systems and improve quality of optimization models solutions by using convenient interfaces.

The paper was carried out under financial support of the RFBR grant (Russian Foundation for Basic Research) No. 18-07-00495 A.

REFERENCES

Belyaev, L., Chudinova, L.,Khamisov, O., Kovalev, G., Lebedeva, L., Podkovalnikov, S., Savelyev, V. (2002). Studies of Interstate Electric Ties in Northeast Asia. *International Journal of Global Energy Issues*, 17 (2), 228-249.

Chudinova, L., Podkovalnikov, S., & Trofimov, I. (2018). Multilateral cooperation for power interconnection in Northeast Asia. In *Proceedings of the 10th International Conference on Asian Energy Cooperation (AEC 2017)*, Irkutsk: E3S Web of Conf., 27, 20-28.

Energy Pedia Portal: Power Plants. (2017, November 15). Retrieved from https://enipedia.org/wiki/Portal:Power_Plants/

Energy Statistical Analytical Service: ESAS. World Energy on interactive maps. (2017, December 10). Retrieved from http://esas.com.ru/en/

International Energy Agency: Data Services. (2016, February 24). Retrieved from http://wds.iea.org/WDS/Common/Login/login.aspx

Podkovalnikov, S., Trofimov, I., & Trofimov, L. (2018). Data processing and optimization system to study prospective interstate power interconnections. In *Proceedings of the 10th International Conference on Asian Energy Cooperation (AEC 2017)*, Irkutsk: E3S Web of Conf., 27, 32-40.

Trofimov, I., & Trofimov, L. (2017). Modern problems of data search and data verification on power plants in China and other countries. *Information and mathematical technologies in science and management*, 4 (8), 120-128.

Trofimov, L., & Trofimov, I. (2015). Internet-servis dlja analiza statisticheskih dannyh v oblasti jenergetiki na primere mirovogo rynka nefti. In *Proceedings of the conference "Jenergetika Rossii v XXI veke. Innovacionnoe razvitie i upravlenie"*, Irkutsk, 170-174.

Trofimov, I. (2014). Using metadata to query the database on thermal economy of Russia through the Internet. In *Proceedings of the 10th International Conference Mathematical and Informational Technologies (MIT-2013)*, Vrnjacka Banya, 706-714.

Xuming, L. (2018). Application and research of global grid database design based on geographic information. *Global Energy Interconnection Development and Cooperation Organization*, 1 (1), 87-95.

B6
Game Theory

XIII Balkan Conference on Operational Research
BALCOR 2018

RECIPROCITY IN GROUPS: AN EMPIRICAL ANALYSIS IN GREECE

ANTONIOS AVGERIS[1], PANAGIOTA SERGAKI[2], ACHILLEAS KONTOGEORGOS[3], ANTONIOS TIGANIS[1]

[1]MSc Researcher, Dpt. of Agricultural Economics, Aristotle University of Thessaloniki, Thessaloniki, Greece, antoavge@gmail.com
[2]Assistant Professor, Dpt. of Agricultural Economics, Aristotle University of Thessaloniki, Thessaloniki, Greece
[3]Assistant Professor, Dpt. of Business Administration of Food and Agricultural Enterprises, University of Patras, Agrinio, Greece

---

Abstract: *According to Theory of Expected Utility (EUT) and Classical Decision Theory (CDT), making decisions involve choosing a course of action among a fixed set of alternatives with a specific goal in mind. These theories define the way that individuals make decisions under emergency and risk situations. On contrast, there are many researches which indicate that not only people are not self-seeking but also care about the welfare of others having the concept of reciprocity as a motivational behavior. Based on the above mentioned, this study aims to highlight the behavior of individuals as far as their transactions between common groups. More specifically, for the purpose of this experiment methodologies of Experimental Economics as well as, of Game Theory (Ultimatum Game & Dictator Game) were adopted and applied into two different Greek Universities in undergraduate level. The analysis of the samples showed that subjects behave with reciprocity in both cases, however, there is a statistically significant difference.*

Keywords: *Experimental Economics, Game Theory, Greece, Reciprocity, Undergraduate Students*

1. INTRODUCTION

Classical Decision Theory (CDT), is the central theory that defines the way individuals making decisions among a wide set of choices with a specific goal in mind. The three components of a decision are: options or courses of action; beliefs and expectancies of the options in achieving the goal and outcome expectancies (negative or positive) (Hastie & Dawes, 2010). CDT focuses on how and why decisions deviate from a certain standard of rationality, which is based on optimality. According to this theory, the aim in making a decision is to maximize the gains, or expected value of the outcome, and use information in a way that would accomplish this goal. (Gutnike et al., 2006).

In today's increasingly complex and changeable world, it is imperative that managerial decision making should take into account the relevant aspects of many disciplines (Harrison, 1993). Accordingly, it is appropriate to acknowledge that:

> "Decision making is a meeting ground for psychologists, economists, sociologists, organizational theorists, statisticians, philosophers and others. It is an exciting field, endowed with a deep formal theory, a rich technology, numerous intriguing observations of individuals and organizations, and a growing body of experiential results" (Tversky, 1982).

However, CDT has failed to explain behavior and decision-making in practical, real world situations. This theory is limited in descriptive power because it treats all decisions as essentially the same, comparing them to a normative standard (Beach & Lipshitz, 1993). The classical economic theory asserts that subjects are self-interested and that the economy is self-regulating, i.e., individuals are motivated by selfish preferences with the aim of maximising their economic welfare (Bicskei, et al., 2016). It treats people as beings who behave completely rationally. Consequently, a number of economic models are based on the assumption of rational and selfish agents (Dufwenberg and Kirchsteiger, 2004). In these models motives such as fairness and reciprocity are automatically excluded from human behavior (Fehr, Kirchsteiger, & Riedl, 1998). In contrast, many empirical and field experiments as well as a board set of evidence indicate that concerns for fairness and reciprocity motivate a substantial number of people. Moreover, the presence of fair-minded people is likely to have important economic effects (Kahneman et al., 1986; Camerer and Thaler, 1995; Fehr and Gächter, 2000; Falk, Fehr & Fischbacher, 2008). This has led to the development of several fairness and reciprocity models. These models share the property that some people are assumed to have a preference for fairness and reciprocity—in addition to their preference for material payoffs (Rabin, 1993; Levine, 1998; Bolton and Ockenfels, 2000; Charness and Rabin, 2002).

Experimental studies in economics and the psychology of individual choice have identified numerous cognitive, informational, temporal and other limitations which bound human rationality, often producing behavior which differs from the predictions of rational models (Simon, 1979; Plott 1986; Smith, 1986). So, an important level of evidence, mainly of the previous two decades, refers that agents have a behavior not only driven by material motivation but also by concerns of equity, altruism and reciprocity (Fischbacher et al., 2001; Charness and Haruvy, 2002; Falk, 2007; Carpenter, 2017). Reciprocity is the act of response in which subjects reward and praise good deeds and punish unkind as well as hostile ones (positive and negative). Numerous experiments constitute the evidence of this behavioral disposition of people as well as a large body of evidence from various scientific fields such as sociology, psychology, anthropology, biology etc. which verifies the existence of this emotional motive, either positive or negative (Fehr and Gächter, 2000; Gintis, 2000; McCabe et al., 2003; Falk and Fischbacher, 2006; Cox et al., 2007).

Thus, as mentioned at the abstract this empirical paper aims to investigate the existence of reciprocity or not between groups of undergraduate student. In other words, it attempts to identify whether students behave with reciprocity or rationality.

2. SCOPE

Standard economic models of human decision-making (such as utility theory) have typically minimized or ignored the influence of emotions on people's decision-making behavior, idealizing the decision-maker as a perfectly rational cognitive machine. However, in recent years this assumption has been challenged by behavioral economists, who have identified additional psychological and emotional factors that influence decision-making (Camerer, Loewenstein & Rabin, 2008; Loewenstein & Lerner, 2003). So, at this point a general question that arises is the following:

Why do rational models such as those used in economics and the classical decision-making theory not always accurately predict an individual's behavior?

Based on this question, this project tries to verify and measure the reciprocity level between human (economic) relations. Other words, whether the agents act solely on the basis of increasing their individual - economic interest, i.e they have a selfish behaviour or they are interested about the others' well-being, i.e., they have a philanthropic behaviour. In particular, tries to give valuable answers if people are possessed by feelings of reciprocity or opportunism. Finally, the results are going to provide us a clear view about the behavior of students, i.e. which kind of economic man are identified with, Homo Economicus vs Homo Reciprocans

For this purpose, two well-known games of Game Theory (ultimatum game and dictator game), were used, whose both the analysis and the description as well as the explanation will be made in the section of methodology.

3. METHODOLOGY

The methodological approach of this research lies in the use of concepts and terms of the Experimental Economics. In particular, two of the most well-known games of Game Theory field were used, combined and interacted, as one called as "Reciprocity Game" (Avgeris, Kontogeorgos & Sergaki, 2018), in order to complete the experiment. Also, a questionnaire with various questions about trust, fairness, collaboration and reciprocity was used. For grading the answers, the five-level scale was used (Likert scale) depending on the degree of agreement or disagreement. After the end of the experimental process the students could, if they wanted, add new answers based on Likert scale. For the statistical analysis a 23.0 version of S.P.S.S. program was used.

In more details now, this paper contains two different investigations concerned students in undergraduate level of two different Greek universities. The first empirical research took place at the University of Patras, in the *Department of Business Administration of Food and Agricultural Enterprises* and lasted two months from May till June of 2015. The second and most recent one, comes from Aristotle University of Thessaloniki and specifically from the *Department of Agricultural Economics*. It carried out during a two-months period, too (November-December of 2017). In both cases the sample is composed exactly of the same number of participants, where it amounts to 100 students for each of them (N=100).

The first game was used is the Ultimatum Game (first developed by Güth, Schmittberger, & Schwarze, 1982). Ultimatum Game is a very famous and multi-use game both in theoretical and in experimental works. In the Ultimatum Game, two players are given the opportunity to split a sum of money. One player is deemed the proposer and the other, the responder. The proposer makes an offer as to how this money should be split between the two. The second player (the responder) can either accept or reject this offer. If it is accepted, the money is split as proposed, but if the responder rejects the offer, then neither player receives anything. In either event, the game is over. After this, the Ultimatum Game was applied. The first who developed and established this game were Kahneman, Knetsch, & Thaler (1986). The Dictator Game is also a famous game similar to the Ultimatum Game. Typically consists of two individuals, one of whom is given some quantity of money. The second individual is given nothing. The participant given the money, known in the experiment as "the dictator," is told that he must offer some amount of that money to the second participant, even if that amount is zero. Whatever amount the dictator offers to the second participant must be accepted. The second participant, should he find the amount unsatisfactory, cannot then punish the dictator in any way. Both in the first and in the second experimental procedure each student had 10€ (not real money) available for playing the game.

As far as ethical issues are concerned, researchers have fully respected them. Each student participated with his / her personal volition while firstly they had been informed about the subject of the survey. Even after this, the participants retained the right not to join in, but if the experimental process had begun, a withdrawal from the part of the student was impossible, because an even number of participants was required.

4. RESEARCH QUESTIONS

i. Is there evidence of reciprocal behavior between fellow students and by which kind of economic man the participants are identified in each experiment?
ii. Is there difference the way the students act in each case, if yes, which is the reason of this?

5. RESULTS

To begin with, the first two questions in the questionnaire, after the demographic ones, were dichotomous: the first, i) I believe it is fair when someone shares something equally between himself and some other person, and the second, ii) If someone did an unfair share against me then I would refuse to accept it. In both of these questions the participants showed their agreement and answered in the affirmative, in a more than 75% rate in each sample.

During the experimental part of the research, various actions and reactions from the participating players were observed. More specifically, as far as the implementation of the Ultimatum Game and the first offer, students from the University of Patras were more generous with their fellow students and made an average offer of 5.65€ while the students from Aristotle University of Thessaloniki offered an average amount of 4.05€. This, as easily understood, led to a statistically significant difference between the two average offers of the students according to a "Independent Samples t-Test analysis" **(2-tailed Sig.=0.000<0.05)**. However, despite the fact of the statistically significant difference these actions showed a clear manifestation tendency of reciprocity behavior and collaboration between students. The latter stems from the fact that both average offers are close or equal to half of the available amount (40% - 50%) that was shared to players and thus it comes in complete agreement with the research findings of many empirical studies (Bornstein, & Yaniv, 1998; Rand et al., 2013; Wang, Chen, & Wang, 2014; Avgeris, Kontogeorgos, & Sergaki, 2017).

On the other hand, the "dictators" despite the advantage of their role decided to respond to the good offers they received and retaliated with equally good allocations, thus, showing elements of good cooperation and positive reciprocity. In particular, the average offer in both experiments is over than 40% of the total sum of money. According to the above, both during the run of the Ultimatum Game and the Dictator Game, the students-players had a reciprocal behavior since they collaborated and exchanged good offers.

6. CONCLUSIONS

First of all, according to the two dichotomous questions, participants wish to share and distribute something worthy and equally straightforward. They, also, react and seem to be negative to accept an unfair offer, failing, thus, to maximize their utility and therefore to verify the phenomenon of "Homo Economicus".

Moreover, as far as the experimental procedures the participants did not adopt a rational behavior in general, instead of this, they behaved with reciprocity and self-denial. So, they failed to maximize their utility and economic prosperity by opposing the Classical economic Theory and the Theory of Expected Utility while were identified with the so-called "Homo Reciprocans".

Last but not least, the statistically significant difference that emerged in the average offers of the Ultimatum Game among the participants of the two Greek universities can be explained by the fact that the first department is much smaller than the second one which increases the probability that the students were known each other

REFERENCES

Avgeris, A., Kontogeorgos, A., & Sergaki, P. (2018). THE 'RECIPROCITY' GAME: A THEORETICAL BASIS FOR MEASURING RECIPROCITY IN HUMAN SOCIO-ECONOMIC INTERACTIONS. *International Journal Of Social Sciences*, VII, 13-33.

Avgeris, A., Kontogeorgos, A., & Sergaki, P. (2017). Reciprocity in trades: an experimental game approach. *International Journal Of Sustainable Agricultural Management And Informatics*, 3, 298-313.

Beach, L. R., & Lipshitz, R. (1993). Why classical decision theory is an inappropriate standard for evaluating and aiding most human decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision making in action: Models and methods* (pp. 21-35). Westport, CT, US: Ablex Publishing.

Bicskei, M., Lankau, M., & Bizer, K. (2016). Negative reciprocity and its relation to anger-like emotions in identity-homogeneous and -heterogeneous groups. *Journal of Economic Psychology*, 54, 17–34.

Bolton, G., & Ockenfels, A. (2000). ERC: A Theory of Equity, Reciprocity, and Competition. *American Economic Review, 90,* 166-193.

Bornstein, G., & Yaniv, I. (1998). Individual and group behavior in the ultimatum game: Are groups more "rational" players?. *Experimental Economics*, 1, 101-108.

Camerer, C., Loewenstein, G., & Rabin, M. (2008). *Advances in Behavioral Economics*. Princeton: Princeton University Press.

Camerer, C., & Thaler, R. (1995). Anomalies: Ultimatums, Dictators and Manners. *Journal Of Economic Perspectives*, 9, 209-219.

Charness, G., & Haruvy, E. (2002) . Altruism, equity, and reciprocity in a gift-exchange experiment: an encompassing approach, *Games and Economic Behavior*, 40, 203–231.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*. 117, 817–869.

Carpenter, J. (2017). The sequencing of gift exchange: a field trial, *Journal of Economic Behavior & Organization*, 139, 26–31.

Cox, J., Friedman, D., & Gjerstad, S. (2007). A tractable model of reciprocity and fairness. *Games and Economic Behavior*, 59, 17–45.

Dufwenberg, M. and Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and Economic Behavior*, 47, 268-298.

Falk, A. (2007). Gift exchange in the field. *Econometrica*, 75, 1501–1511.

Falk, A., Fehr, E., & Fischbacher, U. (2008). Testing theories of fairness—Intentions matter. *Games And Economic Behavior*, 62, 287-303.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and Economic Behavior*, 54, 293–315.

Fehr, E., & Gächter, S. (2000). Fairness and Retaliation: The Economics of Reciprocity. *Journal Of Economic Perspectives*, 14, 159-182.

Fehr, E., Kirchsteiger, G. and Riedl, A. (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, 42, 1-34.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, 71, 397–404.

Gintis, H. (2000). Strong reciprocity and human sociality, *Journal of Theoretical Biology,* 206, 169–179.

Güth W., Schmittberger R., & Schwarze B., (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior and Organization*, 3, 367-388.

Gutnik, L., Hakimzada, A., Yoskowitz, N., & Patel, V. (2006). The role of emotion in decision-making: A cognitive neuroeconomic approach towards understanding sexual risk behavior. *Journal Of Biomedical Informatics*, 39, 720-736.

Harrison, E. (1993). Interdisciplinary Models of Decision Making. *Management Decision*, 31, 26-33.

Hastie, R., & Dawes, R. (2010). *Rational choice in an uncertain world*. Los Angeles: SAGE.

Kahneman, D., Knetsch, J.L., & Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *American Economic Review*, 76, 728–741.

Loewenstein, G., & Lerner, J. (2003). The Role of Affect in Decision Making. In R. Davidson, K. Scherer & H. Goldsmith, *Handbook of Affective Science* (1st ed.). Oxford: Oxford University Press.

Levine, D. (1998). Modeling Altruism and Spitefulness in Experiments. *Review Of Economic Dynamics*, 1, 593-622.

McCabe, K., Rigdon, M., & Smith, V. (2003). Positive reciprocity and intentions in trust games. *Journal of Economic Behaviour & Organization*, 52, 267–275.

Rabin, M., (1993). Incorporating fairness into game theory and economics. *American Economic Review,* 83, 1281–1302.

Rand, D., Tarnita, C., Ohtsuki, H., & Nowak, M. (2013). Evolution of fairness in the one-shot anonymous Ultimatum Game. *Proceedings Of The National Academy Of Sciences*, 110, 2581-2586

Plott, C. (1986). Laboratory Experiments in Economics: The Implications of Posted-Price Institutions. *Science*, *232*, 732-738.

Simon, H. A. (1979). Rational Decision Making in Business Organizations. *American Economic Review,* 69, 493-513.

Smith, V. (1986). Experimental Methods in the Political Economy of Exchange. *Science*, 234, 167-173.

Tversky, A. (1982). Remarks on the Study of Decision Making. In G. Ungson, *Decision making: An interdisciplinary inquiry* (1st ed., p. 321). Kent, Boston, MA: Kent Pub. Co.

Wang, X., Chen, X., & Wang, L. (2014). Random allocation of pies promotes the evolution of fairness in the Ultimatum Game. *Scientific Reports*, 4,

BALCOR 2018
XIII Balkan Conference on Operational Research

COOPERATIVE GAMES ON FIXED LENGTH ARRAYS AND THE ASSIGNMENT PROBLEM

Josué Manzanarez Cárdenas[1], William Olvera - Lopez[2]

[1] Dirección Adjunta de Rentas Aduaneras. Tegucigalpa, Honduras, wjmancar@gmail.com
[2] UASLP, School of Economics. SLP, México, william.olvera@uaslp.mx

Abstract: *In this paper we introduce a kind of TU-games where the characteristic function is defined on all the possible ways to fill an array using a subset of agents, allowing empty positions as well as repeated agents. We present several axiomatically characterized solutions for these games, including a closed expression when the length of the array is equal to the number of agents. Finally, we show the possible connection of this kind of games with the classical TU-games definition.*

Keywords: *cooperative games, arrays, generalized efficiency, assignment problems*

1. INTRODUCTION

The cooperative game theory has shown to be a very helpful tool for modeling or solving problems associated to groups of agents earning a profit for working or staying together, in a very general framework. In a transferable utility cooperative game (a TU-game), only subsets (coalitions) of agents (players) are able to cooperate and then, there exists a characteristic function defined on the power set of the set of agents that models the benefits of the cooperation. According to the problem, it could be interesting to study which agent is the most significative one in the game, how to allocate an amount among agents regarding the information on the game or, given a payoff function, which coalitions are formed. There are a lot of works on these topics, but we consider as seminal ones Von Neumann and Morgenstern (2007), Shapley (1953) and Shapley (1971).

If a cooperative game involves other ways of cooperation or players organization, then it is mandatory to extend the classical model of TU-games. So, other concepts for cooperative games arise, as games with coalitional structures Owen (1977), multi-choice games Hsiao and Raghavan (1993), cooperation with graphs Myerson (1977), etc. There is a model introduced by Nowak and Radzik Nowak and Radzik (1994) where the characteristic function is defined over the set of arrangements of players, that they call *games with generalized characteristic functions*: in this kind of games, the order of the players matters, not only which players are cooperating. In our work, we extend this model considering a worth for every possible array of a fixed length filled by subsets of players, allowing repeated players on the array as well as empty positions. This kind of modeling can be useful in situations where the position of the agents into a configuration is important for the worth gained by the players in the array, by example, in problems involving workplaces or land distribution. In this paper we provide some axiomatically characterized solutions for this model and we show that TU-games could be seen as a particular case of our model.

The paper is organized as follows: in Section 2 we introduce the notation as well as the model. In Section 3 we provide several axiomatically characterized solutions for the model and we show the possible link between our model and classical TU-games. Finally, in Section 4 we show the main conclusions of the paper and some ideas of future work for this kind of games. The Appendix contains the proofs of all our results.

2. NOTATION AND DEFINITIONS

Let $N := \{1,\dots,n\}$ be a finite set of players and $M := \{1,\dots,m\}$ a finite set of spaces to be assigned to the players. The first idea behind the model is to position the agents into the spaces, allowing empty spaces and the possible assignation of several spaces to a single player. So, we define an extended set of players $\overline{N} = N \cup \{0\}$, where the player 0 denotes an empty position. Then, the set

$$K_m^n := \underbrace{\overline{N} \times \overline{N} \times \cdots \times \overline{N}}_{m \text{ times}}$$

contains all possible ways to assign the m spaces to the players, that we call *configurations*. We denote by $G_m^n := \{v \mid v : K_m^n \to \mathbb{R}, \text{ with } v(\mathbf{0}) = 0\}$ the set of all possible characteristic functions defined on arrays of fixed length with set of agents N and set of spaces M, where $\mathbf{0}$ is a vector filled with m zeros. It is straightforward to verify that G_m^n is a vector space with the usual operations: for $v_1, v_2 \in G_m^n$ and $\delta \in \mathbb{R}$, $(v_1 + v_2)(k) := v_1(k) + v_2(k)$ and $(\delta v_1)(k) := \delta v_1(k)$, for all $k \in K_m^n$. Given a characteristic function $v \in G_m^n$ and a configuration $S \in K_m^n$, $v(S)$ denotes the worth when the players are positioned into the spaces according to S. Then, a transferable utility cooperative game in an array of fixed length, or an a-game by simplicity, is a tuple (n,m,v) with $v \in G_m^n$. Henceforth, we assume the sets N and M are fixed and we denote an a-game (n,m,v) just by $v \in G_m^n$. A solution for a-games is a mapping $\varphi : G_m^n \to \mathbb{R}^n$ where, for every player $i \in N$, the i-th entry, $\varphi_i(v)$, denotes the solution to player i in the a-game $v \in G_m^n$. The idea of a solution $\varphi : G_m^n \to \mathbb{R}^n$ is to measure the importance of the agents in an a-game. We can use this measurement for ranking them or allocating an amount amongst them.

► Example 1. A company composed of three people has received four pieces of land as a donation, so each person has the same rights over them. Each piece of land can be assigned for working on to any member of the company and each member can own as much land as possible. Due to the properties of the terrain and the ability of the people, it could be beneficial to the company to have unassigned land or to have people without assigned land (for example, a person could be a theoretical physicist without expertise in work of the land). So, the company could know the worth obtained for each assignation of the land, and this could be modeled in an a-game with $m = 4$ and $n = 3$; by example, $v(1,0,1,2) = 20$ means that assigning the pieces of land one and three to Player 1 and the piece of land four to Player 2 represents a revenue for the company of 20 units. So, in this work we are going to provide some methods for measuring the importance of each member of the company in the assignation of the land. We are not solving the problem of how to assign the land to the people.

For $k \in K_m^n$ and every $i \in \overline{N}$, we define $k(i) := \{ j \in M \mid k_j = i\}$, the set of spaces on k where the player $i \in N$ is positioned. Notice that $k(0)$ denotes the set of empty spaces in k. Additionally, given a configuration $k \in K_m^n$ and $i, j \in \overline{N}$ we define a new configuration k_j^i as follows:

$$(k_j^i)_h = \begin{cases} j, & \text{if } h \in k(i); \\ k_h, & \text{otherwise.} \end{cases} \qquad \forall h \in M.$$

k_j^i denotes a configuration where the spaces assigned to the player i on k are now assigned to the player j. Given a player $i \in N$, k_0^i denotes the removal of player i of the configuration k. On addition, we define, for every $S \subseteq N$,

$$K_m^n(S) := \{k \in K_m^n \mid \forall i \in S, k(i) \neq \},$$

the set of all configurations in K_m^n such that every player in S has assigned one space, at least.

For a configuration $k \in K_m^n$ and $S \subseteq N$, we denote by $k_S := \{i \in S \mid k(i) \neq \}$, the subset of players on S positioned into the configuration k. Please, do not mix up $k_{\{i\}}$ with k_i, the i-th position of configuration k.

3. CHARACTERIZATION OF SOLUTIONS

In this section we present axiomatically characterized solutions for a-games.

► Axiom 1. (Linearity) A solution $\varphi : G_m^n \to \mathbb{R}^n$ satisfies the linearity axiom if for every pair of a-games $v_1, v_2 \in G_m^n$ and $c_1, c_2 \in \mathbb{R}$,

$$\varphi(c_1 v_1 + c_2 v_2) = c_1 \varphi(v_1) + c_2 \varphi(v_2).$$

According to this axiom, if we can split an a-game in two a-games, the solution for the original a-game must be equal to the sum of the solutions for the new a-games. This kind of situations arise naturally when a period of time is involved: the solution for the whole period of time must be equal to the sum of the solutions applied to small periods.

We say a player $i \in N$ is a null player in the a-game $v \in G_m^n$ if $v(k) = v(k_0^i)$ for every $k \in K_m^n$.

► Axiom 2. (Nullity) A solution $\varphi : G_m^n \to \mathbb{R}^n$ satisfies the nullity axiom if

$$\varphi_i(v) = 0$$

for every null agent $i \in N$ in the a-game $v \in G_m^n$.

A null player in an a-game is not beneficial or harmful in a problem because there is no difference in the worth of a configuration if we remove him from it. So, his payoff must be equal to zero.

We say that two players $i, j \in N$ are substitute players in an a-game $v \in G_m^N$ if $v(k_j^i) = v(k_i^j)$ for every $k \in K_m^n$.

► Axiom 3. (Substitute players axiom) A solution $\varphi : G_m^n \to \mathbb{R}^n$ satisfies the substitute player axioms if

$$\varphi_i(v) = \varphi_j(v)$$

for every pair of substitute players $i, j \in N$ in $v \in G_m^n$.

If we have an a-game such that when we replace the player i by the player j in every possible configuration the worth does not change, neither when we replace the player j by the player i, we can assume that these players have the same importance in the a-game. So, according to the substitute players axiom, they must have the same payoff.

Given an a-game $v \in G_m^n$ and a subset of players $S \subseteq N$ we define

$$\eta(S,v) = \begin{cases} \dfrac{1}{|K_m^n(S)|} \displaystyle\sum_{k \in K_m^n(S)} v(k), & \text{if } m \geq |S|; \\ 0, & \text{otherwise.} \end{cases}$$

and it denotes the average worth, according to v, of the configurations where every player on S has assigned one space, at least. Notice that if $m < |S|$, $K_m^n(S) = $.

► Axiom 4. (Average efficiency by agents axiom) A solution $\varphi : G_m^n \to \mathbb{R}^n$ satisfies the average efficiency by agents axiom if

$$\sum_{i \in N} \varphi_i(v) = \eta(N,v)$$

for every a-game $v \in G_m^n$.

A solution satisfying the average efficiency by agents axiom establishes an amount to be allocated among the agents: It is assuming that every agent must has assigned one space, at least. But, if $m \geq n$, there are several configurations with this requirement. So, this axiom considers that each configuration where all the players are has the same importance in the a-game. That is why we average the worth of these configurations. Clearly, in situations where $m < n$, the average efficiency by agents axiom is meaningless. In our example, to accept this axiom means that it is important that every member of the company owns a piece of land, at least.

► Theorem 1. There exists a unique solution $\varphi : G_m^n \to \mathbb{R}^n$ satisfying the linearity, nullity, substitute players and average efficiency by agents axiom.

Please refer to Appendix for the proof of the previous and all the ongoing results.

For the next result, we define $\overline{K}_m^n := \{k \in K_m^n \mid |k(i)| \in \{0,1\} \; \forall i \in N\}$, the subset of configurations without repeated players.

► Theorem 2. If $m = n$, the solution $\varphi : G_m^n \to \mathbb{R}^n$ given by

$$\varphi_i(v) = \sum_{\substack{k \in \overline{K}_m^n \\ k(i) \neq}} (|k_N| - 1|)! \left(\frac{(n - |k_N|)!}{n!} \right)^2 (v(k) - v(k_0^i)), \tag{1}$$

for all $i \in N$, satisfies the linearity, nullity, substitute players and average efficiency by agents axioms.

Given a set of players $N = \{1, \ldots, n\}$, a transferable utility cooperative game, a TU-game, is a mapping $v : 2^N \to \mathbb{R}$ such that $v() = 0$. This class of games has been widely studied, and one of the most popular solutions for them is the Shapley value, because it has a lot of good properties. For further information about TU-games and the Shapley value please refer to Shapley (1953).

Because of the formulation given by (1) involves a kind of marginal contribution of each player to every configuration, without repeated players, which contains him, it is natural to think about a relation between this solution and the Shapley value of a TU-game.

► Proposition 1. Given a TU-game (N,v), we construct an a-game $(n, n\overline{v})$ as follows:

$$\overline{v}(T) = v(T_N), \qquad \forall T \in K_m^n.$$

Then, $\varphi(\overline{v}) = \text{Sh}(v)$, with φ given by (1).

According to the previous proposition, a TU-game could be seen as a particular case of an a-game where all the arrays with the same subset of agents has the same worth. That is, a classical TU-game is an a-game where the possible interaction amongst agents, given by a particular array structure, does not matter.

The solution given by Theorem 1 uses the average efficiency by agents axiom. According to this axiom, all the configurations containing the whole set of agents have the same importance for setting the amount to be distributed among them, in a attempt to imitate, in a natural way, the efficiency axiom for TU-games. We provide two other ideas for doing it.

Given an a-game $v \in G^n_m$ and a subset of players $S \subseteq N$, $S \neq$ we define

$$\zeta(S,v) = \frac{1}{|S|^m} \sum_{\substack{k \in K^n_m \\ k_{\overline{N}\setminus S} =}} v(k),$$

and it denotes the average worth, according to v, of the configurations without empty spaces where the players on $N \setminus S$ do not have assigned spaces.

► Axiom 5. (Average efficiency by positions axiom) A solution $\varphi : G^n_m \to \mathbb{R}^n$ satisfies the average efficiency by positions axiom if

$$\sum_{i \in N} \varphi_i(v) = \zeta(N,v)$$

for every a-game $v \in G^n_m$.

A solution satisfying the average efficiency by positions axiom assumes that the goal of the cooperation is not to have empty spaces in the configuration and then, those configurations are taken into account for setting the allocation amount to be distributed among the players. In our example, to accept this axiom means that it is important for the company not to have unassigned land.

For the next axiom we define, for every non-empty subset of players S,

$$\mathring{K}^n_m(S) := \{k \in K^n_m(S) \mid k_{\overline{N}\setminus S} = \},$$

the set of configurations without empty spaces where each element on S has assigned one assigned, at least, but the players on $N \setminus S$ do not have assigned spaces. Then, given an a-game $v \in G^n_m$ and a subset of players $S \subseteq N$, $S \neq$, we define

$$\xi(S,v) = \begin{cases} \dfrac{1}{|\mathring{K}^n_m(S)|} \displaystyle\sum_{k \in \mathring{K}^n_m(S)} v(k), & \text{if } m \geq |S|; \\ 0, & \text{otherwise.} \end{cases}$$

and it denotes the average worth, according to v, of the configurations without empty spaces where all the spaces are assigned to players on S.

► Axiom 6. (Average efficiency by agents and positions axiom) A solution $\varphi : G^n_m \to \mathbb{R}^n$ satisfies the average efficiency by agents and positions axiom if

$$\sum_{i \in N} \varphi_i(v) = \xi(N,v)$$

for every a-game $v \in G^n_m$.

A solution satisfying the previous axiom assumes that the configurations used for establishing the allocation amount to be distributed amongst the players must contain to every single player and it must be without empty spaces. So, this axiom mixes the ideas of the average efficiency by agents and by positions axioms. In our example, to accept this axiom means that it is important for the company not to have unassigned land and that each agent must have assigned a piece of land, at least.

► Theorem 3. There exists a unique solution $\phi : G^n_m \to \mathbb{R}^n$ satisfying the linearity, nullity, substitute players and average efficiency by positions axioms.

► Theorem 4. There exists a unique solution $\psi : G^n_m \to \mathbb{R}^n$ satisfying the linearity, nullity, substitute players and average efficiency by agents and positions axioms.

The two previous theorems allow us to decide which configurations to consider for setting an amount to be distributed amongst the agents, and still to have a solution defined uniquely. It depends on the problem which one has more sense than the others. We could provide additional ideas about how to establish this amount and to have a unique solution axiomatically characterized too, but we think that the axioms we present include the most reasonable ways of generalizing the efficiency axiom for TU-games in our context.

4. CONCLUSIONS AND FUTURE WORK

In this paper we work with a kind of generalization of cooperative TU-games, where the characteristic function is defined on arrays of a fixed length. Also, we provide three solutions for this class of cooperative games. For just one of them, the solution given in Theorem 1, we show an expression to calculate it. In fact, we obtain closed expressions for the solutions of theorems 3 and 4, but they were very complicated. We left the problem of finding friendly-and-manageable expressions for both cases as an open problem. Also, it could be interested to show a comparison of the solutions on particular cases, by example, in a kind of superadditivity background.

Acknowledgment

We acknowledge support from CONACyT grant 240229.

A Proof of the results

Given two configurations $T, Q \in K_m^n$, we say $T \sqsubseteq Q$ if for every $j \in M$ with $T_j \neq 0$, $T_j = Q_j$. Thus, given a configuration $T \in K_m^n$ we define a new a-game $\mathcal{U}_T \in G_m^n$ as follows:

$$\mathcal{U}_T(Q) = \begin{cases} 1, & \text{if } T \sqsubseteq Q; \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

for every $Q \in K_m^n$.

► Lemma 1. The set of a-games $\{\mathcal{U}_T\}_{T \in K_m^N \setminus \mathbf{0}}$ forms a basis of G_m^n.

Proof. We can create a square matrix A where the ij-th entry, a_{ij}, corresponds to $\mathcal{U}_T(Q)$ with T the configuration in the position i considering the lexicographical order on $K_m^n \setminus \mathbf{0}$ and Q the configuration in the j-th position using the same ordering. It is easy to see that A is a lower-triangular matrix with ones in the main diagonal. So, the elements on $\mathcal{U}_T$ form a basis of G_m^n. ◄

► Lemma 2. Given a configuration $T \in K_m^n$, if $i \notin T_N$, then i is a null player in $\mathcal{U}_T$.

Proof. Let $T, Q \in K_m^n$ be configurations such that $i \notin T_N$ and $i \in Q_N$. If $\mathcal{U}_T(Q) = 0$, then $\exists j \in M$ such that $T_j \neq 0$ and $T_j \neq Q_j$. So, $(Q_0^i)_j \neq T_j$. Then, $\mathcal{U}_T(Q_0^i) = 0$. If $\mathcal{U}_T(Q) = 1$ then, $\forall j \in Q(i)$, $T_j = 0$. So, $\mathcal{U}_T(Q_0^i) = 1$. ◄

► Lemma 3. Given a configuration $T \in K_m^n$, if $i, j \in T_N$, then i and j are substitute players on $\mathcal{U}_T$.

Proof. Let $T \in K_m^n$ be a configuration such that $\{i, j\} \subseteq T_N$. Notice that, for any configuration $Q \in K_m^n$ such that $\{i, j\} \cap Q_N \neq$, player i does not have any position assigned in Q_j^i; so $\mathcal{U}_T(Q_j^i) = \mathcal{U}_T(Q_i^j) = 0$; if $\{i, j\} \cap Q_N =$, the result is trivial. ◄

► Theorem. There exists a unique solution $\varphi : G_m^n \to \mathbb{R}^n$ satisfying the linearity, nullity, substitute players and average efficiency by agents axiom.

Proof. Let $\varphi : G_m^n \to \mathbb{R}^n$ be a solution satisfying the axioms of the theorem. Because of Lema 1, any a-game v can be written as follows:

$$v = \sum_{T \in K_m^n \setminus \mathbf{0}} \delta_T \mathcal{U}_T.$$

for some numbers δ_T, $T \in K_m^n$. Because φ is a linear solution, it will be enough to prove uniqueness of the solution for each a-game $\mathcal{U}_T$. Because of Lemma 2, we have that $\varphi_i(\mathcal{U}_T) = 0$ if $i \notin T_N$. Lemma 3 establishes that $\varphi_i(\mathcal{U}_T) = \varphi_j(\mathcal{U}_T)$ if $\{i, j\} \subseteq T_N$. Now, we are going to use the fact that φ satisfies the average efficiency by agents axiom. If T is a configuration such that $|T(0)| < (n - |T_N|)$, we have that $\varphi_i(v) = 0$ for every $i \in N$ and then, the solution is defined uniquely; if $|T(0)| \geq (n - |T_N|)$, we need to count the number of configurations Q containing all the players such that $\mathcal{U}_T(Q) = 1$. Let $F(x, y, z)$ be the function counting the number of arrays of length x filled by y elements (it could be repeated elements and elements without assigned spaces, but not empty spaces) such that z ($z \leq y$, $z \leq x$) of those elements must have assigned one space, at least. This function can be computed by several ways, including the use of Stirling numbers of the second kind (denoted by the funcion $\mathbb{S}$) and the inclusion-exclusion principle:

$$F(x,y,z) = \sum_{i=0}^{z} \binom{z}{i} (-1)^i (y+1-i)^x = z!\ \mathbb{S}(x,z;y-z+1). \tag{3}$$

Notice that, using function (3), $|K_m^n(N)| = F(m,n+1,n)$. The second entry of the function F is equal to $n+1$ because the set of available elements for the configuration is $\overline{K}$. So, we have

$$\varphi_i(\mathcal{U}_T) = \begin{cases} \dfrac{F(|T(0)|,n+1,n-|T_N|)}{F(m,n+1,n)\ |T_N|}, & \text{if } |T(0)| \geq (n-|T_N|),\ i \in T\ ; \\ 0, & \text{otherwise.} \end{cases}$$

and this shows that the solution is uniquely determined by these four axioms. ◀

▶ Theorem. If $m = n$, the solution $\varphi : G_m^n \to \mathbb{R}^n$ given by

$$\varphi_i(v) = \sum_{\substack{k \in \overline{K}_m^n \\ k(i) \neq}} (|k_N| - 1|)! \left(\frac{(n-|k_N|)!}{n!} \right)^2 (v(k) - v(k_0^i)), \tag{4}$$

for all $i \in N$, satisfies the linearity, nullity, substitute players and average efficiency by agents axioms.

Proof. To prove that Solution (4) satisfies the linearity and nullity axioms is straightforward.

Assume that $i, j \in N$ are substitute players on the a-game v. Notice that for every $T \in \overline{K}_m^n$ such that $i \in T_N$ but $j \notin T_N$, there exists a configuration Q such that $v(Q) = v(T_j^i) = v(T)$ with $i \notin Q_N$ and $j \in Q_N$, and viceversa. So, $\varphi_i(v) = \varphi_j(v)$.

For proving that Solution (4) satisfies the average efficiency by agents axioms, we check how many times the term $v(T)$, with $T \in K_m^n$, appears in $\sum_{i \in N} \varphi_i(v)$. If $|T_N| \neq n$, the term $v(T)$ appears for every player $i \in T_N$; on the other hand, the term $-v(T)$ appears for every $i \notin T_N$, but it appears when the solution for them is adding over a configuration Q such that $Q_0^i = T\ \ (|Q_N| = |T_N| + 1)$. Notice that there are $|T(0)| = n - |T_N|$ of those configurations. Then, when we add over all the players, there is a term

$$\left((|T_N| - 1)! \left(\frac{(n-|T_N|)!}{n!} \right)^2 |T_N| - (|T_N| + 1 - 1)! \left(\frac{(n-|T_N|-1)!}{n!} \right)^2 (|n-|T_N|)!^2 \right) v(T)$$

for every T such that $|T_N| \neq N$. It is easy to see the previous coefficient for $v(T)$ is equal to zero. If $|T_N| = N$, there are not any negative term in the coefficient for $v(T)$, and, again, it is easy to simplify it to $1/n!$. ◀

▶ Proposition. Given a TU-game (N,v), we construct an a-game $\overline{v}$ as follows:

$$\overline{v}(T) = v(T_N), \qquad \forall T \in K_m^n.$$

Then $\varphi(\overline{v}) = \mathrm{Sh}(v)$, with φ given by (1).

Proof. Given a coalition $S \subseteq N$, we need to count the number of configurations $k \in \overline{K}_m^n$ containing all the agents on S, and this number is

$$\binom{n}{|S|} (|S|)! = \frac{n!}{(n-|S|)!}$$

and, considering the formulation given by (1), the result holds. ◀

▶ Theorem. There exists a unique solution $\phi : G_m^n \to \mathbb{R}^n$ satisfying the linearity, nullity, substitute players and average efficiency by positions axioms.

Proof. Let $\phi : G_m^n \to \mathbb{R}^n$ be a solution satisfying the axioms of the theorem. The same arguments used in the proof of Theorem 1 regarding to the linearity, nullity and substitute players axioms apply to this theorem. In this case, for every $T \in K_m^n$, we need to count the number of configurations Q with no empty spaces such that $\mathcal{U}_T(Q) = 1$. That is,

$$\phi_i(\mathcal{U}_T) = \begin{cases} \dfrac{n^{|T(0)|}}{n^m\ |T_N|}, & \text{if } i \in T\ ; \\ 0, & \text{otherwise,} \end{cases}$$

and then, the solution is uniquely determined by these four axioms. ◀

▶ Theorem. There exists a unique solution $\psi : G_m^n \to \mathbb{R}^n$ satisfying the linearity, nullity, substitute players and average efficiency by agents and positions axioms.

Proof. Let $\psi : G_m^n \to \mathbb{R}^n$ be a solution satisfying the axioms in the statement of the theorem. The same arguments used in the proof of Theorem 1 regarding to the linearity, nullity and substitute players axioms apply to this theorem. For every $T \in K_m^n$, we need to count the number of configurations Q without empty spaces such that $\mathcal{U}_T(Q) = 1$ and $Q(i) \neq$ for every $i \in N$ (this is possible if $n - |T_N| \leq |T(0)|$), as well as the total number of configurations without empty spaces and all the agents on them. It is necessary to use the function F given by 3. So,

$$\psi_i(\mathcal{U}_T) = \begin{cases} \dfrac{F(|T(0)|, n, n - |T_N|)}{F(m, n, n)\ |T_N|}, & \text{if } |T(0)| \geq (n - |T_N|), i \in T\ ; \\ 0, & \text{otherwise.} \end{cases}$$

and this shows that the solution is uniquely determined by these four axioms. ◀

REFERENCES

Hsiao, C. R. and Raghavan, T. (1993). Shapley value for multichoice cooperative games. *I. Games and Econ. Behav.*, 2(3):240–256.

Myerson, R. B. (1977). Graphs and cooperation in games. *Math. of Oper. Res.*, 2(3):225–229.

Nowak, A. S. and Radzik, T. (1994). The shapley value for n-person games in generalized characteristic function form. *Games and Economic Behavior*, 6(1):150–161.

Owen, G. (1977). Values of games with a priori unions. In *Mathematical economics and game theory*, pages 76–88. Springer.

Shapley, L. S. (1953). A value for n-person games. In Kuhn, H. W.; Tucker, A. W., editor, *Contributions to the Theory of Games. Annals of Mathematical Studies. Vol 2. No 28.*, pages 307–317. Princeton University Press.

Shapley, L. S. (1971). Cores of convex games. *International journal of game theory*, 1(1):11–26.

Von Neumann, J. and Morgenstern, O. (2007). *Theory of Games and Economic Behavior (commemorative edition)*. Princeton university press.

XIII Balkan Conference on Operational Research

ASYMMETRIC BARGAINING POWER AND MULTI-LAYERED COALITION STRUCTURE

HARUO IMAI[1],

[1] Taisho University, Department of Regional Development, Tokyo, Japan, h_imai@mail.tais.ac.jp

---

Abstract: *Based on a sequential coalition formation game, we consider a role of coalition in a pure bargaining problem. When a difference in bargaining power creates a non-symmetric outcome, and joining a coalition including a player with stronger power, a player may receive a benefit. Furthermore, if this bargaining power is common to both the bargaining situation for the grand coalition and the bargaining situation within a sub-coalition, then there would be an incentive for a player with stronger power may get benefit for joining a coalition with weaker players, as there may be a possibility that a stronger player's power is magnified by conferring its power to other members at the level of bargaining in the grand coalition, while within a coalition, stronger player would exert its power straightforwardly to obtain its share directly but there may be enough left for a weaker player if the coalition is successful to obtain at the level of the bargaining in the grand coalition. Thus there is a possibility that a stronger player and a weaker player to form a coalition.*

Looking at the limit of stationary subgame perfect equilibrium of multi-stage game, we exemplify that this is indeed the case under some set of assumptions and list some properties of equilibria. Finally we show that one could extend this to the coalition formation within a coalition to yield a nested coalition structure by menas of a example.

Keywords: *Bargaining, Coalition Formation, Nested Coalition Structure.*

1. INTRODUCTION

The effect of binding agreement reached among a subgroup of participants of negotiation prior to bargaining has been expected to be both positive and negative. Harsanyi's joint bargaining paradox (Harsanyi (1977), Chae and Heidhue (2004), and Vidal-Puga (2010)), received some attention recently which dictates that forming a coalition would worsen the position of the members based on the prediction of the n-person Nash bargaining solution, on the one hand. On the other hand, in some realistic occasions, coalition formation can be observed and the insight given by the coalition formation theory tells that such coalition is formed because it must be beneficial to all of the members, backed by the intuition that a coalition can do whatever its sub-coalitions can do.

We try to investigate this problem combining sequential bargaining game to split a dollar a la Binmore and Rubinstein with the coalition formation game utilized by Bloch (1996) and Ray and Vohra (1999) for the class of bargaining problems to split money. The game is in two (or more correctly three) stages, and in the first stage, players form coalitions with coalitional agreement, and in the second bargaining stage, players play ordinary sequential bargaining game but their incentives are governed by the agreement they committed to earlier. (We also consider the third stage in which players bargain over the division of the coalitional gain within each coalition.) The solution we adopt here is the limit outcomes of the stationary subgame perfect equilibria when the interval between offers are made vanishes. As is well known, the result often depends upon the protocol of the game in each stage (cf. Baron and Ferejohn (1989), Montero (1999)) as well as the coverage of the coalitional agreement. Especially, we consider the case of the random proposer rule, with asymmetric recognition probability to represent a kind of bargaining power (cf. Okada (1996)). We may compare the result here with the one under the reject-and-offer rule (where the player who rejects the standing offer becomes the next proposer. (cf. Britz et. al. (2010))). Even though we do not consider the option to let a single representative to bargain on behalf of the members, still we get a possibility of the joint bargaining paradox under the latter rule.

To consider a possibility of a scale merit in bargaining situation, some sort of difference in bargaining power when players bargain individually and when they bargaining jointly. Here, we consider the situation with differing recognition probabilities representing one kind of bargaining power inspired by the legislative bargaining. Suppose that a larger recognition probability stems from visibility of a player or a special accessibility to the institution which determines the next proposer, like through the able lobbyist in an institution. Then supposing that forming a coalition allows the same accessibility to the most capable lobbyist among those who coalition members have access to for all members, then one could stretch the strongest bargaining power of the member possessing the highest recognition probability to all the members. Of course, the other players outside the coalition could form other coalitions and so the resulting recognition probabilities of each coalition depends upon the entire coalition structure formed.

Coalition formation literatures often assume that the choice of actions is determined by noncooperative equilibria under a coalition structure as in Ray and Vohra (1997) and so there is essentially choice of strategies is not much of an issue (which is criticized by proponents of other approaches). Focus is mainly put on the redistribution of payoffs among coalition members. With the assumption of utilities linear in money, the analysis goes well with the restriction of the contracts to those specifying constant share to each member, and in fact Imai and Watanabe (2005) found that players are neutral in forming a coalition. Here we assume that signing contract simply implies that they are going to divide the total earning later, and division of the earning is determined by ex post bargaining. (Although ex post bargaining often creates inefficiency as is known in the incomplete contract literature, here efficiency loss is not of a problem because there is no investment decision involved.) As a result, we have a point estimate of the resulting payoffs once a coalition structure is given. The resulting reduced form is known as a hedonic game (especially after Bogolomonaia and Jackson (2001), and so the problem boils down to find equilibrium coalition structure in a hedonic partition game.

One concomitant issue is the fact that the agreement to form a coalition provides an incentive to maximize coalitional gains, and this gives a perfect correlation of interests among members. As is discussed in Imai and Salonen (2000), this causes a de fact delegation to the "toughest" player in the coalition in the bargaining process, whilehere the toughness is represented by the recognition probability. This provides a room for the positive effect of coalition formation in a pure bargaining model. However, under the reject-and-offer rule, this entails in the Harsanyi's paradox so that no coalition forms. By contrast, under the random proposer rule, a nontrivial coalition formation problem arises. Earlier, we investigated this problem via nonlinear utility and fthe proposer selection rule with a fixed order and as a modification of the core or its farsighted extension in the sequential coalition formation game, we investigate the order independent equilibrium (first introduced by Mordovanu and Winter (1995) and received more attention these days, cf. Koscy (2009) for example). Here we only appeal to the stationary equilibrium which essentially correspond to that refinement of order independent equilibrium, and we derive an equilibrium coalition structure, exhibiting the properties that for generic situations, the softest player joins the coalition which obtains the highest share (per person) and this coalition consists of all players softer than a certain player plus possibly the toughest player. Further the above mentioned structure is nested, which is a reminscate of the top coalition property (Banerjii, Konishi, and Sonmez (2001)) for the non-empty core.

Finally as one extension, we consider a multi-layer coalition formation problem by means of an example to show the possibility of a nested coalition structure (cf. Myerson (1991)). This suggests that an investigation of political coalition has some aspects different from economic coalitions.

2. BASIC MODEL

Here, we give definition of the game and the solution. The set of players is $N=\{1,...,n\}$.

In the first stage, at the outset, given a probability distribution $p = (p_i)_{i=1}^{n}$ $(p_i > 0)$ each player i is chosen as a proposer and makes a proposal of a formation of the coalition S_i, and members of S_i other than i replies by "Yes" or "No". If everybody says "Yes", then S_i forms. (In particular, if $S_i = \{i\}$, then S_i forms automatically.) If somebody rejects the proposal by saying "No", then a proposer j is chosen randomly with the same distribution and player j makes the next proposal in the next period and the same procedure follows as in the first period. In case S_i forms and is equal to N, then the game moves into the second stage. Otherwise, with the probability distribution restricted to $N\backslash S_i$, the next proposer is chosen and makes a proposal to members within $N\backslash S_i$ in the same period. The rest is the same as in the first period. The game ends if no player remains to make the next proposal and a coalition structure $\Pi = (S^j)_{j=1}^{k}$ results. If the game continues without an end at this stage, then impasse outcome (D) realizes and each player recieves 0 utility.

In the second stage, players engage in a sequential and pure bargaining game with a coalition structure Π. Let p_S be $max \{p_{i'} : i' \in S^i \in \Pi\}$ for each i and $q_j = P_i/(\Sigma|S^i|P_i)$. The next proposer is chosen from N with probability distribution $q = (q_i)_{i=1}^n$ ($q_i > 0$) and the chosen player j makes an offer $x = (x_i)_{i=1}^n$ with x_i and $\Sigma x_i = 1$. The rest of players replies by "Yes" or "No" in this order and if everybody says "Yes", then this stage ends with x in period 0. If some player rejects, then in period 1, The same procedure is repeated. This stage ends either with everybody agreeing on the player i's offer x in period t $((x,t))$ or with a perpetual disagreement (D).

Provided the stage 2 ends with agreement (x,t), the game moves into the third stage where within each $S \in \Pi$, essentially the same game as in the second stage is played to decide a division of the coalitional surplus $x(S)= \Sigma x_i$ among members, given the initial probability distribution p restricted to S. We assume that agreement within a coalition and the consumption of agreed shares are made independent of whether agreement is reached or not in other coalitions.

We consider a stationary subgame perfect equilibrium of this game. That is, in an equilibrium strategy, the action choice in stage 3 is independent of the history except for the coalition each player belongs to, the coalitional surplus, and the offer made in that period (if the action is the choice of "Yes" or "No"); and the action choice of stage 2 is independent of the history except that the coalition the player belongs to and the offer made in that period (if the player is choosing between "Yes" and "No").

As mentioned above, we shall investigate the solution when δ tends toward 1. It is known that for the third stage, for each S $(\neq \phi) \subset N$ and coalitional surplus $x(S)$ $(\in[0,1]$), there is a unique stationary subgame perfect equilibrium, and its limit outcome is given by $argmax\{ \Pi x_i : \Sigma x_i = x(S), \ x_i \geqq 0 \}$, .i.e. proportional to p restricted to that coalition.

Note that each player is interested in an increase in $x(S)$ with $i \in S \in \Pi$. Therefore for each $i \in S$ $(\in \Pi)$, decision over whether to accept an offer x or not depends solely on $\Sigma x_i = x(S)$. This implies that an offer x is accepted by members of S only if the player in S whose minimum acceptance level is the greatest. Consequently, the second stage becomes equivalent to a bargaining game with $s(=|S|)$ players with the same recognition probabilities.

Note that the limit outcome in the third stage depends upon the way how they negotiate within the coalition. In fact, the recognizability in the grand coalition does not necessarily guarantee a similar power inside a coalition. One might envision that a bargaining inside a coalition could be totally unrelated to the bargaining power for the bargaining at the level of the grand coalition. Further, one could elaborate on a theory which might explain the bargaining power for different coalition altogether. Here, on a trial basis, we simplistically assume that the same recognizability carries to both levels.

As we mentioned above, that the limit outcome in the third stage for $S \in \Pi$ with $x(S)$ expressed in terms of the share is given by $(p_i/P_S)_{i \in S}$ where $P_S = \Sigma p_i$. In the second stage, the limit outcome is given by $[s_i p_{S_i} p_i / (P_{S_i}(\sum_{S\in\prod} sP_S))$, where $i \in S_i \in \Pi$, $p_S = max\ p_j$, and $s_i=|S_i|$. This determines the payoffs for the first stage (reduced) game. What makes analysis easier is the fact that the payoffs are proportional to the weights p_j's, and hence players' incentives to form a coalition is aligned at least in some respect. In other words, if we define the coalitional coefficients by $C(S,\Pi) = (sp_S/(P_S \cdot (\sum_{S'\in\prod} s'p_{S'}))$, then the preference order of player i over coalition together with a coalition structure is completely reflected in $C(S,\Pi)$(with $i \in S$).

For the analysis of the first stage outcome, we assume the generic p_i's, i.e. all p_i's are distinct. In the first stage, in general the limit outcome depends on the ordering and so unlike the second and the third stage the outcome, in particular the equilibrium coalition structure may not be unique in this sense. By the backward induction procedure over all possible set of remaining players (which is finite) one can assure the existence of the stationary subgame perfect equilibrium. Below, we shall concentrate on this reduced hedonic game.

2. HEDONIC GAME

Since the reduced game for the first stage is a hedonic partition function game, one may associate the equilibrium coalition structure with the stable coalition structure discussed in the literature, mainly for the hedonic characteristic function game (cf. Banerjii, Konishi, and Sonmez (2001), Bogomolnaia and Jackson (2001)). Several concepts have been proposed by far and one of the major concerns is the non-emptiness of the solution.

Among several criteria proposed, the core is the most basic concept, in which the absence of a coalition (currently not in the structure) which foresees an immediate gain by deviation to form. Nash stability and several derived concepts restrict the deviation to a single player. Farsighted stability requires a lack of deviation by a coalition when taking into account of a chain of reactions by other coalitions.

Compared to these stability concepts, equilibrium coalition structures we consider lies somewhere between the core and the farsighted stable set, in the sense that players foresee the consequence of their deviation but along with a probabilistic expectations and stationarity of equilibria help existence. Backed by this motivation, here we shall focus upon the stationary equilibrium coalition structures.

First we make preliminary observations:

Lemma: (a) $\Pi=\{N\}$ is not an equilibrium coalition structure except for the case where $p_i=p$ for all i (which we preclude here in this section).

(b) Given an equilibrium coalition structure Π, for $i,j \in N$, $\{i\},\{j\} \in \Pi$ only if $p_i = p_j$ (which we preclude in this section)

(c) Over coalitions including two players, those two players' preferences coincide.

Proof: We provide proof for (b). In Π, for generic $(p_i)_N$ at most one $S \in \Pi$ with $s=1$.

(If there are $\{i\}$ and $\{j\}$ with $\{i\},\{j\} \in \Pi$, then defining $\Pi' =(\Pi \setminus \{i\},\{j\} \cup \{i,j\})$, $c(\{i\},\Pi)=c(\{j\},\Pi)=(1/(p_i+p_j+P^C)<((2p_{\{i,j\}})/(2p_{\{i,j\}}+P^C)(p_i+p_j)=c(\{i,j\},\Pi')$.)

These properties show that there is always a benefit from forming a coalition starting from being a single player, but to include all is not beneficial. This can be seen from the formula for C saying that the gain from coalition stems from the size s and the recognition probability of the most recognizable players in the coalition, and the property of the game where players are dividing fixed amount of money. This suggests that the softest player would be popular because it contribute to the size of coalition while demanding modestly inside the coalition. Then the most recognizable player is also popular? Not always, as the next example shows.

Example: Consider 3 person case, with $p_1 > p_2 > p_3$.

The equilibrium coalition structure is either $\Pi=\{\{1,3\},\{2\}\}$ or $\Pi' =\{\{2,3\},\{1\}\}$ depending upon the sign of $[C(\{1,3\},\Pi)-C(\{2,3\},\Pi')]$ or that of $[p_2{}^2(p_1-p_3)\geq p_1{}^2(p_2-p_3)]$. Thus roughly speaking, if p_2 is close to p_3, then $\{1,3\}$ forms, while if p_2 is close to p_1, then $\{2,3\}$ forms. In the former case, player 1's recognizability is unsurmountable, so to have 1 as the most recognizable player to obtain higher share for the coalition while having 3 to inflate the size of coalition is the optimal for the participants. In the latter case, 1 and 2 are close substitutes, and so 2 may be chosen as the representative by 3 to protect its internal share. In either case, the softest player 3 belongs to the formed coalition.

In the case of 4 person problem, by Lemma 1, either one 3 person coalition, or two 2 person coalitions form. Let S be the coalition with the highest C under this restriction. Then all the member may wait to form S in any ordering.

For 5 players game, 2 players coalition is the best to form with the expectation of remaining 3 players partitioned into 2 players and one player.

We list further properties which one derive from the formula for C. Throughout, we shall assume that $p_1 > p_2 \ldots > p_n$.

1. Representation Effect : Suppose $S(\neq \phi, N)$ forms. Then for any $\Pi \ni S$, $c(S,\Pi) sp_S/(sp_S+(n-s)p_{N\setminus S})P_S)$.
(For, $(n-s)p_{N\setminus S}$ is the upper bound for possible representative coefficients for any sub-partition of $N \setminus S$.)

2. Replacement : If $c((s \setminus\{i\}) \cup\{j\},\Pi') > c(S,\Pi)$ where $\Pi' =(\Pi \setminus\{S,S'\}) \cup\{S \setminus\{i\} \cup\{j\},(S' \setminus\{j\}) \cup\{i\}\}$.
(For $c(S,\Pi) = sp_S/(sp_S+sp_{S'}+P^C)P_S) < sp_S/(sp_S +s'p_{S'} +P^C)(P_S-p_i+p_j)) = c((S \setminus\{i\}) \cup\{j\},\Pi')$.)

3. Redundancy of an above average player : If $i \in S$ and $p_{S\setminus\{i\}} = p_S$ and $p_i \geqq P_S/s$), then for $\Pi' =(\Pi \setminus S) \cup\{S \setminus\{i\},\{i\}\}$, $c(S,\Pi) < c(S \setminus\{i\},\Pi')$.
(For $c(S,\Pi) = sp_S/((sp_S+P^C)P_S) < (s-1)p_S/((s-1)p_S+p_i+P^C)(P_S-p_i) = c(S \setminus\{i\},\Pi')$.)

4. Addition of the softest player : Let $p_n = \min_{i\in S\cup S'} p_i$ with $n \in S' \neq\{n\}$ and $p_S > p_{S'}$. Then for $\Pi' = (\Pi \setminus \{S, S'\}) \cup\{S \cup\{n\},S' \setminus\{n\}\}$, $c(S,\Pi) < c(S \cup\{n\},\Pi')$:

(For, $c(S, \Pi)=sp_S/(sp_S+s'p_{S'}+P^C)P_S)$ and $c(S \cup \{n\},\Pi') = (s+1)p_S/((s+1)p_S+(s'-1)p_{S'}\}+P^C)(P_{S \cup \{n\}}))$, and $c(S \cup \{n\},\Pi')=(s+1)p_S/((s+1)p_S+(s'-1)p_S)$, and $c(S, \Pi)<c(S \cup \{n\},\Pi')$ is equivalent to $(P_S-p_n)(sp_S+s'p_{S'}+P^C)>(p_S-p_{S'})P_S$ which is true because $P_S \geqq p_S$, $p_{S'} \geqq p_n$, P^C and $sp_S \geqq P_S$.)

Next, we define the properties which hold for equilibrium coalition structures.

(PS) Popularity of the softest player : If n is the only softest player, then in equilibrium, $n \in S^*$ such that $c(S^*, \Pi) = \max_{S \in \Pi} c(S, \Pi)$.

(HB) Heavy bottom with or without the most recognizable player : S^* given as above consists of either, all players i with $i \geqq k$ for some k, or all players i with $i \geqq k$ for k and player 1.

Theorem : Sspe coalition structures satisfy (PS) and (HB).

Proof : (PS) : Suppose this is not the case and so in Π there are S,S' with $n \in S'$, $c(S', \Pi) < c(S, \Pi)$.

Consider the ordering under which all $S'' \in \Pi \setminus \{S,S'\}$ form and so there is (on-the-path) subgame consisting of only players in $S \cup S'$. In particular an $i \in S$ could be the initial proposer. Since a singleton coalition is always the worst in terms of the weight, $|S|=1$. Suppose that there is $i \in S$ with $p_i \leqq p_S'$. Then by property 3, we have $c(S \cup \{n\} \setminus \{i\},\Pi') > c(S,\Pi)$ where Π' is as in property 3. For any subgame equilibrium Π'' (which is a refinement of Π' on the part of S'), $c(S \cup \{n\} \setminus \{i\},\Pi'') \geqq c(S \cup \{n\} \setminus \{i\},\Pi')$, and n also wishes to join S. Therefore this cannot be an equilibrium.

Next, suppose that for all $i \in S$, $p_i > p_{S'}$. In particular $p_S > p_{S'}$ and so both S and n have an incentive to merge, and hence this cannot be an equilibrium. This establishes (PS).

As for (HB), note that for a sspe, each player proposes the coalition to which that player belongs, which is accepted by the members along the equilibrium path. Thus there would be no delay when $\delta < 1$. So, for any $S_1,S_2 \in \Pi$, there is a subgame reacheable in a SSPE such that $S_1 \cup S_2$ are the remaining players. We refer to such a subgame as a subgame with S_1 and S_2.

(HB): Suppose there is $j \notin S^*$ but $x(S^*,\Pi^*)<x(S^* \cup \{j\},\Pi^{*'})$ where $\Pi^{*'}=(\Pi^* \setminus \{S^*,S^j\}) \cup \{S^* \cup \{j\},S^j \setminus \{j\}\}$ (where if $S^j \setminus \{j\}=\varphi$, cross this term out). We may choose j so that p_j is the smallest among these j's. Then in the subgame with S^* and S^j, players in S can offer $S^* \cup \{j\}$ and the acceptance by j does not hurt the members of S^* because whatever a subpartition of $S^j \setminus \{j\}$ may be, the sum of weights is at most $(s^j-p)P_S^j$.

By (HB), no player whose recognition probability is less than the average of S^* is not left outside of S^*.

Next, suppose that there is $j \in S^*$ with j's recognition probability is more than the average of S^*. Then for the subgame with S^*, a member in $S^* \setminus \{j\}$ can propose $S^* \setminus \{j\}$ which shall be accepted.

Finally, suppose $P_{S^*}<p_1$ and there is $j \notin S^*$ with $p_j < P_{S^*}$. Let $i \in S^*$ be $p_i = P_{S^*}$, and $1 \in S^1 \in \Pi^*$. Consider the subgame with S^* and S^1. A member of $S^* \setminus \{i\}$ can propose $(S^* \setminus \{i\}) \cup \{1\}$, and this shall be accepted because whatever a sub-partition of $(S^1 \setminus \{1\}) \cup \{i\}$ may be, coalitional weight for S^1 was s^1a_1 and that for a sub-partition of $(S^1 \setminus \{1\}) \cup \{i\}$ is at most $s^1P_{\{(S^1 \setminus \{1\}) \cup \{i\}\}}<s^1p_1$.

4. Multistage GAME

Now, let us consider a game in which coalitions could be formed within each coalition when they divide coalitional gain following the preceding bargaining stage. The rule of the game is identical to what we stipulated above for a single layer problem. As to the timing, we may assume that coalitional bargaining proceeds simultaneously without knowing the actions taken in other coalitions, just to avoid the possibility that what is going on in the other coalitions affects the development of the bargaining in the another.

Then we shall have a family of games $G(S; x)$ meaning coalition formation game for the bargaining among members of S which is formed in the previous round, to divide monetary gain x. This process goes on until the coalitions formed are singletons and pairs, or the same coalition structure reappears in a raw. Thus the process ends in finitely many rounds, say K. Let the family of partitions $\{\Pi^k\}$ $(k = 1,...,K)$ be resulting coalition structures. These must satisfy that Π^k is a refinement of Π^{k-1} and if an element $S \in \Pi^k \cap \Pi^{k-1}$, then $S \in \Pi^{k'}$ for all k'. Π^K is the finest coalition structure and a coalition S in some Π^k but not in Π^K such that any coalition strictly contained in S and belongs to some $\Pi^{k'}$ is an element of Π^{K}, a coalition 1-level above Π^{K}.

Not much can be said in general, but we have the following properties inherited from the one round game. (We shall describe everything in terms of the respective limit values.)

Proposition: With generic coefficients, the finest coalition structure includes coalitions of at most size 2, and in a sspe, the softest player in a coalition S which is 1-level above Π^{K} belongs to 2-person coalition.

Proof: Suppose there is a coalition with more than 2 players in sspe. Suppose other players committed to singletons, two players are left. Then if the penultimate player offers to form a pair, the last player gets better off by accepting the proposal, and so the penultimate player has an incentive to offer it, which leads to a contradiction.

As to the softest player, we apply the same argument as above.

4.1 NESTED STRUCTURE: AN EXAMPLE

Nested Structure: Example

To show the possibility of the emergence of a nested coalition structure, we appeal to an example with non-generic probabilities, i.e. two types case. There are m more recognizable players, and n players with a less probability. Say the ratio of two probabilities is given by $\alpha < 1$.

First, take the case where $m = 1$, which we may call the player 0 and $n>1$.

Then, consider a nested sequence $\{S_i\}$ $i=0, ..., n-1$ such that $S_i = \{0,1,...,n-i\}$. Given S_0, if a partition $\{S_1,\{n\}\}$ forms, then the shares of money for the player n, is $(\alpha/(n+\alpha))$. Then if a partition of S_1, $\{S_2,\{n-1\}\}$ forms for the bargaining to split $(n/(n+\alpha))$, then the share of $n-1$ is $(\alpha/(n-1+\alpha))$. Note that $(n/(n+\alpha))(\alpha/(n-1+\alpha)) > (\alpha/(n-1+2\alpha))(> (\alpha/(n+\alpha)))$, where the right-hand-side represents the monetary gain when 2 players remain a singleton in the beginning, i.e. at the partition $\{S_2,\{n-1\},\{n\}\}$. Thus, for the player $n-1$, to join S_1 once and then knocked out is better than remaining singleton to start with. Also for the player 0, monetary gain under $\{S_1,\{n\}\}$ is $(n/(n+\alpha))(1/(1+n\alpha))$, that under $\{\{S_2,\{n-1\}\},\{n\}\}$ is $(n/(n+\alpha))((n-1)/(n-1+\alpha))(1/(1+(n-1)\alpha))$, and that under $\{S_2,\{n-1\},\{n\}\}$ is $((n-1)/(n-1+2\alpha))(1/(1+(n-1)\alpha))$. By the same or similar inequality, we have that $(n/(n+\alpha))((n-1)/(n-1+\alpha))(1/(1+(n-1)\alpha)) > ((n-1)/(n-1+2\alpha))(1/(1+(n-1)\alpha)) > (n/(n+\alpha))(1/(1+n\alpha))$. Incentive for remaining players in terms of monetary gain is proportional to that of the player 0. In general, replacing n by $n-k$, one obtains the same observation. Therefore, players prefer a gradual formation of nested coalitions. Due to the property of this non-generic situation, these comparisons exhaust all the possibilities, and hence we conclude that the nested sequence $\{S_i\}$ $i=0,...,n-1$ emerges.

Second, consider another extreme, where $m > 1$ and $n = 1$. By HB principle, no coalition consisting of 2 most recognizable players and one soft player would form. In fact, 2-player coalition consisting of one more recognizable and one soft players form, and the rest remains (i.e. indifferent between forming a coalition and staying as a singleton).

Next consider the case where 2 more recognizable and 2 soft players. If 3 person coalition consisting of 2 soft and 1 more recognizable player forms, then the first soft player to be kicked out would get monetary share of $(3/4)(\alpha/(2+\alpha))$, while if this soft player forms a pair with another more recognizable player, the monetary share is $(1/2)(\alpha/(1+\alpha))(> (3/4)(\alpha/(2+\alpha)))$, and hence 3 players coalition would not emerge. Instead, two pairs of a soft and moe recognizable players form, in order to prevent others to take advantage of remaining a singleton. Thus if there are many more recognizable players, the power of competition drives the size of a coalition downward.

The above 2 examples exemplify what to expect in general. If $n < m$, then n pairs would form. If $m < n$, then m coalitions would form with the size $[n/m]$ or $[n/m]+1$, where $[x]$ is the greatest integer less than or equal to x. Inside each coalition, a nested coalition structure would emerge when its size is more than 2.

5. Conclusion

Here, we investigated the coalition formation for pure bargaining game under the random proposer rule with asymmetric recognition probabilities which represent bargaining power. In this case, there is a possible gain from formation of a coalition, and in fact a nontrivial coalition structure arises in some equilibrium. When coalition does matter, we found that the softest player enjoys the popularity in forming coalition under the stationary perfect equilibrium coalition structure. One could apply this result to explain several practices in the parliament.

As one extension, we consider the multi-layer problem by means of an example to show that there is a possible nested coalition structure to emerge. This could be one illustration of anecdotal stories on how to control power by a few figures.

Aside from the generalization of the results obtained here, one interesting question is to investigate the origin of the power, like utilization of rent-seeking model to derive recognition probabilities from investment by players endogenously, which is one of the future research agenda.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Numbers JP 16K03550.

REFERENCES

Banerjee, S., Konishi, H., and Sonmez, T. (2001). Core in a Simple Coalition Formation Game. Social Choice and Welfare, 18, 135-153..

Baron, D. and Ferejohn J. (1989). Bargaining in legislatures. American Political Science Review, 83, 1181-1206.

Bloch, F. (1996). Sequential Formation of Coalitions with Externalities and Fixed Payoff Division. Games and Economic Behavior, 14, 90-123.

Bogomolnaia, A., and Jackson, M. O. (2001). The Stability of Hedonic Coalition Structures. Games and Economic Behavior, 38, 201-230.

Britz, V. Herings, P.J. & Preditetchinski, A. (2010). Non-cooperative Support For The Asymmetric Nash Bargaining Solution, Journal of Economic Theory, 145, 1951-67.

Chae, S. and Heidhues P. (2004). A Group Bargaining Solution. Mathematical Social Sciences, 48, 37-53.

Harsanyi, J., (1977). *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*. Cambridge: Cambridge Univ. Press,.

Imai, H. and Salonen H. (2000). Representative Bargaining Solution for Two-Sided Bargaining Problems. Mathematical Social Sciences, 39, 349-365.

Imai, H. and Watanabe N. (2005). On the Neutrality of Coalition Formation In a Pure Bargaining Model. Japanese Economic Review, 56, 352-362.

Koszy, L. (2009). Sequential Coalition Formation and the Core in the Presence of Externalities. Games and Economic Behavior, 66, 559-565.

Montero, M. (1999) "Coalition Formation in Games with Externalities", mimeo., Tilburg University.

Moldovanu B, Winter E. (1995) Order independent equilibria. Games and Economic Behavior, 9, 21-34.

Myerson, R. (1991) Game Theory. Cambridge: Harvard University Press.

Okada, A. (1996). A Noncooperative Coalitional Bargaining Game with Random Proposers. Games and Economic Behavior 16, 97-108.

Ray, D. and Vohra R. (1997). Equilibrium Binding Agreement. Journal of Economic Theory, 73, 30-78..

Ray, D. and Vohra R. (1999). A Theory of Endogenous Coalition Structures. Games and Economic Behavior, 26, 286-336.

Vidal-Puga, J. J. (2012). The Harsanyi Paradox and the "Right to Talk" in the Bargaining Model among Coalitions, Mathematical Social Sciences, 64, 214-224.

MODELLING PROBLEMS DYNAMICS BY DIFFERENTIAL GAMES

SALTANAT AMIRGALIYEVA[1], VALENTIN OSTAPENKO[2]

[1]Institute of Information and Computational Technologies, Almaty, Kazakhstan, saltanat_amirgal@mail.ru
[2]Institute for applied system analysis, Kiev, Ukraine

Abstract: *The research investigates differential games, the dynamics of which is described by ordinary differential equations. These game models consider problems of pursuit-evasion that are defined by terminal set and the set of phase constraints or terminal functionality. Strategies of different players are described by - strategy of Pshenichnyi B. (1992), its modifications, and connections between are established. Structure of differential games is described using one-parameter semigroups of operators of Pshenichnyi B., Ostapenko V.et al. (1992, 2008) on the basis of which - strategies can be built and operators describe the set of initial positions, favorable for a particular player in the game models with terminal set. Two classes of games are considered: games with terminal sets and with terminal functionality.*

Keywords: *Differential games, pursuit-evasion, terminal set, set of phase constraints, strategy*

1. INTRODUCTION

In article the differential games which dynamics is described by ordinary differential equations are considered. In these gaming models tasks of convergence-evasion which are described by terminal set, set of phase constraints or a terminal functional are researched. Also various strategies of players are described: ε-strategies, its modifications connection in between are established.

2. THE DYNAMICS OF THE GAME PROBLEMS AND OPERATOR CONSTRUCTIONS IN DIFFERENTIAL GAMES

2.1 Games, described by ordinary differential equations

We consider the dynamic system specified by the differential equation

$$\dot{z} = f(z,u,v), \tag{2.1}$$

where is $z \in E^n$, $u \in U$, $v \in V$, U and V – compacts in Euclidean spaces.

Parameters U and V are managed accordingly by the players P (catching up) and E (escaping). Under valid controls of the players P and E will be understood functions $u(t)$ and $v(t)$ with values in the U and V, accordingly. The set of all admissible controls of players P and E, defined on the interval $[a,b]$ (in half interval $[a,b)$),will be denote, respectively, through the $U[a,b]$ and $V[a,b]$ ($U[a,b)$ and $V[a,b)$).

We assume that in the future function f and sets U and V satisfies the following assumptions.

Assumption 1. Function $f(z,u,v)$ – is continuous on set of variables and locally Lipschitz on z (i.e. satisfies the condition of Lipchitz on z on each compact $K \subset E^n$ with constant L_K, depending on the K).

Assumption 2. There exists a constant $C \geq 0$ such that for all $z \in E^n$, $u \in U$, $v \in V$

$$\left|\langle z, f(z,u,v)\rangle\right| \leq C(1+\|z\|^2).$$

Assumption 3. The set $f(z,U,v)$ – is convex for all $z \in E^n$, $v \in V$.

Assumptions 1 and 2 guarantee the existence, uniqueness and continuity of solutions $z(t)$ equation (2.1) on all axis $[0,+\infty)$ for an arbitrary initial condition $z(0) = z_0$ and by substituting in (2.1) instead of parameters u and v any valid controls $u(t)$ and $v(t)$ of players P and E, respectively. We shall denote a solution $z(t)$ of equation (2.1), the corresponding $u(t)$, $v(t)$ and initial condition $z(0) = z_0$ through

$$z(t \mid u(\cdot), v(\cdot), z_0).$$

Consider an arbitrary interval $[0,\theta]$, $\theta < +\infty$. Assumption 3 guarantees in the topology of uniform convergence on a segment $[0,\theta]$ the compactness of multitude solutions that correspond to various admissible controls of player *P*, $u(\cdot)$ and the initial position z_0. Said remains in force, if the initial position z_0 is not fixed and runs through a compact set of $K \subset E^n$.

From the described properties follows that, if $u_k(\cdot) \in U[0,\theta]$, $x_k \in K$, $k = 1,2,...$ – some sequences, and

$$z_k(t) = z(t \mid u_k(\cdot), v(\cdot), x_k)$$

the sequence of corresponding solutions to the equation (2.1), then there exists a subsequence $\{ z_{k_m}(\cdot) \}$ sequence of $\{ z_k(\cdot) \}$, which evenly on $[0,\theta]$ converges to the function $z_0(\cdot)$. And there are such $u(\cdot) \in U[0,\theta]$, $x \in K$, that $z_0(t) = z(t \mid u(\cdot), v(\cdot), x)$.

Consider two classes of gaming models: game models with terminal set and game models with terminal functionality. In the first case, the objective of the players is described using the terminal set $M \subset E^n$ and set of phase constraints $N \subset E^n$. Sets *M* and *N* are assumed to be closed, and $M \subset N$.

We fix the moment $\theta > 0$. The player P`s goal is to achieve inclusions $z(\theta) \in M$, $z(t) \in N$, for all $t \in [0,\theta]$, i.e. output trajectory $z(t)$ to *M* at the moment of θ, holding it in a variety of *N*. The goal of the player *E* –is the opposite, i.e. to achieve conditions: either $z(\theta) \notin M$, or for some $t < \theta$ $z(t) \notin N$.

Game models with a terminal functionality the goals of players are described using $\Phi : E^n \to E^1$. The player *P*'s goal is to minimize the functionality $\Phi(z(\theta))$, depending on the end of the trajectory. The player *E*'s goal is the opposite, i.e. is to maximize this functionality. Game models with the terminal sets *M* and *N* are selected not arbitrary, but closed subsets in E^n. This is done for the convenience of building an appropriate mathematical apparatus. In the same aim, we impose some conditions on function $\Phi(z)$. We believe that $\Phi(z)$ satisfies Lipchitz condition with constant L_K on each compact K.

The above viewed gaming models have a relationship. Functionality Φ can represent the distance to the set M. In this case, the player P's goal – to approach at the moment θ as close as possible to the set M. Formally, the first game can be reduced to the second, thinking that $\Phi(z) = 0$, $z \in M$ and $\Phi(z) = 1$, $z \notin M$. However, the specified function does not satisfy the required above Lipchitz criteria condition and mathematical apparatus developed for studies of these classes of games varies greatly.

A characteristic feature of differential games is that players don't know the enemy's actions in the future. The paper used different strategies of players, which use some information about the current position and the actions of the enemy. The player E will choose his current control, using mainly the knowledge of current position. For the player P different strategies are used. This is ε - strategies [1], in which the greatest enemy's informational discrimination is supposed: player E reports its control to player P for a while $\varepsilon > 0$ forward. In addition, the player P uses information about the current position. Since the parameter ε manages player E, then ε - strategies are equivalent to strategies , in which the player P selects its current control, knowing the initial position and the entire background to enemy action. These strategies are based on certain Voltaire maps. A special case of the latter strategies are strategies in which the player P chooses its current control, knowing the starting position and the current control of the enemy. Such a strategy will be called a counter-strategy.

2.2. Operators on sets

Consider a dynamical system described by equation (2.1) and satisfies the assumptions 1-3.

Definition 1. Through P_ε, $\varepsilon \ge 0$,denote the operator which puts into each closed set in compliance with $M \subset E^n$ the set of $P_\varepsilon M$ of all points $z_o \in E^n$ such that for any admissible control $v(t)$, $t \in [0,\varepsilon]$, of the player E there is a valid control $u(t)$, $t \in [0,\varepsilon]$, of the player P, such that the corresponding solution $z(t) = z(t \mid u(\cdot), v(\cdot), z_0)$ of equation(2.1) with initial z_0 runs $z(\varepsilon) \in M$, switch, i.e. path $z(t)$ with z_0 gets on M at the moment ε.

Formally, with the help of union and intersection operations, the operator P_ε can be described as

$$P_\varepsilon M = \bigcap_{v(\cdot)\in V[0,\varepsilon]} \bigcup_{u(\cdot)\in U[0,\varepsilon]} \{z_0 \in E^n : z(\varepsilon \mid u(\cdot), v(\cdot), z_0) \in M\}. \tag{2.2}$$

Remark 1. In the definition of 1 can be considered that the controls $u(\cdot)$ and $v(\cdot)$ are determined only at the half-open interval $[0,\varepsilon)$ since the change values controls $u(t)$ and $v(t)$ at one point does not change the trajectory. The solution $z(t)$, defined on $[0,\varepsilon)$ can always be uniquely continuously extend to the interval $[0,\varepsilon]$, putting $z(\varepsilon)=\lim_{t\to\varepsilon} z(t)$. This fact will be used later.

Remark 2. The set of $P_\varepsilon M$ can be interpreted as the set of initial positions z_0, starting from which the player P can output trajectory $z(t)$ to M at the moment of ε, knowing the control $v(t)$ of the player E in advance over the whole interval $[0,\varepsilon]$. If $z_o \notin P_\varepsilon M$, then there exists control of player E, that for all admissible controls of the player P holds $z(\varepsilon)\notin M$. In this case, the players strategy, i.e. they select their controls once in the entire interval $[0,\varepsilon]$. Here the player E knows z_0, and the player P uses information about z_0 and chosen control $v(t)$, $t\in[0,\varepsilon]$.

Lemma 1. The set of $P_\varepsilon M$ is closed for the set M.

Consider $P_{N,\varepsilon}M=(P_\varepsilon M)\bigcap N$. Obviously, that $P_{N,\varepsilon}M$ is a closed set, if M and N are closed. Thus, lemma 1 allows to reuse operators $P_\varepsilon M$ and $P_{N,\varepsilon}M$.

Let $\omega=\{\tau_0=0\le\tau_1\le...\le\tau_k=t\}$ – the final partition of the interval $[0,t]$. Suppose,

$$P_N^\omega M=P_{N,\delta_1}P_{N,\delta_2}...P_{N,\delta_k}M,$$

where $\delta_i=\tau_i-\tau_{i-1}$, $i=1,...,k$.

Remark 3. If $N=E^n$, then we write that $P^\omega M=P_{\delta_1}P_{\delta_2}...P_{\delta_k}M$. Suppose $z_0\in P^\omega M$ and at the initial time player P knows the player `s control forward for the time δ_1. Then P can aim and get on $P_{\delta_2}...P_{\delta_k}M$ at the moment $\delta_1=\tau_1$. If he got hit to $P_{\delta_2}...P_{\delta_k}M$, he learns control of player E on the time δ_2, then the player P can aim and get to the set $P_{\delta_3}...P_{\delta_k}M$ in $\tau_2=\delta_1+\delta_2$ time. Continuing this process further, the player P will achieve inclusion $z(t)\in M$ .Meanwhile, the player P chooses his control at the points τ_{i-1}, $i=1,...,k$, on the interval $[\tau_{i-1},\tau_i)$, knowing $z(\tau_{i-1})$ and future control of the player E on the interval $[\tau_{i-1},\tau_i)$. Similarly, if $z_0\notin P^\omega M$, then the player E can at the time $\tau_0=0$ choose such a control that for any control of player P the corresponding trajectory does not fall to the set $P_{\delta_2}...P_{\delta_k}M$ at the moment $\delta_1=\tau_1$. Continuing this process further, we find that $z(t)\notin M$ .Thus, the player E chooses his control at the points τ_{i-1} for the next interval $[\tau_{i-1},\tau_i)$, knowing $z(\tau_{i-1})$.

Using formula (2.2) $P_N^\omega M$ can be represented as follows

$$P_N^\omega M=\bigcap_{v_1(\cdot)\in V[0,\tau_1)}\ \bigcup_{u_1(\cdot)\in U[0,\tau_1)}\ \bigcap_{v_2(\cdot)\in V[\tau_1,\tau_2)}\ \bigcup_{u_2(\cdot)\in U[\tau_1,\tau_2)}\ ...$$
$$...\bigcap_{v_k(\cdot)\in V[\tau_{k-1},t]}\ \bigcup_{u_k(\cdot)\in U[\tau_{k-1},t]}\{z_0\in N: z(t)=z(t\,|\,\{u_1(\cdot),u_2(\cdot),...,u_k(\cdot)\},$$
$$\{v_1(\cdot),v_2(\cdot),...,v_k(\cdot)\},z_0)\in M, z(\tau_i)\in N, i=1,...,k-1\}. \quad (2.3)$$

Definition 2. $\tilde{P}_{N,t}M=\bigcap_{|\omega|=t}P_N^\omega M$.

Theorem 2.1. The equality

$\tilde{P}_{N,t_1+t_2}M=\tilde{P}_{N,t_1}\tilde{P}_{N,t_2}M$ is justly

2.3. Operators on functions

Let $\Phi: E^n\to E^1$ satisfies Lipchitz condition with constant L_K on each compact K and run 1-3 assumptions.

We define the operator R_ε ,that assigns to each continuous function $\phi: E^n\to E^1$ the function

$$\psi(x)=\sup_{v(\cdot)\in V[0,\varepsilon]}\ \min_{u(\cdot)\in U[0,\varepsilon]}\phi(z(\varepsilon\,|\,u(\cdot),v(\cdot),x)). \quad (2.4)$$

Note that due to the continuity ϕ and assumption 3 minimum at (2.4) is achieved.

The operator R_ε can be associated with the operator P_ε. Indeed, if $D_c(\phi) = \{x \in E^n : \phi(x) \le c\}$, then

$$D_c(\psi) = P_\varepsilon D_c(\phi).$$

The proof of this equation follows directly from the definition.

Let $\omega = \{\tau_0 = 0 \le \tau_1 \le ... \le \tau_k = t\}$, $\delta_i = \tau_i - \tau_{i-1}$, $i = 1,...,k$, – a partition. Put $R^\omega \phi = R_{\delta_1} ... R_{\delta_k} \phi$, $\tilde{R}_t \phi = \sup_{|\omega|=t} R^\omega \phi$.

An analogue of formula (2.3) is just:

$$R^\omega \Phi(x) = \sup_{v_1(\cdot)\in V[0,\tau_1)} \min_{u_1(\cdot)\in U[0,\tau_1)} \sup_{v_2(\cdot)\in V[\tau_1,\tau_2)} \min_{u_2(\cdot)\in U[\tau_1,\tau_2)} ...$$

$$... \sup_{v_k(\cdot)\in V[\tau_{k-1},\tau_k]} \min_{u_k(\cdot)\in U[\tau_{k-1},\tau_k]} \Phi(z(t \mid \{u_1(\cdot),...,u_k(\cdot)\},\{v_1(\cdot),...,v_k(\cdot)\},x)) \tag{4.9}$$

From the assumption 2 and a local Lipschitz Φ can deduce the following:

Theorem 2.2. For any $x \in E^n$

$\tilde{R}_{t_1+t_2}\Phi(x) = \tilde{R}_{t_1}\tilde{R}_{t_2}\Phi(x)$ runs.

3. ε - STRATEGY IN THE GAME MODELS

3.1. Description of strategies and course of the game

A characteristic feature of ε - strategy [1] is that the player P uses the information on the future controls of the player E for a certain time interval, the length of which is determined by the player E.

It is possible to enter a number of different equivalent definitions of the ε --strategy [1-3]. The following will focus on one of them and explore the game models in these strategies.

In the beginning we give some informal description of the progress of the game. The game can be held on the finite interval $[0,\theta]$ or on the endless semiaxis $[0,+\infty)$. In the first case, the player E at the starting time moment selects a finite partition $\omega = \{\tau_0 = 0 \le \tau_1 \le ... \le \tau_k = \theta\}$ of interval $[0,\theta]$, in the second case, the player E selects a partition $\omega = \{\tau_0 = 0 \le \tau_1 \le ... \le \tau_k \le ...\}$ of semiaxis $[0,+\infty)$ with points τ_i, no having condensation points.

Let that at the time τ_{i-1} the dynamic system is at a point $z(\tau_{i-1})$, $i = 1,2,...$. Using this information player E chooses his controls $v_i(t)$, $t \in [\tau_{i-1},\tau_i)$. We believe that the player P knows, τ_{i-1}, $z(\tau_{i-1})$ and $v_i(t)$, $t \in [\tau_{i-1},\tau_i)$ and selects his controls $u_i(t)$, $t \in [\tau_{i-1},\tau_i)$. Substituting $v_i(t)$ and $u_i(t)$ to (1.1) we can find solution $z(t)$ of the equation (2.1) with the beginning of the $z(\tau_{i-1})$. By virtue of Lipschitz $z(t)$, this solution can be extended to the interval $[\tau_{i-1},\tau_i]$. Thus, the dynamic system at the moment τ_i is at the point $z(\tau_i)$ and repeats the process further. Since the number of points τ_i neither finite, nor they are not the condensation points, using the process described, solution of the equation (1.1) on the interval $[0,\theta]$, or on the entire semiaxis $[0,+\infty)$ is constructed.

Note that during the game the player P at each moment τ_{i-1} knows the player E 's control at the time $\varepsilon_i = \tau_i - \tau_{i-1}$ forward; i.e. there is an information discrimination of the player E.

We give a formal definition of ε - strategy.

Definition 3. ε - strategy of the player P will be called the mapping $\Gamma_P(z,\tau,v(\cdot),\varepsilon)$, that to the point $z \in E^n$, moment $\tau \ge 0$, number $\varepsilon > 0$ and control $v(\cdot) \in V[0,\varepsilon)$ puts into conformity the line control $u(\cdot) \in U[0,\varepsilon)$.

Definition 4. Under ε - strategy of the player E we mean a pair of objects: partition $\omega = \{\tau_0 = 0 \le \tau_1 \le ... \le \tau_k \le ...\}$, the points τ_i, without points of thickening and display $\Gamma_E(z,i)$, that assigns a point $z \in E^n$ and integer i>o. , control $v(\cdot) \in V[0,\varepsilon_i)$, where $\varepsilon_i = \tau_i - \tau_{i-1}$.

We now describe a formal course of the game by using certain above ε - strategy. Thus, we believe that given display Γ_P, Γ_E and partition ω.

Let z_0 – initial position. At the moment $\tau_0 = 0$ the player E chooses valid control

$v(t)=\Gamma_E(z_0,1)(t)$, $t\in[0,\varepsilon_i)$.

The player P , knowing z_0 , $\tau=\tau_0=0$, ε_1 and $v(t)$, $[0,\varepsilon_1)$, builds on an interval $[0,\varepsilon_1)$ control:

$u(t)=\Gamma_P(z_0,0,v(\cdot),\varepsilon_1)(t)$, $t\in[0,\varepsilon_1)$.

Let $z(t)$ is solution of (1.1) corresponding $u(\cdot)$ and $v(\cdot)$ starting with z_0 on an interval $[0,\varepsilon_1)$. This solution can be extended to the interval $[0,\varepsilon_1]$. At the moment τ_1 we take a point $z(\tau_1)$ for initial and repeat the described process.

As a result, to the time τ_i will be built controls $u(t)$ and $v(t)$ on an interval $[0,\tau_i)$ and trajectory $z(t)$ on an interval $[0,\tau_i]$. When τ_i the player *E* chooses control

$v(t+\tau_i)=\Gamma_E(z(\tau_i),i+1)(t)$, $t\in[0,\varepsilon_{i+1})$,

the player *P* – control

$u(t+\tau_i)=\Gamma_P(z(\tau_i),\tau_i,v(t+\tau_i),\varepsilon_{i+1})(t)$, $t\in[0,\varepsilon_{i+1})$.

The result is a control $u(t)$ and $v(t)$ on $[0,\tau_{i+1})$ and solution $z(t)$, $t\in[0,\tau_{i+1}]$.

Continuing this process further, we construct controls $u(t)$, $v(t)$ and solution $z(t)$ on interval $[0,\theta]$ or semiaxis $[0,+\infty)$.

We describe modified ε -strategy, which, as will be seen from the under written, are equivalent to the defined in the definitions of 3 and 4.

1) The player *E* may inform in advance the player *P* its partition ω , selected at the initial moment. In this case, the information discrimination of the player *E* is enhanced, the player *P* can build its control differently, but it will not improve its situation.

2) The player *E* may not choose in advance at the time $\tau_0=0$, partition ω , and specify the points τ_i in the process of game. The choice τ_i may depend on τ_{i-1} and $z(\tau_{i-1})$. If the set $\{\tau_i, i=1,2,...\}$ has no condensation points, the game runs in the same manner that was described above. Controls $u(t)$, $v(t)$ and solution of (2.1), $z(t)$ are determined on all $[0,\theta]$ or on $[0,+\infty)$. If there are points of condensation, it is necessary to introduce new concepts, which would allow to continue the solution to address the entire segment $[0,\theta]$ or to semiaxis $[0,+\infty)$. In the above case, the set of strategies of the player *E* is wider than that which has been described in the definition of 4. However, the new strategies do not extend the capabilities of the player *E*.

4. THEOREMS ON ALTERNATIVES

Consider the game models with the terminal set, as described in paragraph 2.1.

Theorem 4.1 1) Let $z_0\in\tilde{P}_\theta M$. Then exists ε - strategy of player *P* such that for any ε - strategy of player *E* for the corresponding trajectory $z(t)$ with initial z_0 , runs $z(\theta)\in M$.

2) Let $z_0\notin\tilde{P}_\theta M$. Then exists ε - strategy of player *E* such that for any ε - strategy of player *P*, for the corresponding trajectory $z(t)$, with initial z_0 runs $z(\theta)\notin M$.

Theorem 4.1 is a theorem about the alternative. According to this theorem, the entire space of the game E^n is divided into two subsets $\tilde{P}_\theta M$ and $E^n\setminus\tilde{P}_\theta M\mathbb{S}$. The first subset describes all the initial position favourable for the player *P*, the second - for the player *E*.

Theorem 4.1 describes the structure of the game without phase restrictions, ... Consider the case when *N* is a proper subset of the E^n space.

Theorem 4.2 1) Let $z_0\in\tilde{P}_{N,\theta}M$. Then exists ε - strategy of the player *P* such that for any ε - strategy of the player *E* for the corresponding trajectory $z(t)$ with initial z_0 runs $z(\theta)\in M$ and $z(t)\in N$ for all $t\in[0,\theta]$.

2) Let $z_0\notin\tilde{P}_{N,\theta}M$. Then exists ε - strategy of the player *E* such that for any ε - strategy of the player *P* for the corresponding trajectory $z(t)$ with initial z_0 runs: either $z(\theta)\notin M$, or exists such $t\in[0,\theta]$, that $z(t)\notin N$.

Consider the game models with terminal functionality and the task of minimizing the terminal functionality, supplied in 2.1.

Theorem 4.3. 1) Exists ε - strategy of the player *P* such that for any ε - strategy of the player *E* for the corresponding trajectory $z(t)$ with initial z_0 runs:

$$\Phi(z(\theta)) \le \tilde{R}_\theta \Phi(z_0) .$$

2) For any $\delta > 0$ exists ε - strategy of the player *E* such that for any ε - strategy of the player *P* for the corresponding trajectory $z(t)$ with initial z_0 runs:

$$\Phi(z(\theta)) \ge \tilde{R}_\theta \Phi(z_0) - \varepsilon .$$

5. CONCLUSION

Structure of differential games described using one-parameter semigroups of operators on the basis of which you can build ε-strategy and many operators describe the initial positions, enabling for a player in game models with terminal sets. Game models with functional terminal operators describe the price game.

REFERENCES

Pshenichnyi, B.N, and Ostapenko, V.V. (1992). *Differential games*. Kyiv, p. 260.

Ostapenko, V.V., Amirgalieva, S.N., and Ostapenko E.V. (2005). *Convex analysis and differential games*. Almaty, p. 392.

Ostapenko, V.V. (2008). Convexity in differential games, in: P. Pardalos (Ed.). *Pareto-Optimality. Game Theory and Equilibria*, Springer Book.

B7
Heuristics, Metaheuristics & Nature Inspired Optimization Algorithms

BALCOR 2018
XIII Balkan Conference on Operational Research

Vehicle Scheduling Problem in Sugar Beet Transportation: A General Variable Neighborhood Search Approach

ANA ANOKIĆ[1], ZORICA STANIMIROVIĆ[2], TATJANA DAVIDOVIĆ[3], ĐORĐE STAKIĆ[4]

[1] Independent researcher, anokicana@gmail.com
[2] University of Belgrade, Faculty of Mathematics, zoricast@matf.bg.ac.rs
[3] Mathematical Institute, Belgrade, tanjad@mi.sanu.ac.rs
[4] University of Belgrade, Faculty of Mathematics, djordjes@matf.bg.ac.rs

---

Abstract: *A variant of Vehicle Scheduling Problem (VSP) arising from the sugar beet transportation in a sugar factory in Serbia is presented. The objective of the considered VSP is to minimize the required transportation time under problem-specific constraints. The problem is first formulated as a Mixed Integer Quadratically Constrained Program (MIQCP) and then transformed to a Mixed Integer Linear Program (MILP). The proposed MILP model was used within the framework of CPLEX solver, which produced optimal solutions only for small-size problem instances. Therefore, a General Variable Neighborhood Search (GVNS) is designed to solve problem instances of larger dimensions. GVNS is evaluated and compared against CPLEX on the set of real-life and generated problem instances. Obtained computational results show that GVNS is a promising solution approach to VSP, as it is able to reach high-quality (mostly optimal) solutions within very short running times.*

Keywords: *Vehicle Scheduling Problem, Mixed Integer Quadratically Constrained Programming, Mixed Integer Linear Programming, General Variable Neighborhood Search*

1. INTRODUCTION

This study considers a variant of Vehicle Scheduling Problem (VSP) that arises from a real-life problem of optimizing sugar beet transportation in a sugar factory in Serbia. The low price of sugar beet on the market has two main consequences regarding the organization of transport. First, transportation costs represent significant percentage of total production costs and therefore, savings in this early production stage are of great importance. Second, a company needs to keep individual farmers as suppliers of raw material for a longer period. For this reason, a company usually organizes the transport of sugar beet from each of many producers to the factory on its own expense by renting vehicles and hiring workers for transporting the goods, including loading and unloading. Therefore, from the company's point of view, it is necessary to have an efficient transport organization, which will satisfy all problem-specific constraints with minimum expenditure of time and money.

In the literature, there are several studies dealing with organizing the transport of agriculture raw materials. Sethanan and Pitakaso (2016) considered a vehicle routing problem (VRP) for raw milk collection, with the goal to minimize the sum of fuel costs and the costs of cleaning raw milk tanks on vehicles. The considered problem was solved by five solution approaches based on differential evolution algorithm. The problem of collecting olive oil in Tunisia was studied by Lahyani et al. (2015) and formulated as a Multi-Compartment Vehicle Routing Problem. The authors proposed an integer linear programming formulation and used branch-and-cut algorithm to provide optimal solutions. A case study of Australian sugar mill was presented by Higgins (2006), dealing with vehicle scheduling problem in sugar cane transportation. Two metaheuristic methods, Tabu search and Variable neighborhood search, were proposed as solution approaches to the VSP considered in Higgins (2006). Milan et al. (2006) integrated the problems of rail and road sugar cane transportation in Cuba, in order to reduce the total transportation costs. As the resulting problem showed to be difficult to solve, Milan et al. (2006) first solved a smaller-size subproblem by HyperLINDO solver and than used the obtained solution to construct a feasible solution of the integrated problem. Thuankaewsing et al. (2015) investigated the problem of maximizing the estimated sugar cane yield in Thai sugar cane industry, assuming fair benefits for all producers. The considered problem was solved by a tabu search heuristic proposed in Thuankaewsing et al. (2015). A review of applications of vehicle scheduling problems and vehicle routing problems together with solution approaches can be found in Bunte and Kliewer (2009).

Note that the variant of VSP considered in this study differs from the ones considered in Higgins (2006), Milan et al. (2006), and Thuankaewsing et al. (2015), due to the differences between sugar cane and sugar beet related to the sustainability in the open, the type of vehicles used for transport, and the available resources of the sugar factory. The considered VSP is first formulated as a Mixed Integer Quadratically Constrained Program

(MIQCP), with the objective and specific constrains that arise from the real-life situation. The proposed MIQCP is further reformulated as a Mixed Integer Linear Program (MILP) and tested on real-life and generated problem instances within the framework of the commercial CPLEX solver. As CPLEX provides optimal solutions only for small-size instances, a variant of Variable Neighborhood Search (VNS) metaheuristic, known as General Variable Neighborhood Search (GVNS) is designed to solve problem instances of larger dimensions. The choice of VNS method is motivated by its successful applications to various vehicle routing and scheduling problems, such as: inventory routing and scheduling problem in Liu and Chen (2012), location routing scheduling problem in Macedo et al. (2015), Vehicle Routing Problem with multiple trips in Cheikh et al. (2015), multiple trips VRP with backhauls in Wassan et al. (2017), Heterogeneous Fleet VRP in Bula et al. (2016), etc. Constructive elements of the proposed GVNS implementation are adapted to the problem characteristics. GVNS metaheuristic is tested and compared against the CPLEX solver on the set of real-life instances obtained from a sugar company in Serbia, as well as on the set of generated instances. The obtained computational results clearly indicate the potential of GVNS for the considered variant of VSP.

2. PROBLEM DESCRIPTION AND MATHEMATICAL FORMULATIONS

In the harvesting season of the sugar beet, a precise transportation plan has to be made on the daily basis. For each day, the list of locations with the amounts of goods to be transported is prepared in advance. As the quantities of collected goods at each location significantly exceed the capacity of vehicles, each location has to be visited several times in order to be emptied. It is assumed that all vehicles have same technical characteristics (i.e., average velocity, capacity). The factory area is the starting and the finishing point for each tour, meaning that each vehicle serves only one location in a tour before it returns to the factory. Tours with more than one location are not allowed, as it is important to prevent a bad quality mixture of sugar beet from different locations. Therefore, it is possible that last vehicle that serves some location is not fully loaded. When a vehicle loaded with sugar beet arrives to the factory area, a certain time is needed for unloading the sugar beet including the analysis of samples. After the unloading is finished, a vehicle can start a new tour. It is assumed that there is enough space and labor at each location and in the factory area to complete the loading and unloading of unlimited number of vehicles at the same time.

Having in mind specific sustainability of sugar beet, it is important that the collected quantities of goods do not stand in the open for too long, otherwise they will lose quality. In the case that on some location the goods are standing longer than a predetermined number of days, this location is considered as an urgent one. It is important that urgent locations are emptied during the working day. The total amount of goods transported to the factory within the working day should not be lower than a predetermined constant, which ensures continuous work of factory machines. Once the factory machines start to work, they should not be stopped, because the starting process is very expensive.

By taking into account all problem specific constrains mentioned above, the goal of the considered VSP is to find the optimal set of vehicle schedules which minimizes the maximum working time among all vehicles, i.e., the moment of time when all vehicles finish their tours. In our problem, a vehicle schedule is defined as the array of tours and the corresponding departure times from the factory.

In order to present mathematical formulations of the considered VSP, the following notation is introduced:

J : The set of locations; I : The set of vehicles; K : The set of tours;
n : The total number of locations; m : The total number of vehicles;
k_{max} : The maximum number of tours a vehicle can make during working day;
c_j : The quantity of goods collected at location $j \in J$;
d_j : The distance between the factory and location $j \in J$;
C : Capacity of vehicle; D : Daily factory needs; v : The average speed of a vehicle;
u : The average time that a vehicle spends in factory area between two tours, (the time needed for unloading with analyzing the samples);
w : The average time needed for loading a vehicle;
S_j : The time needed for serving location $j \in J$, calculated as the sum of driving time in both directions and loading and unloading time, i.e., $S_j = 2\frac{d_j}{v} + u + w$;
t_j: The number of days that the goods are kept in the open at location $j \in J$;
t_0: The maximum number of days that the goods can stay in the open without losing quality;
U : The set of urgent locations, i.e., $U = \{j \in J : t_j > t_0\}$;
T_j : Binary value assigned to location $j \in J$, defined as $T_j = 1$ if $j \in U$, and $T_j = 0$ otherwise;
ε: Small positive constant, satisfying $0 < \varepsilon < c_j/C$ for each $j \in J$;
t_{start}: Starting time; t_{end}: The end of working day.

Note that vehicles make different number of tours during the working day, depending on the visited locations. However, the number of tours for each vehicle is limited by k_{max}. In order to equalize the number of tours for all vehicles, we introduce the concept of *virtual tour*. The duration of a virtual tour is equal to zero, as it is assumed that during this tour a vehicle stays in the factory area. Therefore, by adding virtual tours to vehicles (if necessary), the objective function value remains unchanged. On the other hand, by setting the number of tours for each vehicle to k_{max}, we are able to simplify problem formulations without affecting its characteristics.

The following decision variables are used in mathematical formulations:

- Binary variables x_{ik}^{j} are equal to 1 if a vehicle $i \in I$ visits location $j \in J$ in the tour $k \in K$, and 0 otherwise. If a vehicle $i \in I$ has virtual tour $k \in K$, then $\sum_{j \in J} x_{ik}^{j} = 0$ holds;
- Real variables t_{ik} represent the departure time of a vehicle $i \in I$ from the factory in the tour $k \in K$;
- Binary variables y_j are set to 0 if location $j \in J$ is emptied, and 1 otherwise. The role of variables y_j is to keep the track on the total amount of goods delivered to the factory from location $j \in J$. If $y_j = 1$, the corresponding amount is obtained by multiplying the vehicle capacity C with the number of tours visiting location j (by all vehicles). Otherwise, in case that $y_j = 0$, whole amount of goods c_j collected at $j \in J$ is transported to the factory;
- Real variable T stands for the objective function value, i.e., the very last moment of time when all vehicles finish their last tours.

Using the above notation and decision variables, the considered VSP can be formulated as a Mixed Integer Quadratically Constrained Program (MIQCP) as follows:

$$\min \quad T \tag{1}$$

subject to

$$\sum_{j \in J} x_{ik}^{j} \le 1, \qquad \forall i \in I,\ \forall k \in K \tag{2}$$

$$\sum_{i \in I} \sum_{k \in K} C x_{ik}^{j} - C + \varepsilon \le c_j, \qquad \forall j \in J \tag{3}$$

$$T_j \le 1 - y_j, \qquad \forall j \in J \tag{4}$$

$$\sum_{j \in J} (1 - y_j) c_j + \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} C y_j x_{ik}^{j} \ge D, \tag{5}$$

$$\sum_{i \in I} \sum_{k \in K} x_{ik}^{j} \ge (c_j / C) \cdot (1 - y_j), \qquad \forall j \in J \tag{6}$$

$$\sum_{i \in I} \sum_{k \in K} y_j x_{ik}^{j} + \varepsilon \le c_j / C, \qquad \forall j \in J \tag{7}$$

$$t_{ik} + \sum_{j \in J} S_j x_{ik}^{j} \le t_{i,k+1}, \qquad \forall i \in I,\ \forall k \in K \setminus \{k_{max}\} \tag{8}$$

$$t_{i,k_{max}} + \sum_{j \in J} S_j x_{ik_{max}}^{j} \le T, \qquad \forall i \in I \tag{9}$$

$$x_{ik}^{j} \in \{0,1\}, \qquad \forall i \in I,\ \forall j \in J,\ \forall k \in K \tag{10}$$

$$y_j \in \{0,1\}, \qquad \forall j \in J \tag{11}$$

$$t_{start} \le t_{ik} \le t_{end}, \qquad \forall i \in I,\ \forall k \in K \tag{12}$$

The objective function (1) together with constraints (9) minimizes the moment of time when all vehicles finish their last tours. Each vehicle in each tour serves only one location or make a virtual tour, which is ensured by constraints (2). Constraints (3) provide that the total amount of goods transported from a location $j \in J$ is not greater than the total amount c_j collected at the same location. As the last vehicle visiting a location $j \in J$ may not be full with goods, it is necessary to subtract C on the left hand side of constraints (3). The role of small positive constant ε that is added to the left hand side of constraints (3) is to provide a strict inequality between the quantities of goods transported from a location $j \in J$ diminished by vehicle capacity C and the quantities c_j. A strict inequality prevents a vehicle to make additional unnecessary tour to location $j \in J$ in the case when c_j/C is an integer.

All goods collected at an urgent location $j \in J$ must be transported during the working day, which is ensured by constraints (4). If $t_j \le t_0$ holds, the value of T_j is set to 0, and the constraint (4) is satisfied regardless of the transported amount of goods. Otherwise, if $t_j > t_0$, T_j is equal to 1, meaning that location $j \in J$ is urgent and it must be emptied during the working day, which further implies that variable y_j takes the value of 0. Constraint (5) provides that the total amount of goods transported to the factory is at least D. The transported amount is obtained by summing the amount of goods delivered to the factory from emptied locations (the first sum on the left hand side of (5)) and the quantities of goods transported from locations that are still not emptied (the second sum on the left hand side of (5)).

If a location $j \in J$ is emptied, variable y_j is equal to 0 and the constraints (6) provide that location $j \in J$ is visited at least c_j/C times, which is ensured by constraints (7). In the case that location $j \in J$ is not emptied, variable y_j is set to 1, and in this case, constraints (6) are obviously satisfied. The role of constraints (8) is to provide that a vehicle can not start a new tour before finishing the previous one. More precisely, the starting time of each tour must not be less than the starting time of the previous one, increased by the serving time for a location visited in the previous tour. If a tour $k \in K \setminus \{k_{max}\}$ of a vehicle $i \in I$ is virtual, the sum on the left hand side of constraints (8) is equal to zero. The upper limit on the finishing time for each vehicle is set by the constraints (9). Finally, constraints (10)-(12) indicate the type of decision variables x_{ik}^j, y_j, and t_{ik}, respectively.

In order to reformulate the proposed MIQCP model (1)-(12) to a linear program, it is necessary to transform the product of binary variables y_j and x_{ik}^j that occurs in constraints (5) and (7). We introduce binary variables r_{ik}^j, $i \in I, j \in J, k \in K$ to replace the product $y_j x_{ik}^j$. These variables satisfy the following constraints:

$$r_{ik}^j - (y_j + x_{ik}^j)/2 \le 0 \qquad \forall i \in I,\ \forall j \in J,\ \forall k \in K, \tag{13}$$

$$r_{ik}^j - y_j - x_{ik}^j + 1 \ge 0 \qquad \forall i \in I,\ \forall j \in J,\ \forall k \in K, \tag{14}$$

$$r_{ik}^j \in \{0,1\} \qquad \forall i \in I,\ \forall j \in J,\ \forall k \in K. \tag{15}$$

Therefore, constraint (5) is replaced by constraint (5'), while constraints (7) are replaced by constraints (7'):

$$\sum_{j \in J} (1-y_j)c_j + \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} C r_{ik}^j \ge C, \tag{5'}$$

$$\sum_{i \in I} \sum_{k \in K} r_{ik}^j + \varepsilon \le c_j/C \qquad \forall j \in J. \tag{7'}$$

Finally, MIQCP model (1)–(12) is transformed to the MILP model (1)-(4), (5'), (6), (7'), (8)-(15).

3. General Variable Neighborhood Search for the VSP in sugar beet transportation

Variable Neighborhood Search (VNS) is a well-known metaheuristic method based on systematic change of neighborhoods within a possibly randomised local search algorithm. Since the early work on VNS by Mladenović and Hansen (1997), different variants of this metaheuristic have been proposed in the literature: Basic VNS (BVNS), Reduced VNS (RVNS), Variable Neighborhood Descent (VND), General VNS (GVNS), Skewed VNS (SVNS), Variable Neighborhood Decomposition Search (VNDS), Primal-dual VNS, Parallel VNS, etc. An overview of VNS based methods and applications can be found in Hansen et al. (2010) or Hansen and Mladenović (2014). In this section, we describe the GVNS implementation proposed as a solution approach to the considered VSP.

Solution representation and objective function calculation. A solution of the problem is represented as an integer matrix S that consists of m rows of length k_{max}. The i-th row of S corresponds to vehicle $i \in I$, and contains integers $s_{ik} \in \{0,1,\dots,n\}$, where s_{ik} denotes the index of a location visited by vehicle i in tour $k \in K$. If s_{ik} is 0, it means that the k-th tour of vehicle i is virtual. For simplicity, virtual tours (if they exist) are always placed at the end of rows in the solution matrix. Objective function of a solution is calculated as follows. For each vehicle $i \in I$, values t_{ik} representing departure times in each tour $k \in K$ are calculated based on the values s_{ik} in the solution matrix S. As all vehicles start their first tour at the same time (i.e., at the beginning of the working day), t_{i1} is set to starting time t_{start}. The departure time t_{ik} of a vehicle $i \in I$ in a tour $k \in \{2,\dots,k_{max}\}$, is equal to $t_{i,k-1}$ increased by the time needed to serve the location visited in the tour $k-1$. The finishing time for each vehicle is obtained as the departure time in its last non-virtual tour increased by the time required to serve visited location. Finally, the objective function value $T = f(S)$ is calculated as the maximum of finishing times among all vehicles, i.e., $T = \max_{i \in I} t_f(i)$, where $t_f(i) = t_{ik_{max}} + \sum_{j \in J} S_j x_{ik_{max}}^j$. Note that all t_{ik}, for k being virtual tour, are mutually equal and represent the finishing time of the last non-virtual tour of vehicle i. In the case that $T > t_{end}$, the corresponding solution is infeasible and it is discarded.

Generating initial solution. Initial solution for GVNS is generated in a greedy way. First, urgent locations are identified, as these locations have high priority and they must be served during the working day. If the quantity of goods collected at urgent locations can not be transported by the given set of vehicles within the working day, no feasible solution exists for the given data. Otherwise, the procedure continues with sorting the locations to be served according to these two criteria: urgency and distance from the factory. The sorted list of locations is structured as follows: at the beginning of the list are urgent locations sorted in non-decreasing order in respect to their distances from the factory, followed by non-urgent locations that are also sorted in non-decreasing order according to their distances from the factory. Initially, the elements of initial solution's matrix S are set to 0. One by one location from the sorted list is taken to fill in the first column of S (corresponding to the first tours of vehicles), then the second column (corresponding to the second tours of vehicles), etc., until all urgent locations are emptied and the factory needs are satisfied. Note that a vehicle may start its last tour to a non-empty location only if this tour can be finished until the end of working day, otherwise, this location is visited by the next vehicle. After the solution matrix S is filled in, one by one row of S is considered and locations in each row are sorted in non-increasing order in respect to their distances from the factory. In this way, it is ensured that a vehicle will first make longer tours and then the shorter ones. As described above, the values of corresponding departure times t_{ik}, $i \in I, k \in K$ are calculated based on the values s_{ik} in the sorted matrix of initial solution.

Neighborhood structures. The proposed GVNS uses four types of neighborhoods, denoted as N^I, N^{II}, N^{III}, and N^{IV}. N^I-neighbor of solution S is obtained by exchanging a tour to a non-urgent location with a tour to a non-emptied location. Neighborhood N^{II} consists of all solutions S' obtained by replacing virtual tour of a vehicle with a tour to a non-emptied location. N^{III}-neighbor S' of solution S is obtained by exchanging a tour to a non-urgent location with virtual tour in the list of tours of vehicle that has the longest working time. Finally, neighborhood N^{IV} is obtained when a pair of vehicles in S exchanges non-virtual tours. Note that moves in neighborhoods N^{II} and N^{IV} preserve the feasibility of solutions. On the other hand, the transformation that defines neighborhoods N^I and N^{III} may violate feasibility, as the factory needs in the newly obtained solution may not be satisfied. However, the next move within the same neighborhood may produce a feasible solution. It can be proved that the above described neighborhoods are correctly defined for the considered VSP.

Algorithm 1 The proposed GVNS for VSP

```
procedure GVNS(Problem Data, r_max, t_max)
    Generate initial solution S;
    repeat
        λ ← 1; r ← 1;
        while (λ ≤ 2) do
            if r ≤ r_max then
                S' ← ShakeN^I(S, r):                //Shaking in N^I
            else
                S' ← ShakeN^II(S):                  //Shaking in N^II
            if S' is feasible then
                S'' ← VND(S'):                      //Local Search-VND
                if f(S'') < f(S) then               //Move or Not step
                    S ← S''; r ← 1; λ ← 1;
                else   r ← r + 1;
                if r > r_max then λ ← λ + 1;
    until SessionTime ≥ t_max
```

GVNS implementation. The structure of the proposed GVNS is presented by Algorithm 1. GVNS starts with generating initial solution S in a greedy way. Each GVNS iteration starts by setting λ and r to 1 and continues with performing the three main steps *Shaking*, *Local Search*, and *Move or Not* within the neighborhood change loop while $\lambda \leq 2$. GVNS iterations are repeated until time limit t_{max} is reached. Shaking step is performed

Algorithm 2 VND

```
procedure VND(Problem Data, S')
    S'' ← S';
    while (Improvement) do
        S' ← Sort_1(S');
        λ ← 1;
        while (λ ≤ 2) do
            if λ = 1 then
                Find the best neighbor S'' ∈ N^III(S');     //Local Search in N^III
            else
                Find the best neighbor S'' ∈ N^IV(S');      //Local Search in N^IV
            if f(S'') < f(S') then                          //Move or Not step
                S' ← S''; λ ← 1;
            else   λ ← λ + 1;
    return (S');
```

within neighborhoods N^I and N^{II}. Neighborhood N^I of order r, $r = 1, \dots, r_{max}$ used in this step is obtained by repeating r times random move that defines this neighborhood. Note that only non-urgent locations in the vehicle's list of tours are allowed to be replaced by non-emptied locations, and therefore, all urgent locations will remain served and the delivered quantities from each location will not exceed the prepared amounts. As this move can change the amount of goods delivered to the factory, it may affect the feasibility of the solution. For this reason, only feasible solution S' produced by Shaking step in neighborhood N^I is further passed to Local Search, otherwise, this step is repeated for the same value of r. When the value of r exceeds r_{max}, Shaking step continues in neighborhood N^{II} by performing an exchange of virtual tour with a tour to the non-emptied location, resulting in a feasible solution. Neighborhood change in Shaking step is regulated by parameter λ.

Instead of standard Local Search, the proposed GVNS uses VND (presented in Algorithm 2) that explores neighborhoods N^{III} and N^{IV}. First, solution S' is transformed as follows: vehicles are sorted in non-increasing order according to their finishing times, and for each vehicle, the locations are sorted in non-increasing order in respect to their distances from the factory. Note that a permutation of tours for a vehicle in a solution does not affect the objective function value, as finishing time of vehicle remains unchanged. Neighborhood N^{III} of the solution S' is first explored by performing a procedure that tries to replace one by one tour to the non-urgent location of the first vehicle (the one with a longest working time) by a virtual tour. The obtained solution is evaluated only if daily factory needs are satisfied.

Neighborhood structure N^{IV} is explored as follows. The search starts from the first ($i_1 = 1$) and the last ($i_2 = m$) vehicle in the sorted list, and tries to exchange the locations from the first tour ($k_1 = 1$) of vehicle i_1 and the last tour ($k_2 = k_{max}$) of vehicle i_2 (i.e., the longest and the shortest tour of vehicles i_1 and i_2, respectively). If this move leads to a decrease of the maximal finishing time among vehicles i_1 and i_2, the exchange of their tours k_1 and k_2 is performed. In the newly obtained solution, the tours of the two considered vehicles i_1 and i_2 are sorted in non-increasing order according to the distances of the visited locations from the factory. Finally, the list of vehicles is arranged in non-increasing order according to their finishing times. If the exchange of tours k_1 and k_2 produces no improvement of the maximal finishing time among vehicles i_1 and i_2, the Local Search continues through the sorted list of tours in vehicle i_2 and tries to exchange, one by one, its tour k_2, $k_2 < k_{max}$ with the tour k_1 of vehicle i_1. These attempts are performed until an improvement is found or the first tour of vehicle i_2 is reached without an improvement. In the latter case, the search goes forward through the sorted tours of vehicle i_1 trying to exchange, one by one, tour k_1, $k_1 \leq k_{max}$ with the tour k_2 of vehicle i_2 until an improvement is reached or k_1 reaches k_{max} without an improvement. If no improvement is found by exchanging tours of vehicles i_1 and i_2, the described steps are repeated with vehicle i_1 and vehicles $i_2 - 1, i_2 - 2, \dots, i_1 + 1$.

If the best found solution S'' in the Local Search step is better than S', solution S' is replaced with S'' and the search continues within the neighborhood N^{III}, otherwise λ is set to $\lambda + 1$ (Move or Not step). The termination criterion for our VND is completed local search in both neighborhoods without improvement. Note that the size of neighborhoods N^{III} and N^{IV} used in VND step is $O(mk_{max})$ and $O(\binom{m}{2}k_{max}^2)$, respectively. The sorting of tours for each vehicle according to distances from the factory requires $O(mk_{max}^2)$ operations, while sorting the vehicles in respect to their finishing times is performed in $O(m^2)$ steps. Therefore, the overall worst-case complexity of our VND is $O(\binom{m}{2}k_{max}^2 + mk_{max} + mk_{max}^2 + m^2)$.

4. Experimental analysis

All computational experiments presented in this section were carried out on an Intel Core i7-2600 processor on 3.40GHz with 12GB RAM memory under Linux operating system. The experiments were performed on real-life and generated data set. Real-life instances are obtained from the considered sugar factory in Serbia. This data set includes 40 problem instances with up to 15 locations, 40 vehicles and the maximum number of 20 tours during the working day. Generated problem instances involve up to 1000 locations, 400 vehicles, and the maximum number of 400 tours during the working day. This data set includes 28 test examples, which are generated following the structure of real-life instances.

Commercial CPLEX 12.6.2 MIP solver is used to solve MILP model described in section 2. on small and medium-size instances to optimality (if possible). The time limit imposed on CPLEX run is set to 5 hours. The proposed GVNS algorithm is evaluated on all real-life and generated instances. On each instance, GVNS is run 30 times. The values of GVNS stopping criterion parameter t_{max} is set to $t_{max} = 1$ second for small-size real-life instances, $t_{max} = 10$ seconds for medium-size real-life instances, and $t_{max} = 100$ seconds for generated instances. Regarding the parameter r_{max}, eight formulas expressing r_{max} as a function of k_{max} are evaluated through preliminary computational experiments on the subset of real-life and generated test instances. Based on the obtained results, the value of r_{max} is set to $r_{max} = k_{max}/2 + 3$.

Computational results obtained on all considered instances are presented in Table 1. The left part of Table 1

Table 1: Computational results on small-size real-life instances solved to optimality by CPLEX 12.6.2

Instance	CPLEX 12.6.2		GVNS			Instance	CPLEX 12.6.2	GVNS		
$T_{n,m,k_{max}}$	*opt. sol.*	*t(s)*	*best*	*t(s)*	*gap(%)*	$T_{n,m,k_{max}}$	*best*	*best*	*t(s)*	*gap(%)*
$T_{3,2,4}$	**16.874**	0.09	**opt**	0.000	0.000	$T_{5,5,10}$	**17.781**	**17.781**	0.000	0.000
$T_{3,3,3}$	**13.727**	0.13	**opt**	0.000	0.000	$T_{5,10,10}$	**17.014**	**17.014**	0.079	0.005
$T_{3,3,4}$	**16.303**	0.20	**opt**	0.000	0.000	$T_{5,20,20}$	24.648	**24.591**	0.001	0.000
$T_{3,3,5}$	**17.164**	0.09	**opt**	0.000	0.000	$T_{8,40,10}$	26.291	**26.201**	2.075	0.065
$T_{3,4,2}$	**10.580**	0.04	**opt**	0.000	0.000	$T_{10,10,10}$	**16.243**	**16.243**	0.007	0.000
$T_{3,4,3}$	**13.727**	0.17	**opt**	0.000	0.000	$T_{15,20,15}$	26.214	**25.990**	0.255	0.061
$T_{3,4,4}$	**16.303**	0.08	**opt**	0.000	0.000	$T^r_{10,20,20}$	58.600	**58.486**	5.144	0.062
$T_{3,5,2}$	**12.294**	0.17	**opt**	0.000	0.000	$T^r_{10,30,15}$	/	**49.093**	0.489	0.000
$T_{4,2,4}$	**13.446**	0.05	**opt**	0.000	0.000	$T^r_{10,50,10}$	36.896	**36.667**	5.785	0.002
$T_{4,3,2}$	**10.580**	0.04	**opt**	0.000	0.000	$T^r_{15,30,25}$	/	**63.580**	7.020	0.001
$T_{4,4,2}$	**11.151**	0.07	**opt**	0.000	0.000	$T^r_{15,40,20}$	/	**53.453**	0.256	0.003
$T_{4,4,3}$	**14.299**	0.12	**opt**	0.000	0.000	$T^r_{20,30,10}$	/	**22.214**	0.007	0.000
$T_{4,4,4}$	**16.874**	0.57	**opt**	0.000	0.000	$T^r_{20,40,20}$	/	**41.163**	0.129	0.000
$T_{4,5,7}$	**15.287**	307.93	**opt**	0.172	0.037	$T^r_{25,50,15}$	/	**36.060**	1.508	0.001
$T_{5,2,4}$	**13.156**	0.06	**opt**	0.000	0.000	$T^r_{30,30,25}$	/	**46.674**	8.405	0.008
$T_{5,3,2}$	**10.580**	0.03	**opt**	0.000	0.000	$T^r_{30,60,15}$	/	**29.493**	1.840	0.006
$T_{5,3,3}$	**13.156**	0.44	**opt**	0.000	0.000	$T^r_{40,60,55}$	/	**103.680**	19.199	0.035
$T_{5,3,4}$	**16.589**	0.36	**opt**	0.000	0.000	$T^r_{45,65,70}$	/	**109.407**	16.911	0.047
$T_{5,4,2}$	**10.866**	0.10	**opt**	0.000	0.000	$T^r_{50,70,70}$	/	**116.669**	7.056	0.044
$T_{5,4,4}$	**17.446**	0.65	**opt**	0.000	0.000	$T^r_{55,65,65}$	/	**153.287**	12.581	0.012
$T_{5,4,5}$	**18.879**	0.21	**opt**	0.000	0.000	$T^r_{60,80,80}$	/	**126.453**	12.420	0.004
$T_{5,5,2}$	**11.437**	0.22	**opt**	0.000	0.000	$T^r_{70,60,60}$	/	**171.906**	28.357	0.017
$T_{5,5,5}$	**18.879**	40.74	**opt**	0.000	0.000	$T^r_{80,65,65}$	/	**181.413**	28.391	0.025
$T_{6,2,4}$	**13.160**	0.08	**opt**	0.000	0.000	$T^r_{90,80,80}$	/	**158.011**	31.732	0.016
$T_{6,3,2}$	**10.294**	0.14	**opt**	0.000	0.000	$T^r_{100,85,85}$	/	**184.046**	34.671	0.009
$T_{6,3,3}$	**12.870**	0.48	**opt**	0.000	0.000	$T^r_{120,90,90}$	/	**216.260**	38.895	0.007
$T_{6,4,2}$	**10.009**	0.13	**opt**	0.000	0.000	$T^r_{150,100,100}$	/	**240.303**	36.428	0.027
$T_{6,5,2}$	**10.866**	0.18	**opt**	0.000	0.000	$T^r_{300,120,120}$	/	**254.463**	56.403	0.031
$T_{6,6,6}$	**16.569**	258.00	**opt**	0.184	0.158	$T^r_{400,150,150}$	/	**373.000**	81.898	0.042
$T_{6,7,6}$	**11.479**	53.00	**opt**	0.000	0.000	$T^r_{500,200,200}$	/	**370.121**	95.446	0.034
$T_{7,3,3}$	**12.299**	0.36	**opt**	0.000	0.000	$T^r_{600,220,220}$	/	**428.366**	93.223	0.025
$T_{7,5,2}$	**10.580**	0.29	**opt**	0.000	0.000	$T^r_{800,300,300}$	/	**417.210**	83.698	0.009
$T_{7,5,6}$	**16.140**	24.43	**opt**	0.172	0.085	$T^r_{900,350,350}$	/	**650.036**	10.868	0.008
$T_{8,6,5}$	**14.246**	269.84	**opt**	0.049	0.000	$T^r_{1000,400,400}$	/	**948.847**	80.059	0.002
Average	**13.170**	28.22	**opt**	0.017	0.008	**Average**	/	**141.788**	23.566	0.018

contains experimental results on small-size real-life instances, while results on 6 medium-size real-life and all generated instances are presented on the right part of Table 1. Instance's name is presented in column $T_{n,m,k_{max}}$. Columns related to CPLEX contain the objective function value of the optimal solution *opt. sol.* and the corresponding runtime *t(s)*. In the case when CPLEX reached only feasible solution, the column $t(s)$ is omitted, as CPLEX run until time or memory limit was reached. The first column related to GVNS presents the objective function value of the best GVNS solution, with mark *opt* when it coincides with *opt. sol.* Remaining columns related to GVNS contain the average time *avg. t(s)* in which GVNS reaches *best* solution, the average percentage gap *avg. gap (%)* of the GVNS solution with respect to *opt. sol.* or *best*, all calculated through 30 runs. The last row of Table 1, denoted with **Average** contains average values of the presented results. The objective function values of the optimal or best-known solutions are bolded in columns *best*, as well as the best values in the last row **Average**.

As it can be seen Table 1, the proposed GVNS method reached all optimal solutions on small-size instances solved to optimality by CPLEX. On these instances, the average running time of GVNS was 0.017 seconds, which is significantly shorter compared to the average running time of CPLEX (28.22 seconds). In addition, GVNS was very stable, as it produced solutions with low average gap calculated in respect to known optimal solutions (0.008%). Table 1 also shows that CPLEX solver produced feasible solutions only for 6 real-life and two generated instances within the given time limit of 5 hours. These feasible solutions are presented in a column *best* related to CPLEX. On the other hand, the proposed GVNS reached all upper bounds provided by CPLEX and in case of five instances ($T_{5,20,20}$, $T_{8,40,10}$, $T_{15,20,15}$, $T^r_{10,20,20}$, and $T^r_{10,50,10}$), the corresponding upper bounds were improved. For medium-size real-life and generated problem instances, GVNS was also stable in providing the best-known solutions, as its average gap was very low (0.018%). The average running time that GVNS needed to obtain its best solutions on medium and generated instances was 23.566 seconds.

5. Conclusion

A variant of Vehicle Scheduling Problem that arises from optimizing the transport of sugar beet in Serbia is considered. The problem is formulated as Mixed Integer Quadratically Constraint Program (MIQCP) that is further linearized, and the obtained MILP formulation was used within the framework of CPLEX 12.6.2 MIP solver. As optimal solutions are obtained only for small-size real-life problem instances, a General Variable Neighborhood Search (GVNS) metaheuristic is designed to solve problem instances of larger dimensions, Computational results on small-size real-life data set show that GVNS quickly reaches all optimal solutions obtained by CPLEX solver. For medium-size real-life and generated test examples, GVNS either reaches or improves upper bounds provided by CPLEX within the given time limit. Based on the presented results, it may be concluded that proposed GVNS represents a promising solution approach to the considered VSP. As future work, the proposed VSP model can be extended by including non-homogenous vehicles or more than one factory. In addition, the designed GVNS may be combined with other optimization methods to improve the obtained best solutions of the considered VSP or to solve its extensions.

Acknowledgement

This research was partially supported by Serbian Ministry of Education, Science and Technological Development under the grants nos. 174010 and 174033.

REFERENCES

Bula, G., Prodhon, C., Gonzalez, F. A., Afsar, H., and Velasco, N. (2016). Variable neighborhood search to solve the vehicle routing problem for hazardous materials transportation. *Journal of Hazardous Materials*, 324(part B):472–480.

Bunte, S. and Kliewer, N. (2009). An overview on vehicle scheduling models. *Public Transport*, 1(4):299–317.

Cheikh, M., Ratli, M., Mkaouar, O., and Jarboui, B. (2015). A variable neighborhood search algorithm for the vehicle routing problem with multiple trips. *Electronic Notes in Discrete Mathematics*, 47:277–284.

Hansen, P. and Mladenović, N. (2014). Variable neighborhood search. In Burke, E. K. and Graham, R. D., editors, *Search methodologies: Introductory Tutorials in Optimization and Decision Support Techniques*, pages 313–337. Springer-Verlag, New York.

Hansen, P., Mladenović, N., and Pérez, J. A. (2010). Variable neighbourhood search: methods and applications. *Ann. Oper. Res.*, 175(1):367–407.

Higgins, A. (2006). Scheduling of road vehicles in sugarcane transport: A case study at an Australian sugar mill. *European Journal of Operational Research*, 170(3):987–1000.

Lahyani, R., Coelho, L., Khemakhem, M., Laporte, G., and Semet, F. (2015). A multi-compartment vehicle routing problem arising in the collection of olive oil in Tunisia. *Omega*, 51:1–10.

Liu, S. and Chen, A. (2012). Variable neighborhood search for the inventory routing and scheduling problem in a supply chain. *Expert Systems with Applications*, 39(4):4149–4159.

Macedo, R., Alves, C., Hanafi, S., Jarboui, B., Mladenović, N., Ramos, B., and de Carvalho, J. (2015). Skewed general variable neighborhood search for the location routing scheduling problem. *Computers & Operations Research*, 61:143–152.

Milan, E., Fernandez, S., and Aragones, L. (2006). Sugar cane transportation in Cuba, a case study. *European Journal of Operational Research*, 174(1):374–386.

Mladenović, N. and Hansen, P. (1997). Variable neighborhood search. *Computers & Operations Research*, 24(11):1097–1100.

Sethanan, K. and Pitakaso, R. (2016). Differential evolution algorithms for scheduling raw milk transportation. *Computers and Electronics in Agriculture*, 121:245–259.

Thuankaewsing, S., Khamjan, S., Piewthongngam, K., and Pathumnakul, S. (2015). Harvest scheduling algorithm to equalize supplier benefits: A case study from the Thai sugar cane industry. *Computers and Electronics in Agriculture*, 110:42–55.

Wassan, N., Wassan, N., Nagy, G., and Salhi, S. (2017). The Multiple Trip Vehicle Routing Problem with Backhauls: Formulation and a Two-Level Variable Neighbourhood Search. *Computers & Operations Research*, 78:454–467.

BALCOR 2018
XIII Balkan Conference on Operational Research

VNS-based Solution Approaches to the Maximal Covering Location Problem with Customer Preference Ordering

LAZAR MRKELA[1], ZORICA STANIMIROVIĆ[2]

[1] Belgrade Metropolitan University, lazar.mrkela@metropolitan.ac.rs
[2] University of Belgrade, Faculty of Mathematics, zoricast@matf.bg.ac.rs

Abstract: *This study considers Maximal Covering Location Problem with Customer Preference Ordering. The goal of the problem is to find optimal locations for establishing p facilities and optimal allocations of customers to facilities that are located within given coverage radius, such that the demand of customers is maximized. It is assumed that a set of facilities belonging to other firms are already present at the market and that customers are free to choose facilities based on their preferences. As the considered problem is NP-hard, metaheuristic approaches are required to solve problem instances of large dimensions. In this study, two meteheuristic metods are proposed: Variable Neighborhood Search (VNS) and a hybridization of Reduced Variable Neighborhood Search (RVNS), Tabu Search (TS) and VNS, denoted as RVNS-TS-VNS. The performance of VNS and RVNS-TS-VNS is evaluated on the set of 60 test instances and compared with the existing solution approach from the literature. The obtained computational results show the potential of both proposed VNS and RVNS-TS-VNS approaches when solving the considered variant of MCLP.*

Keywords: *Maximal Covering Location Problem, Customer Preference Ordering, Variable Neighborhood Search, Tabu Search*

1. INTRODUCTION

Maximal Covering Location Problem (MCLP) is one of the well-known facility location problems, which was introduced by Church and ReVelle (1974). MCLP considers the set of customers with demands assigned and the set of potential locations for establishing service providers (facilities). It is assumed that a customer can be served only by facilities located within a given distance (maximal covering radius). The goal of MCLP is to choose optimal locations for a fixed number of facilities such that the sum of covered population is maximized. MCLP is one of the most studied location problems in the literature, due to its numerous applications. Most of them are related to the design of emergency service networks (ambulances, police stations, fire stations, etc) and optimization of fast delivery systems (DHL, Fedex, food delivery, etc).

There are numerous studies in the literature dealing with MCLP and its variants. Various solution approaches have been proposed to MCLP, such as: greedy heuristic in Church and ReVelle (1974), Lagrangian relaxation in Galvão and ReVelle (1996), linear programming relaxation in Daskin (2011), Genetic Algorithm in Arakaki and Lorena (2001), etc. White and Case (1974) considered variant of the MCLP with equal customer demands and proposed a heuristic based on steepest-descent method. Planar case of the MCLP was studied by Church (1984), and it was showed that this MCLP variant can be solved over a small finite set of points located at the intersections of circles around locations of customers. Marianov and Serra (1998) imposed a lower limit on the quality of resources based on the number of covered nodes, resulting in Queueing Maximal Covering Location-Allocation Problem. A variant of MCLP with negative weights on customer locations was proposed in Berman et al. (2009), while Berman and Krass (2002) introduced Generalized Maximal Covering Location Problem (GMCLP). Colombo et al. (2016) considered different types of service providers and imposed a limit on the number of available service providers at each location. The authors also proposed a Variable Neighborhood Search heuristic for solving this variant of MCLP. Zarandi et al. (2013) involved multiple time periods in the classical MCLP model, leading to Dynamic Maximal Covering Location Problem (DMCLP). Variants of MCLP with different types of uncertainties were also considered in the literature. Rajagopalan et al. (2008) proposed a variant of MCLP for locating ambulance stations that involves time periods in which customer demands may vary significantly. A stochastic MCLP was introduced in ReVelle and Hogan (1989), where each facility location has a certain reliability assigned. Takači et al. (2012) considered the fuzzy maximal covering location problem and develop a particle swarm optimization as solution approach. MCLP with fuzzy travel times was studied in Davari et al. (2011) and Simulated Annealing was used to solve this variant of MCLP. A greedy variable neighborhood search heuristic for the MCLP with fuzzy coverage radii was proposed in Davari et al. (2013). A robust variant of the DMCLP with uncertain customer demands was introduced in Mišković (2017). A hybrid solution approach to deterministic and robust variants of the DMCLP was developed, based on VNS method and a linear programming technique.

In this study, we consider MCLP with Customer Preference Ordering, which was introduced by Díaz et al. (2017). This problem addresses the following situation on a market: a company intends to open exactly p facilities within a predefined set of potential sites with the aim to maximize the captured customers' demand. Customers can choose to be served by any of the p newly established facilities or by some of the existing facilities opened by competitors. A facility can only serve customers that are located within the given coverage radius. Customers choose a facility that will provide service for them, based on customer preferences and the coverage radius. Díaz et al. (2017) model this situation as a bilevel mathematical program, in which the leader seeks for locations of p facilities to maximize the captured demand, while the follower allocates each customer to its most preferred facilities among those selected by the leader and existing facilities belonging to competitors. The proposed bilevel model is further transformed to single-level formulation that replaces the follower's decision problem with a set of valid inequalities. Díaz et al. (2017) proposed a Greedy Randomized Adaptive Search Procedure (GRASP) heuristic and a hybrid GRASP-Tabu Search method (GRASP-TS) as solution approaches to MCLP with Customer Preference Ordering.

In this study, we propose two metaheuristic methods for the considered problem: Variable Neighborhood Search (VNS) and a hybridization of Reduced Variable Neighborhood Search (RVNS), Tabu Search (TS) and VNS, denoted as RVNS-TS-VNS. The hybrid RVNS-TS-VNS combines two variants of VNS metaheuristic and Tabu Search to efficiently provide solutions for large-scale instances of the considered MCLP with Customer Preference Ordering. To evaluate the performance of the proposed VNS and RVNS-TS-VNS methods, we used 60 randomly generated instances from Díaz et al. (2017). The results of the proposed two approaches on this data set are compared with the results of the GRASP-TS, which showed to be better than GRASP. It is shown that both VNS and RVNS-TS-VNS reach optimal or best-known solutions presented in Díaz et al. (2017). Regarding stability and CPU times, VNS and RVNS-TS-VNS are competitive with or better than GRASP-TS method.

2. MATHEMATICAL FORMULATION

Let $I = I_1 \cup I_2$ denote the set of facility locations at the market, where I_1 denotes the set of locations for establishing new facilities of the considered company, while I_2 stands for existing facility locations of other firms. The set of customers is $J = J_1 \cup J_2$, with J_1 representing the set of costumers that are not covered by existing facilities and J_2 denoting the set of customers covered by an existing facility. For each customer $j \in J$, a demand D_j is associated, as well as the set of facility locations $I(j)$ that can cover its demand. If facility location i belongs to $I(j)$, it means that the distance from customer j to location i does not exceed a predetermined coverage radius. For each facility location $i \in I$, the set of customers that can be covered by i is denoted as $J(i)$. The number of new facilities to be established from the set I_1 is fixed to p. Customer preferences are given by preference matrix $G = [g_{ij}]$, where g_{ij} represents the preference of customer $j \in J$ towards a facility at location $i \in I$. More precisely, $g_{i_1 j} > g_{i_2 j}$ implies that customer j prefers to be served by facility at location i_1 over facility at i_2.

Two sets of binary decision variables are used. Variables y_i take the value of 1 if facility is established at location $i \in I$, and 0 otherwise. Allocation variables x_{ij} are set to 1 if customer $j \in J$ is allocated to facility established at location $i \in I$, and 0 otherwise.

Using the above notation and decision variables, the considered variant of MCLP can be formulated as a Single-Level Integer Linear Program, as proposed in Díaz et al. (2017):

$$\max \quad \sum_{i \in I_1} \sum_{j \in J(i)} D_j x_{ij} \tag{1}$$

subject to

$$\sum_{i\in I_1} y_i = p \;, \tag{2}$$

$$\sum_{i\in I(j)} x_{ij} = 1 \quad \forall j \in J_2 \;, \tag{3}$$

$$\sum_{i\in I(j)} x_{ij} \leq 1 \quad \forall j \in J_1 \;, \tag{4}$$

$$x_{ij} \leq y_i \quad \forall i \in I, j \in J(i) \;, \tag{5}$$

$$y_i = 1 \quad \forall i \in I_2 \;, \tag{6}$$

$$\sum_{s=k+1}^{|I(j)|} x_{i_s,j} + y_{i_k} \leq 1 \quad \forall j \in J, k \in 1,...,|I(j)|-1 \;, \tag{7}$$

$$y_i \in \{0,1\} \quad \forall i \in I \;, \tag{8}$$

$$x_{ij} \in \{0,1\} \quad \forall i \in I, j \in J \;. \tag{9}$$

The objective function (1) maximizes total demand of customers allocated to the newly established locations. Constraint (2) ensures that exactly p new facilities from the set I_1 are opened. Customers covered by an existing facility must be allocated to exactly one facility (3), while customers not covered by existing facilities can be allocated to at most one new facility (4). By constraints (5), it is provided that customers are allocated only to locations with previously established facilities. Constraints (6) mean that existing facilities belonging to competitors remain opened. Each customer is allocated to its most preferred location with established facility, which is ensured by (7). Finally, constraints (8)–(9) indicate that variables y_i and x_{ij} are binary.

3. VNS-BASED SOLUTION APPROACHES TO THE MCLP WITH CUSTOMER ORDERING PREFERENCES

Variable neighborhood search (VNS) is a metaheuristic proposed by Mladenović and Hansen (1997), based on a systematic change of neighborhood during the search, applied to both descent phase and perturbation phase. The role of the descent phase is to find a local optimum, while perturbation phase tries to escape from the local optimum trap and enable the algorithm to reach global optimum. The basic variant of VNS iteratively applies perturbation and descent phase within a neighborhood change step until a stopping criterion is met. The perturbation phase, also denoted as *shaking*, is performed by random selection of one solution from the neighborhood of the current solution. The descent phase is implemented by *local search*, which moves from one solution to another within a neighborhood of the current solution, trying to find a local optimum. Usually, the steepest descent direction, known as best improvement, is used in local search phase. If the found local optimum is better than the current best solution, the current best solution is updated and the next iteration starts from the first neighborhood. Otherwise, the search continues in the next neighborhood from the given set of neighborhood structures. Besides basic VNS, numerous variants of variable neighborhood search have been proposed for solving various combinatorial and global optimization problems. An overview of VNS based methods and applications can be found in Hansen et al. (2010).

The choice of VNS approach is motivated by its successful applications to other variants of maximal covering problems, see Davari et al. (2013), Colombo et al. (2016), Mišković (2017), etc. In the literature, there are also examples of efficient VNS based approaches for different coverage problems, such as Schmid and Doerner (2010), Lust and Tuyttens (2014), Colombo et al. (2015), Janković et al. (2017), etc. In this study, we have designed basic VNS metaheuristic to solve the considered MCLP with Customer Preference Ordering. Adequate neighborhood structures are defined and the elements of the proposed VNS are designed in accordance to problem characteristics. In addition, a hybrid metaheuristic method, denoted as RVNS-TS-VNS, is developed by combining the proposed VNS with Reduced VNS and Tabu Search heuristic. Starting from a randomly generated feasible solution, the hybrid method iteratively applies RVNS, TS, and VNS, and returns best solution found through all iterations. The role of RVNS is to provide a good quality initial solution for TS, which returns an improved solution that is further used as input for the basic VNS. The details of the proposed VNS and RVNS-TS-VNS implementations are given in the following subsections.

3.1. VNS for the MCLP with customer ordering preferences

A solution is represented by a binary array of length $|I_1|$, where each bit corresponds to one candidate location. If the bit on the i-th position in the array has the value of 1, it means that facility is established at location $i \in I_1$. A solution is feasible if exactly p bits in the array are equal to 1, meaning that exactly p facilities are opened.

Objective function of a solution is calculated as follows. After locations of established facilities are obtained from the solution's code, customers are allocated to opened facilities in a greedy way, based on customer preferences. The objective function value is then calculated as the total demand of customers allocated to established facilities.

Neighborhood structures used in the proposed VNS implementation are defined by *swap* operation, i.e., closing one facility and opening another one. This operation is performed by inverting two bits with different values in the solution's code. Note that swap operation does not affect solution's feasibility, as the number of established facilities remains unchanged. Neighborhood $N_k(x)$ of solution x consists of solutions x' obtained from x by using swap operation k times, which is denoted as $k-swap$. Therefore, a solution x and its neighbour $x' \in N_k(x)$ differ in exactly $2k$ bits.

The proposed VNS starts with an initial feasible solution, which is randomly generated. Shaking step uses neighborhoods N_k of size $k = 1, \ldots, k_{max}$. In this step, a solution x' from the neighborhood $N_k(x)$ of the current solution x is chosen randomly. Local Search step tries to improve the current solution by exploring neighborhood $N_1(x')$ and returns the first found solution x'' that is better than x' (the first improvement strategy). The algorithm moves to the newly found solution x'' only if it is better than the current best one, which is controlled by Move or Not step. These three steps are repeated until the maximal number of iterations I_{max} without improving the best solution is reached (stopping criterion). The structure of the proposed VNS is shown in Algorithm 1.

Algorithm 1 The proposed VNS for MCLP with customer ordering preferences

```
Procedure: VNS(x)
Input: initial solution x
Output: best solution found x_best
x_best ← randomly generated solution
I_count ← 0
while I_count ≤ I_max do
    k ← 1
    improvement ← false
    while k ≤ k_max do
        x' ← randomly generated solution from neighborhood N_k(x)      ▷ Shaking step
        while true do
            x'' ← the first solution from neighborhood N_1(x') better than x'      ▷ Local Search step
            if x'' exists then
                x' ← x''
            else
                break
        if f(x') > f(x) then      ▷ Move or Not step
            x ← x'
            k ← 1
            improvement ← true
        else
            k ← k + 1
    if improvement then
        I_count ← 0
    else
        I_count ← I_count + 1
return x_best
```

During the VNS run, *swap* operation (opening one and closing another facility) is performed many times. Calculation of the objective function after every *swap* is time consuming. Therefore, updating objective function after *swap* operation can significantly reduce VNS running times.

When closing facility at location i, objective function is reduced in two cases:

- There are customers allocated to i, but no other location can cover them,
- There are customers allocated to i, covered by other facilities, but based on preferences, they are allocated to the competitive facilities.

When opening facility at location i, objective function is increased in two cases:

- There are customers not allocated to some facility, but they are covered by newly opened facility at location i,

- There are customers allocated to competitive facilities, but based on preferences, they are now allocated to the newly opened facility at location i.

Specific data structures are used to enable efficient objective function update. For each location i, the list of customers that can be covered by facility at i is stored. For each customer j, the list of candidate locations that are able to cover j is created and sorted based on preferences, starting with the location of the highest preference for customer j. In addition, for each customer j, the index of the currently allocated opened facility from the sorted list of facilities is stored.

When opening a facility at location i, only the list of customers that can be covered by i is processed. If a customer j from that list was covered by a location with lower preference, customer j is being re-allocated to newly opened facility at i. If customer j was not covered by location with established facility, or it was covered by competitive one, objective function is increased. When closing a facility at i, only customers that were allocated to this facility are observed. For each customer j from this set, the procedure goes through the sorted list of the candidate locations that are able to cover j. Customer j is assigned to the first location k with established facility from the list. If location k with established facility does not exists or location k is competitive, objective function is decreased.

3.2. RVNS-TS-VNS for the MCLP with customer ordering preferences

The proposed VNS method is further combined with Reduced Variable Neighborhood Search (RVNS) and Tabu search heuristic (TS). Reduced VNS is a variant of VNS, obtained by omitting local search phase, which is the most time-consuming part of VNS algorithm. Therefore, RVNS is suitable for providing a good-quality initial solution for other variants of VNS or other optimization methods. Pseudocode of the proposed RVNS for the considered variant of MCLP is given in Algorithm 2.

Algorithm 2 RVNS method

```
    Procedure: RVNS(x)
    Input: initial solution x
    Output: best solution found x_best
x_best ← randomly generated solution
I_count ← 0
while I_count ≤ I_max do
    k ← 1
    improvement ← false
    while k ≤ k_max do
        x' ← randomly generated solution from neighboorhood N_k(x)
        if f(x') > f(x) then
            x ← x'
            k ← 1
            improvement ← true
        else
            k ← k + 1
    if improvement then
        I_count ← 0
    else
        I_count ← I_count + 1
return x_best
```

The best solution obtained by RVNS phase is further improved by Tabu Search heuristic. Unlike simple local search, TS continues the search even when there is no better solution in the considered neighborhood. During the TS, neighborhoods of recently visited solutions can not be explored again. This restriction is based on the structure, denoted as *tabu list*, which contains forbidden solutions. In each iteration, the best solution from the neighborhood of the current solution is accepted only if it is not stored in tabu list. The size of tabu list is determined by parameter $size_{max}$. If the list is full, the first element is removed, and the new element is added at the end of the list. In many TS applications, instead of storing complete solutions in the tabu list, only attribute values are stored. Therefore, an element of tabu list corresponds to a set of solutions having the same attribute value. Our TS implementation uses this idea, which results in more efficient TS algorithm. Each time when *swap* operation is performed, the index of the location from which facility is removed is placed in the tabu list. During the TS run, while this index is in the tabu list, it is forbidden to open facility at this location. Stopping criteria for TS is the maximal number of iterations without improvement of the best solution. Pseudocode of the proposed TS heuristic is shown in Algorithm 3.

Proposed RVNS, TS, and VNS are combined in the hybrid RVNS-TS-VNS method as follows. Initial feasible solution for the hybrid method is randomly generated and set as the current best solution. In each

Algorithm 3 TS method

Procedure: TS(x)
Input: initial solution x
Output: best solution found x_{best}

```
x_best ← randomly generated solution
tabu_list ← empty list
I_count ← 0
while I_count ≤ I_max do
    improvement ← false
    x ← random solution from neighboorhood N_1(x)
    for all solution x' from neighborhood N_1(x) do
        if tabu_list does not contains index of newly established facility in x' and f(x') > f(x) then
            x ← x'
    if f(x) > f(x_best) then
        x_best ← x
        improvement ← true
    add index of the last closed facility in x to the end of tabu_list
    if size of tabu_list > size_max then
        remove first element from tabu_list
    if improvement then
        I_count ← 0
    else
        I_count ← I_count + 1
return x_best
```

iteration of the proposed hybrid method, a randomly generated solution x is first subject to RVNS, resulting in possibly improved solution x'. Then, TS heuristic is applied to x' and the best solution found by TS phase is stored in x''. Finally, solution x'' is used as initial solution for the basic VNS heuristic. The main loop of the hybrid RVNS-TS-VNS method is repeated J_{max} times, and the best found solution is returned. The structure of the proposed RVNS-TS-VNS is given in Algorithm 4.

Algorithm 4 Hybrid RVNS-TS-VNS method for MCLP with customer ordering preferences

```
J_count ← 0
x_best ← generate a random solution
while J_count ≤ J_max do
    x ← generate a random solution
    x' ← RVNS(x)
    x'' ← TS(x')
    x''' ← VNS(x'')
    if f(x''') > f(x_best) then x_best ← x'''
    J_count ← J_count + 1
return x_best
```

4. EXPERIMENTAL RESULTS

Proposed VNS and RVNS-TS-VNS methods are implemented in C++ programming language. All computational experiments were performed on a machine with AMD A6-3670 APU 2.7GHz and 8GB of RAM. The experiments were carried out on 60 random generated instances from Díaz et al. (2017), which are divided in 6 groups of 10 instances having the same size. Instance size is determined by number of customers and the number of potential facility locations: 225×25, 450×50, 675×75, 900×100, 1350×150, and 1800×200. Each set of instance has different value of parameter p and uses different coverage radius. For more detailed information on test instances, we refer to Díaz et al. (2017).

Parameter values in the proposed VNS and RVNS-TS-VNS methods were obtained through the set of preliminary experiments. In VNS implementation, parameter k_{max} used in VNS is set to $min(p, 10)$, while I_{max} takes the value of 30. When used within RVNS-TS-VNS approach, the VNS parameter I_{max} is set to 1. The value of parameter J_{max} in RVNS-TS-VNS is equal to 5, while parameters $size_{max}$ and I_{max} used in TS are set to 10 and 70, respectively. Parameters k_{max} and I_{max} in RVNS take the value of $p/3+1$ and 100, respectively.

Table 1 gives the summary of results obtained by GRASP-TS from Díaz et al. (2017) and the proposed VNS and RVNS-TS-VNS on test instances of the same dimension. As metaheuristics are essentially stochastic approaches, each method was run 5 times on each considered instance. The first column of Table 1 shows the size of the instances, while next columns contain the results of the GRASP-TS, VNS, and RVNS-TS-VNS, respectively. For each metaheuristic method, the following data are given: the average percentage deviation of the best solutions obtained instance with respect to the optimal or best known solution (*best avg*), the average percentage deviation of the solution's mean with respect to the optimal or best known solution (*mean avg*), the average percentage deviation of the worst solutions, obtained with respect to the optimal or best known solution (*worst avg*), the average CPU time (in seconds) required by the considered method. Note that GRASP-TS from Díaz et al. (2017) and the proposed VNS and RVNS-TS-VNS are tested on different computing platforms, Intel Xeon E3-1220 3.10GHz and AMD A6-3670 APU 2.7GHz, respectively. In order to provide fair comparisons of GRASP-TS with VNS and RVNS-TS-VNS running times, we have normalized CPU times of GRASP-TS by using data from www.cpubenchmark.net. The GRASP-TS running times are normalized as follows:

$$CPU^* = CPU \times \frac{passmark(\text{Intel Xeon E3-1220 3.10GHz})}{passmark(\text{AMD A6-3670 APU 2.7GHz})}. \qquad (10)$$

For this reason, Table 1 involves additional column CPU* related to GRASP-TS, which contains the corresponding normalized running times.

Table 1: Computational results of the proposed VNS and RVNS-TS-VNS and comparisons with the results of GRASP-TS

Size	GRASP-TS					VNS				RVNS-TS-VNS			
	best avg %	*mean avg %*	*worst avg %*	CPU *sec*	CPU* *sec*	*best avg %*	*mean avg %*	*worst avg %*	CPU *sec*	*best avg %*	*mean avg %*	*worst avg %*	CPU *sec*
225x25	0.000	0.000	0.000	0.0	0.0	0.000	0.000	0.000	0.6	0.000	0.000	0.000	1.7
450x50	0.000	0.000	0.000	1.4	2.67	0.000	0.000	0.000	5.2	0.000	0.000	0.000	4.4
675x75	0.000	0.006	0.031	7.6	14.5	0.000	0.000	0.000	25.0	0.000	0.000	0.000	12.8
900x100	0.000	0.000	0.000	13.9	26.5	0.000	0.000	0.000	54.0	0.000	0.000	0.000	25.8
1350x150	0.000	0.000	0.000	58.8	112.3	0.000	0.002	0.003	149.0	0.000	0.000	0.000	72.8
1800x200	0.000	0.001	0.001	172.8	330.0	0.000	0.006	0.012	318.6	0.000	0.000	0.000	168.3

As it can be seen from Table 1, each metaheuristic method reach optimal or best-known solution for each instance in at least one of 5 runs. For the instances of size 675×75, VNS is more stable when compared to GRASP-TS, however, for instances of size 1350×150 and 1800×200, GRASP-TS has better stability than VNS. Computational results related to RVNS-TS-VNS show that this hybrid approach is more stable than GRASP-TS in reaching optimal or best-known solutions. For the instances of size 675×75, percentage deviations *mean avg* and *worst avg* of RVNS-TS-VNS solutions are equal to 0%, while the values *mean avg* and *worst avg* of GRASP-TS solutions are 0.006% and 0.031 %, respectively. Similar situation is for the instances of size 1800×200: *mean avg* and *worst avg* are equal to 0.001% for GRASP-TS and 0% for RVNS-TS-VNS. The proposed RVNS-TS-VNS approach has slightly longer CPU time than GRASP-TS for instances of size 225×25 and 450×50, but the advantages of RVNS-TS-VNS are obvious for larger problem dimensions. For instances 1350×150, RVNS-TS-VNS provides solutions around 1.5 times faster than GRASP-TS, while for the largest set 1800×200, RVNS-TS-VNS is approximately two times faster than GRASP-TS. In general, VNS method is slower compared to both RVNS-TS-VNS and GRASP-TS. Detailed computational results can be found at http://www.matf.bg.ac.rs/p/files/10-Detailed-comp-results-MCLP.pdf.

5. CONCLUSION

A variant of Maximal Covering Location Problem that includes Customer Preference Ordering is considered in this study. In order to solve problem instances of large dimensions, two metaheuristic methods are designed: Variable Neighborhood Search (VNS) and a hybrid RVNS-TS-VNS method that combines Reduced Variable Neighborhood Search (RVNS), Tabu Search (TS) and VNS. Basic variant of VNS is used in both proposed methods. The role of RVNS and TS within the RVNS-TS-VNS is to ensure high-quality initial solution for VNS. The proposed VNS and RVNS-TS-VNS methods are benchmarked on 60 test instances from the literature, and compared with the GRASP-TS, which is state-of-the-art method for the considered problem. Based on obtained computational results, it can be concluded that both VNS and RVNS-TS-VNS reach optimal or best-known solutions from the literature. In the sense of stability and running times, the proposed VNS-based methods are competitive with or better than GRASP-TS. The presented results indicate that proposed VNS and RVNS-TS-VNS methods are good choice as solution approaches for MCLP with Customer Preference Ordering. The future work will be directed to developing extensions of the considered problem and implementing appropriate modifications of the VNS and RVNS-TS-VNS.

Acknowledgement

This research was partially supported by Serbian Ministry of Education, Science and Technological Development under the grant no. 174010.

REFERENCES

Arakaki, R. G. I. and Lorena, L. A. N. (2001). A constructive genetic algorithm for the maximal covering location problem. In *Proceedings of metaheuristics international conference*, pages 13–17. Citeseer.

Berman, O., Drezner, Z., and Wesolowsky, G. O. (2009). The maximal covering problem with some negative weights. *Geographical analysis*, 41(1):30–42.

Berman, O. and Krass, D. (2002). The generalized maximal covering location problem. *Computers & Operations Research*, 29(6):563–581.

Church, R. and ReVelle, C. (1974). The maximal covering location problem. In *Papers of the Regional Science Association*, volume 32, pages 101–118. Springer.

Church, R. L. (1984). The planar maximal covering location problem. *Journal of Regional Science*, 24(2):185–201.

Colombo, F., Cordone, R., and Lulli, G. (2015). A variable neighborhood search algorithm for the multimode set covering problem. *Journal of Global Optimization*, 63(3):461–480.

Colombo, F., Cordone, R., and Lulli, G. (2016). The multimode covering location problem. *Computers & Operations Research*, 67:25–33.

Daskin, M. S. (2011). *Network and discrete location: models, algorithms, and applications*. John Wiley & Sons.

Davari, S., Zarandi, M. H. F., and Hemmati, A. (2011). Maximal covering location problem (mclp) with fuzzy travel times. *Expert Systems with Applications*, 38(12):14535–14541.

Davari, S., Zarandi, M. H. F., and Turksen, I. B. (2013). A greedy variable neighborhood search heuristic for the maximal covering location problem with fuzzy coverage radii. *Knowledge-Based Systems*, 41:68–76.

Díaz, J. A., Luna, D. E., Camacho-Vallejo, J.-F., and Casas-Ramírez, M.-S. (2017). GRASP and hybrid GRASP-tabu heuristics to solve a maximal covering location problem with customer preference ordering. *Expert Systems with Applications*, 82:67–76.

Galvão, R. D. and ReVelle, C. (1996). A lagrangean heuristic for the maximal covering location problem. *European Journal of Operational Research*, 88(1):114–123.

Hansen, P., Mladenović, N., and Pérez, J. A. (2010). Variable neighbourhood search: methods and applications. *Ann. Oper. Res.*, 175(1):367–407.

Janković, O., Mišković, S., Stanimirović, Z., and Todosijević, R. (2017). Novel formulations and vns-based heuristics for single and multiple allocation p-hub maximal covering problems. *Annals of Operations Research*, 259(1-2):191–216.

Lust, T. and Tuyttens, D. (2014). Variable and large neighborhood search to solve the multiobjective set covering problem. *Journal of Heuristics*, 20(2):165–188.

Marianov, V. and Serra, D. (1998). Probabilistic, maximal covering location—allocation models for congested systems. *Journal of Regional Science*, 38(3):401–424.

Mišković, S. (2017). A VNS-LP algorithm for the robust dynamic maximal covering location problem. *OR Spectrum*, 39(4):1011–1033.

Mladenović, N. and Hansen, P. (1997). Variable neighborhood search. *Computers & Operations Research*, 24(11):1097–1100.

Rajagopalan, H. K., Saydam, C., and Xiao, J. (2008). A multiperiod set covering location model for dynamic redeployment of ambulances. *Computers & Operations Research*, 35(3):814–826.

ReVelle, C. and Hogan, K. (1989). The maximum availability location problem. *Transportation Science*, 23(3):192–200.

Schmid, V. and Doerner, K. F. (2010). Ambulance location and relocation problems with time-dependent travel times. *European journal of operational research*, 207(3):1293–1303.

Takači, A., Marić, M., and Drakulić, D. (2012). The role of fuzzy sets in improving maximal covering location problem. In *Intelligent Systems and Informatics (SISY), 2012 IEEE 10th Jubilee International Symposium on*, pages 103–106. IEEE.

White, J. A. and Case, K. E. (1974). On covering problems and the central facilities location problem. *Geographical Analysis*, 6(3):281–294.

Zarandi, M. H. F., Davari, S., and Sisakht, S. A. H. (2013). The large-scale dynamic maximal covering location problem. *Mathematical and Computer Modelling*, 57(3-4):710–719.

B9
Logistics & Supply Chain Management

XIII Balkan Conference on Operational Research

CONTAINERS DRAYAGE PROBLEM WITH SIMULTANEOUS ROUTING OF VEHICLES AND HANDLING EQUIPMENT

DRAŽEN POPOVIĆ[1], NENAD BJELIĆ[2], MILORAD VIDOVIĆ[3]

[1] University of Belgrade, Faculty of Transport and Traffic Engineering, d.popovic@sf.bg.ac.rs
[2] University of Belgrade, Faculty of Transport and Traffic Engineering, n.bjelic@sf.bg.ac.rs
[3] University of Belgrade, Faculty of Transport and Traffic Engineering, m.vidovic@sf.bg.ac.rs

Abstract: *Containers drayage involves the delivery of a full container from an intermodal terminal to a receiver and the following collection of an empty container, as well as the provision of an empty container to the shipper and the subsequent transportation of a full trailer or container to the intermodal terminal. Most of the practical problems as well as published researches deal with the routing and scheduling of container vehicles only, where it is implicitly assumed that the customer nodes, both pickup and delivery (P/D), are equipped with appropriate container handling equipment able to load or unload container. However, in real world systems some customer nodes, usually smaller companies, may have container P/D requests, although they are not equipped with appropriate handling equipment. For such a customer nodes service provider may leave containers on trailers until they are loaded or unloaded, or customer may rent appropriate handling device to perform container loading or unloading operation. In the second case, to avoid waiting of P/D vehicles it is needed to synchronize moments when vehicle and rented handling device arrive at customer nodes. When the P/D vehicle in a single route have to visit few nodes which are not equipped with appropriate handling devices, arises the problem of simultaneous routing both, vehicles and handling devices so that their arrivals at customer nodes are synchronized. In this paper we address the problem and propose mixed integer linear model to determine optimal synchronized routes of vehicles and handling devices performing containers' P/D operations.*

Keywords: *containers drayage; syncronized routing; vehicle routing; MILP.*

1. INTRODUCTION

Intensive worldwide container transportation flows are the result of significant increase in global trade, as well as of the competitive prices and carbon footprint levels compared to other transportation alternatives. Usually, container transport is divided into three phases. In the first one, containers are transferred from a shipper to the originate terminal. In the second phase containers are transferred through the network of terminals to the destination terminal, while in the third phase containers are transferred from the destination terminal to the receiver. Because locations of shippers and receivers are scattered over a terminal gravity zone the first and the third phase are executed by road vehicles in the activity known as container drayage. Although distances traversed in the drayage operations are significantly shorter compared to inter-terminal distances, according to Morlok and Spasovic (1994) and Escudero et al. (2013) drayage costs take from 22 to 40% (depending on the drayage to inter-terminal distance ratio) of overall container transportation costs. Beside economical reasons for the drayage optimization, the fact that great majority of the drayage is realized by (semi-)trailer trucks sets a reduction of carbon foot print as another driver for researchers to improve efficiency of drayage operations. In that sense, taking into consideration as much as possible real life drayage features contributes to higher usefulness of provided models. More detailed insight in the container drayage problems (CDP) and modeling approaches in this area can be found in recent papers of Lai et al., (2013), Zhang et al. (2015), and Vidović et al. (2016). Beside the fact that the literature related to drayage problem is very extensive, lot of real world problems still remain out of the research focus.

Most of the practical problems as well as published researches deal with the routing and scheduling where it is implicitly assumed that customer nodes, both pickup and delivery (P/D), are equipped with appropriate container handling equipment able to load or unload container. However, in real world systems some customer nodes, usually smaller companies, may have container P/D requests, although they are not equipped with appropriate handling equipment. For such a customer nodes service provider may leave containers on trailers until they are loaded or unloaded, or customer may rent appropriate handling device to perform container loading or unloading operation. In the second case, to avoid waiting of P/D vehicles it is needed to synchronize moments when vehicle and rented handling device arrive at customer nodes. Since the new technology in vehicles construction enables transportation of at least two fully loaded 20ft containers, this increases the possibility that the vehicle in a single route have to visit few nodes which are not equipped with appropriate handling devices. In other words, this arises the problem of simultaneous routing of both, vehicles and handling devices so that their arrivals at customer nodes are synchronized. In classical vehicle routing problems vehicle routes are insensitive on changes of other routes. However, in VRP with synchronization that is not the case because routes are interdependable, and therefore problems are much harder to solve. According to Drexl (2012) vehicle routing problems with multiple synchronization constraints are emerging field in the vehicle routing and a “hot” topic for researchers. Different kinds of routes interdependability cause different types of synchronization. Beside the classification of synchronization on task, operation, movement, load and resource classes, Drexl (2012) gives detailed literature overview of papers dealing with given classes.

Therefore, this work is an extension of the modeling approach presented in Vidović et al. (2016) in a way that it includes the possibility that some end customers require handling equipment in order to manipulate the container from/to a semi-trailer chassis. In this paper we address the problem and propose mixed integer linear model to determine optimal synchronized routes of vehicles and handling devices performing containers’ P/D operations.

2. PROBLEM DESCRIPTION

The problem considered in this paper is realization of container drayage operations based on modular concept vehicles with additional synchronization of P/D vehicles and handling devices. The modular concept vehicles are accepted by European Union (EU) regulation as a vehicle offering the possibility of transporting two fully loaded 20-ft containers using a special combined chassis.

This vehicle concept generates numerous node matching possibilities which include a large number of combinations for 20ft and 40ft container pickup delivery tasks. On the other side, handling device routing must be synchronized with P/D vehicles for those nodes that does not have handling equipment. Our approach to solving this CDP with synchronizing constraints is based on the CDP formulation proposed by Vidović et al. (2016) including the assumption that some nodes does not have handling devices for unloading/loading of containers. We extended proposed matching MILP formulation by introducing set of linear constraints for routing and synchronization of handling devices, defining the CDP with Simultaneous Routing of Vehicles and Handling Equipment.

It is assumed that any node may simultaneously have both demand and supply move requests (outbound and inbound containers). The number of requests which correspond to 20ft containers demand (20-) and supply (20+), and 40ft containers demand (40-) and supply (40+) are known in advance for all nodes. Therefore, in our mathematical formulation we observe a graph where each request for inbound/outbound container represents one task node. This means that when the network node has a request to send one container to the terminal and to receive two containers from the terminal, it is replaced with three task nodes, each with exactly one request. P/D vehicles and handling devices start from the terminal and travel with known speed. For task nodes that does not have handling equipment, P/D vehicle and handling device must have synchronized time of arrival in their separate routes. Task nodes can have time windows. Overall objective of the observed problem is to minimize total travel distance of both P/D vehicles and handling devices.

3. MATHEMATICAL FORMULATION

To define routes of P/D vehicles we use matching formulation. Defining feasible matchings of task nodes (possible vehicle routes) can be realized in two steps. First, we must identify all possible task nodes matching patterns for modular vehicle that can carry up to two 20 ft or single 40 ft container (given in Table 1). Then, for each matching pattern we must generate variables for all feasible task node sequences (one sequence represents one route of P/D vehicle). For example, matching pattern "-20→-20→+20→+20" denotes delivery of two 20ft containers and pickup of two 20ft containers. The order of tasks' type is fixed in this pattern, but which task node exactly will be in what position of that pattern defines possible P/D vehicle route. For example, one possible matching variable from that pattern can be delivery of 20 ft container to the node 1, then delivery of 20 ft container to the node 4, followed by the pickup of 20ft container from the node 6, and then pickup of 20ft container from the node 2. Another possible matching variable from the same pattern can describe delivery of 20ft container to the node 4, delivery of 20ft container to the node 3, followed by the pickup of 20ft container from the node 2, and then pickup of 20ft container from the node 7, etc.

Table 1: Set SMAX of all possible task nodes matching patterns when the modular concept vehicle is used

Four nodes matchings	Three nodes matchings	Two nodes matchings	Direct pickup or delivery
- 20 → - 20 → +20 → +20	- 20 → - 20 → +20	- 40 → +40	+20
- 20 → +20 → - 20 → +20	- 20 → +20 → +20	- 20 → +40	- 20
	- 20 → +20 → -20	- 40 → +20	+40
	+20 → - 20 → +20	- 20 → +20	- 40
	- 20 → - 20 → +40	+20 → - 20	
	- 40 → +20 → +20	+20 → +20	
		- 20 → - 20	

For the routing of handling devices, we use classical mTSP routing formulation with time windows where multiple handling devices must visit each task node that does not have handling equipment. For both P/D vehicles and handling devices, time of arrival at the task node is represented by variable w. Synchronization is achieved by these variables, where arrival time at task node without handling equipment is constrained by the routes from both P/D vehicle and handling device.

To formulate the problem, we use following notation:

N	set of all task nodes ($N = N^{20-} \cup N^{20+} \cup N^{40-} \cup N^{40+}$)
H	set of all task nodes that have handling equipment (additional handling device is not needed for these tasks)
E	set of all task nodes without handling equipment (E=N/H)
S_{MAX}	set of all possible task nodes matching patterns when the modular concept vehicle is used
γ	index of task node $\gamma \in N$
i	matching pattern $i \in S_{MAX}$
k	feasible task nodes combination for some pattern *i* ($k \in K_i$)
A_{ik}	sequence (ordered set) of task nodes
$\delta^{\gamma}_{A_{ik}}$	binary coefficient which is equal to 1 if $\gamma \in A_{ik}$, and 0 otherwise (if task γ is in the route A_{ik})
$y_{A_{ik}}$	matching decision variable which is equal to 1 if the sequence of nodes $a_{1k}, a_{2k}, \ldots, a_{\lvert S_i \rvert k}$ is merged in the same route in MILP solution, and 0 otherwise
$l_{a_{jk}}, r_{a_{jk}}$	the left and the right bound of the time window [$l_{a_{jk}}, r_{a_{jk}}$] at the node a_{jk} indicating the time interval when the node is available for the servicing (where *j* represents index of task node γ in the route *k*)
$w_{a_{jk}}$	continuous decision variables indicating the time at which vehicle starts servicing task node a_{jk}
$h_{a_{jk}}$	service time at the task node a_{jk}
$t_{a_{jk}, a_{j+1,k}}$	travel time between task nodes a_{jk} and $a_{j+1,k}$
$M_{a_{jk}, a_{j+1,k}}$	constant used to linearize time windows constraint, $M_{a_{jk}, a_{j+1,k}} = \max\lfloor 0, r_{a_{jk}} + h_{a_{jk}} + t_{a_{jk}, a_{j+1,k}} - l_{a_{j+1,k}} \rfloor$

$c_{A_{ik}}$ — costs of visiting sequence of task nodes ($a_{1k}, a_{2k}, \ldots, a_{|S_i|k}$) in a single route, including costs from/to the terminal (0). For example, in the case of four task nodes sequence, costs are: $c_{a_{1k},a_{2k},a_{3k},a_{4k}} = c_{0,a_{1k}} + c_{a_{1k},a_{2k}} + c_{a_{2k},a_{3k}} + c_{a_{3k},a_{4k}} + c_{a_{4k},0}$

γ_1, γ_2 — task nodes $\gamma_1, \gamma_2 \in \mathrm{E}$ (corresponds to a task node a_{jk} in the matching part of mathematical formulation)

c_{γ_1,γ_2} — handling device travel cost from task node γ_1 to task node γ_2

p — index of handling device ($p \in \mathrm{P}$)

$x_{\gamma_1}^{p}$ — decision variable which is equal to 1 if the handling device *p* serves task γ_1, otherwise equals 0

x^{p} — decision variable which is equal to 1 if the handling device *p* is used for any task, otherwise equals 0

$z_{\gamma_1,\gamma_2}^{p}$ — decision variable which is equal to 1 if the handling device *p* serves task γ_2 after task γ_1 (handling device *p* is travelling on arc $\gamma_1 - \gamma_2$), otherwise equals 0

Based on the given notation, MILP formulation of Containers Drayage Problem with Simultaneous Routing of Vehicles and Handling Equipment is as follows:

$$\min \rightarrow \sum_{i=1}^{S_{MAX}} \sum_{k=1}^{K_i} y_{A_{ik}} \cdot c_{A_{ik}} + \sum_{\gamma_1 \in E} \sum_{\gamma_2 \in E} \sum_{p \in P} z_{\gamma_1,\gamma_2}^{p} \cdot c_{\gamma_1,\gamma_2} \tag{1}$$

$$\sum_{i=1}^{S_{MAX}} \sum_{k=1}^{K_i} \delta_{A_{ik}}^{\gamma} \cdot y_{A_{ik}} = 1 \quad \forall \gamma \in \mathrm{N} \tag{2}$$

$$w_{a_{j+1,k}} \geq w_{a_{jk}} + h_{a_{jk}} + t_{a_{jk},a_{j+1,k}} - M_{a_{jk},a_{j+1,k}} \cdot (1 - y_{A_{ik}}) \\ \forall a_{jk} \in A_{ik}, \ i = 1,\ldots,S_{MAX}, \ j = 1,\ldots,|S_i| - 1, \ k = 1,\ldots,K_i \tag{3}$$

$$l_{a_{jk}} \leq w_{a_{jk}} \leq r_{a_{jk}} \\ \forall a_{jk} \in A_{ik}, \ i = 1,\ldots,S_{MAX}, \ j = 1,\ldots,|S_i|, \ k = 1,\ldots,K_i \tag{4}$$

$$y_{A_{ik}} \in \{0,1\}, \ w_{a_{jk}} \geq 0 \tag{5}$$

$$x^{p} \geq \frac{\sum_{\gamma_1 \in E} x_{\gamma_1}^{p}}{|E|} \quad \forall p \in \mathrm{P} \tag{6}$$

$$\sum_{\gamma_1 \in E} z_{0,\gamma_1}^{p} = x^{p} \quad \forall p \in \mathrm{P} \tag{7}$$

$$\sum_{\gamma_1 \in E} z_{\gamma_1,\gamma_2}^{p} = x_{\gamma_1}^{p} \quad \forall \gamma_2 \in E / \gamma_1, \ p \in \mathrm{P} \tag{8}$$

$$\sum_{\gamma_2 \in E} z_{\gamma_1,\gamma_2}^{p} = x_{\gamma_2}^{p} \quad \forall \gamma_1 \in E / \gamma_2, \ p \in \mathrm{P} \tag{9}$$

$$\sum_{p \in P} x_{\gamma_1}^{p} = 1 \quad \forall \gamma_1 \in E \tag{10}$$

$$w_{\gamma_2} \geq w_{\gamma_1} + h_{\gamma_1} + t_{\gamma_1,\gamma_2} - \mathrm{M}_{\gamma_1,\gamma_2} \cdot \left(1 - z_{\gamma_1,\gamma_2}^{p}\right) \quad \forall \gamma_1 \in E, \ \gamma_2 \in E, \ \gamma_1 \neq \gamma_2, \ p \in \mathrm{P} \tag{11}$$

$$x^{p}, x_{\gamma_1}^{p}, z_{\gamma_1,\gamma_2}^{p} \in \{0,1\}, \quad w_{\gamma_1}, w_{\gamma_2} \geq 0 \qquad (12)$$

Objective function (1) tries to minimize total transportation costs that consists of two segments: (Term I) all routes that are used for serving all pickup/delivery nodes by solving the set of nodes matching problems; (Term II) all routes of handling devices that are servicing nodes without handling equipment. Sets of constraints (2) ensure that each node must be visited exactly once by one route (sequence of task nodes). Constraints (3) and (4) are time windows constraints which allow nodes to be visited only during the interval when the nodes are available for servicing. Constraints (3) do not consider time windows availability of the first node in the sequence because servicing a vehicle in this node may start immediately after the beginning of observed time interval. We assume that the terminal node (denoted by 0) is always open. Constraints (5) and (12) define domain of decision variables. Constraints (6) define the number of used handling devices. This number of used handling devices determines number of outbound arcs from handling devices by constraints (7). Constraints (8) and (9) define that one task node without handling equipment can have one inbound and one outbound arc from handling devices. Constrains (10) define that one task node without handling equipment must be serviced by single handling device. Constraints (11) are time windows constraints where notation for task nodes without handling equipment γ_1, γ_2 corresponds with notation of task nodes $a_{jk}, a_{j+1,k}$ (both of these notations represents task nodes, but in two different segment of objective function).

4. COMPUTATIONAL RESULTS

The proposed MILP formulation was tested on instances that are generated according to the Solomon VRPTW benchmark problems (http://web.cba.neu.edu/~msolomon/problems.htm). Solomon problems include six different sets of problem instances (in total 56 instances) different by nodes position, clustering and length of time windows. We modified these instances by randomly transforming pickup and delivery demands to reflect two container types (20 and 40ft). Additionally, we observed two general cases of instances: (I) without synchronization where each node has handling equipment; (II) with synchronization where some nodes don't have handling equipment (randomly generated, where each node has chance of 50% to have handling equipment). Input data for one instance with synchronization is presented in Table 2, and graphical solution of that instance (for both general cases) is given in Figure 1. Each instance has 20 nodes, and depending on the number of requests in each node, even higher number of task nodes (one node can have more than one request).

Results from all 56 modified Solomon instances for two general cases are given in Table 3, where d1, n1 and t1 respectively represents results for P/D vehicles: total travel distance, total number used, total travel time. Results for handling devices are denoted with d2, n2 and t2. The CPU represents solution time in seconds obtained from CPLEX 12.4 on the Intel(R) Core(TM) i3 CPU M380 2.53 GHz with 6 GB RAM.

Comparing results from two general cases we can see the increase of the main parameters in the solutions obtained for the case with synchronization: (1) travel distance of P/D vehicles is increased in average by 1.02%; (2) number of routes for P/D vehicles is increased by 2.06%; (3) and the most significant increase is in the travel time of P/D vehicles in average by 8.51%. As expected, for the case with synchronization, number of routes for handling devices is considerable lower than for the P/D vehicles (more than 4 times). This is due the fact of lesser number of task nodes without handling equipment and possibility of servicing more than 4 task nodes in a single route by handling device. Regarding computational time to obtain the solution, synchronization has significant impact on the increase of that time (in average from 0.14 sec to 235.44 sec). Additionally, for 3 instances CPLEX could not find optimal solution in the time limit of 1800 sec. Therefore, instances with significantly more nodes (e.g. 25 or more) cannot be solved to optimality in reasonable computational time with the used resources and formulation.

Table 2: Input data for instance R2_03 with vehicles synchronization (some nodes don't have container handling equipment)

Node	x-coor	y-coor	TW start	TW end	Service time	20-	20+	40-	40+	Handling equipment
0	3500	3500								
1	4100	4900	0	97400	1000	0	1	0	0	yes
2	3500	1700	0	97200	1000	1	0	0	0	no
3	5500	4500	0	96700	1000	0	0	0	1	yes
4	5500	2000	67800	80100	1000	0	0	0	1	yes
5	1500	3000	0	96900	1000	0	0	1	1	yes
6	2500	3000	41500	51400	1000	0	0	0	1	yes
7	2000	5000	0	96800	1000	1	0	0	0	yes
8	1000	4300	40400	48100	1000	0	0	0	1	yes
9	5500	6000	40000	49700	1000	0	1	0	0	no
10	3000	6000	57700	63200	1000	0	1	0	0	no
11	2000	6500	20600	32500	1000	0	0	1	0	yes
12	5000	3500	0	97500	1000	0	0	0	1	no
13	3000	2500	69000	82700	1000	1	1	0	0	yes
14	1500	1000	0	95700	1000	0	0	1	0	no
15	3000	500	17500	30000	1000	0	1	0	0	no
16	1000	2000	0	96000	1000	0	1	0	0	no
17	500	3000	73300	87000	1000	0	0	1	0	yes
18	2000	4000	0	97400	1000	0	1	0	0	no
19	1500	6000	0	95700	1000	0	0	1	0	no
20	4500	6500	0	95800	1000	1	0	0	0	yes

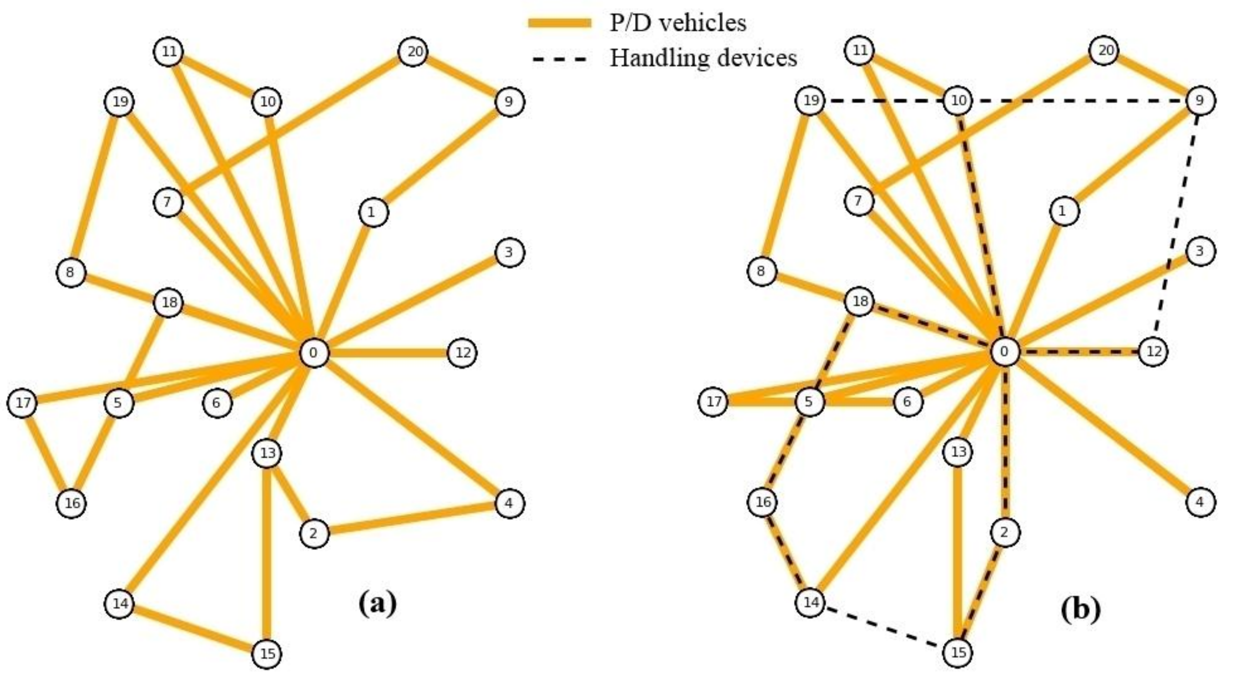


Figure. 1: Illustration of solution for instance R2_03: (a) without vehicles synchronization; (b) with vehicles synchronization

Table 3: Optimal solution results for 56 modified Solomon's instances

		WITHOUT SYNCHRONIZATION				WITH SYNCHRONIZATION						
Instance	*|N|*	***d1***	*n1*	*t1*	*CPU*	***d1***	*n1*	*t1*	*d2*	*n2*	*t2*	*CPU*
R1_01	22	**76534**	14	201844	0.01	**76534**	14	202664	171480	4	58445	0.1
R1_02	22	**65202**	9	155874	0.06	**65733**	11	169804	293885	5	87747	996.3
R1_03	22	**68352**	10	144980	0.07	**68352**	10	180987	188168	4	80785	1337.4
R1_04	22	**61856**	9	110944	0.08	**61856**	9	128472	72543	2	31351	2.1
R1_05	22	**66226**	12	171080	0.02	**68855**	13	194043	128971	3	42393	0.1
R1_06	22	**63785**	10	145210	0.08	**65383**	11	146719	122914	3	53408	4.0
R1_07	22	**62857**	11	140730	0.07	**62857**	11	152371	130556	3	53420	225.2
R1_08	22	**56388**	8	111246	0.29	**57992**	10	138298	66247	2	39584	25.1
R1_09	22	**62843**	11	143861	0.03	**68902**	12	156049	77509	2	40980	0.9
R1_10	22	**55469**	9	105021	0.05	**59653**	10	137903	122528	3	52444	28.5
R1_11	22	**64949**	11	142181	0.10	**66908**	11	155635	79213	2	36168	1.6
R1_12	22	**56157**	9	125220	0.44	**57528**	9	113934	71165	2	35002	79.2
R2_01	22	**72609**	12	603164	0.01	**74422**	13	684847	95130	2	151705	0.3
R2_02	22	**57879**	9	490039	0.13	**59333**	9	460849	62298	2	150085	0.5
R2_03	22	**60364**	10	433809	0.13	**61182**	10	457097	73724	2	89265	2.2
R2_04	22	**54801**	8	323693	0.53	**54835**	8	381178	76802	2	153726	271.9
R2_05	22	**67486**	10	444428	0.18	**68498**	11	507386	119593	3	207829	27.7
R2_06	22	**61274**	10	413837	0.17	**61274**	10	481804	169678	3	185811	*1800.0
R2_07	22	**56406**	9	449669	0.17	**58491**	10	522473	77472	2	141940	28.7
R2_08	22	**53085**	9	263652	0.35	**53680**	9	358467	135168	3	119114	577.9
R2_09	22	**69886**	12	436964	0.09	**69886**	12	515840	77551	2	122025	1.2
R2_10	22	**65388**	12	521020	0.05	**65483**	12	578684	68600	2	151550	6.6
R2_11	22	**59165**	8	326182	0.30	**59165**	8	410445	79498	2	117850	95.6
C1_01	24	**66882**	12	613101	0.04	**66882**	12	619201	204067	4	219142	0.1
C1_02	24	**63936**	11	528925	0.04	**63936**	11	574947	93908	2	175092	0.1
C1_03	24	**62172**	10	403310	0.27	**62189**	10	489528	183233	3	190899	759.3
C1_04	24	**70220**	14	494312	0.06	**70220**	14	729831	164996	3	203610	31.6
C1_05	24	**62632**	10	616903	0.03	**62632**	10	640058	102843	2	194810	0.2
C1_06	24	**63813**	11	634469	0.03	**63813**	11	635269	123297	3	232098	0.1
C1_07	24	**67780**	13	708796	0.03	**67780**	13	718882	153063	3	195658	1.1
C1_08	24	**63847**	11	587329	0.13	**63969**	11	624833	157798	3	267960	50.5
C1_09	24	**68265**	13	630019	0.06	**69086**	13	719188	147526	3	234118	*1800.0
C2_01	24	**80913**	13	2357528	0.03	**80913**	13	2357528	94471	2	570540	0.3
C2_02	24	**73607**	13	1791253	0.20	**73607**	13	1799001	63256	2	264132	0.6
C2_03	24	**65863**	10	1638355	0.21	**65863**	10	2015978	100898	2	570942	71.5
C2_04	24	**64876**	10	1861967	0.16	**65245**	10	2147599	127150	3	872388	871.7
C2_05	24	**78918**	13	2373780	0.04	**78918**	13	2382878	96836	2	383279	0.2
C2_06	24	**75574**	13	2462276	0.04	**75574**	13	2462276	123241	3	820206	0.9
C2_07	24	**83633**	14	2202146	0.03	**83662**	14	2375879	186186	3	731856	*1800.0
C2_08	24	**62037**	9	1942390	0.46	**62037**	9	1979390	98669	2	503356	35.4
RC1_01	25	**120879**	15	226474	0.02	**120879**	15	230804	279622	3	56673	0.4
RC1_02	25	**109351**	11	201537	0.05	**116285**	13	228065	173293	2	44901	36.6
RC1_03	25	**105842**	14	167702	0.07	**105842**	14	187436	240756	3	44115	5.3
RC1_04	25	**101100**	13	145842	0.19	**101100**	13	157173	263271	3	46092	21.6
RC1_05	25	**92807**	12	189360	0.04	**93832**	12	191385	347713	4	72342	8.2
RC1_06	25	**112227**	14	195114	0.07	**116806**	14	207033	204840	3	54375	0.8
RC1_07	25	**86374**	10	143802	0.28	**86374**	10	146440	67334	2	29477	0.7
RC1_08	25	**96241**	11	152093	0.34	**96241**	11	169800	268297	3	54024	1455.2
RC2_01	25	**104512**	13	557646	0.10	**105555**	13	608373	161685	2	141147	0.5
RC2_02	25	**84968**	11	507138	0.37	**84968**	11	552458	160890	2	143022	2.8
RC2_03	25	**82143**	10	382283	0.24	**82143**	10	463049	161473	2	161090	3.9
RC2_04	25	**86147**	11	281340	0.39	**86147**	11	363990	250178	3	104572	410.5
RC2_05	25	**95369**	12	593755	0.15	**95382**	12	613694	316714	3	183468	289.6
RC2_06	25	**106522**	13	565079	0.11	**106522**	13	605428	164935	2	115231	3.1
RC2_07	25	**111254**	15	506327	0.11	**111439**	15	592317	166304	2	131866	4.2
RC2_08	25	**107742**	14	416370	0.18	**107742**	14	508308	153836	2	114111	4.2

*- CPLEX time limit is set to 1800 seconds (suboptimal or optimal solution found)

5. CONCLUSIONS

By solving two general cases we wanted to examine the impact of synchronization on vehicle routing solutions, which is reflected in somewhat increased travel distance and number of used routes, and larger increase of travel time (depending on the instance set). Vehicles synchronization has significant impact on the increase of computational time for obtaining optimal solution due to the fact of increased complexity and size of the solution space.

Further research should include more elaborate testing of instances parameters impact on the solutions for Containers Drayage Problem with Simultaneous Routing of Vehicles and Handling Equipment: percentage of nodes without handling equipment, different travel speed for two vehicle types, number of nodes, etc. To be able to solve larger real-life size instances, the development of heuristics approach is another favorable research direction in this area.

Acknowledgement

This work was supported by the Ministry of Education, Science and Technical development of the Government of the Republic of Serbia through the project TR36006, for the period 2011-2018.

REFERENCES

Drexl, M., (2012. Synchronization in vehicle routing–A survey of VRPs with multiple synchronization constraints. Transportation Science 46 (3) 297–316

Escudero, A., Munuzuri, J., Guadix, J, Arango, C., (2013). Dynamic approach to solve the daily drayage problem with transit time uncertainty. *Computers in Industry* 64, 165–175.

Lai, M., Crainic, T.G., Di Francesco, M., Zuddas, P., (2013). An heuristic search for the routing of heterogeneous trucks with single and double container loads. *Transportation Research Part E: Logistics and Transportation Review* 56, 108–118.

Morlok, E.K., Spasovic, L.N., (1994). Redesigning rail-truck intermodal drayage operations for enhanced service and cost performance. *Journal of the Transportation Research Forum* 34, 16–31

Vidović, M., Popović, D., Ratković, B., Radivojević, G., (2016). Generalized mixed integer and VNS heuristic approach to solving the multisize containers drayage problem, *International Transactions in Operational Research* 24 (3), 583-614.

Zhang, R., Yun, W.Y., Kopfer, H., (2015). Multi-size container transportation by truck: modeling and optimization. *Flexible Services and Manufacturing Journal* 27, 2, 403–430.

INTERPRETATION OF STATIC TIME-CONTINUOUS INVENTORY MODEL WITH STOCK-LEVEL DEPENDENT DEMAND RATE AND VARIABLE HOLDING COST AS A DYNAMIC DISCRETE-TIME SYSTEM CONTROL PROCESS

SLOBODAN ANTIĆ[1], LENA ĐORĐEVIĆ[1], UROŠ JEREMIĆ[1]

[1] University of Belgrade, Faculty of Organizational Sciences, antic.slobodan@fon.bg.ac.rs, djordjevic.lena@fon.bg.ac.rs, jeremicuros7@gmail.com

Abstract: *This paper presents a static time-continuous inventory control problem with stock-level dependent demand rate and variable holding cost, published by Alfares (2007), which is modelled as a combinatorial optimization problem of the corresponding dynamic discrete-time system control process. The paper analyzes inventory systems with a continuously changing state, but changes of state are registered at the ends of the defined time periods (discrete-time processes). Mathematical apparatus, used for solving of mentioned inventory management problem, is optimal control of the discrete system. The discrete controlled object is represented with clearly separated: the law of dynamics, control domain and performance criterion. The main objective of this study is development of dynamic discrete spreadsheet simulation model for inventory control. The model is aimed to obtain solutions of the problem in acceptable simulation time. The model is preliminarily tested on several numerical experiments.*

Keywords: *Inventories, discrete controlled object, simulation model, spreadsheets.*

1. INTRODUCTION

Dynamic discrete simulation models implemented in a spreadsheet can be used as a quite reliable and relatively simple tool for presenting static inventory models with a complex mathematical apparatus. These models can be easily implemented in real systems, e.g. companies. Static inventory models belong to the group of infinite time horizon inventory control models, which assume that the rate of the annual demand is known and constant over a continuous time period. In order to notate the difference between these types of models, the paper considers dynamic discrete inventory model, which implies more than a one-time period, more precisely, the finite time horizon is divided into t discrete time periods and the annual demand is known and constant over t discrete time periods. The number of replenishments obtained by the fixed time horizon is always an integer number, which defines demand through t discrete time periods. Discretization of continuous infinite time horizon to more than one of finite time periods is a more natural manner to analyse dynamic of real systems. In this way continuous time inventory models are interpreted as discrete time inventory models. The main characteristic of the inventory control in the case of the fixed order quantity for the finite time horizon is that the replenishment quantity is constant and performed throughout several replenishments which occur at the beginning of the equal portions of the time horizon. The sum of replenishment quantities over the time horizon is equal to the demand in the time horizon (D). In accordance to Kostic (2009) this type of flows are called “discrete input and continuous output”. Inventory flows may occur with or without allowed shortages.

The Economic order quantity (EOQ) model represents classical inventory model with known total product demand. The model is aimed at determination of order quantity, while the total cost of the production, ordering and inventory holding should be minimized. It was originally developed by Harris (1915), though Wilson (1934) is credited for his early in-depth analysis of the model. It was a time when it wasn’t easy to afford computers and the simple useful mathematical models were mostly used (see Erlencotter 1989 for the history of EOQ). Even more recent books which are considering inventory control (Axsäter, 2006; Russell and Taylor, 2006; Vollmann et al. 2005; Jacobs, Chase and Aquilano, 2004; Barlow, 2003; Muller, 2003; Wild, 2002) describe the classical EOQ model and its variants, as a starting point for further understanding of inventory dynamics.

A discrete time system control may be considered as very convenient for inventory dynamics describing, as it is stated in Kostic (2009). A model of the discrete system control could be both a simulation model of inventory dynamics and an optimization model which can give the optimal control according to a defined performance criterion.

There are a lot of articles describing the usage of the discrete-time system control for dynamic deterministic inventory problems. Most of them address lot-sizing problems, beginning with Wagner (1958) and Scarf (1959). In order to find an optimal inventory control for various variants of the dynamic lot-sizing problems, dynamic programming algorithms can be applied (Bertsekas 1987) and Anily (2005). Additionally, different special heuristics are developed in order to solve such problems (Jans & Degraeve, 2007).

Starting from this point, the paper is organized as follows. A static time-continuous inventory control model with stock-level dependent demand rate and variable holding cost, published in International Journal of Production Economics (Alfares, 2007) is described in Section 2. Section 3 presents previously described problem modelled as a combinatorial optimization problem of dynamic discrete-time system control process which is implemented in a spreadsheet. The model is developed in accordance with basic elements of the discrete controlled object, according to Kostic (2009) and Antic et al. (2015). Furthermore, this section contains some preliminary results of several numerical experiments. The final section addresses future work and gives summary of this paper.

2. STATIC TIME-CONTINUOUS INVENTORY MODEL WITH STOCK-LEVEL DEPENDENT DEMAND RATE AND VARIABLE HOLDING COST

According to Alfares (2007) inventory models with stock-level dependent demand rate are based on a real-life observation, where higher product availability on the market tends to stimulate higher sales. The same author examines stock policy for a single product with stock-level dependent demand rate and variable holding cost depending on the time of holding the products on stock. The holding cost per unit is represented as an increasing step function of the time spent on stock. Two time-dependent increase functions of holding cost were considered: a retroactive increase of holding cost and an incremental increase of holding cost. The paper (Alfares, 2007) presents procedures for optimal order quantity and optimal cycle time determination for both of these functions. The main objective of the model is to minimize total holding cost by calculating the optimal order quantity with the appropriate cycle time of inventory renewal. Alfares stated that the step structure of the holding cost is unique case that has not been analyzed so far in the other papers. This structure is representative for a large number of real-life situations in practice in which storage time could be classified into different ranges. For each of time ranges there is an appropriate different unit holding cost. This is particularly prevalent in the food products warehouses that have a high degree of deterioration (loss of stock quality). The holding cost of these products increasing as longer as they are kept in the warehouse, because more advanced preconditions are required in order to maintain the usability of supplies. As an example, it can be stated that there are three different holding costs, which can be applied to short, medium and long-term storage time. The storage time is divided into a number of delimited periods with successively increasing holding cost. If the storage time overstep into a new time period, the new holding cost may be retroactively applied to all previous periods, or in the case of an incremental increase of holding cost, the new cost will only be applied for a new period of time.

The model, developed in the previously mentioned paper, is aimed at defining of optimal inventory policy (with minimal holding cost), i.e inventory control system with stock-level dependent demand and time dependent holding cost. Considering that demand depends on stock level, it can be stated that higher demand influence higher inventory level. Assuming that holding costs per item unit in time unit are time-dependent, unit holding costs are higher for longer periods of storing. Holding costs represent variables over different storage periods. The holding cost per item unit is increased only when the storage time of the item unit exceeds a certain discrete unit of time, i.e. the holding cost per unit in time unit is an increasing step function of the storage time. Two types of the holding cost step functions are given as follows (Alfares, 2007):

1. Retroactive holding cost increase – the holding cost per unit in the last storage period is applied to all previous periods.
2. Incremental holding cost increase - the holding cost per unit for each period (including the last one) applies only to items that has been stored in that particular period.

The basic assumptions of the model (Alfares, 2007) are:

1. The demand rate R is an increasing function of the inventory level q (relation 1):

$$R(q) = Dq^{\beta}, \quad D > 0, \quad 0 < \beta < 1, \quad q \geq 0 \tag{1}$$

2. The holding cost is varying and represents an increasing step function of time in storage.
3. There is no delay in inventory replenishments, they are immediate.
4. Shortages are not allowed.
5. A single item is considered.

In order to explain the model, author uses the following notations (Table 1).

Table 1: Notations in stock management model presented in (Alfares, 2007)

Notation in Alfares (2007)		Notation in spreadsheet model
q(t)	The quantity on-hand at time *t*	
D	Constant (base) demand rate	**D**
n	Number of distinct time periods with different holding cost rates	
t	Time from the start of the cycle (t=0)	
t_i	End time of period *i*, where *i* =1, 2, 3..., *n*, t=0 and $t_n = \infty$.	
k	Ordering cost per order	**S**
h_i	Holding cost of the item in period *i*	
h(t)	Holding cost of the item at time *t* , $h(t)=h_i$ if $t_{i-1} \le t \le t_i$	**H(t)**
T	Cycle time	**T**
β	Demand parameter indicating elasticity in relation to the inventory level	**β**

The main objective of the inventory control model from (Alfares, 2007) is minimization of total holding costs in a unit of time and this cost includes two components:

- The ordering cost k/T, because there is only one order per cycle.
- The holding cost per cycle is obtained as integral of product of holding costs *h(t)* and stock level *q(t)* during the entire cycle.

The total inventory costs are represented by equation (2).

$$TIC = \frac{k}{T} + \frac{1}{T}\int_0^T h(t)q(t)dt. \tag{2}$$

In the case of retroactive holding cost increase, the holding cost is an increasing step function of storage time and correspondingly $h_1 < h_2 < h_3 < ...h_n$. In this case holding cost depends on the length of storage time. Holding cost in last period is applied retroactively to all previous periods. If the cycle ends in period *e* ($t_{e-1} \le T \le t_e$), then the holding cost rate is applied to all previous periods (*1, 2, ..., e*). Total holding costs per unit time could be represented by relation (3):

$$TIC = \frac{k}{T} + \frac{h_i}{T}\int_0^T q(t)dt, \qquad t_{i-1} \le T \le t_i \tag{3}$$

In the case of incremental holding cost increase, cost function implies higher holding cost in later periods of the time cycle. Thus, if the cycle ends in period *e* ($t_{e-1} \le T \le t_e$), then holding cost rate h_1 is applied to period 1, rate h_2 is applied to period 2, all the way to the rate h_e which is applied to period *e*. In this case, firstly it is needed to set the value $t_e = T$ and then total holding cost is represented as:

$$TIC = \frac{k}{T} + \frac{h_1}{T}\int_0^{t_1} q(t)dt + \frac{h_2}{T}\int_{t_1}^{t_2} q(t)dt + \cdots + \frac{h_e}{T}\int_{t_{e-1}}^{t_e=T} q(t)dt. \tag{4}$$

In the paper (Alfares, 2007), the author presents numerical experiments for both types of functions with parameters presented in Table 2.

Table 2: Input data for numerical experiments from (Alfares, 2007)

D = 400 units per year,
k = $300 per order,
β = 0.1,
h_1 = $5/unit/year, $0<T\le 0.2$, t_1 = 0.2 year,
h_2 = $6/unit/year, $0.2<T\le 0.4$, t_2 = 0.4 year,
h_3 = $7/unit/year, $0.4<T\le N$, t_3 = N.

Table 3: Solutions for examples listed in Alfares (2007)

1)	2)
Q* = 243 units,	Q* = 250 units,
T* = 0.39 year,	T* = 0.4 year,
TIC* = $1460.43/year.	TIC* = $1369.96/year.

The solutions obtained by applying algorithms developed in (Alfares, 2007), for solving the problem in both cases, over the input data shown in Table 2 are given in Table 3.

Based on all of the above stated, it can be concluded that the observed model has the following characteristics:

- It represents a variation of the basic economic order quantity (EOQ) model with the minimal cost and has all the characteristics of the classical EOQ model (Antic et al., 2015).
- Single product inventory model.

- Fixed order quantity model.
- Model with a fixed ordering interval, there is no delay in inventory replenishments, they are immediate.
- The model belongs to a group of models with discrete inflow and continuous outflow (Kostic, 2009).
- It is a static inventory control model and dynamic of flow subject accumulation is observed for only one period of time.
- Shortages are not allowed.

3. DYNAMIC DISCRETE INVENTORY CONTROL MODEL WITH STOCK-LEVEL DEPENDENT DEMAND RATE AND VARIABLE HOLDING COST

Considering that the model described is a variation of the basic EOQ model, the mathematical relations of the law of behaviour, control domain and flow regulators can be considered the same as the ones in the discrete dynamic EOQ model presented in (Antic et al., 2015). Considering a time period of one year, the whole time horizon can be divided into 365 days, which means that T = 365 days (t=1, 2..., T). Since there is only one product, one material flow with appropriate phases is analysed, where:

- X_t – quantity of the flow subject in accumulation at the end of time period t,
- Y_t^I or (Y_t^1) - quantity of the flow subject that represents inflow of accumulation during a period t.
- Y_t^O or (Y_t^2) - quantity of the flow subject that represents outflow of accumulation during a period t.
- $u_t = N_0^I$- number of deliveries in time horizon T, i.e. control variable $u_t = N_0^I \ \forall \ t = 1, 2, \dots, T$

The law of behaviour of the discrete controlled object can be represented by relations (5):

$$X_o = known$$
$$X_t = X_{t-1} + Y_t^1 - Y_t^2, t = 1, 2, \dots, T \tag{5}$$

Based on the equation of the law of behaviour of a discrete controlled object, the state (quantity) of accumulation at the end of the current period is equal to the state of the accumulation from the previous period, increased for all inflows into the accumulation in the current period and decreased by all outflows from the accumulation in the current period. All costs are constant except for the holding cost that is growing, as the quantity of products in the accumulation increases.

It is assumed that there is discrete inflow with one stage received at the beginning of time period t. The quantity of received items in one time period can be zero (no inflow) or the u-part of the total demand D (6).

$$Y_1^1 = \begin{Bmatrix} D/u_t, & X_{t-1} + 0{,}0001 < D/T \\ 0, & otherwise \end{Bmatrix}, t = 1 \tag{6}$$

i.e. for t = 2,...,T the flow regulator equation is represented by relation (7):

$$Y_t^1 = \begin{Bmatrix} \min(D/u_t, D - \sum_{t=2}^{T} Y_{t-1}^1), \ X_{t-1} + 0{,}0001 < D/T \\ 0, \quad otherwise \end{Bmatrix}, t = 2, \dots, T \tag{7}$$

Incensement of the flow subject quantity in the accumulation is a discrete, while the outflow from the accumulation is continuous. The inflow and outflow are realised at the very beginnings of the time periods, so that the time required for the forming of a discrete quantity of the flow subject is equal to zero (LT = 0), consequently the delay of the flow regulator is equal to zero (8).

$$Y_1^2 = D/T, t = 1 \tag{8}$$

i.e. for t = 2,...,T the equation of flow regulator is represented by relation (9):

$$Y_t^2 = \min(D/T, \ D - \sum_{t=2}^{T} Y_{t-1}^2), t = 2, \dots, T \tag{9}$$

As daily demand is constant, the outflow of inventory will be D/T per day, until the entire demand isn't met. The amount of inflow D/N_0^I will be available at the moment when the quantity of stock falls below the amount of daily demand in order to avoid a lack of inventory (negative stock). The item will be ordered until the demand is fully met. Inventory replenishment will be realized N_0^I times, in equal amounts of D/N_0^I.

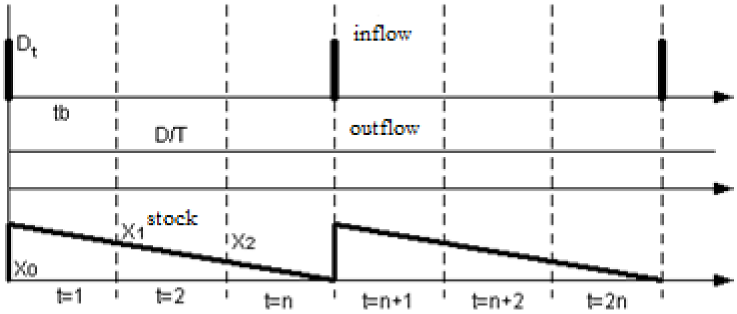


Figure 1: Inflows, outflows and inventory state over time horizon, in accordance with (Kostic, 2009)

Control domain is defined by relation which assures non-negativity of the stock and by relation which assures non-negativity of control variable, as well as integer value (10).

$$0 \leq u_t = N_0^I \leq T$$
$$u_t = integer$$
$$0 \leq X_{t-1} + Y_t^1 - Y_t^2$$
$$t = 1, 2, \ldots, T \tag{10}$$

In order to implement the model in a spreadsheet and to perform a simulation the basic EOQ model was used, whereby the law of behaviour and flow regulators are unchanged, but it is necessary to:

1. Develop a function for demand variable;
2. Complement the relations for the holding cost in accordance with the author's assumptions;
3. Adjust the time periods from the author's (Alfares, 2007) to the time periods in the discrete time model in order to determine the periods in which ordering should be realised;
4. Make an order counter.

Based on the basic assumptions of the classical EOQ model, it is possible to define the following assumptions of the dynamic discrete model:

$$D \neq const,$$
$$S = const,$$
$$C = const = 0.$$
$$H = H(t) \neq const. \tag{11}$$

The total inventory cost (*TIC*= (min) *J*) at the end of the time horizon *T*, can be expressed by performance criterion (12) that should be minimized.

$$(min)\ J = \sum_{t=1}^{T} (S \cdot \begin{Bmatrix} 1,\ Y_t^1 > 0 \\ 0,\ Y_t^1 = 0 \end{Bmatrix} + H(t) \cdot (X_{t-1} + Y_t^1 - Y_t^2/2) + C \cdot Y_t^1) \tag{12}$$

Where:

- T - number of days of the time horizon ($T \leq 365$ days),
- C - unit price of the item, which is equal to 0.
- S - delivery cost of one delivery stage in period t, if delivery is realised. This cost is constant for each delivery. If there is no delivery, then this cost is equal to 0 (relation 13).

$$S = \begin{Bmatrix} S = const, & Y_t^1 > 0 \\ 0, & Y_t^1 = 0 \end{Bmatrix} \tag{13}$$

- *H(t)* – The holding cost per unit depends on storage time and for different duration of the time spent on storage, there are different holding cost per unit. The defined cost is multiplied by the average amount of stock in the storage (relation 14).

$$H = \begin{bmatrix} h_1 = \$5/\text{unit/year},\ 0 < T \leq 0.2,\ t_1 = 0.2\ \text{year}, \\ h_2 = \$6/\text{unit/year},\ 0.2 < T \leq 0.4,\ t_2 = 0.4\ \text{year}, \\ h_3 = \$7/\text{unit/year},\ 0.4 < T \leq N,\ t_3 = N. \end{bmatrix} \tag{14}$$

The discrete controlled inventory model is defined by law of behaviour and control domain. The initial state (inventory level) is known as X_o=0 and the problem of the optimal control of discrete system can be describes as follows: For the given time horizon (*T*) under the given circumstances (relation 11) of the demand (*D*) considered as variable, the price of items (*C*), the ordering costs(*S*) and the holding cost per unit (*H*) depending on the time that the items spend on inventory, it is necessary to determine a control variable u_t. Control variable should be determinate in the manner that for a given time (*T*) guide a discrete control object from the initial state X_0 to a set of ending states X_T, under the defined constraints, while the performance criterion (relation 12) obtains minimum value.

In order to solve this problem by total search method it is necessary to evaluate values of the performance criterion for all of the possible values of control variable u_t. The control variable represents unknown order quantity *Q* for single item, so that $u_t=Q$, for *t* =1, 2, …, *T* (*T*=365 days). The number of orders is counted based on the simulation model. The model of discrete controlled object is suitable for development in spreadsheet. Based on the mathematical model of discrete controlled object and its performance criterion, the following formulas are implemented in a spreadsheet.

Table 4: Spreadsheet formulas of the dynamic discrete EOQ inventory model based on (Alfares, 2007):

Cell	Formula or initial data	Copy range
D14:	=$D12/$D$13	None
D17:	=IF($D$13/$D$24<=0,2*$D$13;5/$D$13;IF($D$13/$D$24<=0,4*$D$13 ;6/$D$13;7/$D$13))-D20	None
D24:	=F393	None
D27:	=M26+H27-I27	D27:D391
F27:	=1	None
F28:	=IF(H28>0;F27+1;F27)	D28:D391
H27:	=D22	None
H28:	=IF(M27+0,0001<V27;$D$22;0)	H28:H391
I27:	=V27	I27:I391
M27:	=M26+H27-I27	M27:M391
Q27:	=Q26+R27+S27+T27	Q27:Q391
R27:	=IF($H27>0;$D$15;0)	R27:R391
S27:	=$D$17*($M26+$H27-$I27/2)	S27:S391
T27:	=$D$16*$H27	T27:T391
V27:	=(M26+H27)^$D$18*$F$3/$D$13	V27:V391

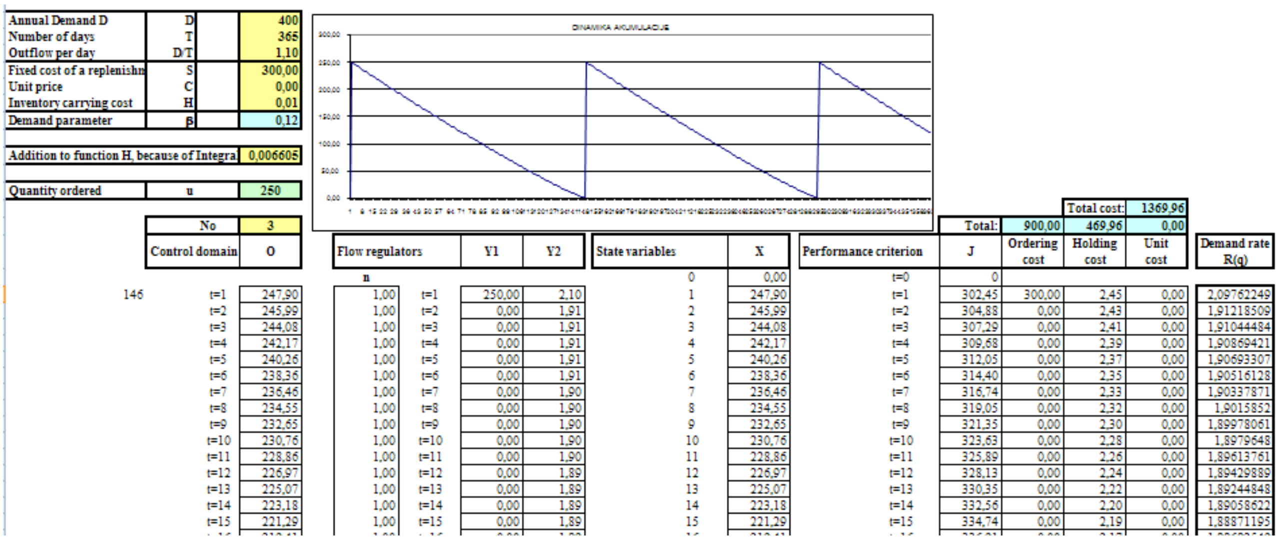
3.1. Numerical results and discussion

Created spreadsheet model (Figure 2) is suitable for simulation and "What-if..." analysis. For different control variable values it is possible to define a set of different decision-making scenarios in spreadsheet. Testing of the results of the simulation model in the spreadsheet is done for the case of an incremental holding cost increase, according to the input data in Table 2. In accordance to the presented static model and defined circumstances, ordering of the optimal quantity of items (250 units) results in the minimal annual cost *TIC*=$1.369,00, with a duration of replenishment cycle of 146 days, or 0.4 year (Table 3).

By setting the same starting values in the spreadsheet model, it can be concluded that optimal order quantity of 250 units results in 3 orders per year, while the minimal annual cost is *TIC*=$1.369,00 at the end of the observed time horizon *T*, with a duration of replenishment cycle of 146 days, or 0.4 year. Although the results of a dynamic discrete simulation model coincide with the results of a static model, it is necessary to specify certain differences that are conditioned by the nature of discrete and continuous models.

The basic differences between the continuous and the discrete model are:

- In order to simulate the exact result it was necessary to correct the demand function, which is an increasing function depending on the inventory stock-level. The parameter β represents the demand parameter and affects the sharpness of the demand curve. Changing of β parameter (from 0.10 to 0.12) in the simulation spreadsheet model results in the correct total costs (from $1369.83 to $1.369.96). It can also be noticed that the surfaces of the triangle are of irregular shape (Figure 3) and can be calculated only by the integral of the function. In that case the discretization of the period in which we observe the demand will lead to minor differences in discrete values of demand for a particular period, unlike the value which would be obtained by applying an integral function to a particular period. The integral calculates the demand in each part of the moment ***t***, while the simulation model calculates the demand in discrete points (t_1, t_2, t_3, ..., T).



Control domain	O	Flow regulators n		Y1	Y2	State variables	X	Performance criterion	J	Ordering cost	Holding cost	Unit cost	Demand rate R(q)
						0	0,00	t=0	0				
t=1	247,90	1,00	t=1	250,00	2,10	1	247,90	t=1	302,45	300,00	2,45	0,00	2,09762249
t=2	245,99	1,00	t=2	0,00	1,91	2	245,99	t=2	304,88	0,00	2,43	0,00	1,91218509
t=3	244,08	1,00	t=3	0,00	1,91	3	244,08	t=3	307,29	0,00	2,41	0,00	1,91044484
t=4	242,17	1,00	t=4	0,00	1,91	4	242,17	t=4	309,68	0,00	2,39	0,00	1,90869421
t=5	240,26	1,00	t=5	0,00	1,91	5	240,26	t=5	312,05	0,00	2,37	0,00	1,90693307
t=6	238,36	1,00	t=6	0,00	1,91	6	238,36	t=6	314,40	0,00	2,35	0,00	1,90516128
t=7	236,46	1,00	t=7	0,00	1,90	7	236,46	t=7	316,74	0,00	2,33	0,00	1,90337871
t=8	234,55	1,00	t=8	0,00	1,90	8	234,55	t=8	319,05	0,00	2,32	0,00	1,9015852
t=9	232,65	1,00	t=9	0,00	1,90	9	232,65	t=9	321,35	0,00	2,30	0,00	1,89978061
t=10	230,76	1,00	t=10	0,00	1,90	10	230,76	t=10	323,63	0,00	2,28	0,00	1,8979648
t=11	228,86	1,00	t=11	0,00	1,90	11	228,86	t=11	325,89	0,00	2,26	0,00	1,89613761
t=12	226,97	1,00	t=12	0,00	1,89	12	226,97	t=12	328,13	0,00	2,24	0,00	1,89429889
t=13	225,07	1,00	t=13	0,00	1,89	13	225,07	t=13	330,35	0,00	2,22	0,00	1,89244848
t=14	223,18	1,00	t=14	0,00	1,89	14	223,18	t=14	332,56	0,00	2,20	0,00	1,89058622
t=15	221,29	1,00	t=15	0,00	1,89	15	221,29	t=15	334,74	0,00	2,19	0,00	1,88871195

Figure 2: Dynamic discrete inventory model in the spreadsheet for Q = 250, according to solutions from (Alfares, 2007)

- While simulating the results, it is necessary to correct the function of the holding cost, which is a step function of the storage time. In order to obtain the most accurate result of total costs, the relation for calculating the holding cost needs to be expanded by the difference that arises as a result of the discretization of the time in which the continuous function of holding cost is observed. In the presented dynamic simulation model, the discrete value of the holding cost function should be reduced by the difference that would be obtained if the same continuous function was solved by the integral of the function. As in the previous case, the curve is a non-linear shape and the area under the curve can be calculated only by the integral of the function, but in the case of a discrete simulation model, calculation of the holding cost is done in discrete time points.

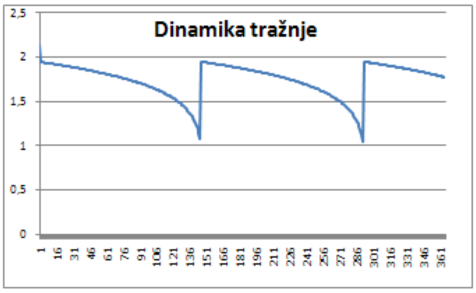


Figure 3: Dynamic of demand for β=0.12

- In the initial data of the static model, the annual demand is 400 units and it meets the order quantity of Qo = 250, with total ordering costs of TIC = \$1.369,96. The author of the paper (Alfares, 2007) does not specify the number of deliveries in which the annual demand is satisfied. In order to satisfy this demand in (Alfares, 2007), it is necessary to order an optimal quantity of supplies twice, which amounts 500 units of stock (which is higher than the annual demand of 400 units of products). However, the total cost (if we consider that the holding cost for all periods is greater than 7\$/day/ product), to satisfy the total demand of 400 units, will be \$1207.7, which is considerably less then optimal in comparison to the established total cost of the TIC = \$1,369.96 from Alfares (2007). This means that although the maximum holding cost are considered in order to obtain the optimal result for the TIC, there is a difference of \$162.26, which cannot be compensate in any way.
- By simulating the initial input data in the spreadsheet model and using the proposed holding cost function (relation 14), it can be seen that in order to achieve the total cost of TIC = \$1.369,96, it is necessary to order Qo = 250 units, 3 times. In this case, the demand of 400 units of products would be fully satisfied, and the stock would remain in excess of 350 units, which seems extremely non-logical, as the total cost could be lower than the proposed optimal cost.

4. CONCLUSION

In this paper a static time-continuous inventory model with stock-level dependent demand rate and variable holding cost, published by author Alfares (2007), is modelled as a combinatorial optimization problem of the corresponding dynamic discrete time system control process. For solving this problem a dynamic discrete simulation model is developed and implemented in a spreadsheet. Preliminary numerical results show that this model could be efficiently applied on problems of smaller dimensions, which can be solved through numerous simulations steps.

A future research could be directed toward more systematic investigations of algorithms appropriate for solving of presented combinatorial optimization problem. In the case of real-life problems with larger dimensions (i.e. models with an increased number of items) some special, metaheuristics, or hybrid heuristic approaches for solving the described problem could be developed. Example of special heuristic developed for dynamic discrete EOQ based inventory control model, which can be used as a starting point for new approach, is presented in (Djordjevic et al., 2017).

REFERENCES

Alfares, H. K. (2007). Inventory model with stock-level dependent demand rate and variable holding cost. *International Journal of Production Economics, 108*(1-2), 259-265.

Axsäter S. (2006). *Inventory Control*. International Series in Operations Research & Management Science, Springer Science + Business Media, New York, pp. 51–60.

Anily, S., & Tzur, M. (2006). Algorithms for the multi- item multi- vehicles dynamic lot sizing problem. *Naval Research Logistics (NRL), 53*(2), 157-169.

Antic, S., Djordjevic, L., Kostic, K. & Lisec. A. (2015). Dynamic discrete simulation model of an inventory control with or without allowed shortages. *University Politehnica of Bucharest Scientific Bulletin-series A-applied mathematics and physics, 77*(1), 163-176.

Barlow J. (2003). *Excel Models for Business and Operations Management*, Wiley, New York, pp. 244–258.

Bertsekas D.P. (1987). *Dynamic Programming: deterministic and stochastic models*, Prentice-Hall, Englewood Cliffs, New Jersey.

Djordjevic, L., Antic, S., Cangalovic, M., & Lisec, A. (2017). A metaheuristic approach to solving a multiproduct EOQ-based inventory problem with storage space constraints. *Optimization letters, 11*(6), 1137-1154.

Erlenkotter, D. (1989). Note—An Early Classic Misplaced: Ford W. Harris's Economic Order Quantity Model of 1915. *Management Science*, *35*(7), 898-900.

Harris F.W. (1915). How many parts to make at once, factory, *Magazine Management*, *10*(2), 135–136

Harris, F. W., (1915). *Operations and costs.* New York, pp. 48-52.

Jacobs, F. R., Chase, R. B., & Aquilano, N. (2004). *Operations management for competitive advantage*. Boston: Mc-Graw Hill, 64, 70.

Jans, R., & Degraeve, Z. (2007). Meta-heuristics for dynamic lot sizing: A review and comparison of solution approaches. *European Journal of Operational Research*, *177*(3), 1855-1875.

Kostic, K. (2009). Inventory control as a discrete system control for the fixed-order quantity system. *Applied Mathematical Modelling, 33*(11), 4201-4214.

Muller M. (2003). *Essentials of Inventory Management,* AMACOM, New York, pp. 115–129.

Russell R., Taylor B. (2006). *Operations Management: Quality and Competitiveness in a Global Environment*, Wiley, New York, pp. 529–552.

Scarf H. (1959). *The optimality of (s; S) policies in the dynamic inventory problem*. In: Arrow, Karlin, Patrick (Eds.), Proceedings of the First Stanford Symposium

Vollmann T., Berry W., Whybark D., Jacobs R. (2005). *Manufacturing Planning and Control Systems for Supply Chain Management*. The Definitive Guide for Profession.

Wilson, R.H., (1934). A scientific routine for stock control. *Harvard Business Rev.* 13, 116–128.

Wild T. (2002). *Best Practice in Inventory Management*, Elsevier Science, London, pp. 112–148.

Wagner H.M., Whitin T. (1958). Dynamic version of the economic lot size model, *Management Science, 5*(1), 89–96.

XIII Balkan Conference on Operational Research

INVENTORY LOADING PROBLEM WITH A HETEROGENEOUS VEHICLE FLEET

Türkay Umut Yılmaz[1], Ali Ekici[2]

[1] Department of Industrial Engineering, Ozyegin University, Istanbul, Turkey, turkay.yilmaz@ozu.edu.tr

[2] Department of Industrial Engineering, Ozyegin University, Istanbul, Turkey, ali.ekici@ozyegin.edu.tr

Abstract: *In this paper, we study an inventory loading problem with multiple customers and a heterogeneous vehicle fleet. In this problem, a single supplier satisfies the demand at multiple customer locations over a given planning horizon using vehicles with compartments each of which can carry the demand of a single customer. Supplier's goal is to develop a delivery plan (including delivery routes, delivery times and delivery quantities) with minimum transportation cost while avoiding stock-outs at the customer locations. We develop a mathematical model for the problem. Since the mathematical model can only handle small instances, we propose a two-phase solution approach to solve large instances. The proposed approach first generates several clusters (possibly overlapping) of customers. Then, we solve a set partitioning problem to select the best subset of clusters so that each customer is included in one cluster only. While solving the set partitioning problem, we use the cost of serving each cluster which is determined with the help of a heuristic algorithm. We test the proposed two-phase approach on instances with up to 25 customers and 6 compartments on a vehicle and observe that the proposed heuristic can find solutions with less than 1% optimality gap on average.*

Keywords: *inventory loading problem, heterogeneous vehicles, multi-compartment vehicles, clustering*

1. INTRODUCTION

Vendor-managed inventory is a business model that passes the responsibility of managing the inventory at the customer locations from customers to the supplier. In this business model, supplier plans the replenishment at the customer locations considering the demand at different locations and the supply. It provides significant advantages over the traditional vendee-managed inventory systems including increased visibility of stock-levels, increased customer service and reduced inventory and stock-outs and eventually reduced logistics costs. However, in order to exploit such benefits, one has a develop efficient replenishment policies including when and how much to replenish each customer location.

Motivated by the vendor-managed inventory business model, we analyze the replenishment operations of a supplier which is responsible for replenishing the inventory of multiple customers over a planning horizon. We assume that the delivery vehicles have multiple compartments each of which can be used to transport the demand of one customer only at a time. That is, no two customer's deliveries can share a compartment. Such a compartment structure for delivery vehicles is applicable for the transportation of oil products and chemicals in real life.

We formulate the problem as an integer programming problem. However, since the initial model is nonlinear, we implement linearization techniques to convert it to a linear model. Since the mathematical model can solve small instances only, we propose a two-phase heuristic approach to solve large instances. In the proposed heuristic approach, we first generate several (possibly overlapping) clusters of customers, and assuming each cluster of customers is served by one vehicle we construct a delivery plan for each cluster. Then, in the second phase, we choose the best subset of clusters so that each customer is included in one cluster by solving a set partitioning problem.

The rest of the paper is organized as follows. In Section 2, we review the related studies in the literature and discuss the differences between the ones in the literature and our study. In Section 3, we provide a formal problem definition, explain the problem with a simple example and present the proposed mathematical model. We discuss the two-phase heuristic approach in Section 4. The proposed heuristic approach is tested on a set of randomly generated instances, and the results are presented in Section 5. We conclude the paper in Section 6.

2. LITERATURE REVIEW

The problem studied in the paper is related to the well-known *Inventory Routing Problem* (IRP). IRP is first introduced in the seminal-paper by Bell et al. (1983) with an application in the distribution of industrial gases. They look for a solution with minimum transportation cost while avoiding stock-outs at the customer locations.

Since then, different variants of IRP have been studied by several authors. We refer the reader to Coelho et al. (2013) for a recent survey of IRP.

Variants of IRP differ in terms of the objective function, replenishment policy, demand structure, planning horizon and number of products. Objective function may include transportation costs (Desaulniers et al., 2016; Raa and Aghezzaf, 2009) and/or inventory holding costs (Archetti et al., 2012; Van Anholt et al., 2016). Depending on the application, certain replenishment policies (pre-established rules to replenish customers) including *order-up-to-level* (OU), *maximum level* (ML) and *zero-inventory ordering* (ZIO) can be implemented. In OU policy, a customer's inventory is raised to its maximum level each time it is visited (Bertazzi et al., 2002; Solyalı and Süral, 2011). In ML policy, any quantity up to a customer-specific maximum level can be delivered (Campbell and Savelsbergh, 2004; Coelho and Laporte, 2013; Desaulniers et al., 2015). In ZIO policy, a customer's inventory is replenished only when it drops to zero (Bertazzi et al., 2007). Demand at the customer locations can be either deterministic (Ekici et al., 2015; Jung and Mathur, 2007) or stochastic (Kleywegt et al., 2004). Variants of IRP with different planning horizon assumptions have also been studied in the literature. Federgruen and Zipkin (1984) and Golden et al. (1984) consider a single-period planning horizon whereas Bertazzi et al. (2002) and Campbell and Savelsbergh (2004) study a multi-period planning horizon setting. Moreover, some authors analyze IRP when the planning horizon is infinite (Bell et al., 1983; Chan and Simchi-Levi, 1998). Finally, both the single-product (Campbell and Savelsbergh, 2004; Solyalı and Süral, 2011) and multi-product (Chandra, 1993; Parthanadee and Logendran, 2006) variants of IRP have been studied in the literature.

Both heuristic algorithms and exact approaches have been proposed in the literature to address different variants of IRP. For example, Archetti et al. (2007), Coelho and Laporte (2013) and Solyalı and Süral (2011) propose a branch-and-cut algorithm, and Desaulniers et al. (2016) develop a branch-and-price-and-cut algorithm. Campbell and Savelsbergh (2004) and Ekici et al. (2015) propose a two-phase decomposition heuristic, and Archetti et al. (2012) and Coelho et al. (2012) solve the problem using metaheuristics.

IRP with multi-compartment delivery vehicles has been studied by Popovic et al. (2012) and Vidovic et al. (2014) with an application in fuel delivery. They assume a homogeneous fleet of delivery vehicles with multiple compartments. Different from our study, they assume that compartments have same capacity and focus on only full compartment deliveries. The demand at a petrol station is assumed to be deterministic but not necessarily constant. Moreover, each petrol station has a certain storage capacity for each fuel type. Their goal is to minimize total transportation and inventory holding cost. Due to compartment structure and the number of fuel types, each vehicle visits up to three (Popovic et al., 2012) or four (Vidovic et al., 2014) stations per route, and they allow using a compartment to deliver the same product to more than one station in a route. Since the proposed mixed integer model can only handle small instances, Popovic et al. (2012) develop a variable neighborhood search heuristic, and Vidovic et al. (2014) solve the two phases (inventory management and route construction) of the problem sequentially. Vidovic et al. (2014) first solve an MIP model for the inventory decisions, and then implement a heuristic to construct delivery routes. Finally, they improve the solution using a variable neighborhood descent search.

In addition to IRP, another problem related to our study is the multi-product inventory loading problem (Yuceer, 1997, 1999; Yuceer and Ozakca, 2010) where the delivery of multiple products to a single customer location (Yuceer, 1997) or the delivery of a single product to multiple customer locations (Yuceer and Ozakca, 2010) is planned using a single vehicle that has multiple compartments.

Although Yuceer (1997) and Yuceer and Ozakca (2010) study a related problem, the main difference is the routing aspect. Yuceer (1997) study the single customer case, and hence, assume direct delivery between the depot and customer location. Yuceer and Ozakca (2010) ignore the routing aspect of the problem and focus how to assign the compartments of the vehicle to different products. However, in our setting we consider multiple vehicles and allow multi-stop tours. Although it has the advantage of improving the efficiency of the delivery plan, routing aspect combined with compartment assignment decisions brings additional complexity to the problem. We call the problem under consideration *Inventory Loading Problem with a Heterogeneous Vehicle Fleet* (ILP-HV) since the single vehicle version of the problem is first called *Inventory Loading Problem* (Yuceer, 1997). Finally, the main differences between our setting and the studies by Popovic et al. (2012) and Vidovic et al. (2014) are the compartment and fleet structure. We assume a heterogeneous fleet of vehicles with different compartment structures and plan the delivery to multiple customer locations using these delivery vehicles where one compartment of a vehicle can be used to deliver to one customer only. Moreover, the compartments are not necessarily identical in our case.

3. PROBLEM DEFINITION

ILP-HV is defined on an undirected graph $G = \{V,E\}$ which consists of a vertex set $V = \{0,1,...,N\}$ including one depot (node 0) and N customers, and an edge set $E = \{(i, j) : i, j \in V\}$. Each edge (i, j) has an associated traversal cost c_{ij}. Customer i has a demand rate of d_i units per unit time. We assume that we have K heterogenous vehicles with different compartment capacity structure. We use M_k to denote the number of compartments of vehicle k ($k \in \{1,2,\ldots,K\}$) and q_{mk} to denote the capacity of compartment m of vehicle k ($k \in \{1,2,\ldots,K\}, m \in \{1,2,\ldots,M_k\}$). Deliveries to the customers are made via routes that start and end at the depot. We assume that each vehicle serves a cluster of customers such that the vehicle visits all the customers in the cluster in a single tour and performs this single tour with a certain frequency to satisfy the demand at the customer locations. Such a replenishment policy is called *static policy* and has the practicality advantage over a random replenishment policy. Such a replenishment policy can result in underutilization of trucks (hence a nonoptimal solution) due to inconsistency between the compartment capacities and the demand rates of the customers assigned to those compartments. However, these kind of replenishment policies including order-up-to-level, powers-of-two, zero-inventory order policies (Coelho et al., 2013) and full compartment deliveries (Popovic et al., 2012) are studied in the literature in order to simplify the problem and make the solution more practical. Each compartment can be used to satisfy the demand at one customer location. However, depending on the demand rate more than one compartment can be assigned to one customer. Our objective is to minimize total transportation cost per unit time while avoiding stock-out at the customer locations.

Next, we present a simple example to further explain the problem. In the example provided in Figure 1, we have three customer locations and the demand rate of each customer location is provided next to each node. For simplicity, the cost of traversing each edge is set to 1. We have two vehicles to perform the delivery to these customers. First vehicle has two compartments each of which has a capacity of 40. Vehicle 2 has three compartments. Two of them have a capacity of 100, and the last one has a capacity of 20.

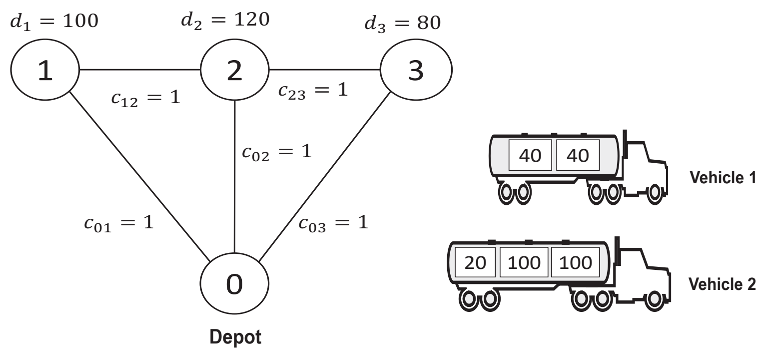


Figure 1 An illustrative example.

A feasible delivery plan for this example would be assigning customers 1 and 3 to vehicle 2, and customer 2 to vehicle 1. Both compartments of vehicle 1 is used to make delivery to customer 2, and the total delivery amount would be 80 in this case. The cost of this tour is 2, and since the demand rate of customer 2 is 120, this tour have to be performed with a frequency of 1.5 ($= 120/80$). That is, this delivery is going to be performed every $2/3$ time unit. The per unit time cost of satisfying this demand is 3 ($= 2 \times 1.5$). For vehicle 2, if we assign one of the compartments with capacity 100 to customer 3, and the other two compartments to customer 1, then the delivery frequency of this tour is going to be $5/6$ ($= \max\{100/120, 80/100\}$) and the per unit time cost of this plan is going to be $10/3$ ($= 5/6 \times 4$). The per unit time total cost of this plan is $19/3$. Note that in this solution customer 3 will be delivered 96 ($= 1.2 \times 80$) units in each tour due to static policy although we can deliver 100 units in each tour.

Finally, we provide an integer programming formulation for ILP-HV. We use $\mathcal{N}$ and $\mathcal{K}$ to denote the set of customers and vehicles respectively: $\mathcal{N} = \{1,2,\ldots,N\}$, $\mathcal{K} = \{1,2,\ldots,K\}$. $\mathcal{M}_k$ denotes the set of compartments of vehicle k ($k \in \mathcal{K}$): $\mathcal{M}_k = \{1,2,\ldots,M_k\}$. The following decision variables are used in the formulation:

x_{ikm}	: 1, if customer i is assigned to compartment m of vehicle k; 0, otherwise	$i \in \mathcal{N}, m \in \mathcal{M}_k, k \in \mathcal{K}$
y_{ik}	: 1, if customer i is served by vehicle k; 0, otherwise	$i \in \mathcal{N}, k \in \mathcal{K}$
w_{ijk}	: 1, if edge (i, j) is used by vehicle k; 0, otherwise	$i, j \in V, k \in \mathcal{K}$
f_k	: Delivery frequency of vehicle k	$k \in \mathcal{K}$
u_i	: Auxiliary variable defined for customer i to eliminate subtours	$i \in \mathcal{N}$

Using B to denote a sufficiently large number, the proposed mathematical model is as follows:

$$\text{Min} \quad \sum_{k\in\mathcal{K}}\sum_{i\in V}\sum_{j\in V} c_{ij}w_{ijk}f_k \tag{1}$$

$$\text{s.t.} \quad \sum_{k\in\mathcal{K}} y_{ik} = 1 \qquad \forall i\in\mathcal{N} \tag{2}$$

$$x_{ikm} \leq y_{ik} \qquad \forall i\in\mathcal{N}, m\in\mathcal{M}_k, k\in\mathcal{K} \tag{3}$$

$$\sum_{m\in\mathcal{M}_k} x_{ikm} \geq y_{ik} \qquad \forall i\in\mathcal{N}, k\in\mathcal{K} \tag{4}$$

$$\sum_{i\in\mathcal{N}} x_{ikm} \leq 1 \qquad \forall m\in\mathcal{M}_k, k\in\mathcal{K} \tag{5}$$

$$B(1-y_{ik}) + \frac{\sum_{m\in\mathcal{M}_k} q_{mk}x_{ikm}}{d_i} f_k \geq 1 \qquad \forall i\in\mathcal{N}, k\in\mathcal{K} \tag{6}$$

$$\sum_{i\in V} w_{ijk} = y_{jk} \qquad \forall j\in\mathcal{N}, k\in\mathcal{K} \tag{7}$$

$$\sum_{j\in V} w_{ijk} = \sum_{j\in V} w_{jik} \qquad \forall i\in V, k\in\mathcal{K} \tag{8}$$

$$\sum_{j\in\mathcal{N}} w_{0jk} \leq 1 \qquad \forall k\in\mathcal{K} \tag{9}$$

$$u_i - u_j + (N+1)\sum_{k\in\mathcal{K}} w_{ijk} \leq N \qquad \forall i,j\in\mathcal{N} \tag{10}$$

$$x_{ikm} \in \{0,1\} \qquad \forall i\in\mathcal{N}, k\in\mathcal{K}, m\in\mathcal{M}_k \tag{11}$$

$$y_{ik} \in \{0,1\} \qquad \forall i\in\mathcal{N}, k\in\mathcal{K} \tag{12}$$

$$w_{ijk} \in \{0,1\} \qquad \forall i,j\in V, k\in\mathcal{K} \tag{13}$$

$$f_k \geq 0 \qquad \forall k\in\mathcal{K} \tag{14}$$

$$1 \leq u_i \leq N+1 \qquad \forall i\in\mathcal{N} \tag{15}$$

In this model, the objective function (1) minimizes per unit time logistics costs which can be calculated as the multiplication of transportation cost and delivery frequency for each vehicle. Constraints (2) ensure that each customer is served by exactly one vehicle. Constraints (3-4) guarantee that each customer is assigned to at least one compartment of the vehicle serving that customer. Constraints (5) make sure that each compartment is assigned to at most one customer. Constraints (6) determine delivery frequency for each vehicle. When a customer is visited by a vehicle, then this vehicle has to go from one of the nodes to this customer. This is enforced by Constraints (7). Constraints (8) guarantee continuity of the delivery routes. Constraints (9) make sure that each vehicle is assigned to at most one route. Constraints (10) are the classical subtour elimination constraints. Remaining constraints are the restrictions on the values each decision variable can take.

Note that the proposed model is nonlinear because of the objective function (1) and Constraints (6). However, these nonlinear expressions include only multiplication of binary and continuous decision variables which can be easily linearized using two auxiliary decision variables:

s_{ijk} : Auxillary variable for $w_{ijk}f_k$ $\quad \forall i,j\in V, k\in\mathcal{K}$
t_{ikm} : Auxillary variable for $x_{ikm}f_k$ $\quad \forall i\in\mathcal{N}, m\in\mathcal{M}_k, k\in\mathcal{K}$

In the linearized version of the mathematical model, the new objective function is as follows:

$$\text{Min} \quad \sum_{k\in\mathcal{K}}\sum_{i\in V}\sum_{j\in V} c_{ij}s_{ijk} \tag{16}$$

Constraint (6) is removed from the model, and the following constraints are added to linearize the nonlinear terms in the model:

$$B(1-y_{ik}) + \sum_{m\in\mathcal{M}_k} q_{mk}t_{ikm} \geq d_i \qquad \forall i \in \mathcal{N}, k \in \mathcal{K} \tag{17}$$

$$t_{ikm} \leq f_k \qquad \forall i \in \mathcal{N}, m \in \mathcal{M}_k, k \in \mathcal{K} \tag{18}$$

$$t_{ikm} \leq Bx_{ikm} \qquad \forall i \in \mathcal{N}, m \in \mathcal{M}_k, k \in \mathcal{K} \tag{19}$$

$$t_{ikm} \geq f_k - B(1-x_{ikm}) \qquad \forall i \in \mathcal{N}, m \in \mathcal{M}_k, k \in \mathcal{K} \tag{20}$$

$$s_{ijk} \leq f_k \qquad \forall i,j \in V, k \in \mathcal{K} \tag{21}$$

$$s_{ijk} \leq Bw_{ijk} \qquad \forall i,j \in V, k \in \mathcal{K} \tag{22}$$

$$s_{ijk} \geq f_k - B(1-w_{ijk}) \qquad \forall i,j \in V, k \in \mathcal{K} \tag{23}$$

$$t_{ikm} \geq 0 \qquad \forall i \in \mathcal{N}, m \in \mathcal{M}_k, k \in \mathcal{K} \tag{24}$$

$$s_{ijk} \geq 0 \qquad \forall i,j \in V, k \in \mathcal{K} \tag{25}$$

4. SOLUTION APPROACH

In the proposed solution approach, we first form clusters of customers. At this step, one customer can be included in more than one cluster. Then, in the second phase we solve a set partitioning problem to choose the best subset of clusters among the generated ones so that each customer is included in one cluster only.

4.1. Phase I: Clustering

In this phase, we generate several overlapping clusters of customers. Without loss of generality, we assume that the customer locations are given on a map of $D \times D$ where the centroid of the map is the origin of the two dimensional space. Then, we divide the two dimensional space into regions of equal size using rays emanating from origin. An example is given in Figure 2. In Figure 2(a), we divide the region into six regions of equal size starting from the x-axis. Here, we choose $\beta = 60$ degrees to form six regions. Then, in Figure 2(b), we perform this step by choosing a starting ray which makes α degrees with the x-axis. We implement this region forming step for different β and α values to form several regions. More specifically, for β we try 3, 5, 15, 30, 45, 60, 90, 120, 180 and 360 degrees. Note that β determines the number of regions to be formed. Then, for each $\beta_1 \in \{3,5,15,30,45,60,90,120,180,360\}$ value we try all α values such that $\alpha = t\beta_0$ for some positive integer t and $\beta_0 \in \{3,5,\ldots,\beta_1\}$.

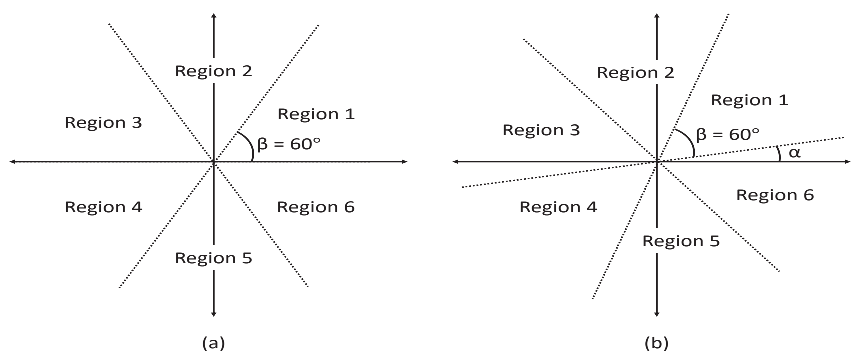


Figure 2 An illustration of clustering phase.

Then, within each region formed using the procedure above, we construct clusters as follows. We order the customers in the region under consideration in a descending order of their distances from the depot. Without loss of generality, assume that we have an order of customers $1,2,\ldots,N'$ in the region. We use M_{max} to denote the maximum number of compartments on a vehicle. For each integer k ($\leq M_{max}$) and $i \in \{1,2,\ldots,N'-k+1\}$, we form a cluster which includes customers $i, i+1,\ldots,i+k-1$. More specifically, for each k value, we form all subsets of $\{1,2,\ldots,N'\}$ with size k such that k consecutive customers are included in the subset. These subsets are the clusters of customers generated in the first phase.

By trying different values for β and α, we form several regions each of which are used to generate several clusters of customers. By forming diversified regions and then clusters, we expect to choose a good final combination at the end of the second phase.

4.2. Phase II: Set Partitioning

After generating several clusters of customers, we estimate the cost of serving each cluster by each vehicle. For some of the clusters, some of the vehicles may not be able to serve them if the number of customers in the cluster is greater than the number of compartments on the vehicle. In that case, we take the cost of serving as infinity.

For each cluster and vehicle pair, we solve the single vehicle version of the problem heuristically. There are two main decisions here: (i) how to route the vehicle, and (ii) how to assign the customers to the compartments. We use V_j to denote the set of customers in the cluster under consideration and k to denote the vehicle. Since we assume a static delivery policy, the route of each vehicle is a *Hamiltonian tour* visiting the customers in V_j. Hence, one has to solve the *Traveling Salesman Problem* (TSP) to find the best delivery route. We implement the Christofides algorithm followed by 1-0 insertion, 1-1 exchange and 2-opt improvement steps to find a "good" route heuristically. Since we solve the single vehicle version of the problem several times for each cluster, we choose to implement an efficient and effective heuristic for solving the TSP. For the assignment of customers to the compartments of a vehicle, we propose a fast heuristic called the *Greedy Assignment Algorithm*. We sort the compartments and the customers in the decreasing order of their capacities and demand rates, respectively. Then, we assign one compartment to each customer starting from the top of the list. Then, the remaining compartments are assigned to customers one by one by choosing the customer with the largest frequency each time. The frequency of a customer is calculated as ratio of its demand rate to the total capacity of the compartments assigned to it. Finally, we implement an improvement procedure where the customer with the largest delivery frequency is determined, and one of the compartments from other customers is reassigned to this customer as long as there is improvement (reduction) in the maximum delivery frequency among all customers.

Using F to denote the maximum delivery frequency among all customers and L to denote the cost of the delivery route, the per unit time cost of serving cluster V_j by vehicle k is calculated as: $\gamma_{jk} = FL$.

At the end of the second phase, we solve a set partitioning problem to choose the best subset of clusters among the generated ones. We first define the additional notation needed for the model:

$\mathcal{S}$	: Set of all generated clusters	
$\mathcal{S}_i$	: Set of generated clusters that include customer i	$\forall i \in \mathcal{N}$
γ_{jk}	: Cost of serving cluster j by vehicle k	$\forall j \in \mathcal{S}, k \in \mathcal{K}$

Note that γ_{jk} is set to a very large number if vehicle k cannot serve cluster j. We define one type of decision variable for the set partitioning model:

z_{jk} : 1, if cluster j is served by vehicle k; 0, otherwise $\quad \forall j \in \mathcal{S}, k \in \mathcal{K}$

We solve the following set partitioning model to choose the final set of clusters:

$$\text{Min} \quad \sum_{j\in\mathcal{S},k\in\mathcal{K}} \gamma_{jk} z_{jk} \tag{26}$$

$$\text{s.t.} \quad \sum_{j\in\mathcal{S}_i,k\in\mathcal{K}} z_{jk} = 1 \quad \forall i \in \mathcal{N} \tag{27}$$

$$\sum_{j\in\mathcal{S}} z_{jk} \leq 1 \quad \forall k \in \mathcal{K} \tag{28}$$

$$z_{jk} \in \{0,1\} \quad \forall k \in \mathcal{K}, j \in \mathcal{S} \tag{29}$$

Objective (26) is to find the minimum cost partition from $\mathcal{S}$. Constraints (27) ensure that each customer is included in exactly one of the clusters chosen. Each vehicle can serve at most one cluster. This is enforced by Constraints (28). Remaining constraints are the binary restrictions. After solving this problem, the objective function value gives the per unit time cost of serving the chosen clusters.

5. COMPUTATIONAL STUDY

We conduct a preliminary computational study to see the performance of the proposed heuristic. We compare its performance against the optimal solution. We first explain how we find the optimal solution of each instance. For the generated instances, we generate all the possible clusters and calculate the cost of serving each cluster by solving the single vehicle version of the problem optimally. For each cluster, we solve the corresponding TSP to find the best route and solve another integer programming model to find the best assignment compartments to the customers for each cluster-vehicle pair. Then, we solve the set partitioning problem presented above.

Since we generate all the possible clusters and solve the single vehicle version of the problem optimally, this takes a lot of time, and it is not possible to do this for large instances. We generate instances with up to 25 customers and find the optimal solution as explained above. Note that the proposed mathematical model can also be used to find the optimal solution. The procedure implemented above to find the optimal solution shows how the computational time of the proposed heuristic increases when it is implemented with all clusters generated and the optimal solution found for each cluster by solving a single vehicle version of the problem.

For each instance, the locations of the customers and the depot is randomly generated on a map of 1000×1000. The cost of travelling from one location to another one is determined by calculating the Euclidean distance between the locations. The demand rate of each customer and the capacity of each compartment is uniformly generated between 1000 and 3000. Number of compartments on each vehicle is randomly generated between 2 and 6. The number of customer locations and the number of vehicles for each instance is provided in Table 1. The proposed heuristic is coded using C# language in Visual Studio 2015 Community platform and executed on a laptop with Intel (R) Pentium Dual CPU at 2.33 GHZ, 2GB of RAM on Windows 7. IBM CPLEX 12.6 is used to find the optimal solution and solve the final set partitioning problem in the proposed heuristic.

In Table 1, we provide the optimality gap of the solutions found by the proposed heuristic. We observe that for 14 out of 25 instances, the proposed heuristic finds the optimal solution. For the remaining instances, the largest optimality gap is just above 2%. The average optimality gap of the solutions found is 0.42%. The average computational time of the proposed heuristic is 0.28 minutes whereas the optimal solution is found in around 8.5 hours.

We observe that the proposed heuristic finds optimal or near-optimal solutions in a reasonable amount of time, and can easily be implemented for larger instances as well.

Table 1: Optimality gaps of the solutions found by the proposed heuristic

Ins. #	K	N	Opt. Gap	Ins. #	K	N	Opt. Gap	Ins. #	K	N	Opt. Gap	Ins. #	K	N	Opt. Gap	Ins. #	K	N	Opt. Gap
1	5	10	0.11%	**6**	10	15	0.00%	**11**	14	15	1.23%	**16**	16	20	0.00%	**21**	17	25	0.85%
2	5	10	0.20%	**7**	11	15	0.00%	**12**	11	20	0.00%	**17**	18	20	2.12%	**22**	17	25	0.00%
3	9	10	0.00%	**8**	13	15	0.54%	**13**	13	20	0.14%	**18**	19	20	0.00%	**23**	19	25	1.23%
4	9	10	0.00%	**9**	13	15	0.00%	**14**	15	20	0.00%	**19**	13	25	0.00%	**24**	22	25	1.31%
5	9	10	0.00%	**10**	14	15	0.00%	**15**	16	20	0.71%	**20**	16	25	0.00%	**25**	23	25	2.01%

6. CONCLUSION

In this study, we analyze a vendor-managed inventory system where a supplier is replenishing the inventories of its customers using vehicles that have multiple compartments to transport the products. Products of different customers have to transported in different compartments, and no customer can be served by more than one vehicle. We assume a heterogeneous fleet of vehicles in terms of the compartment structure and focus on a static delivery policy where each vehicle is assigned to a set of customers and replenishes the customers using the same delivery route every time dispatched.

We develop a mathematical model for the problem, and linearize it using linearization techniques. Since it can handle only very small instances, we propose a two-phase heuristic approach. In the first phase, we generate several clusters of customers. Then, in the second phase we choose the best subset of clusters among the generated ones by solving a set partitioning problem. In the set partitioning problem, we use the cost of serving each cluster by each vehicle. We estimate this cost of serving by finding the best route to serve these customers and the best assignment of compartments to the customers. We solve the *Traveling Salesman Problem* for each cluster using the well-known Christofides algorithm followed by some improvement ideas to find the delivery route. Then, we find a “good” assignment of compartments to the customers implementing a greedy assignment algorithm followed by an improvement step to determine the delivery frequency.

In the preliminary computational study, we observe that the proposed heuristic finds solutions with less than 0.5% optimality gap on average. We test the proposed heuristic on instances with up to 25 customers and compare the solution found against the optimal solution. However, for larger instances it is not possible to find the optimal solution. Hence, in order to evaluate the performance of the proposed heuristic on larger instances a possible future direction is developing an exact method that can handle relatively larger instances or finding a tight lower bound on the total cost.

REFERENCES

Archetti, C., Bertazzi, L., Hertz, A., and Speranza, M. G. (2012). A hybrid heuristic for an inventory routing problem. *INFORMS Journal on Computing*, 24(1):101–116.

Archetti, C., Bertazzi, L., Laporte, G., and Speranza, M. G. (2007). A branch-and-cut algorithm for a vendor-managed inventory-routing problem. *Transportation Science*, 41(3):382–391.

Bell, W. J., Dalberto, L. M., Fisher, M. L., Greenfield, A. J., Jaikumar, R., Kedia, P., Mack, R. G., and Prutzman, P. J. (1983). Improving the distribution of industrial gases with an on-line computerized routing and scheduling optimizer. *Interfaces*, 13(6):4–23.

Bertazzi, L., Chan, L. M. A., and Speranza, M. G. (2007). Analysis of practical policies for a single link distribution system. *Naval Research Logistics*, 54(5):497–509.

Bertazzi, L., Paletta, G., and Speranza, M. G. (2002). Deterministic order-up-to level policies in an inventory routing problem. *Transportation Science*, 36(1):119–132.

Campbell, A. M. and Savelsbergh, M. W. (2004). A decomposition approach for the inventory-routing problem. *Transportation Science*, 38(4):488–502.

Chan, L. M. A. and Simchi-Levi, D. (1998). Probabilistic analyses and algorithms for three-level distribution systems. *Management Science*, 44(11):1562–1576.

Chandra, P. (1993). A dynamic distribution model with warehouse and customer replenishment requirements. *Journal of the Operational Research Society*, 44(7):681–692.

Coelho, L. C., Cordeau, J.-F., and Laporte, G. (2012). Consistency in multi-vehicle inventory-routing. *Transportation Research Part C: Emerging Technologies*, 24:270–287.

Coelho, L. C., Cordeau, J.-F., and Laporte, G. (2013). Thirty years of inventory routing. *Transportation Science*, 48(1):1–19.

Coelho, L. C. and Laporte, G. (2013). A branch-and-cut algorithm for the multi-product multi-vehicle inventory-routing problem. *International Journal of Production Research*, 51(23-24):7156–7169.

Desaulniers, G., Errico, F., Irnich, S., and Schneider, M. (2016). Exact algorithms for electric vehicle-routing problems with time windows. *Operations Research*, 64(6):1388–1405.

Desaulniers, G., Rakke, J. G., and Coelho, L. C. (2015). A branch-price-and-cut algorithm for the inventory-routing problem. *Transportation Science*, 50(3):1060–1076.

Ekici, A., Özener, O. Ö., and Kuyzu, G. (2015). Cyclic delivery schedules for an inventory routing problem. *Transportation Science*, 49(4):817–829.

Federgruen, A. and Zipkin, P. (1984). A combined vehicle routing and inventory allocation problem. *Operations Research*, 32(5):1019–1037.

Golden, B., Assad, A., and Dahl, R. (1984). Analysis of a large scale vehicle routing problem with an inventory component. *Large Scale Systems*, 7(2-3):181–190.

Jung, J. and Mathur, K. (2007). An efficient heuristic algorithm for a two-echelon joint inventory and routing problem. *Transportation Science*, 41(1):55–73.

Kleywegt, A. J., Nori, V. S., and Savelsbergh, M. W. (2004). Dynamic programming approximations for a stochastic inventory routing problem. *Transportation Science*, 38(1):42–70.

Parthanadee, P. and Logendran, R. (2006). Periodic product distribution from multi-depots under limited supplies. *IIE Transactions*, 38(11):1009–1026.

Popovic, D., Vidovic, M., and Radivojevic, G. (2012). Variable neighborhood search heuristic for the inventory routing problem in fuel delivery. *Expert Systems with Applications*, 39(18):13390–13398.

Raa, B. and Aghezzaf, E.-H. (2009). A practical solution approach for the cyclic inventory routing problem. *European Journal of Operational Research*, 192(2):429–441.

Solyalı, O. and Süral, H. (2011). A branch-and-cut algorithm using a strong formulation and an a priori tour-based heuristic for an inventory-routing problem. *Transportation Science*, 45(3):335–345.

Van Anholt, R. G., Coelho, L. C., Laporte, G., and Vis, I. F. (2016). An inventory-routing problem with pickups and deliveries arising in the replenishment of automated teller machines. *Transportation Science*, 50(3):1077–1091.

Vidovic, M., Popovic, D., and Ratkovic, B. (2014). Mixed integer and heuristics model for the inventory routing problem in fuel delivery. *International Journal of Production Economics*, 147:595–604.

Yuceer, U. (1997). A multi-product loading problem: a model and solution method. *European Journal of Operational Research*, 101(3):519–531.

Yuceer, U. (1999). A multi-product inventory loading problem: a model and solution method. *International Journal of Mathematical Algorithms*, 1:107–132.

Yuceer, U. and Ozakca, A. (2010). A truck loading problem. *Computers & Industrial Engineering*, 58(4):766–773.

A BI-OBJECTIVE APPROACH TO LOCATING FACILITIES IN SOLID WASTE MANAGEMENT SYSTEMS

BRANISLAVA RATKOVIĆ[1], BRANKA DIMITRIJEVIĆ[2], VLADIMIR MOMČILOVIĆ[3]

[1] University of Belgrade, Faculty of Transport and Traffic Engineering, b.ratkovic@sf.bg.ac.rs
[2] University of Belgrade, Faculty of Transport and Traffic Engineering, b.dimitrijevic@sf.bg.ac.rs
[3] University of Belgrade, Faculty of Transport and Traffic Engineering, v.momcilovic@sf.bg.ac.rs

Abstract: *This paper presents a bi-objective mixed integer mathematical model for siting landfills and transfer stations in waste management systems. The first objective minimizes total costs of facility establishing and entire demand satisfaction while the second objective minimizes total number of end users undesirably influenced by located facilities. Hence, we are locating landfills and transfer stations, so as to minimize overall costs (landfills and transfer stations establishing and entire demand satisfaction) while keeping located facilities on a certain predefined distance from each other simultaneously minimizing their impact on end users. The model performances are tested on a small scale illustrative example.*

Keywords: *landfill location, transfer station location, bi-objective mixed integer modeling, Pareto optimal solutions.*

1. INTRODUCTION

Solid waste management (SWM) represents one of the major challenges of modern society. Namely, current global solid waste generation levels are approximately 1.3 billion tons per year, and are expected to increase to approximately 2.2 billion tons per year by 2025, which represents a significant increase in per capita waste generation rates, from 1.2 to 1.42 kg per person per day in the next fifteen years (World Bank 2012). Hence, an integrated approach in dealing with these quantities of waste is needed, not only because of decreasing number of landfills and their capacities, but also because waste represents lost resources. For example, in terms of household waste alone, each person in Europe is producing, on average, half of tone of such waste, of which only 40% is reused or recycled and in some countries more than 80% still goes to landfills (EU 2018).

In order to deal with solid waste quantities, local authorities must establish efficient logistics networks in terms of logistics costs, but also include other factors like environmental risk associated with facilities in SWM networks e.g. landfills and transfer stations when designing these networks. Usually, basic SWM network consists of waste generators (end users, municipalities, cities, etc.), which represent the initial nodes on the network, transfer stations (TSs) as intermediate nodes, and landfills as final nodes (Figure 1).

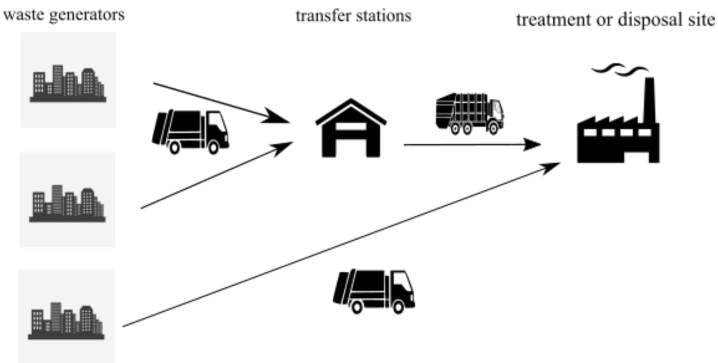


Figure 1. Illustration of solid waste management system

Landfills are facilities that pose environmental risks due to pollution of the local environment manifested in contamination of groundwater or soil, generation of dangerous gases, air pollution, etc., hence they fall in the category of undesirable facilities. As mentioned, the number of landfills has decreased dramatically throughout the last decades: there were close to 8000 landfills and dumps in the United States in 1988, but only 1908 in 2010 (Eiselt and Marianov, 2014). This means that existing or/and new landfills will have larger capacities (which increase their level of undesirability) and many population centers will no longer have a landfill in their direct vicinity, thus necessitating long transportation routes (Eiselt and Marianov, 2014). So, in SWM systems intermediate nodes must be included, that is TSs, to which waste collected from end users is hauled in vehicles with smaller capacity(collection vehicles).

For waste transporting to a landfill or a processing site, collection vehicles are unloaded within TSs and then waste is re-loaded into larger vehicles (transfer vehicles). Frequently, waste transfer is accompanied by some removal, separation, or handling. From transportation point of view, TS represents an option to obtain shorter vehicle routes in waste collection, and after unloading quick return to their primary task of waste collection (UNEP 1996). Additionally, larger landfills, incinerators or other treatment facilities are designed to serve a number of communities or an entire region, resulting in being sited at a considerable distance from the collection service areas (UNEP 1996).

Like landfills, TSs also fall into the category of undesirable facilities, which makes designing SWM system involving multiple objectives often conflicted. One of them is certainly related to “not in my backyard” philosophy. However, landfill/TS locations tend to be close to highly populated areas as important waste generation sites to minimize transportation costs.

In the literature, a number of papers are dealing with SWM system design problem, which resulted in review paper about operations research applications in SWM systems by Ghiani et al. (2014). Since, this paper addresses the problem of locating landfills and TSs as a multi-criteria decision model, short literature review of SWM system designing problems, formulated and solved as multi-criteria models, is given.

Eiselt and Marianov (2014) examined problem of locating landfills and transfer stations in the network, and formulated problem as a bi-objective mixed integer optimization problem. First objective minimizes costs of establishing facilities as well transportation costs, while the second to minimize pollution caused by established facilities. The model was tested on real data collected from a region of Chile. Xi et al. (2010) proposed an inexact chance-constrained mixed-integer linear programming model for supporting long-term planning of solid waste management in the City of Beijing, China. They examined three scenarios for waste management in Beijing and applied fuzzy multi-criteria decision analysis (MCDA) for analyzing the optimal solutions among the three alternatives. Erkut et al. (2008) proposed a multi-objective mixed integer linear model for the location–allocation municipal problem at the regional level in North Greece. The multi-objective problem was formulated as a lexicographic minimax problem aiming at finding a non-dominated solution with all objectives in balance. Silva, Alçada-Almeida and Dias (2017) presented multi-objective mixed-integer linear programming approach to identify locations and capacities of biogas plants for animal waste from dairy farms, and assign each farm to a subset of the opened biogas plants. They considered three objectives in the mathematical model: minimizing initial investment, operation and maintenance costs; minimizing transportation cost; and minimizing social rejection. The proposed model was applied to the Entre-Douro-e-Minho Region in Portugal.

This paper is an extension of the research presented in Dimitrijevic, Ratkovic and Selmic (2017), in which landfills are located with the aim to minimize costs of their establishing and the total number of undesirably influenced end users. In this paper model based on two objectives focuses on TS locations determining, as well. . The first objective function, as in the classical Fixed Charge Location Problem (FCLP) (Balinski, 1964), aims to minimize total facility and transportation costs, while the second, inspired by Minimum Covering Location Problem with Distance Constraint (MCLPDC) introduced by Oded and Rongbing (2008), minimizes the total number of end users undesirably influenced by landfills and TSs.

The paper is structured as follows. Description of the problem as well as mathematical formulation is presented in Section 2. In Section 3 an illustrative example is presented, while Section 4 summarizes our findings and provides some thoughts regarding future research.

2. DESCRIPTION OF THE PROBLEM AND MATHEMATICAL FORMULATION

We considered a problem, where end users are represented by population centers located at known sites i. Typically, population is aggregated at these centers, and we assume v_i residents are located at site i, each one generating q_i kg of waste per day. That means that $w_i=q_iv_i$ kg of waste will be generated at site i. All generated waste from end users should be collected and transported to the closest TS or directly to a landfill location. Waste from TSs is also transported to landfills by larger vehicles than those used for its transportation during the collection process. We made an assumption that waste quantities transported from TSs to landfills are not reduced due to the primary waste processing, which is not often, but still present, as for example in Serbia.

In order to formulate mathematical model for selecting landfill and transfer station locations, we defined five distances R_1, R_2, R_3, R_4 and R_5 (Figure 2). For locating landfills we defined R_1 representing their minimal acceptable separation distance. That is, landfills are dispersed over observed territory according to some predefined separation distance and their common influence on end users intended to be minimized. The same logic is applied for locating TSs by defining minimal acceptable separation distance between them (R_2), as well as between TSs and landfills (R_3). Distances R_4 and R_5 represent radiuses of areas negatively influenced by landfill and TS locations, respectively.

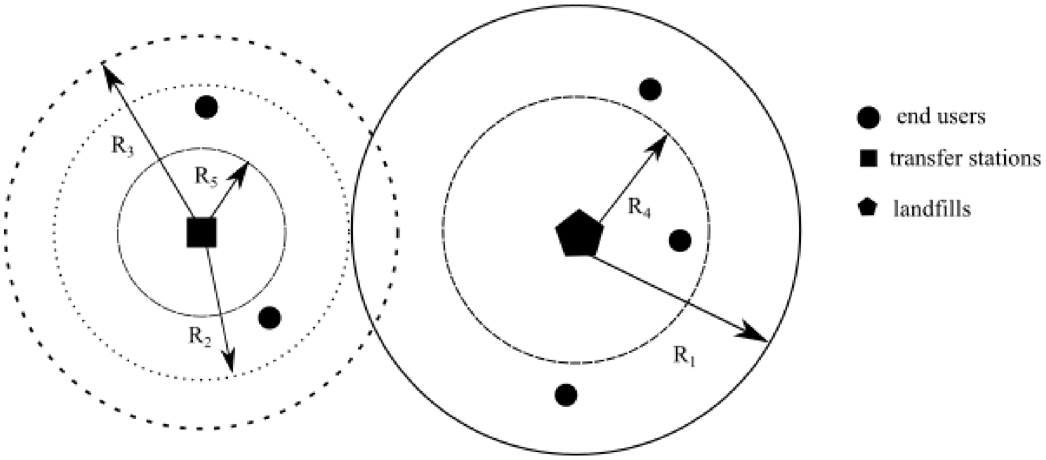


Figure 2. Illustration of separation distances

The following notation is used for mathematical formulation of the problem:

Sets and parameters:

J - set of nodes representing potential landfill locations
I - set of nodes representing end users
L - set of nodes representing potential transfer station locations
i - index used to represent end user
j,p - indices used to represent potential landfill locations
l,k - indices used to represent potential transfer station locations
d_{ij} - the shortest traveling distance between end user i and potential landfill location j
d_{il} - the shortest traveling distance between end user i and potential transfer station location l
d_{lj} - the shortest traveling distance between potential transfer station location l and landfill location j
d_{ij}^E - the Euclidean distance between end user i and potential landfill location j
d_{il}^E - the Euclidean distance between end user i and potential transfer station location l
f_j - fixed cost of locating a landfill at potential location j
f_l - fixed cost of locating a transfer station at potential location l
v_i - number of residents located at site i
q_i - quantity of waste generated per person at end user i
c_{ij} - unit waste transportation cost from end user i to landfill j
c_{il} - unit waste transportation cost from end user i to transfer station l
c_{lj} - unit waste transportation cost from transfer station l to landfill j
Q_l - capacity of a transfer station at potential location l
Q_j - capacity of a landfill at potential location j
R_1 - minimal acceptable separation distance between any two located landfills
R_2 - minimal acceptable separation distance between any two located transfer stations
R_3 - minimal acceptable separation distance between any located landfill and transfer station
R_4 - radius that represents landfills negative impact on end users
R_5 - radius that represents transfer stations negative impact on end users
M - large positive number
$N_j = \left\{ p \middle| d_{jp} < R_1, j \neq p \right\}, \forall j \in J$ - set of potential landfill locations that are on distance less than R_1 from particular landfill location j, excluding itself

$N_l = \left\{ k \middle| d_{lk} < R_2, l \neq k \right\}, \forall l \in L$ - set of potential transfer station locations that are on distance less than R_2 from particular transfer station location l, excluding itself

$N_j^l = \left\{ l \middle| d_{lj} < R_3 \right\}, \forall j \in J$ - set of potential transfer station locations that are on distance less than R_3 from particular landfill location j

$\Pi_i^j = \left\{ j \middle| d_{ij}^E < R_4 \right\}, \forall i \in I$ - set of potential landfill locations that cover end user i within radius R_4

$\Pi_i^l = \left\{ l \middle| d_{il}^E < R_5 \right\}, \forall i \in I$ - set of potential transfer station locations that cover end user i within radius R_5

Variables:

y_{ij} - continuous variable that measures waste quantity shipped from end user i to landfill j
y_{il} - continuous variable that measures waste quantity shipped from end user i to transfer station l
y_{lj} - continuous variable that measures waste quantity shipped from transfer station l to landfill j
z_i - integer variable which counts number of landfills that cover end user i
s_i - integer variable which counts number of transfer stations that cover end user i

$$x_j = \begin{cases} 1, & \text{if landfill is located at potential location } j \\ 0, & \text{otherwise} \end{cases}$$

$$x_l = \begin{cases} 1, & \text{if transfer station is located at potential location } l \\ 0, & \text{otherwise} \end{cases}$$

Formulation of the problem:

$$\min \quad OF_1 = \sum_j f_j x_j + \sum_l f_l x_l + \sum_i \sum_j c_{ij} d_{ij} y_{ij} + \sum_i \sum_l c_{il} d_{il} y_{il} + \sum_l \sum_j c_{lj} d_{lj} y_{lj} \quad (1)$$

$$\min \quad OF_2 = \sum_i v_i z_i + \sum_i v_i s_i \quad (2)$$

s.t.

$$Mx_j + \sum_{p \in N_j} x_p \leq M, \forall j \in J \quad (3)$$

$$Mx_l + \sum_{k \in N_l} x_k \leq M, \forall l \in L \quad (4)$$

$$Mx_j + \sum_{l \in N_j^l} x_l \leq M, \forall j \in J \quad (5)$$

$$\sum_j y_{ij} + \sum_l y_{il} = w_i, \forall i \in I \quad (6)$$

$$\sum_i y_{il} - \sum_j y_{lj} = 0, \forall l \in L \quad (7)$$

$$\sum_i y_{il} \leq Q_l x_l, \forall l \in L \quad (8)$$

$$\sum_i y_{ij} + \sum_l y_{lj} \leq Q_j x_j, \forall j \in J \quad (9)$$

$$y_{il} \leq Q_l x_l, \forall i \in I, l \in L \quad (10)$$

$$y_{ij} + y_{lj} \leq Q_j x_j, \forall i \in I, j \in J, l \in L \quad (11)$$

$$\sum_{j \in \Pi_i^j} x_j = z_i, \forall i \in I \quad (12)$$

$$\sum_{l \in \Pi_i^l} x_l = s_i, \forall i \in I \quad (13)$$

$$x_j \in \{0,1\}, x_l \in \{0,1\}, y_{ij} \geq 0, y_{il} \geq 0, y_{lj} \geq 0, z_i \in N_0, s_i \in N_0, \forall i \in I, j \in J, l \in L \quad (14)$$

In the first objective function (1), the aim is to minimize total facility and transportation costs, i.e. the total cost of facility establishing and entire demand satisfaction. Second objective function (2), minimizes the total number of end users undesirably influenced by landfills as well TSs. This objective function is created to account for multiple landfill and TS locations covering an end user, e.g. if some end user's node is covered by two landfills, two transfer stations or by their combination, then its population will be counted twice in the objective function, because the negative impact on that population is doubled. Constraints (3), (4) and (5) are characteristic for Anti-covering location problem (ACLP) (Moon and Chaudhry, 1984). They are referred as Neighborhood Adjacency Constraints. If, for example, in constraints (3) location j is selected for landfill placement (i.e. $x_j = 1$), then the term Mx_j equals the right hand side term M and forces $\sum_{p \in N_j} x_p = 0$. Thus, if location j is used, then all potential landfill locations p in its neighborhood defined by R_1, are restricted from use. The very similar explanation stands for constraints (4) and (5). Those constraints are practically aiming at landfills' and TSs' dispersion over observed region, thus preventing their common negative impact on the environment. Constraints (6) and (7) are balance flow constraints, ensuring that waste quantities generated at end user i are transported either to TS locations or landfill locations (6) and that all waste quantities generated at end users finally end at landfills (7). Constraints (8), (9), (10) and (11) ensure that the capacities of TSs and landfills are not exceeded and prevent users from being allocated to non-open facilities. Number of landfills and TSs that negatively affect each particular end user are obtained by constraints (12) and (13). In other words, those constraints determine which particular end users are covered and by how many landfills (12) and TSs (13). Constraints (14) describe nature of variables appearing in the model. Note that if distances R_1, R_2, R_3, R_4 and R_5 are set in the following way: $R_1 \geq 2R_4$, $R_2 \geq 2R_5$ and $R_3 \geq R_4 + R_5$, than z_i and s_i become binary 1-0 variables.

3. ILLUSTRATIVE EXAMPLE

In this section, we tested the proposed bi-objective model for landfill and TSs locating on one small scale illustrative example. The observed example consists of 7 nodes which are simultaneously potential TS locations and end users, while two nodes (1 and 4) are potential landfill locations. All inputs are illustrative and time based costs are normalized at daily level. Daily quantity of waste generated per person at end user i (q_i) is adopted to be 0.8 kg for all $i \in I$, while the capacity of landfill and TS locations are sufficient enough and cannot be exceeded. Unit transportation costs are calculated as average vehicle cost per ton-kilometer including both, loaded and empty runs. Let unit waste transportation cost for 5 t collection vehicle be 0.78 €/km, then the cost per unit transport volume is $c_{ij} = c_{il} = 0.156$ €/ t·km. Waste transportation cost per unit transport volume for 20 t transfer vehicle is fixed at $c_{jl} = 0.0715$ €/ t·km. Transportation costs do not involve vehicle procurement and depreciation. Fixed costs of locating landfills and transfer stations are observed for two time periods, ten and twenty years of operation. So, fixed costs (f_j) of locating landfills is uniform for all $j \in J$ and reduced to a daily amount of cca 330 € for the period of ten years (case 1) and cca 160 € for the period of twenty years (case 2). Fixed costs (f_l) of locating TSs are uniform for all $l \in L$, of cca 150 € for the period of ten years and cca 70 € for the period of twenty years. These costs are estimated from Serbian practice including construction and equipment without operational costs.

Input parameters d_{ij} (d_{il}, d_{lj}), q_i and v_i are presented in Table 1, while Euclidean distances between facility locations and end users are given in Table 2.

Table 1: Input parameters for the proposed model: d_{ij} (d_{il}, d_{lj}), q_i, v_i

	1	2	3	4	5	6	7	q_i	v_i
1	0	41.5	46.13	58.38	74.93	60.5	87.41	113243.2	141554
2	41.5	0	12.29	46.93	63.5	26.62	53.54	26656.8	33321
3	46.13	12.29	0	57.66	74.22	37.14	64.05	9624.8	12031
4	58.38	46.93	57.66	0	17	44.42	38.59	68722.4	85903
5	74.93	63.5	74.22	17	0	59.08	41.83	23143.2	28929
6	60.5	26.62	37.14	44.42	59.08	0	29.68	34480.8	43101
7	87.41	53.54	64.05	38.59	41.83	29.68	0	24123.2	30154

Table 2: Euclidean distances

i/j,l	1	2	3	4	5	6	7
1	0	21.4562	32.72938	49.94716	64.98274	45.10271	62.5071
2	21.456198	0	11.48212	40.79031	53.33917	23.95296	43.60143
3	32.729383	11.48212	0	42.63184	52.56099	14.66329	36.89197
4	49.947159	40.79031	42.63184	0	15.60175	37.34238	31.36767
5	64.982738	53.33917	52.56099	15.60175	0	43.38411	28.36918
6	45.102713	23.95296	14.66329	37.34238	43.38411	0	22.8965
7	62.5071	43.60143	36.89197	31.36767	28.36918	22.8965	0

Values for separation distances are set as follows: R_1 = 70 km, R_2 = 30 km, R_3=25, R_4 = 15 km and R_5 = 20 km. Big M is set to 10.

Problem was developed using Python 2.7 programming language and solved by CPLEX 12.6 software.

Results for numerical example for case 1, including all Pareto optimal solutions, are presented in Figure 3. Marginal solution for OF_1 is: x_j = (0,1), x_l = (1,0,0,0,0,0,0) and objectives' values are OF_1 = 1499.96 €, OF_2 = 256,386 residents. Next Pareto solution is: x_j = (1,0), x_l = (0,0,0,0,1,0) for which OF_1 = 1803.50 € and OF_2 = 196,686 residents. Marginal solution for OF_2 is: x_j = (0,1), x_l = (0,0,0,0,0,0,0), while OF_1 = 1980.70 € and OF_2 = 114,832 residents. It is usual that the decision maker has preferences for certain objectives (e.g. costs), so one solving approach could be relaxed lexicographic method to support such preference expressed by their order. If the decision maker choses in this model OF_1 as preferable objective, then if for example he/she allows the total cost to increase for 20.24 % from the optimal OF_1 (from 1499.6 to 1803.5 €), OF_2 could decrease for 23.28 % (from 256,386 to 196,686 residents).

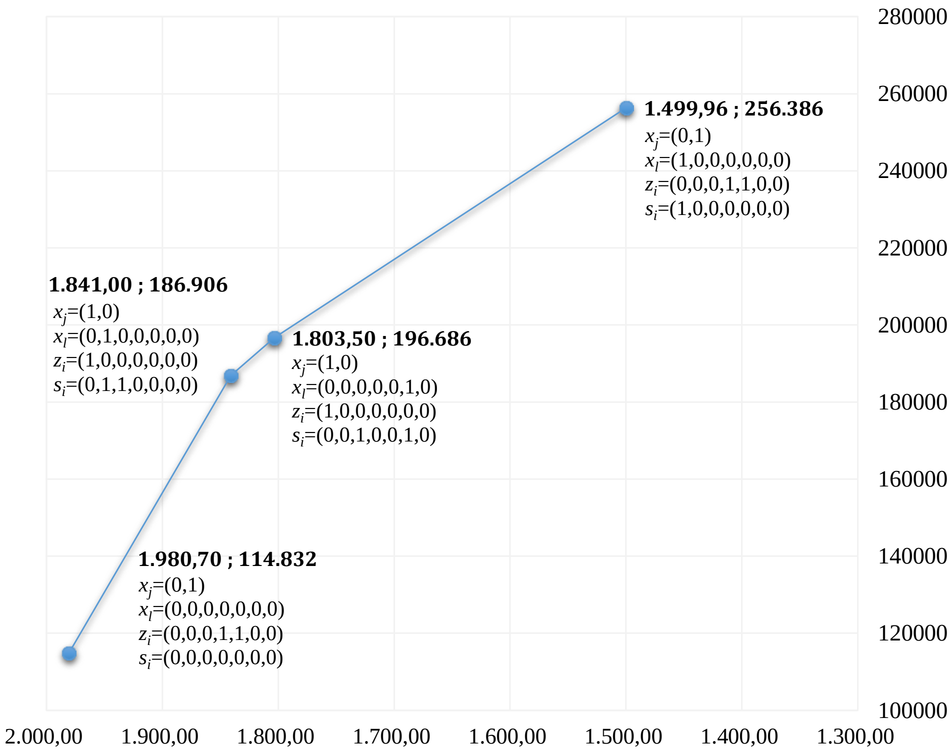


Figure 3. Pareto optimal solutions for illustrative example (case 1)

Results for numerical example for case 2, including all Pareto optimal solutions, are presented in Figure 4. Marginal solution for OF_1 is: x_j = (1,0), x_l = (0,1,0,1,0,0,1) and OF_1 = 1247.22 €, OF_2 = 331,892 residents. Next Pareto solution is: x_j = (1,0), x_l = (0,0,0,1,0,1,0) with objectives' values: OF_1 = 1261.9 €, OF_2 = 311,518 residents, and so on until marginal solution for OF_2 which is: x_j = (0,1), x_l = (0,0,0,0,0,0,0) and OF_1 = 1828.8 €, OF_2 = 114,832 residents. Again, if the decision maker choses OF_1 as preferable objective, then if for example he/she allows the total cost to increase for 1.18 % from the optimal OF_1 (from 1247.22 to 1261.9 €), OF_2 could decrease for 6.14 % (from 331,892 to 311,518 residents), or if total cost is increased for 6.74% from the optimal OF_1, OF_2 could decrease for 13.66 %.

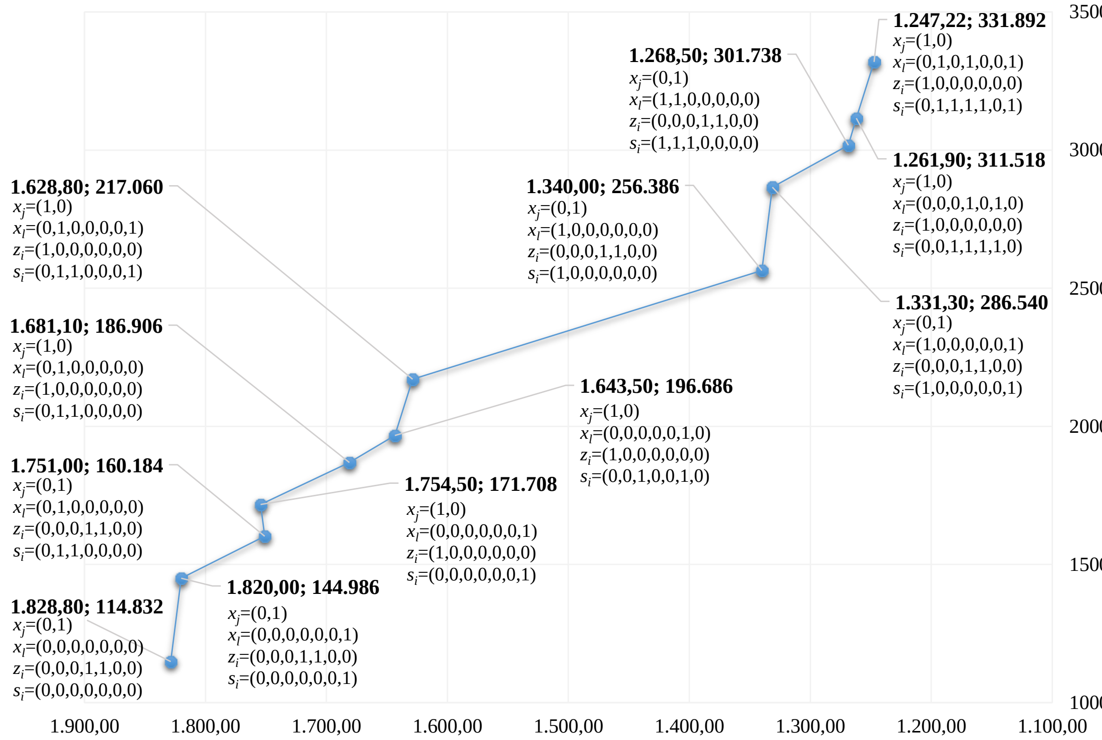


Figure 4: Pareto optimal solutions for illustrative example (case 2)

In case of landfill and TS opening costs spread over ten years period (case 1) up to one TS and one landfill is opened in all Pareto optimal solutions. Meanwhile, in case of landfill and TS opening costs spread over twenty years period (case 2) different combinations of opened facilities are obtained as Pareto optimums. Marginal solution for OF_2 is the same for both observed time periods (OF_2 = 114,832) while only one landfill is opened and minimal number of residents are negatively influenced.

All presented solutions are valuable for the decision maker because they provide complete insight into the necessary information in finding the most preferred solution when deciding on landfill's/TS's locations. In this small scale example it was possible to find all Pareto solutions, which is difficult for larger and almost impossible for real world data cases. Decision maker have to decide which solution satisfies his/hers preferences making appropriate trade-off.

5. CONCLUSION

This paper presents bi-objective model for determining landfill and transfer station locations in SWM systems. Model combines approaches and ideas from several location problems, such as FCLP, ACLP and MCLPDC. The first objective minimizes total facility (landfills and transfer stations) and transportation costs. The second objective minimizes the total number of end users undesirably influenced by landfills and transfer stations.

We presented one numerical example, with all Pareto optimal solutions and an illustration of the relaxed lexicographic method application. All obtained solutions are valuable for the decision maker because they provide complete insight into the necessary information in finding the most preferred solution when deciding on landfills'/TSs' locations. In order to test the model sensitivity on projected costs, two cases of landfill and TS opening costs are adopted. It is obvious that data accuracy is of great importance and their thorough estimation in Serbian circumstances will be part of authors' future research, especially regarding missing cost components and vehicle capacity variations.

Future research will also focus on real world data problem testing, involving different approaches for problem solving. Model will be upgraded so as to include consideration of uncertainty in generated waste quantities and negative impact radiuses. Another possibility is to include a third objective which minimizes number of end users influenced by routing collection and transfer vehicles.

Acknowledgement

This paper was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, through the projects TR 36006 and TR 36010.

REFERENCES

Balinski, M.L. (1965). Integer Programming: Methods, Uses, Computation, *Management Science*, 12, 253-313.

Dimitrijevic, B., Ratkovic, B., & Selmic, M. (2017). A multi-objective model for undesirable facility location, *Procedings of the 3rd Logistics International Conference Belgrade*, Serbia 25 - 27 May 2017, 19-24

Eiselt, H.A., & Marianov V. (2014). A bi-objective model for the location of landfills for municipal solid waste, *European Journal of Operational Research*, 235, 187–194.

Erkut, E., Karagiannidis, A., Perkoulidis, G., & Tjandra, S.A. (2008). A multicriteria facility location model for municipal solid waste management in North Greece. *European Journal of Operational Research*, 187, 1402–21.

EU (2018). Retrieved from http://ec.europa.eu/environment/waste/).

Ghiani, G., Laganà, D., Manni, E., Musmanno, R., & Vigo, C. (2014). Operations research in solid waste management: A survey of strategic and tactical issues G. *Computers & Operations Research*, 44, 22–32.

Oded, B., & Rongbing, H.J. (2008). The minimum weighted covering location problem with distance constraints, *Computers & Operations Research* 35, 356–372.

Silva, S., Alçada-Almeida,L., & Dias, L.C. (2017). Multiobjective programming for sizing and locating biogas plants: A model and an application in a region of Portugal, *Computers and Operations Research*, 83, 189–198.

UNEP (1996). Retrieved from http://www.unep.or.jp/Ietc/ESTdir/Pub/MSW/SP/SP3/SP3_3.asp

World Bank (2012). WHAT A WASTE: A Global Review of Solid Waste Management. Retrieved from http://www.ifc.org/wps/wcm/connect/1e5ca7004c07698db58eb7d8bd2c3114/What-A-Waste-Report.pdf?MOD=AJPERES.

Xi, B.D., Su, J., Huang, G.H., Qin, X.S., Jiang, Y.H., Huo, S.L., Ji, D.F., & Yao, B. (2010). An integrated optimization approach and multi-criteria decision analysis for supporting the waste-management system of the City of Beijing, China. *Engineering Applications of Artificial Intelligence*, 23, 620–31.

AN APPROACH TO LOCATING TRANSFER STATIONS IN WASTE MANAGEMENT SYSTEMS

MILORAD VIDOVIĆ[1], BRANISLAVA RATKOVIĆ[1], NENAD BJELIĆ[1], DRAŽEN POPOVIĆ[1]

[1] University of Belgrade, Faculty of Transport and Traffic Engineering, mvidovic@sf.bg.ac.rs
[1] University of Belgrade, Faculty of Transport and Traffic Engineering, b.ratkovic@sf.bg.ac.rs
[1] University of Belgrade, Faculty of Transport and Traffic Engineering, n.bjelic@sf.bg.ac.rs
[1] University of Belgrade, Faculty of Transport and Traffic Engineering, d.popovic@sf.bg.ac.rs

Abstract: *This paper addresses the problem of locating transfer stations in waste management systems. We considered problem where location of landfill and end users, as waste generators, are known. Proposed mixed integer linear problem minimizes costs of establishing transfer stations as well transportation costs. We tested proposed model on the case study for the south-western region of the Republic of Serbia.*

Keywords: *waste management systems, transfer stations, location modeling, MILP*

1. INTRODUCTION

Transfer stations (TS) in waste management systems (WMSs), represent sites for temporary storage, preparation and handling of municipal solid waste (MSW) intended for transport in the regional waste management center or landfill. That is, the transfer station is a place where MSW is unloaded from the collection vehicles (vehicles with smaller capacity), temporary stored and if no planed activities such as sorting/testing, loaded into transfer vehicles (vehicles with larger capacity) and transported for further treatment in a regional center or landfill (Figure 1).

Figure 1. WMS with TS

They can be of different sizes offering vary additional services like sorting, testing, can be owned by public, private or public/private entities, but the main function o TS is the same: consolidating waste from multiple collection vehicles into larger, high-volume transfer vehicles for more economical shipment to distant disposal sites (USEPA 2002). The main reason for using TSs is to reduce the cost of transporting waste to disposal facilities. Namely, consolidating smaller loads from collection vehicles into larger transfer vehicles reduces hauling costs by enabling collection crews to spend less time traveling to and from distant disposal sites and more time collecting waste (USEPA 2002).

In the literature, a number of papers addressed a problem of locating TS within WMSs systems and while using different approaches. Eiselt and Marianov (2015), Eiselt (2006), and Eiselt (2007) used the discount factor (1-α) that expresses the discount for the transportation from TSs to landfills as compared to that from the end users to TSs or landfills. As mentioned, the reason for the discount is justified by the fact that the transportation from end users involves smaller collection vehicles as opposed to the larger transport trucks that haul waste from TS to landfills (Eiselt 2007). Other authors (Yadav et al. 2016a; Yadav et al. 2016b) used approach presented in Phelps et al. (1996) as well USEPA 2002, where the distance between the landfill and collection area is one of the principal variables in deciding whether to use a transfer station or to haul the solid wastes directly from the collection area to the landfill site.

The main objective of this paper is to propose an approach that can be used for TS establishing in WMS system, in which location of landfill is known so only locations of TS are determined. Our idea was to consider the problem as two level location problem in which the first level facility (first layer) - landfill has predefined location. The second level (second layer) corresponds to TSs to be located. Because of discount factor related to the transport through TS, the modeling approach proposed here is similar to p-hub median problem formulation (Campbell 1996) but our approach differ because of only one (second layer) instead two hubs. Also, the model corresponds to multi-level facility location models which occupied research attention more than fourthy years (Geoffrion and Graves 1974). However, our approach differs since flows not necessarily must be routed through the TSs, and also, because our approach considered variable TSs capacities corresponding to waste flows being assigned.

Paper is organized as follows. In section 2 description of the problem as well as mathematical formulation of the problem are presented, while in section 3 numerical results of the case is presented. Finally, in section 4 some concluding remarks are given.

2. DESCRIPTION OF THE PROBLEM AND MATHEMATICAL FORMULATION

We considered a problem, in which logistics network consists of three layers (Figure 2). On first layer there are municipalities (i, i ∈ I) as waste generators who generate estimated quantities of waste Qi. These quantities are collected and transported by collection vehicles either to a TSs at site j, j ∈ J (second layer) or to landfill located at site l, l ∈ L (third layer). Also, from TSs, the collected waste can be hauled by transport truck to the landfill.

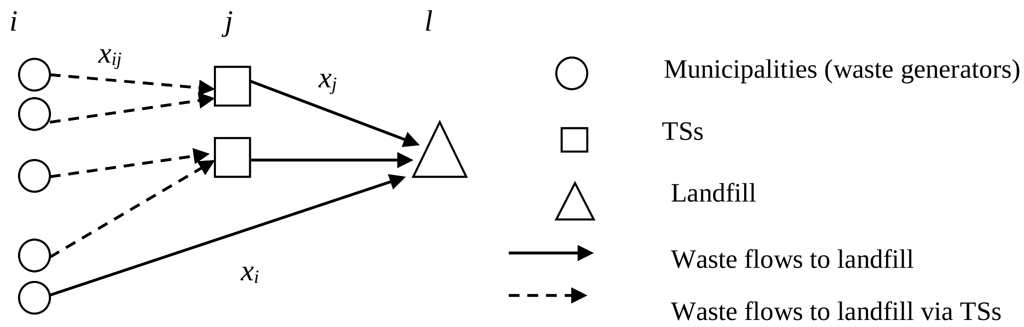


Figure 2. Waste flows in modeled problem

Following notation is used for mathematical formulation of the problem.

Sets:
i, $i \in I$ - locations of end users (municipalities)
j, $j \in J$ - potential locations of transfer stations
l, $l \in L$ - locations of landfills

Parameters:
Q_i - quantities of generated waste at site $i \in I$
Q_s - capacity of collection vehicles
Q_l - capacity of transfer vehicles
c_i - tour costs of collection vehicles with from municipalities i, $i \in I$ and location of landfill l, $l \in L$
c_j - tour costs of transfer vehicles from potential locations of Tss j, $j \in J$ and location of landfill l, $l \in L$
c_{ij} - tour costs of collection vehicles from municipalities i, $i \in I$ and potential locations of TSs j, $j \in J$
M - large enough positive number
s_k - *k-th* capacity of transfer station
f_{kj} - costs of opening transfer station with capacity s_k at location j

Variables:

$$y_{kj} = \begin{Bmatrix} 1, if\ transfer\ station\ of\ capacity\ \ s_k\ is\ located\ at\ site\ j \\ 0, otherwise \end{Bmatrix}$$

x_{ikj} - number of tours with collection vehicles from municipalities i, $i \in I$ and potential locations of TSs j, $j \in J$ with capacity s_k
x_i - number of tours with collection vehicles from municipalities i, $i \in I$ and location of landfill l, $l \in L$

x_{kj} - number of tours with transfer vehicles from potential locations of TSs j, $j \in J$ with capacity s_k to location of the landfill l, $l \in L$
z_{ikj} - waste flow quantities between municipalities i, $i \in I$ and potential locations of TSs j, $j \in J$ with capacity s_k
z_i - waste flow quantities between municipalities i, $i \in I$ and location of landfill l, $l \in L$
z_{kj} - waste flow quantities between potential locations of TSs j, $j \in J$ with capacity s_k and and location of landfill l, $l \in L$

Mathematical formulation of the problem:

$$\min \sum_j \sum_k f_{kj} y_{kj} + \sum_i \sum_j c_{ikj} x_{ikj} + \sum_i c_i x_i + \sum_j c_{kj} x_{kj} \quad (1)$$

$$Q_i \le Q_s \sum_j x_{ikj} + Q_s x_i, \forall i \quad (2)$$

$$\sum_i Q_i x_{ikj} - Q_l x_{kj} \le 0, \forall j, \text{k} \quad (3)$$

$$x_{ikj} \le M y_{kj}, \forall i, j, k \quad (4)$$

$$x_{kj} \le M y_{kj}, \forall j, k \quad (5)$$

$$Q_i = \sum_j z_{ikj} + z_i, \forall i \quad (6)$$

$$z_{ikj} \le Q_s x_{ikj,} \forall i, j, \text{k} \quad (7)$$

$$z_i \le Q_s x_i, \forall i, j, k \quad (8)$$

$$\sum_i z_{ikj} \le s_k y_{kj}, \forall j, k \quad (9)$$

$$\sum_k y_{kj} \le 1, \forall j \quad (10)$$

$$\sum_i z_{ikj} - z_{kj} = 0, \forall j, k \quad (11)$$

$$Q_l x_{kj} \ge z_{kj}, \forall j, k \quad (12)$$

$$x_{ikj}, x_i, x_{kj} \in N \cup \{0\}, \forall i, j, k \quad (13)$$

$$y_{kj} = 0 \vee 1, \forall j, k \quad (14)$$

$$z_{ikj}, z_i, z_{kj} \ge 0, \forall i, j, k \quad (15)$$

Objective function (1) minimizes total costs of transporting required quantities of waste either to TSs or directly to known location of landfill and costs of establishing TSs of certain capacity. Constraint (2) ensures that generated waste quantities are hauled with collection vehicles from end users either to locations of TSs or landfill location. Constraint (3) ensures that waste quantities delivered to TSs with collection vehicles are hauled to landfill with transfer vehicles. Constraints (4) and (5) ensure that no tours to TS exist if TS is not opened. Constraint (6) ensures that quantities generated at node i are delivered either to TSs or to landfill location. Constraints (7) and (8) ensures that generated quantities are hauled with collection vehicles if tours between end users and TSs as well landfill location exists, respectively. Constraints (9) are capacity constraints of TSs with *k-th* capacity, while constraints (10) ensures that on one potential location for TS only one TS with *k-th* capacity can be opened. Constraints (11) ensures that all quantities of waste sent from end users to TS with *k-th* capacity are quantities sent from TS with *k-th* to landfill location. Constraint (12) ensures that waste quantities from TSs with *k-th* capacities are hauled with transfer vehicles to landfill location. Constraints (13),(14) and (15) define nature of the variables included in model.

3. NUMERICAL RESULTS

We tested proposed model on real size problem for the region in south-western part of Republic of Serbia (Figure 3) for the period of one week. Regional Waste Management Plan (RWMP) "Duboko" covers a region which consists of nine local authorities: Arilje, Bajina Bašta, Čačak, Čajetina, Ivanjica, Kosjerić, Lučani, Požega and Užice with aproximatelly 375000 inhabitants. Location of landfill "Duboko" is known, and it is located near Užice (cca 13 km from Uzice). According to RWMP, the construction of eight TSs in the municipalities of Ivanjica, Lučani, Požega, Arilje, Kosjerić, Čajetina, Bajina Bašta and Čačak is planned. In our model we included Užice as well for potential location for TS, in order to see how logistics network for WMS would look like and how system will behave if we include location that is close to location of landfill. So, we have 9 potential locations for locating TSs.

For this region, there are total 335 waste generators who generates approximately 279 t of waste daily. We assumed that all generated quantities are collected and transported to designated locations, although in real WMS not all of the end users are covered with collection and transport services. But according to RWMP in the future it is planed that all end users in this region will be covered with this services.

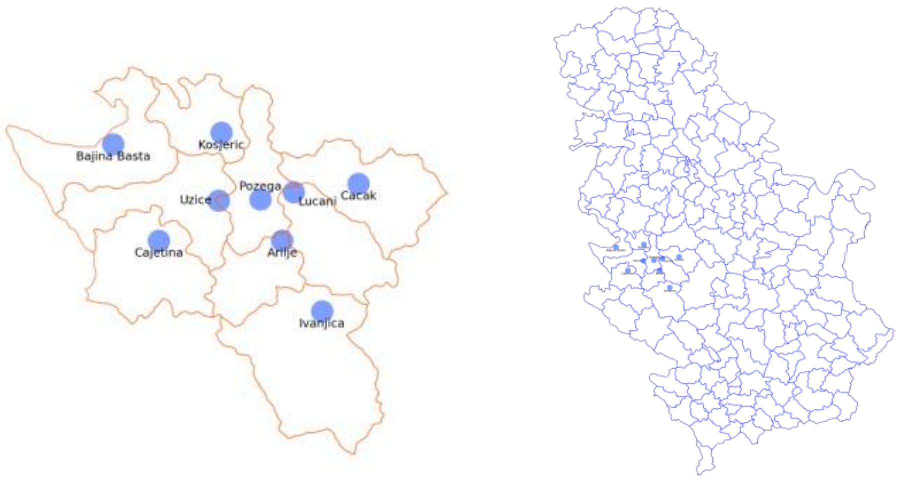


Figure 3. Territory of Regional Waste Management Plan "Duboko"

Unit transportation costs are calculated as follows:

$c_i(€) = d_i * 0.6 * 1.3$

$c_j(€) = d_j * 1.1 * 1.3$

$c_{ij}(€) = d_{ij} * 0.6 * 1.3$

where:

dj-distances between potential locations of TSs $j \in J$ and location of landfill $l \in L$

di-distances between municipalities $i \in I$ and location of landfill $l \in L$

dij-distances between municipalities i, $i \in I$ and potential locations of TSs j,$j \in J$

and constant 1.3 represent costs of empty tours of vehicle.

Fixed costs of opening TSs differs and they are dependent of the capacity of TSs. According to Regional Waste Management Plan (RWMP) "Duboko costs of establishing TS at each location are cca 200€ per day. So, in our model we used following capacities as well fixed cost of opening TSs, respectively for the period of one week [250t,700t,2000t] and [500€,1400€,2000€].

Large positive number M is calculated as ∑Qi/Qs. Capacities of collection and transfer vehicles are 5t and 21t, respectively. Problem solving procedure was implemented using Python 2.7 programming language and solved by Cplex 12.6 software (IBM 2012). Numerical results are presented in the Table 1.

Table 1. Numerical results

Objective function value	16108.91				
Opened locations	Landfill	Požega (capacity 1)	Čačak (capacity 1)	Ivanjica (capacity 1)	Bajina Bašta (capacity 1)
x_i	314				
x_{ij}		102	72	60	45
x_j		18	14	7	5

Also, we tested our model for different values of fixed costs of establishing TSS, which we varied in range +/- 10% and +/-20%, to see how and if results change with different input parameters. These results are presented in Table 2.

Table 2. Numerical results of sensitivity analysis

Objective function value with 10% smaller fixed costs					15909.79
Opened locations	Landfill	Požega (capacity 1)	Čačak (capacity 1)	Ivanjica (capacity 1)	Bajina Bašta (capacity 1)
x_i	314				
x_{ij}		103	72	59	45
x_j		18	14	7	5
Objective function value with 20% smaller fixed costs					15708.91
Opened locations	Landfill	Požega (capacity 1)	Čačak (capacity 1)	Ivanjica (capacity 1)	Bajina Bašta (capacity 1)
x_i	314				
x_{ij}		102	72	60	45
x_j		18	14	7	5
Objective function value with 10% larger fixed costs					16308.91
Opened locations	Landfill	Požega (capacity 1)	Čačak (capacity 1)	Ivanjica (capacity 1)	Bajina Bašta (capacity 1)
x_i	314				
x_{ij}		102	72	60	45
x_j		18	14	7	5
Objective function value with 20% larger fixed costs					16497.84
Opened locations	Landfill	Požega (capacity 1)	Čačak (capacity 1)	Ivanjica (capacity 1)	Bajina Bašta (capacity 1)
x_i	359				
x_{ij}		102	58	60	/
x_j		18	14	7	/

As can be seen from the results, based on input parameters obtained from Regional Waste Management Plan, our solution approach gives four TSs as needed for this region of Serbia. From the results of sensitivity analysis reported in Table 2, it can be seen that different number of TSs are opened in case when fixed costs of establishing TSs are 20% higher than the original input parameters. It can be seen also, that only TSs with smallest capacity are opened and that the number of tours with transfer vehicles differs in all other cases except in the case of 20% higher fixed costs, when they are the same. Number of tours with collection vehicles differs as a consenqence of different allocation of end users either to TS location or landfill.

4. CONCLUSION

In recent years, managing of SWM systems became very important due to quantities of generated waste and environmental concerns and consequences when solid waste treated inadequately. This paper present a possible approach for locating transfer stations within waste management systems, where locations of waste generators and landfill location are known.

Proposed model determines optimal locations of transfer stations as well number of tours needed to haul generated waste either to transfer station locations or landfill locations with collection and transfer vehicles, respectively. This approach should be understood as the beginning of the research in this area, because not all realistic factors and constraints are included. Possible extensions of the proposed model could include stochastic nature of waste quantities, routes of vehicles, etc.

Acknowledgements: This paper was supported by the Ministry of Science, Education and Technological Development of Republic of Serbia, through the project TR 36006.

REFERENCES

Campbell J.F, (1996). Hub Location and the p-Hub Median Problem. *Operations Research*, 44 , 923-935.

Eiselt, H.A. & Marianov, V., (2015). Location modeling for municipal solid waste facilities. *Computers & Operations Research,* 62, 305–315.

Eiselt, H.A. (2006). Locating Landfills And Transfer Stations In Alberta. *INFOR: Information Systems and Operational Research*, 44(4), 284-298.

Eiselt, H.A. (2007). Locating Landfills-Optimization vs.reality. *European Journal of Operational Research,* 179, 1040–1049.

Geoffrion, A.M., Graves, G.W., (1974). Multicommodity distribution system design by Benders decomposition. *Management Science,* 20, 822–844.

Yadav,V. V., Karmakar, S., Dikshit, A.K., &Vanjari, S. 2016a). A feasibility study for the locations of waste transfer stations in urban centers: a case study on the city of Nashik, India. *Journal of Cleaner Production,* 126, 191-205.

Yadav, V., Karmakar, S., Dikshit, A.K., &Vanjari, S. (2016b). Transfer stations siting in india: a feasibility demonstration. *Waste Management*, 47, I–IV.

Phelps, H.O., Heinke, G.W., Jonker, J.P., Ouano, E.A.R. & Vandecasteele, C. (1996). *Management of Solid Wastes*. Technical Report. UNESCO: Series of Learning Materials in Engineering Sciences, Paris. http://unesdoc.unesco.org/images/0021/002187/218722eo.pdf

Regional Waste Management Plan (RWMP) "Duboko", http://www.sepa.gov.rs/download/UpravOtpad/RPUO_Duboko.pdf

USEPA (2002). United States Environmental Protection Agency (EPA*). Waste Transfer Stations: A Manual for Decision-Making*. https://www.epa.gov/sites/production/files/2016-03/documents/r02002.pdf

B11
Mathematical Modeling & Decision Making

SYSTEM MAINTENANCE DECISION MAKING BASED ON REPAIR RATE STATISTICS

NATAŠA KONTREC[1], STEFAN PANIĆ[2], MILENA PETROVIC[3]

[1] Faculty of Sciences and Mathematics, University of Pristina in Kosovska Mitrovica, natasa.kontrec@pr.ac.rs
[2] Faculty of Sciences and Mathematics, University of Pristina in Kosovska Mitrovica, stefan.panic@pr.ac.rs
[3] Faculty of Sciences and Mathematics, University of Pristina in Kosovska Mitrovica, milena.petrovic@pr.ac.rs

Abstract: *This paper presents a novel approach for estimation of repair rate in an observed system comprised of two or more components. The presented approach is based on calculation of probability density function of maximal and minimal system's repair time by observing the repair rates of its components. Based on the obtained information it can be concluded in which time interval the repair or replacement should be completed in order to achieve the desired level of availability. The model can be further used for planning of maintenance activities, inventory, servicing capacities and dynamic forecast of system characteristics.*

Keywords: *Repairable system, Repair rate, Availability, Maintenance*

1. INTRODUCTION

In maintenance theory, two types of systems have been considered – non-repairable and repairable systems. When studying a non repairable system we usually observe how failure times are distributed. There are a lot of interesting articles in this specific area of research, such as (Chung 1989, Liu and Kapur 2007, Kontrec et al. 2015 and Li et al. 2016). In this paper we are considering only repairable systems. According to (Ascher and Feingold, 1984) a repairable system is a system which, after failing to perform one or more of its functions satisfactorily, can be restored to fully satisfactory performance by any method, other than replacement of the entire system. That means that after a failure occurs, the repair or replacement of components will be conducted and the system will continue to perform its function.

We only take into consideration perfect repairs, which means that after repairs the related system behaves as brand new, i.e. perfect repair action has been carried out. Actually, the system alternates between two states, operational states and failure state. Let's denote the successive length of time that system is in operational state as $T = (T_1, T_2, ..., T_n)$, and the successive lengths of time that the system is in failure state as $R = (R_1, R_2, ..., R_n)$. Starting state of the system is operational, and T and R are completely unknown in a sense that we do not assume any functional form for their distribution functions. Renewal process described under these conditions is an alternating renewal process. Several important applications of alternating renewal processes have been described in (Dickey 1991, Mortensen 1990, Di Crescenso 2001, Chen & Yuan 2003, Kontrec et al 2018).

2. MATHEMATICAL METHOD FOR REPAIR RATE ANALYSIS

Kontrec et al. (2018) proposed a model for the determination of the annual repair rate for critical aircraft components in order to achieve their desired availability. The availability is defined as a probability that a system will perform its function in a period of time (Barlow & Proschan 1965). The repair rate has been observed as stochastic process and the following assumptions have been made:

- The repairable system is modeled with alternating renewal process.
- Mean time to failure (MTTF) has Rayleigh's distribution.
- After repairs, the system returns to its original state, i.e. performs as brand new.

Further, the probability density function of failure time with Rayleigh distribution is: $p(t) = {}^{2t}\!/\!_{x} \exp\left({}^{-t^2}\!/\!_{x}\right)$, where the distribution parameter x is determined by relation $E(t^2) = x$, so the MTTF can be calculated as:

$$MTTF = \int_0^\infty tp(t)\,dt = \int_0^\infty {2t^2}/{x} \exp\left(-{t^2}/{x}\right)dt \tag{1}$$

Also, the renewal theorem (Ross, 2013) has been used to calculate the availability A of the system as:

$$A = \frac{E(T)}{E(T)+E(R)} \tag{2}$$

which can further be reduced to the well known formula for availability

$$A = \frac{MTTF}{MTTF + MTTR}. \tag{3}$$

MTTR is mean time to repair, i.e. expected time needed to repair a failed component. In the afore mentioned paper (Kontrec et al. 2018), the authors provided the PDF expression of repair rate in a case when the MTTF is known and the desired level of availability is set as:

$$p(\mu) = {8A^2}/{\left(1-A^2\right)\mu^3\pi x_0} \exp\exp\left({-4A^2}/{\left(1-A^2\right)\mu^2\pi x_0}\right), \tag{4}$$

where x is a random variable of Rayleigh's distribution, $x_0 = E(x)$ is the expected value of x and μ is repair rate of the observed component. Eq. (4) presents the exact mathematical characterization of the MTTR random process and exact modeling of repair rate process can be obtained by generating exact repair rate sample values for corresponding value of availability and the MTTF. In such way, simulation of repair rate process through generating its samples could serve to a dynamical prediction of the system performances.

3. DETERMINATION OF REPAIR RATE FOR SYSTEM COMPRISED OF TWO OR MORE SUBSYSTEMS OR COMPONENTS

The above presented Eq. (4) is the PDF of the repair rate for a single component. In this paper, the main goal is to determine the maximum and minimum repair rate of the system comprised of two or more subsystems or components. If $p_1(\mu)$ is the PDF function of the first component presented with Eq. (5):

$$p_1(\mu) = {8A_1^2}/{\left(1-A_1^2\right)\mu^3\pi x_{0_1}} \exp\left({-4A_1^2}/{\left(1-A_1^2\right)\mu^2\pi x_{0_1}}\right), \tag{5}$$

where A_1 is set level of availability of the first part and $x_{0_1} = E(x)$, i.e. the mathematical expectation of Rayleigh distributed parameter for that component. The cumulative density function (CDF) of this component is:

$$F_1(\mu) = 1 - \exp\left({-4A_1^2}/{\left(1-A_1^2\right)\mu^2\pi x_{0_1}}\right). \tag{6}$$

Using the same equation we can determine the PDF of the second part $p_2(\mu)$ with availability A_2 and $x_{0_2} = E(x)$ as in (7):

$$p_2(\mu) = {8A_2^2}/{\left(1-A_2^2\right)\mu^3\pi x_{0_2}} \exp\left({-4A_2^2}/{\left(1-A_2^2\right)\mu^2\pi x_{0_2}}\right) \tag{7}$$

and the CDF as in Eq. (8):

$$F_2(\mu) = 1 - \exp\left(\frac{-4A_2^2}{\left(1 - A_2^2\right)\mu^2 \pi x_{0_2}} \right). \tag{8}$$

For a system composed of two parts we can calculate maximum repair rate as $\mu = \max(\mu_1, \mu_2)$. In that case the PDF is:

$$p(\mu) = p_1(\mu)F_2(\mu) + p_2(\mu)F_1(\mu) \tag{9}$$

while the CDF is:

$$F(\mu) = F_1(\mu)F_2(\mu). \tag{10}$$

Further, when a system is comprised of *n* parts than the repair rate can be calculated as $\mu = \max(\mu_1, \mu_2, \ldots, \mu_n)$. The general form of the repair rate's PDF is then:

$$p(\mu) = \sum_{\substack{i=1 \\ i \neq j}}^{n} p_i(\mu) \prod_{j=1}^{n} F_j(\mu) \tag{11}$$

and the general form of the CDF is:

$$F(\mu) = \prod_{i=1}^{n} F_i(\mu). \tag{12}$$

Similarly, we can calculate minimal repair rate as $\mu = \min(\mu_1, \mu_2)$, so the PDF is:

$$p(\mu) = \left(1 - p_1(\mu)\right)\left(1 - F_2(\mu)\right) + \left(1 - p_2(\mu)\right)\left(1 - F_1(\mu)\right) \tag{13}$$

while the CDF is:

$$F(\mu) = 1 - \left(1 - F_1(\mu)\right)\left(1 - F_2(\mu)\right) \tag{14}$$

The general forms of the PDF and CDF equations, when system is comprised of *n* parts, and the repair rate is $\mu = \min(\mu_1, \mu_2, \ldots, \mu_n)$ are:

$$p(\mu) = \sum_{\substack{i=1 \\ i \neq j}}^{n} \left(1 - p_i(\mu)\right) \prod_{j=1}^{n} \left(1 - F_j(\mu)\right), \tag{15}$$

and the general form of the CDF for this system is:

$$F(\mu) = 1 - \prod_{i=1}^{n} \left(1 - F_i(\mu)\right). \tag{16}$$

4. NUMERICAL RESULTS

The model presented in the previous section will be verified by data from paper (Kang 2005, Mirzahosseinian & Piplani 2011). The observed system is an unmanned aerial vehicle (UAV). The concept of unmanned aerial vehicle (UAV) is not new but it has not been utilized in civilian sector due to the insufficient level of reliability of current solutions that leads to high probability of failure occurrence (Krawczyk 2013).

The UAV is comprised of three critical components: aircraft engine, propeller and avionics. According to known data such as:

- Each aircraft has 120 flight hours per month, i.e. 1440 (120*12) flight hours per year.
- Mean time between failures (MTBF is 750 flight hours for the aircraft engine, 500 for the propeller and 1000 for avionics).

Based on that it is possible to determine the MTTF as follows:

- for the aircraft engine $MTTF_e = 750/1400$
- for the aircraft propeller $MTTF_P = 500/1400$
- for the avionics $MTTF_e = 1000/1400$

According to the model presented in the previous section, a numerical analysis was conducted with the goal to calculate the annual expected time for maximum and minimum repair rate of the UAV system comprised of three components in order to acquire availability of A=0.80, A=0.85, A=0.9, A=0.95 by emphasizing the stochastic nature of this process. A similar analysis can also be conducted for other values of the parameter A.

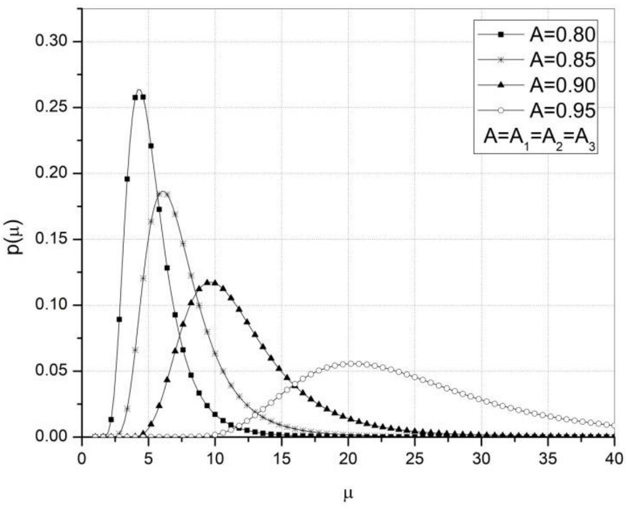


Fig 1. PDF of repair rate for $\mu = \max(\mu_1, \mu_2, \mu_3)$.

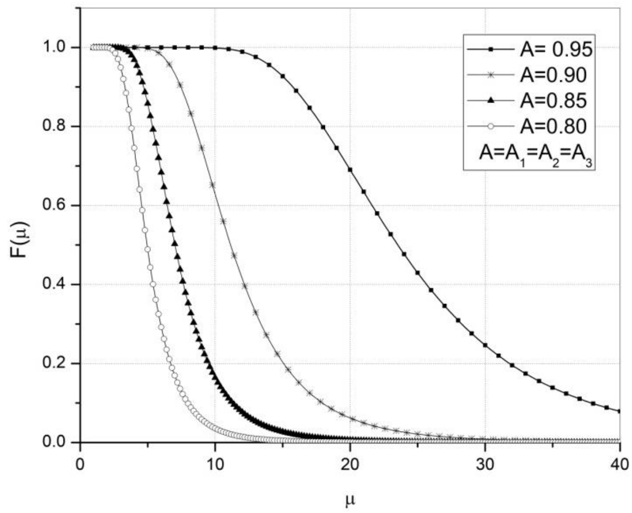


Fig 2. CDF of repair rate for $\mu = \max(\mu_1, \mu_2, \mu_3)$.

Fig 1 and Fig 2 represent the PDF and CDF, retrospectively, of the UVA's repair rate depending on time and the repair rate is calculated as maximum of its components' repair rate. Desired level of availability is set on 80%, 85%, 90% and 95% .

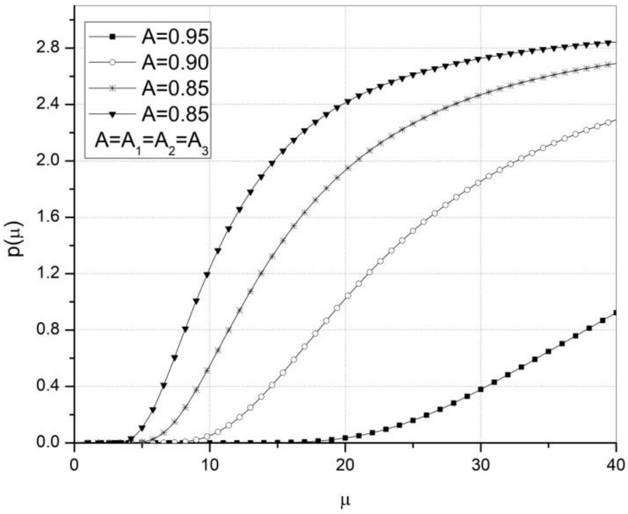


Fig 3. CDF of repair rate for $\mu = \min(\mu_1, \mu_2, \mu_3)$

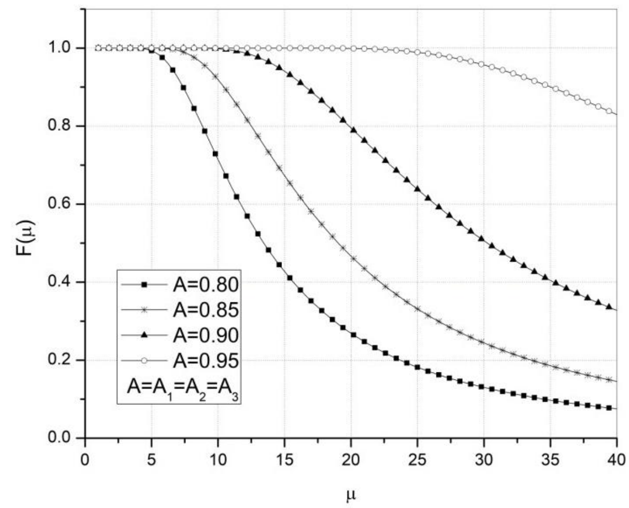


Fig 4. CDF of repair rate for $\mu = \min(\mu_1, \mu_2, \mu_3)$

Fig 3 and Fig 4 represent the PDF and CDF, retrospectively, of the UVA's repair rate depending on time and the repair rate is calculated as minimum of its components' repair rate. The desired level of availability is set on 80%, 85%, 90% and 95% as in the previous example.

The presented figures show probability that the repairs conducted in certain time frame will provide the desired level of system availability. That information can be useful for planning of system's maintenance activities, number of service stations, spare parts and manpower required for maintenance.

5. CONCLUSION

The previous researches showed that reliability and repair rate are crucial for any system's availability. In this paper we extended the previous research (Kontrec et al. 2018) where a novel model for determination of repair rate's characteristics was presented along with a precise mathematical characterization of repair as stochastic process. But, the proposed model was referring to a single subsystem or component of the system and that raised the question how to determine the repair rate's PDF and CDF for the system comprised of two or more subsystems/components. This paper provided an answer to that question. Actually, we calculated probability density function of maximum and minimum repair rate of the system by observing repair rates of its components. In Numerical section, the proposed model was applied to the UVA system comprised of three components and graphical presentation of the PDF and CDF of maximum and minimum repair rate has been provided. Based on this information we can conclude in which time interval maintenance action should be successfully completed in order to achieve the desired level of availability. Even though we set availability on certain levels, the numerical analysis can be repeated with different values of availability. This model can be applied in the same manner to other repairable systems with the alternating renewal process. The obtained results can be used in planning of maintenance activities, inventory, service systems and number of required employees, in the process of system maintenance.

REFERENCES

Ascher, H. and Feingold, H. (1984). *Repairable Systems Reliability. Modeling, Inference, Misconceptions and Their Causes.* Dekker, New York

Barlow R.E. & Proschan, F. (1975) *Statistical theory of reliability and life testing. Probability Models*. Holt, Rinehart and Winston, New York.

Chen C-T, Yuan J. (2003) Throughput analysis using alternating renewal process in balanced serial production lines with no interstage buffers. J Chin Insst Ind Eng 20(5):522–532

Chung W.K. (1989) Reliability analysis of repairable and non-repairable systems with common-cause failures, Microelectronics Reliability, 29(4): 545-547, https://doi.org/10.1016/0026-2714(89)90343-0

Di Crescenso A. (2001) On random motions with velocities alternating at Erlang-distributed random times. Adv Appl Probab 33(3):690–701

Dickey J.M. (1991) The renewal function for an alternating renewal process, which has a Weibull failure distribution and a constant repair time. Reliab Eng Syst Saf 31:321–343

Kang K., Doerr K.H., Boudreau M., Apte U. (2005) A decision support model for valuing proposed improvements in component reliability. Working paper, Naval Postgraduate School.

Kontrec N., Milovanović G., Panić S., & Milošević H., (2015) A Reliability-Based Approach to Nonrepairable Spare Part Forecasting in Aircraft Maintenance System, *Mathematical Problems in Engineering*. 2015:1-7.

Kontrec N., Panić S., Petrović M., Milošević H. (2018). A stochastic model for estimation of repair rate for system operating under performance based logistics. *Eksploatacja I Niezawodnosc – Maintenance and Reliability*. 20 (1): 68– 72.

Krawczyk M. (2013). Conditions for Unmanned Aircraft Reliability Determination. Eksploatacja i Niezawodnosc – Maintenance and Reliability; 15 (1): 31–36.

Li Y, Cui L., Yi H. (2016) Reliability of non-repairable systems with cyclic-mission switching and multimode failure components,Journal of Computational Science,Volume 17(1) , Pages 126-138, https://doi.org/10.1016/j.jocs.2016.05.014

Liu Y-W & Kapur K.C. (2007) Reliability measures for dynamic multistate nonrepairable systems and their applications to system performance evaluation, IIE Transactions, 38(6):511-520, DOI: 10.1080/074081705003412

Mirzahosseinian H., Piplani R.(2011) A study of repairable parts inventory system operating under performance-based contract. European Journal of Operational Research. 2011;214(2):256-61.

Mortensen R.E. (1990) Alternating renewal process models for electric power system loads. IEEE Trans Autom Control 35(11):1245–1249

Ross S.M. Applied probability models with optimization applications. Courier Corporation; 2013 Apr 15.

XIII Balkan Conference on Operational Research

DEAHP MODEL BASED ON THE SPECTRAL PROPERTIES OF PAIRWISE COMPARISON MATRIX

RADE LAZOVIĆ[1]

[1] University of Belgrade, Faculty of Organizational Sciences, lazovic@fon.bg.ac.rs

Abstract: *This paper proposes a new DEAHP model entitled as the DEA/RAR model which employs the Data Envelopment Analysis (DEA) methodology for generating local weights in the Analytic Hierarchy Process (AHP). The model is described by two parameters whose feasible regions are determined by a spectral radius of pairwise comparison matrix and two heuristics are developed to obtain desirable values for these parameters. The obtained values form the restrictive assurance region (RAR) of the variables of the DEA/RAR model which is a subset of the assurance region (AR) of the existing DEA/AR model, Wang et al. (2008). It is proved that the new model calculates true weights when it applies to perfectly consistent pairwise comparison matrices. Some advantages of the DEA/RAR model over DEA/AR model are presented by several illustrative examples. Besides, a numerical example shows that the model parameters can be determined so that the proposed model produces local weights that are very close to the ones obtained by the famous Saaty's method.*

Keywords: *DEAHP, DEA/AR, assurance region, restrictive assurance region, spectral radius*

1. INTRODUCTION

The analytic hierarchy process (AHP) is a well known method of multi-criteria decision-making (MCDM) where a problem is structured hierarchicaly as criteria and alternatives. The main goal of this method is to determine the priority weights of alternatives, i.e. global weights.

The AHP was developed by Saaty (1980) to solve complex problems involving multiple criteria. Since then it has been applied to numerous fields of human activity, Vaidya and Kumar (2006) such as planning, manufacturing, web service, medical and health care, logistics, optimization and so on.

The AHP is based on extracting the set of local weights $w_1,\dots,w_n$ from a comparison matrix. This set should be normalized to the sum of one,

$$\sum_{i=1}^{n} w_i = 1.$$

The elements of the comparison matrix are pairwise comparisons between criteria (alternatives), made by a decision maker. In the traditional AHP, pairwise comparisons are elicited as point judgments on a ratio scale and they are assembled in a pairwise comparison matrix of the form

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix},$$

where $a_{ii} = 1$ and $a_{ji} = 1/a_{ij}$ for $i \neq j$. Each of the elements a_{ij} represents the ratio of an estimation of preferences w_i and w_j. If the decision maker is consistent in his estimations, then all the entries a_{ij} have perfect values $a_{ij} = w_i/w_j$. In this case is $a_{ij} = a_{ik}a_{kj}$ for all $i, j, k = 1,\dots,n$ and A is said to be perfectly consistent; otherwise, it is said to be inconsistent.

The weight vector $w = (w_1,\dots,w_n)$ can be estimated as the right-Perron vector for the maximal eigenvalue λ_{max} in an eigenvalue problem

$$Aw = \lambda_{max} w.$$

Such a method for determining the weight vector of a pairwise comparison matrix is known as eigenvector method (EM). Although strongly recommended and preferred by Saaty, other methods for calculation of weights $w_1,\dots,w_n$ are also suggested, including logaritmic least-square technique, goal programming, and others.

The data envelopment analysis is a nonparametric technique developed by Charnes et al. (1978) to evaluate relative efficiency of similar decision making units (DMUs). It uses a variety of identical inputs to produce a variety of identical outputs.

Assume that there are n DMUs producing r outputs using m inputs. Let y_{ij} and x_{il} denote jth output and lth input of the DMUi, respectively. Charnes et al. (1978) proposed the model in which the efficiency of the DMUk was obtained as the maximum of the ratio of weighted outputs to weighted inputs,

$$\max h_k = \frac{\sum_{j=1}^{r} v_j y_{kj}}{\sum_{l=1}^{m} u_l x_{kl}}, \tag{1}$$

s.t.

$$\frac{\sum_{j=1}^{r} v_j y_{ij}}{\sum_{l=1}^{m} u_l x_{il}} \leq 1, \quad i = 1, \ldots, n \tag{2}$$

$$v_j, u_l \geq 0; \quad j = 1, \ldots, r, \ l = 1, \ldots, m.$$

From (1) and (2) it follows that $0 < h_k \leq 1$ for all $k = 1, \ldots, n$. If $h_k = 1$, then the DMUk is relatively efficient, otherwise the DMUk is inefficient.

Since the DEA has been intensively developing and applying in various fields, consequently a lot of DEA models have been developing. The overview of the models is given in details in the paper Cook and Seiford (2009) published on the occasion of 30 years of the DEA method. Here we present the input-oriented CCR model which is equivalent to the model (1), see Cooper et al. (2004):

$$\max h_k = \sum_{j=1}^{r} v_j y_{kj}$$

s.t.

$$\sum_{l=1}^{m} u_l x_{kl} = 1$$

$$\sum_{j=1}^{r} v_j y_{ij} - \sum_{l=1}^{m} u_l x_{il} \leq 0, \ i = 1, \ldots, n$$

$$v_j, u_l \geq 0; \quad j = 1, \ldots, r, \ l = 1, \ldots, m.$$

The paper is organized as follows. Section 2 briefly presents two models which use the DEA approach in generating local weights in the AHP method. Some results for spectral radius of non-negative matrices are also presented in this section. Section 3 presents the main result - a new DEAHP model for calculating local weights in the AHP. Section 4 contains comparative results for the proposed and two well known DEAHP models. Section 5 gives concluding remarks.

2. DEA CONCEPTS IN THE AHP

There are many similarities between the DEA and MCDM techniques. The literature shows that these formulations coincide if inputs and outputs are seen as performance evaluation criteria. Golany (1988) was among the first to attempt to integrate the DEA with MCDM. There have been several attempts to use the principle of the DEA in MCDM literature since then.

2.1. DEAHP model

This method for weight derivation and aggregation in the AHP was proposed by Ramanathan (2006). It views each decision criterion (alternative) in a pairwise comparison matrix as a decision making unit. The row elements of pairwise comparison matrix represent the outputs of the DMUs, and a fictitious input that has a constant value of 1 for all the DMUs is employed. For each DMUk, $k = 1, \ldots, n$ the following input-oriented CCR model was built:

$$\max w_k = \sum_{j=1}^{n} a_{kj} v_j \tag{3}$$

s.t.

$$\begin{aligned} & u_1 = 1 \\ & \sum_{j=1}^{n} a_{ij} v_j - u_1 \leq 0, \ i = 1, \ldots, n \\ & u_1, v_j \geq 0, \quad j = 1, \ldots, n. \end{aligned}$$

The optimum values for w_k, $k = 1, \ldots, n$ form weight vector $w = (w_1, \ldots, w_n)$ whose component w_k represents the DEA efficiency of the DMUk.

Ramanathan proved that the model (3) generates true weights if A is a perfectly consistent matrix.

2.2. DEA/AR model

Wang et al. (2008) show that the DEAHP model has some drawbacks that restrict applications of this model. The main drawback is that the DEAHP may generate counterintuitive local weights for some inconsistent matrices. For example, let

$$A = \begin{bmatrix} 1 & 4 & 1 & 1 & 3 & 4 \\ 1/4 & 1 & 7 & 3 & 1/5 & 1 \\ 1 & 1/7 & 1 & 1/5 & 1/5 & 1/6 \\ 1 & 1/3 & 5 & 1 & 1 & 1/3 \\ 1/3 & 5 & 5 & 1 & 1 & 3 \\ 1/4 & 1 & 6 & 3 & 1/3 & 1 \end{bmatrix} \tag{4}$$

be a pairwise comparison matrix. For this inconsistent matrix the DEAHP generates the weight vector $w = (1, \ldots, 1)$ for all the six criteria (alternatives). It is clear that this vector does not enable the ranking of the comparing elements.

Wang et al. also point out to another DEAHP drawback which refers to over-insensitivity of some pairwise comparison matrices. This drawback was illustrated by the matrix whose weight vector is immune to the changes of one of its entries.

To overcome these drawbacks of the DEAHP, they proposed a DEA model with assurance region (AR), so called AHP/AR model. For all $k = 1, \ldots, n$ it solves the following optimization problem:

$$\max w_k = \sum_{j=1}^{n} a_{kj} v_j \tag{5}$$

s.t.

$$\begin{aligned} & \sum_{j=1}^{n} a_{ij} v_j \leq 1, \quad i = 1, \ldots, n \\ & \frac{w_j}{\beta} \leq v_j \leq \frac{w_j}{n}, \quad j = 1, \ldots, n \qquad (6) \\ & v_j \geq 0, \quad j = 1, \ldots, n \end{aligned}$$

where

$$\beta = \min \left\{ \max_{1 \leq i \leq n} \frac{1}{r_i} \sum_{j=1}^{n} a_{ij} r_j, \max_{1 \leq j \leq n} \frac{1}{c_j} \sum_{i=1}^{n} a_{ij} c_i, \right\}, \tag{7}$$

while $r_1, \ldots, r_n$ and $c_1, \ldots, c_n$ are the row and column sums of the matrix A, respectively. Constraints (6) in this model determine the assurance region that restricts feasible region of the variables of the DEAHP model. These constraints are obtained by the characteristic equation $Aw = \lambda_{max} w$ using the following lemma result.

► **Lemma 1.** *Let $A = (a_{ij})_{n\times n}$ be a non-negative matrix with nonzero row sums $r_1,\dots,r_n$ and maximal eigenvalue λ_{max}. Then*

$$\min_{1\le i\le n}\frac{1}{r_i}\sum_{j=1}^{n}a_{ij}r_j \le \lambda_{max} \le \max_{1\le i\le n}\frac{1}{r_i}\sum_{j=1}^{n}a_{ij}r_j.$$

Similarly, if $c_1,\dots,c_n$ are nonzero column sums of the matrix A, then

$$\min_{1\le j\le n}\frac{1}{c_j}\sum_{i=1}^{n}a_{ij}c_i \le \lambda_{max} \le \max_{1\le j\le n}\frac{1}{c_j}\sum_{i=1}^{n}a_{ij}c_i.$$

Applying this model to the given matrix A they obtained weight vector $w = (1, 0.734, 0.260, 0.581, 0.942, 0.704)$.

The inequalities (6) for the variables v_i of the model (5) do not use the lower constraint for the maximal eigenvalue in this lemma. Instead, they use a well known inequality $\lambda_{max} \ge n$ that holds for all the positive matrices.

However, there are some more general results related to the maximal eigenvalue of positive matrices. To present them we need a definition of the spectral radius of a given matrix.

► **Definition 2.** Let $\mathcal{M}_n$ denotes set of all square matrices of order n. The spectral radius of $A \in \mathcal{M}_n$ is a non-negative real number

$$\rho(A) = \max\{|\lambda| : \lambda \in \sigma(A)\},$$

where $\sigma(A)$ is a set of all eigenvalues, i.e. spectrum of the matrix A.

The following two theorems, Horn and Johnson (1990) refer to the spectral radius of non-negative matrices.

► **Theorem 3.** *Let $A \in \mathcal{M}_n$ and suppose that $A \ge 0$. Then for any positive vector $x \in \mathbb{R}^n$ it holds*

$$\min_{1\le i\le n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j \le \rho(A) \le \max_{1\le i\le n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j. \tag{8}$$

► **Theorem 4.** *Let $A \in \mathcal{M}_n$ and suppose that $A \ge 0$. If A has a positive eigenvector, then*

$$\rho(A) = \max_{x>0}\Big(\min_{1\le i\le n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j\Big) = \min_{x>0}\Big(\max_{1\le i\le n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j\Big). \tag{9}$$

The next theorem unifies some of the well known results for positive matrices.

► **Theorem 5.** *If $A \in \mathcal{M}_n$ and $A > 0$, then the following statements are true:*

1. $\rho(A) > 0$;
2. $\rho(A) \in \sigma(A)$;
3. *there exists an eigenvector $x > 0$ such that $Ax = \rho(A)x$;*
4. $\rho(A)$ *is the only eigenvalue on the spectral circle of A.*

From Theorem 4 and Theorem 5 we conclude that for all positive matrices A it holds $\rho(A) = \lambda_{max}$. Consequently, relations (8) and (9) hold for the maximal eigenvalue of all the pairwise comparison matrices. These relations play a crucial role in constructing the following DEAHP model.

3. DEA/RAR MODEL

Now we can introduce a new DEA model which is generated by an arbitrary positive vector of the space $\mathbb{R}^n$. From the characteristic equation $Aw = \lambda_{max}w$ we have

$$\sum_{j=1}^{n}a_{ij}w_j = \lambda_{max}w_i \tag{10}$$

for all $i = 1,\dots,n$. Relation (10) can be rewritten in the form

$$w_i = \sum_{j=1}^{n}a_{ij}v_j$$

where $v_j = w_j/\lambda_{max}$. Let x be some positive vector of the space $\mathbb{R}^n$ and

$$\alpha = \min_{1\le i\le n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j, \quad \beta = \max_{1\le i\le n}\frac{1}{x_i}\sum_{j=1}^{n}a_{ij}x_j.$$

According to the Theorem 3 and (8) we have $\alpha \leq \lambda_{max} \leq \beta$, and from (10) it follows that

$$\alpha w_i \leq \sum_{j=1}^{n} a_{ij} w_j \leq \beta w_i,$$

or

$$\alpha v_i \leq \sum_{j=1}^{n} a_{ij} v_j \leq \beta v_i$$

for all $i = 1, \ldots, n$. Based on the previous considerations we will create the following DEA model. For all $k = 1, \ldots, n$ we solve optimization problem

$$\max z_k = \sum_{j=1}^{n} a_{kj} v_j \tag{11}$$

s. t.

$$\sum_{j=1}^{n} a_{ij} v_j \leq 1, \quad i = 1, \ldots, n \tag{12}$$

$$\sum_{j=1}^{n} a_{ij} v_j - \alpha v_i \geq 0, \quad i = 1, \ldots, n \tag{13}$$

$$\sum_{j=1}^{n} a_{ij} v_j - \beta v_i \leq 0, \quad i = 1, \ldots, n \tag{14}$$

$$v_j \geq 0, \quad j = 1, \ldots, n.$$

Note that inequalities (13) and (14) can be written more concisely in the matrix form as $(A - \alpha I)v \geq 0$ and $(A - \beta I)v \leq 0$.

▶ **Theorem 6.** *Let A be a perfectly consistent matrix with $a_{ij} = w_i/w_j$ and $w_{max} = \max_{1 \leq i \leq n} w_i$. If $(v_1, \ldots, v_n)$ is any point that belongs to the part of hyperplane*

$$\frac{v_1}{w_1} + \cdots + \frac{v_n}{w_n} = \frac{1}{w_{max}}$$

in the first octant, then $v = (v_1, \ldots, v_n)$ is the optimal vector of the model (11).

Proof. From (11) it follows that

$$z_k = w_k \sum_{j=1}^{n} \frac{v_j}{w_j}. \tag{15}$$

We may rewrite constraints (12) as

$$\sum_{j=1}^{n} \frac{v_j}{w_j} \leq \frac{1}{w_i}, \quad i = 1, \ldots, n$$

which is equivalent to

$$\sum_{j=1}^{n} \frac{v_j}{w_j} \leq \min_{1 \leq i \leq n} \frac{1}{w_i} = \frac{1}{w_{max}}. \tag{16}$$

Now from (15) and (16) we get

$$z_k \leq \frac{w_k}{w_{max}}$$

for any fixed k. If $(v_1, \ldots, v_n)$ is an arbitrary point that belongs to part of hyperplane $v_1/w_1 + \cdots + v_n/w_n = 1/w_{max}$ in the first octant, then from (15) it follows that

$$z_k = \frac{w_k}{w_{max}} = z_k^*, \tag{17}$$

where z_k^* denotes optimal values of the target function.

We still have to prove that vector $v = (v_1, \ldots, v_n)$ satisfies inequalities (13) and (14) of the model (11). Since $v = w/\lambda_{max}$ and

$$Av = \frac{1}{\lambda_{max}} Aw = \frac{1}{\lambda_{max}} \cdot \lambda_{max} w = \lambda_{max} v,$$

vector v is an eigenvector of the matrix A with a corresponding eigenvalue λ_{max}. As $v > 0$ and $\alpha \leq \lambda_{max}$, we have

$$(A - \alpha I)v \geq (A - \lambda_{max} I)v = Av - \lambda_{max} v = 0.$$

The second inequality can be proved in a similar way. ◀

▶ **Corollary 7.** *Normalized vector $z^* = (z_1^*, \ldots, z_n^*)$ gives weights that coincide with the true weights of the eigenvector method.*

Proof. From (17) it follows that

$$z_k^* \Big/ \sum_{i=1}^{n} z_i^* = w_k \Big/ \sum_{i=1}^{n} w_i.$$

◀

It is not difficult to verify that the proposed model becomes an AHP/AR model when $\alpha = n$ and parameter β is chosen by relation (7).

The model

$$\max z_k = \sum_{j=1}^{n} a_{kj} v_j \tag{18}$$

s. t.

$$\sum_{j=1}^{n} a_{ij} v_j \leq 1, \quad i = 1, \ldots, n$$

$$\sum_{j=1}^{n} a_{ij} v_j - a v_i \geq 0, \quad i = 1, \ldots, n \tag{19}$$

$$\sum_{j=1}^{n} a_{ij} v_j - b v_i \leq 0, \quad i = 1, \ldots, n \tag{20}$$

$$v_j \geq 0, \quad j = 1, \ldots, n$$

in which parameters a and b satisfy $n \leq a \leq \lambda_{max} \leq b < \beta$, we will refer to as the DEA model with a restrictive assurance region or the DEA/RAR model. If $z^* = (z_1^*, \ldots, z_n^*)$ is optimal vector of the DEA/RAR model, then the local weights of the AHP method are the components of the vector

$$w^* = z^* \Big/ \sum_{i=1}^{n} z_i^*.$$

Restrictive assurance region (RAR) is determined by constraints (19) and (20) of this model. These constraints are stronger compared to the corresponding constraints determining the assurance region of the AHP/AR model. Consequently, the feasible region for variables of the model (18) is a subset of the feasible region for variables of the AHP/AR model.

Based on equalities (9), we developed two heuristics for determination of the parameters a and b. First of them employs a procedure *minmax* for the minimization of the maximum of n functions

$$f_i(x) = \frac{1}{x_i} \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, \ldots, n \tag{21}$$

in the region $\mathbb{R}_+^n = \{(x_1, \ldots, x_n) : x_i > 0\}$. Supposing that $r = (r_1, \ldots, r_n)$ and $\beta = \max_{1 \leq i \leq n} \frac{1}{r_i} \sum_{j=1}^{n} a_{ij} r_j$, we will describe this procedure in several steps.

Step 1 Put $x = r$
Step 2 Choose k so that $f_k(x) = \max_{1 \le i \le n} f_i(x)$
Step 3 Put $b = f_k(x)$
Step 4 Choose $h > 0$ so that $x - h\nabla f_k(x) > 0$
Step 5 Put $x = x - h\nabla f_k(x)$
Step 6 Return to the Step 2.

To ensure that vector $x - h\nabla f_k(x)$ in Step 4 belongs to the region $\mathbb{R}^n_+$, it is necessary and sufficiently to choose h so that

$$0 < h < \min\left\{\frac{x_k x_1}{a_{k1}}, \dots, \frac{x_k x_{k-1}}{a_{k,k-1}}, \frac{x_k x_{k+1}}{a_{k,k+1}}, \dots, \frac{x_k x_n}{a_{kn}}\right\} \tag{22}$$

holds for the current k.

Parameter b is being obtained in an iterative process by the consecutive use of procedure *minmax*. Since $r = (r_1, \dots, r_n)$ is an initial value for x, it is clear that $b = \beta$ in the first iteration. Choosing h small enough and which satisfies the condition (22), numerical experiments show that $b < \beta$ occurs in the next iterations.

The second heuristic for determining parameter a uses procedure *maxmin* for the maximization of the minimum of n functions (21). This procedure can be described similarly to the previous one. If $\tilde{a}$ is a value for a obtained by the second heuristic, we choose $a = \max\{n, \tilde{a}\}$.

By choosing parameters a and b according to these heuristics, it is clear that the model (18) becomes the DEA/RAR model if parameter b satisfies condition $b < \beta$.

4. COMPARATIVE RESULTS

We have made a software implementation of the described heuristics in MATLAB environment. Table 1: shows obtained values for the parameters a and b for pairwise comparison matrix (4) in 500 iterations. The current values of a and b are listed after each $100th$ iteration, which gives an insight into this iterative process. The obtained values $a = 8.008$ and $b = 8.1497$ in the last iteration are very close and appropriate for the DEA/RAR model.

Table 1: Iterative process for parameters a and b

n	1	100	200	300	400	500
a	6.0696	7.3125	7.6982	7.8164	8.0827	8.0080
b	11.9650	8.4377	8.1651	8.1800	8.1761	8.1497

Table 2: shows the solution of the AHP/RAR model with $a = 8.008$ and $b = 8.1497$.

Table 2: Solution of the DEA/RAR model

k	v_1	v_2	v_3	v_4	v_5	v_6	z_k^*	w_k^*
1	0.122	0.062	0.025	0.054	0.101	0.062	1.000	0.287
2	0.122	0.063	0.025	0.054	0.100	0.061	0.514	0.148
3	0.127	0.063	0.025	0.054	0.100	0.061	0.201	0.058
4	0.127	0.063	0.025	0.054	0.100	0.061	0.445	0.128
5	0.122	0.063	0.025	0.054	0.100	0.061	0.820	0.235
6	0.122	0.063	0.024	0.057	0.100	0.062	0.503	0.503

Table 3: gives comparative results for DEA/AR, DEA/RAR and EM models.

Table 3: Comparative results for the three methods

Model	w_1^*	w_2^*	w_3^*	w_4^*	w_5^*	w_6^*
DEA/AR	0.237	0.174	0.062	0.138	0.223	0.166
DEA/RAR	0.287	0.148	0.058	0.128	0.235	0.144
EM	0.289	0.147	0.057	0.128	0.235	0.144

We can see that all of the three methods rank criteria (alternatives) in the order $1 \succ 5 \succ 2 \succ 6 \succ 4 \succ 3$. But precise numerical values of the local weights can be of a great importance. Moreover, in some cases the decision maker is certain about the ranking order of comparison elements but uncertain about the precise numerical values of his judgements, Sevkli et al. (2007). We ca also see that, compared to the DEA/AR method, the weight vector of the DEA/RAR method is considerably closer to the weight vector of the EM method. In fact, we have

$$\|w_{DEA/AR} - w_{EM}\|_1 = 0.128, \quad \|w_{DEA/RAR} - w_{EM}\|_1 = 0.004,$$

where $\|\cdot\|_1$ denotes absolute norm in the space $\mathbb{R}^6$.

Another illustrative example reflects the advantage of DEA/RAR method over the DEA/AR method. Let

$$A = \begin{bmatrix} 1 & 2 & 7/29 \\ 1/2 & 1 & 59/20 \\ 29/7 & 20/59 & 1 \end{bmatrix}$$

be a pairwise comparison matrix. Applying DEA/AR, with $\beta = 4.8055$, we obtain the ranking vector

$$z^* = (0.8034,\ 1,\ 1). \tag{23}$$

On the other hand, the DEA/RAR model, with $a = 4.2429$ and $b = 4.2495$, gives a ranking vector

$$z^* = (0.6896,\ 1,\ 0.9844). \tag{24}$$

It is clear that ranking criteria (alternatives) by the vector (23) is impossible. However, the vector (24) ranks comparing elements in the order $2 \succ 3 \succ 1$.

5. CONCLUSION

This paper has proposed a new DEAHP model for generating local weights in the AHP method. The main idea was to reduce the assurance region of the DEA/AR model to a restrictive assurance region (RAR) of a new DEA/RAR model. Contrary to the DEA/AR model, the restrictive assurance region is determined by two parameters which are independent of the row (column) sums of the pairwise comparison matrix. These parameters are obtained by two heuristics that allow a simple implementation. The first of the two numerical examples shows that a weight vector of the DEA/RAR model can be considerably closer to the weight vector of the EM method than corresponding vector of the DEA/AR model. The second shows that the DEA/RAR model can be applied to the comparison matrices for which the DEA/AR model does not work.

REFERENCES

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6):429–444.

Cook, W. D. and Seiford, L. M. (2009). Data envelopment analysis (dea)–thirty years on. *European journal of operational research*, 192(1):1–17.

Cooper, W. W., Seiford, L. M., and Zhu, J. (2004). Data envelopment analysis. In *Handbook on data envelopment analysis*, pages 1–39. Springer.

Golany, B. (1988). An interactive molp procedure for the extension of dea to effectiveness analysis. *Journal of the Operational Research Society*, 39(8):725–734.

Horn, R. A. and Johnson, C. R. (1990). *Matrix analysis*. Cambridge university press.

Ramanathan, R. (2006). Data envelopment analysis for weight derivation and aggregation in the analytic hierarchy process. *Computers & Operations Research*, 33(5):1289–1307.

Saaty, T. L. (1980). Analytic heirarchy process. *Wiley StatsRef: Statistics Reference Online*.

Sevkli, M., Lenny Koh, S., Zaim, S., Demirbag, M., and Tatoglu, E. (2007). An application of data envelopment analytic hierarchy process for supplier selection: a case study of beko in turkey. *International Journal of Production Research*, 45(9):1973–2003.

Vaidya, O. S. and Kumar, S. (2006). Analytic hierarchy process: An overview of applications. *European Journal of operational research*, 169(1):1–29.

Wang, Y.-M., Chin, K.-S., and Poon, G. K. K. (2008). A data envelopment analysis method with assurance region for weight generation in the analytic hierarchy process. *Decision Support Systems*, 45(4):913–921.

B12
Mathematical Programming (Linear & Nonlinear Programming)

Global Optimization : A new covering method combined with a stochastic local search

Raouf Ziadi[1], Abdelatif Bencherif-Madani[2]

[1] Department of Mathematics, University Ferhat Abbas Setif 1, Setif 19000, Algeria, ziadi.raouf@gmail.com
[2] Department of Mathematics, University Ferhat Abbas Setif 1, Setif 19000, Algeria, lotfi_madani@yahoo.fr

Abstract: *In this paper, a method for solving global optimization problems where the objective function is only continuous (not necessarily differentiable or Lipschitzian) is presented. It is based on the generation of parametrized curves combined with the Evtushenko algorithm and a new stochastic local search. It is established that this method converges in a finite number of iterations to the global minimum within a prescribed accuracy* $\varepsilon > 0$. *Numerical experiments are performed on some typical test problems and the preliminary results show that the algorithm is promising.*

Keywords: *Global optimization, Stochastic local optimization, Reducing transformation, Evtushenko's algorithm, Hooke-Jeeves algorithm.*

1. INTRODUCTION

Global optimization problems are nowadays important in numerous applications. To mention just of few, these are applied to signal processing, telecommunication, finance and operation research, network and transportation, engineering design and control, hardware and software design, chemical engineering and so on (Pintér (2006); Ziadi et al.(2017)). The problems in these disciplines may be formulated as optimization problems that involve objective functions which are only continuous and do not possess strong mathematical properties (such as convexity, differentiability, Lipschitz continuity etc.)

In this paper we consider the following bound constrained global optimization problem of the form:

$$f^* = \min_{x \in X} f(x). \tag{P}$$

where the objective function $f(x) : \mathbb{R}^n \to \mathbb{R}$ is only continuous and $X = \prod_{i=1}^{n} [a_i, b_i] \subset \mathbb{R}^n$ with a_i, b_i are real numbers for $i = 1, \ldots, n$.

We suggest here an algorithm which converges in a finite number of evaluation points and less demanding in terms of properties of the objective function. This is of course a challenge and efforts have been made to reduce to the minimum this number of evaluations, see the comments at the end of the paper. Our method is based on the generation of curves that suitably scan the domain X and combined with two optimization techniques: a deterministic global optimization method and a stochastic local search one. The main component of the algorithm is a coupling of the Alienor reducing transformation technique (Cherruault(1999)) with the one-dimensional Evtushenko algorithm (Evtushenko (1985)). However, due to the fact that f is only continuous, thus apt to violent variations with a large number of local extrema, we add an ingredient to quicken our algorithm: a new Stochastic Local Search method (SLS). The latter is inspired by the Hooke and Jeeves deterministic local optimization method (Hooke (1961)). This allows to explore promising regions with a moderate generation of evaluation points. The proposed method will be called RTESLS (Reducing -Transformation-Evtushenko-Stochastic Local Search). To the best of our knowledge, the use of several different α-dense curves is new compared to the classical Alienor method. Recall that if a function g is evaluated at the points $x_1, x_2, \ldots, x_k$, we call the value $g_k^* = \min\{g(x_1), g(x_2), \ldots, g(x_k)\}$ a record and a point x_k^* where $g(x_k^*) = g_k^*$ a record point.

Our approach is essentially based on the following result.

► **Theorem 1.** *A real function f defined on a compact set X of $\mathbb{R}^n$ is continuous if and only if $\forall \varepsilon > 0$ there exists a constant $K > 0$ such that $\forall x, y \in X$*

$$|f(x) - f(y)| \le K \|x - y\| + \varepsilon. \tag{1}$$

For the proof one can see (Rahal and Ziadi (2008)). The idea is that the space $L(X, \mathbb{R})$ of (real) Lipschitz functions on X, i.e. the set of functions f such that $\exists K > 0$ for which $|f(x) - f(y)| \le K \|x - y\|$ $\forall x, y \in X$, is dense in the space $C(X, \mathbb{R})$ of continuous functions on X endowed with the sup-norm $\|f - g\| = \sup_{x \in X} |f(x) - g(x)|$.

A drawback of the above theorem is that there are no specific means to recover the constant K. If K in (1) were explicitly known, with slight modification, all global optimization methods used when f is Lipschitzian carry over to this case. The idea in our approach is to construct a suitable strictly increasing sequence of real numbers $(K_j)_{j\in\mathbb{N}^*}$, tending to infinity with $K_1 > 0$, such that for $\varepsilon > 0$, $\exists\, j_0 \in \mathbb{N}^*$ with $|f(x) - f(y)| \leq K_{j_0}\|x-y\| + \frac{\varepsilon}{4}$, $\forall x,y \in X$. The constant K_{j_0} guaranties the convergence of the algorithm towards its global minimum in a finite number of iterations (within the prescribed accuracy $\varepsilon > 0$).

The RTESLS method performs a series of applications of the coupled Alienor-Evtushenko algorithm by changing successively the parameters $K_1, K_2, \ldots$, until obtaining the global minimum. At the step j, the Alienor method generates a simple curve (without double points) well spread over X and α_j-dense with parameter α_j depending on K_j. Then the Evtushenko algorithm, which has a variable steplength that also depends on the constant K_j and on the record obtained during the $(j-1)$ preceding steps, improves the value of the objective function on this curve. The stochastic local search comes in when a new record obtained by Alienor-Evtushenko is well lower than the preceding one. This allows further improvements of the values of f and consequently increasing the steplength in the combined algorithm.

It is clear that this theoretical approach can be inefficient if the constant

$$K_\varepsilon = \inf\left\{K > 0 \mid |f(x) - f(y)| \leq K\|x-y\| + \frac{\varepsilon}{4}, \forall x,y \in X\right\} \tag{2}$$

is too large, since in this case the algorithm will generate an excessive number of evaluation points of f before reaching the global minimum. But if f possesses a global minimiser in a region of X where the rate of variation of f is moderate, the global minimum can be obtained (by the SLS method) with an acceptable number of evaluation points even if K_ε takes a large value. Indeed, the algorithm approaches the solution well before the sequence $K_1, K_2, \ldots$ reaches K_ε. Note that this is anyway the case for Lipschitzian optimization with too big a constant where the Lipschitzian methods become inefficient, see (Zhigljavsky,A. A. et al (2008)), due to the important number of evaluation points. In this situation, we can call upon probabilistic methods, but the minimum can only be detected with a certain probability (Pintér(2006), Zhigljavsky and Zilinskas (2008)).

There are also certain metaheuristic methods which can be used if K_ε in relation (2) is too large, but this class of algorithms does not always guarantee the convergence to the global minimum in a finite number of iterations (Zhigljavsky and Zilinskas (2008)).

The RTESLS algorithm begins by generating curves which may not at the start be sufficiently dense in X, but given α each one passes uniformly through the regions of X. As the parameters $K_1, K_2, \ldots$ increase, the corresponding curves become progressively denser and therefore enhancing the possibility to pass by an attraction zone of a global minimiser. As soon as the Alienor-Evtushenko algorithm descends appreciably when recording a point belonging to a curve in an attraction zone, the local stochastic search comes in to try to reach the global minimiser.

One advantage in the mixed Alienor-Evtushenko algorithm is that it generates a sequence of points which is dense enough in the regions where f takes small values and much less dense elsewhere. This can be performed since the algorithm steplength is not fix but depends on the region containing the feasible point to evaluate: it increases with the value of the function at this point.

1.1. Notation

Standard notations concerning function spaces are adopted, for example $C^k(\mathbb{R}^n)$ is the space of k times continuously differentiable functions on $\mathbb{R}^n$. The Lebesgue measure is denoted by μ. The number α is strictly positive and assumed small compared to the dimensions of the rectangle $X = \prod_{i=1}^{n}[a_i, b_i]$ where $n \geq 2$. The Euclidean distance in $\mathbb{R}^n$ is noted $d(.,.)$. $\lceil . \rceil$ stands for "Integer part of".

2. PRELIMINARIES

Let use first recall some definitions and results that will be needed below.

► **Definition 2.** We say that a subset S of X $(X \subset \mathbb{R}^n)$ is $\alpha - dense$ in X, if for all $x \in X$, there exists a point $y \in S$ such that $d(x,y) \leq \alpha$, where d is the Euclidian distance in $\mathbb{R}^n$.

► **Definition 3.** A curve $h_\alpha : [0, \theta_{\max}] \to X, \theta_{\max} > 0$, is called $\alpha - dense$ in X, if for all $x \in X$, there exists $\theta \in [0, \theta_{\max}]$ such that $d(x, h_\alpha(\theta)) \leq \alpha$. Where $h_\alpha(\theta) = (h_1^{(\alpha)}(\theta), ..., h_n^{(\alpha)}(\theta))$.

2.1. The Alienor reducing transformation method

This method has been essentially developed for the global optimization but can be extended to many other problems such as approximation of functions with several variables, numerical solution of functional equations and to many other multidimensional problems (Cherruault (1999)). The basic idea consists in using a reducing transformation allowing to simplify the above problems into problems depending on a single variable.

Briefly, for solving the problem (P) we first associate an $\alpha - dense$ curve h_α in X defined by:

$$\begin{aligned} h_\alpha : [0, \theta_{\max}] &\rightarrow X \\ \theta &\rightarrow (h_1^{(\alpha)}(\theta), \ldots, h_1^{(\alpha)}(\theta)). \end{aligned}$$

The minimization problem (P) is then approximated by a problem depending on a single variable θ:

$$\min_{\theta \in [0, \theta_{\max}]} f_\alpha(\theta). \tag{P'}$$

Where

$$f_\alpha(\theta) = f(h_\alpha(\theta)) = f(h_1^{(\alpha)}(\theta), ..., h_n^{(\alpha)}(\theta)).$$

and $\theta_{\max}$ depends on the reducing transformation.

In other words, the objective function f depending on several variables is approximated by a function of a single variable and then apply simple optimization methods adapted to the case of single variable.

The following result was established by Ziadi and Cherruault (2000). They propose a method to generate a curve $\alpha - dense$.

► **Theorem 4.** *Let* $h_\alpha(\theta) = (h_1^{(\alpha)}(\theta), ..., h_n^{(\alpha)}(\theta)) : [0, \theta_{\max}] \rightarrow \prod_{i=1}^{n} [a_i, b_i]$ *be a continuous function; and let* $\xi_1, \xi_2, ..., \xi_{n-1}, \alpha$ *be strictly positive numbers such that:*

(a) h_n *is surjective;*

(b) for any $i = 1, 2, ..., n-1, h_i$ *reaches its bounds* a_i *and* b_i *in every closed interval of lenght* ξ_i;

(c) for any $i = 1, 2, ..., n-1$ *and for any interval* I *of* $[0, \theta_{\max}]$, *we have*

$$\mu(I) < \xi_i \Longrightarrow \mu\left(h_{i+1}^{(\alpha)}(I)\right) < \frac{\alpha}{\sqrt{n-1}}.$$

Then the curve defined by $h_\alpha(\theta) = (h_1^{(\alpha)}(\theta), ..., h_n^{(\alpha)}(\theta))$, *for* $\theta \in [0, \theta_{\max}]$, *is* $\alpha - dense$ *in* $\prod_{i=1}^{n} [a_i, b_i]$.

► **Example 5.** (Ziadi and Cherruault (2000)) This curve will be used in our numerical applications. We consider the function $h_\alpha(\theta)$ defined from $[0, \frac{\pi}{\xi_n}] \rightarrow \prod_{i=1}^{n} [a_i, b_i]$ by:

$$\begin{aligned} h_1^{(\alpha)}(\theta) &= \frac{b_1 + a_1}{2} - \frac{b_1 - a_1}{2} \cos(\xi_1 \theta), \\ h_2^{(\alpha)}(\theta) &= \frac{b_2 + a_2}{2} - \frac{b_2 - a_2}{2} \cos(\xi_2 \theta), \\ &\vdots \\ h_n^{(\alpha)}(\theta) &= \frac{b_n + a_n}{2} - \frac{b_n - a_n}{2} \cos(\xi_n \theta), \end{aligned}$$

Where $\xi_1, \xi_2, ..., \xi_n$ are given by :

$$\begin{aligned} \xi_1 &= (b_2 - a_2) \ldots (b_n - a_n), \\ \xi_2 &= \frac{\alpha}{\pi} (b_3 - a_3) \ldots (b_n - a_n), \\ \xi_3 &= \left(\frac{\alpha}{\pi}\right)^2 (b_4 - a_4) \ldots (b_n - a_n), \\ &\vdots \\ \xi_n &= \left(\frac{\alpha}{\pi}\right)^{n-1}. \end{aligned}$$

By Theorem (1) the curve $h_\alpha(\theta) = (h_1^{(\alpha)}(\theta), ..., h_n^{(\alpha)}(\theta))$ is $\alpha - dense$ in $\prod_{i=1}^{n}[a_i, b_i]$. On the other hand, this curve is of class $\mathcal{C}^\infty$. It is easy to see that the function h_α is Lipschitzian with constant

$$l_\alpha = \frac{1}{2}\left(\sum_{i=1}^{n}(a_i - b_i)^2 \xi_i^2\right)^{\frac{1}{2}},$$

The figures 1 and 2 represent a densification of the hyper-rectangle $[-1,2]^3$ by the support of the the given curve.

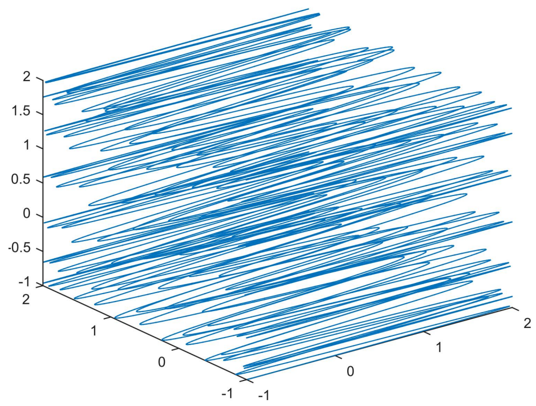


Figure 1 Support of the curve with $\alpha = 0.7$

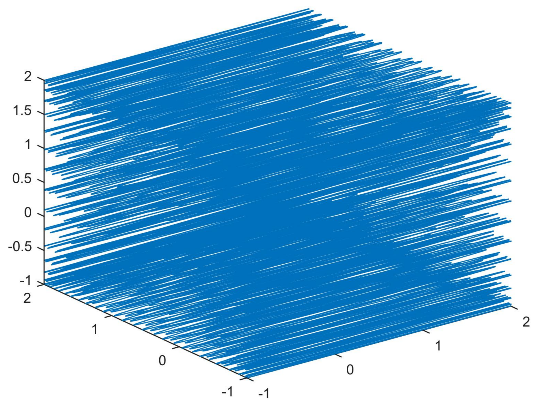


Figure 2 Support of the curve with $\alpha = 0.3$

2.2. The Evtushenko algorithm

It belongs to a class of global optimization algorithms (Evtushenko (1985)) said to be of sequential coverings and applies usually to Lipschitz functions. But as it is easy to see by Theorem 1, it can be applied with certain modifications to continuous functions. Let us first consider the main idea of these methods. Assume that we want to minimize a function g defined on a compact X of $\mathbb{R}^n$ with a desired accuracy $\varepsilon > 0$. Setting $X_k = \{x \in X \mid g_k^* - \varepsilon \leq g(x)\}$, see the introduction, it follows that $g_k^* - \varepsilon \leq \min_{x \in X_k} g(x)$. Therefore, the points of X_k can't improve the values of g and are not interesting for further search. We must go on only within the set $X \setminus X_k$. When for a certain $N \in \mathbb{N}$ we have $X \subset X_N$, the initial problem is solved because in this case we get $g_N^* - g^* \leq \varepsilon$. The value g_N^* can be taken as an approximation of the global minimum. Each covering method has a strategy for supplying the sequence $x_1, x_2, \ldots, x_N$ such that the sequence of subsets $X_1, X_2, \ldots, X_N$ increases, in the inclusion sense, to X. In the Evtushenko algorithm, the subsets X_k are rectangles of $\mathbb{R}^n$. When $X = [a,b] \subset \mathbb{R}$, a variant of the algorithm is as follows: $x_1 = a$ and $x_{k+1} = x_k + \left(g(x_k) - g_k^* + 2\varepsilon\right)/L$ where L is the Lipschitz constant of g. When $x_k > b$ the algorithm stops.

3. The Reducing Transformation -Evtushenko-Stochastic Local Search method

Let us first describe the new stochastic method to be incorporated in the main algorithm.

3.1. The Stochastic Local Search (SLS)

This is a modification of the deterministic local optimization algorithm of Hooke and Jeeves conceived for minimising a function of several variables without using derivatives (Hooke (1961)). It is also well adapted to the main algorithm of our approach. The idea is as follows: one evaluates f at a point generated according to the uniform law on the sphere centred at the last record point with radius a suitable steplength $\lambda > 0$. If this point improves f then it is the new record point; in this case one uses an acceleration step, usually referred to as pattern search. This step allows to seek improvements of f exploring the direction determined by the two last record points. If after a certain number of points generated on the sphere there is no improvement of f, one reduces the steplength λ and repeats the same procedure. The algorithm stops when λ reaches a certain prescribed accuracy $\rho > 0$.

The algorithm is as follows:

Step 1 (Initialization) Choose a scalar $\rho > 0$ to be used for terminating the algorithm and R as maximal number of random directions. Furthermore, choose an initial steplength $\lambda_{init} > \rho$, and an acceleration factor $\eta > 0$. Choose a starting point x_1; let $y = x_1, \lambda = \lambda_{init}$, and let $k = j = 1$.

Step 2 Generate a random direction V_j from the uniform distribution on the unit n-dimensional sphere. If $f(y + \lambda V_j) < f(y)$, the trial is termed a success, set $x_{k+1} = y + \lambda V_j$, and go to step 4. Otherwise, go to step 3.

Step 3 If $j < R$, increase j to $j+1$ and go back to step 2. Otherwise, go to step 5.

Step 4 If $f(x_{k+1} + \eta(x_{k+1} - x_k)) < f(x_{k+1})$ set $x_{k+2} = x_{k+1} + \eta(x_{k+1} - x_k)$, increase k to $k+1$ and repeat step 4. Otherwise, increase k to $k+1$, set $y = x_k, j = 1, \lambda = \lambda_{init}$ and go back to step 2.

Step 5 If $\lambda > \rho$, replace λ by $\lambda/2$, set $j = 1$ and go back to step 2. Otherwise stop: x_k is the prescribed solution.

During the search process, the points that appear may land outside X and this situation is treated as follows. For $y_j = \left(y_1^{(j)}, y_2^{(j)}, ..., y_n^{(j)}\right)$, set for $i = 1, \ldots, n$

$$y_i^{(j)} = \begin{cases} a_i & \text{if } y_i^{(j)} < a_i, \\ b_i & \text{if } y_i^{(j)} > b_i, \\ y_i^{(j)} & \text{otherwise.} \end{cases}$$

3.2. The RTESLS algorithm

As we mentioned earlier on, the principle behind our combined method is quite simple: the Alienor-Evtushenko explores the search space and when it finds a solution well lower than the preceding one, it is used as an initial point for the stochastic local search. In turn, the solution obtained by the latter is used again to quicken the mixed algorithm.

Algorithm RTESLS

Step 1 Initialization

(i) Choose $h_\alpha(\theta)$, a parametrized curve α-dense in X.

(ii) Choose the following components:
$\varepsilon > 0$ is the accuracy of the computation of the global minimum,
$\omega > 1$ is a multiplicative factor of the increasing sequence $(K_j)_{j\geq 1}$,
B_l is the least number of parametrized curves generated by the algorithm before testing the stopping criterion,
B_u is the maximal number of parametrized curves generated by the algorithm with $(B_u > B_l + 2)$.

(iii) Set $j = 1$ and choose $K_1 \geq 1$ as an initial value of the sequence $(K_j)_{j\geq 1}$.
Set $x_\varepsilon = b$, and $f_\varepsilon = f(b)$ as an initial value of the approximate global minimum (see Remark-(b) below).
Set $U = V = W = f(b)$ as initial values of the parameters linked to the stopping criterion.
Set $\alpha = \alpha_1 = \varepsilon/(4K_1)$ the densification parameter of the initial parametrized curve.

Step 2 (i) Set $k = 1$ and $\theta_1 = \varepsilon/(4L_jK_j)$ (where L_j is the Lipschitz constant of the α_j-dense curve $h_j(\theta) = (h_1^{(\alpha_j)}(\theta), h_2^{(\alpha_j)}(\theta), ..., h_n^{(\alpha_j)}(\theta))$.

(ii) Evaluate $f(h_j(\theta_k))$.
If $f(h_j(\theta_k)) + \frac{\varepsilon}{2} < f_\varepsilon$ go to step 5.
If $f(h_j(\theta_k)) < f_\varepsilon$, set $x_\varepsilon = h_j(\theta_k)$ and $f_\varepsilon = f(h_j(\theta_k))$.

(iii) Set $\theta_{k+1} = \theta_k + \frac{2\left(f(h_j(\theta_k)) - f_\varepsilon\right) + \varepsilon}{2L_jK_j}$.
If $\theta_{k+1} < T_j$, set $k = k+1$ and go back to step 2-(ii).

Step 3 (i) If $j \leq B_l$, go to step 3-(ii).
Put $U = V, V = W$ and $W = f_\varepsilon$.
If $U - W \leq \varepsilon$ go to step 4.
If $j \geq B_u$ go to step 4.

(ii) Set $K_{j+1} = \omega Kj$, $\alpha_{j+1} = \alpha_j/\omega, j = j+1$ and go back to step 2.

Step 4 Stop. f_ε is taken as an approximate global minimum and x_ε is a point at which the approximate value is reached.

Step 5 (Stochastic Local Search)
Apply the stochastic local search method, setting $x_1 = f(h_j(\theta_k))$ as an initial point and $\rho_j = \varepsilon/(4K_j)$ as a scalar to terminate the algorithm. Let x^* be the record point obtained; put $x_\varepsilon = x^*, f_\varepsilon = f(x^*)$ and return to Step 2-(iii).

► Remark. (a) From Theorem 1 we deduce that if a curve $h\colon [0,T] \to X$ is Lipschitzian with constant L, then for all $\varepsilon > 0$ the composition $f \circ h$ satisfies the relation

$$|f(h(\theta_1)) - f(h(\theta_2))| \leq K_\varepsilon L|\theta_1 - \theta_2| + \frac{\varepsilon}{4},$$

$\forall \theta_1, \theta_2 \in [0,T]$ (K_ε is the constant in the expression (2)).

(b) The curves $h_j\colon [0,T_j] \to X$ are such that $h_j(0) = a$, $h_j(T_j) = b$. At the step j, the algorithm starts the evaluations of f at a point $h_j(\theta_1)$ ($\theta_1 = \frac{\varepsilon}{4K_jL_j}$) which lies in the neighbourhood of a and ends the same step at a point $h_j(\theta_N)$ which lies in the neighbourhood of b (with $\theta_N < T < \theta_{N+1}$). If a global minimiser is in a neighbourhood of a, the evaluations of f at the points $h_j\big(\frac{\varepsilon}{4K_jL_j}\big)$ (for $j = 1, 2, \dots$) allows approaching the global minimum with the accuracy ε, as will be shown in Theorem 6. On the contrary, if a global minimiser is in the neighbourhood of b, the values $h_j(\theta_N)$ for $j = 1, 2, \dots$ do not allow to approximate, with the desired accuracy, the global minimum as soon as K_j is greater than K_ε: this is the reason why $f(b)$ is to be calculated. Moreover, at the step j the last evaluation of f is $h_j(\theta_N)$ with $\theta_N < T_j$; this strict inequality has been introduced especially to avoid the evaluation of f several times at b since $h_j(T_j) = b$, $\forall j \in \mathbb{N}^*$.

3.2.1 The stopping conditions of RTESLS

As will be seen in Theorem 6 below, the constant K_ε is fundamental for the RTESLS algorithm to reaches the global minimum in a finite number of evaluation points. However, since in most cases this constant is not known, the stopping conditions can't use it. Also we think it is insufficient to stop uniquely in function of the number of evaluations or iterations that must be accomplished by the algorithm without prior information about the solution. In our algorithm there is a stop criterion comprising two conditions for the solution to be accepted. The algorithm must first accomplish a large number of evaluations along the parametrized curves generated successively. When a threshold B_l (minimal condition) is reached, the algorithm calls upon the first stopping test if there are no notable improvements after the generation of a certain number of additional curves. If the constant K_ε is excessively large, the first test might not occur despite an important number of evaluations of f; in this case the algorithm appeals to the second test which stops the algorithm when the number of generated curves reaches a maximal number B_u.

3.2.2 The coordination between the Reducing Transformation -Evtushenko algorithm and the SLS method

The back and forth conditions between the main Reducing Transformation-Evtushenko algorithm and the local method are set so that if the SLS method finishes its search in an attraction zone of a local minimiser, there is a high probability it won't come back. Indeed, in an arbitrary step j, if $K_j < K_\varepsilon$ the SLS method can come back to the same region to affine the local search since at this step the local minimisers detected by SLS have been localized with a precision ρ_j which is not yet sufficient $\big(\rho_j = \frac{\varepsilon}{4K_j} > \frac{\varepsilon}{4K_\varepsilon}\big)$. However if $K_j \geq K_\varepsilon$, the possibility of return is slight; actually this possibility is uniquely linked to the probabilistic aspect of the SLS method (the local minimum is spotted only with a certain probability).

Let us look more closely at the RTESLS algorithm when $K_j \geq K_\varepsilon$. The main algorithm calls upon the stochastic local search when at the iteration k ($k \geq 2$), we have

$$f(h_j(\theta_k)) + \frac{\varepsilon}{2} < f_\varepsilon \tag{3}$$

where f_ε is the smallest value of f up to the iteration $k-1$. Then at the iteration k we have

$$\theta_k - \theta_{k-1} = \frac{2\big(f(h_j(\theta_{k-1})) - f_\varepsilon\big) + \varepsilon}{2L_jK_j},$$

and since $f_\varepsilon \leq f(h_j(\theta_{k-1}))$, it follows from the inequality (3) that

$$f(h_j(\theta_k)) < f(h_j(\theta_{k-1})).$$

Consequently

$$f(h_j(\theta_{k-1})) - f(h_j(\theta_k)) \leq L_jK_j(\theta_k - \theta_{k-1}) + \frac{\varepsilon}{4} = f(h_j(\theta_{k-1})) - f_\varepsilon + \frac{3\varepsilon}{4}$$

and

$$f_\varepsilon - \frac{3\varepsilon}{4} \leq f(h_j(\theta_k)).$$

We see that if f_ε is the record obtained up to the iteration $k-1$, then at the iteration k, $f(h_j(\theta_k))$ can't be less than $f_\varepsilon - \frac{3\varepsilon}{4}$ but can be less than $f_\varepsilon - \frac{\varepsilon}{2}$, which corresponds to condition (3).

Applying the local method with $x_1 = h_j(\theta_k)$ as an initial value, the search stops when the last randomly generated points are at a distance less than $\rho_j = \frac{\varepsilon}{4k_j}$ from the last record. This accuracy ρ_j allows the main algorithm in the subsequent iterations to once again call upon the local search uniquely when it reaches an attraction zone of a local minimum smaller than the local minima obtained so far. Indeed, if y_{ρ_j} is an approximation of a local minimiser y^* obtained by SLS before the iteration k with accuracy $\rho_j = \frac{\varepsilon}{4k_j}$, then

$$\|y_{\rho_j} - y^*\| \leq \frac{\varepsilon}{4K_j}$$

therefore

$$f(y_{\rho_j}) - f(y^*) \leq K_j\|y_{\rho_j} - y^*\| + \frac{\varepsilon}{4} \leq \frac{\varepsilon}{2}.$$

But, up to the iteration $k-1$ we have $f_\varepsilon \leq f(y_{\rho_j})$, hence if at the iteration k the condition (3) is satisfied, we shall have

$$f(h_j(\theta_k)) < f_\varepsilon - \frac{\varepsilon}{2} \leq f(y_{\rho_j}) - \frac{\varepsilon}{2} \leq f(y^*).$$

Since $f(h_j(\theta_k)) < f(y^*)$, the point $h_j(\theta_k)$ does not belong to the attraction zone of the local minimiser y^*. This permits to the stochastic method to go on searching for a new local minimiser of f smaller than the local minimisers obtained till the iteration k.

3.3. The convergence of the RTESLS algorithm

In global optimization, the most common efficiency criterion of a method is the time that is necessary for reaching the global minimum with a given accuracy. This time, in most cases, is linked to the number of the objective function evaluations. Concerning the RTESLS algorithm, the number of evaluation points depends on the length of the α-dense curve used in the combined algorithm. In general, for α fixed, the smaller the curve length is, the better the calculation time will be. We give here a result on the minimal number of evaluation points needed for solving the problem (P) with a given accuracy $\varepsilon > 0$, where the proof is similar to the one given in (Ziadi et al. (2001)).

► **Theorem 6.** *Assume that the RTESLS algorithm uses an α-dense curve satisfying the hypotheses of Theorem 4. Then, under the above notations, the combined algorithm converges in a finite number of evaluation points to the global minimum with accuracy less than or equal to ε, with a minimal number of evaluation points equal to*

$$N = \left\lceil \left(\log_2 \left(\frac{2TK_\varepsilon L_\varepsilon}{\varepsilon} + 1 \right) \right) \right\rceil + 1 \tag{4}$$

where L_ε is the Lipschitz constant of our α-dense curve with $\alpha = \frac{\varepsilon}{4K_\varepsilon}$.

4. Numerical experiments

In this section, we present some numerical results of RTESLS applied to non-smooth test problems found in the literature. The description of these problems is given in tables 1 and 2 below. The parameters used in RTESLS are given as follow: $\varepsilon = 0.1, \omega = 2, K_1 = 0.1, B_l = 2, B_u = 6, \lambda_{init} = 0.6, \eta = 1, R = 3^n - 1$ (n is the dimension of the problem).

Table 3 below displays, the number of function evaluations necessary to reach the stopping criterion using the parametrized curves mentioned in Example 1. The first column denotes the problem number ($Pb-num$), while the second column n stands for the dimension of each test problem. The third and the fourth columns

Table 1: Test functions

Test/benchmark functions	Search region	global minimum
Problem 1		
$f(x) = \max_{i=1,2,3} f_i(x) + \max_{i=4,5,6} f_i(x)$	$x \in [-10,10]^2$	2
Where		
$f_1(x) = x_1^4 + x_2^2,\ f_2(x) = (2-x_1)^2 + (2-x_2)^2,\ f_3(x) = 2\exp(-x_1 + x_2)$		
$f_4(x) = x_1^2 - 2x_1 + x_2^2 - 4x_2 + 4,\ f_5(x) = 2x_1^2 - 5x_1 + x_2^2 - 2x_2 + 4$		
$f_6(x) = x_1^2 + 2x_2^2 - 4x_2 + 1$		
Problem 2		
$f(x) = \max_{i=1,2,3} f_i(x)$	$x \in [-10,10]^n$	-3
Where $f_1(x) = 5x_1 + x_2, f_2(x) = -5x_1 + x_2, f_3(x) = x_1^2 + x_2^2 + 4x_2$		
Problem 3		
$f(x) = \left(\sum_{k=1}^{5} k\lvert\cos((k+1)x_1 + k)\rvert + 5\right)\left(\sum_{k=1}^{5} k\lvert\cos((k+1)x_2 + k)\rvert + 5\right)$	$x \in [-10,10]^2$	44.8872
Problem 4		
$f(x) = -20\exp\left(-0.2\sqrt{n^{-1}\sum_{i=1}^{n} \lvert x_i\rvert}\right) - \exp\left(n^{-1}\sum_{i=1}^{n}\cos(2\pi x_i)\right) + 20$	$x \in [-5,10]^n$	$f(x^*) = -2.7182$
Problem 5		
$f(x) = \sum_{i=1}^{n} \lvert x_i - 0.5\rvert$	$x \in [-5,5]^n$	0
Problem 6		
$f(x) = \sum_{i=1}^{n} \lvert x_i\rvert + \prod_{i=1}^{n} \lvert x_i\rvert$	$x \in [-5,5]^n$	0
Problem 7		
$f(x) = \frac{\pi}{n}\left(10\lvert\sin(\pi y_1)\rvert + \sum_{i=1}^{n-1} \lvert y_i - 1\rvert\left(1 + 10\lvert\sin(\pi y_{i+1})\rvert\right) + \lvert y_n - 1\rvert\right)$	$x \in [-5,5]^n$	0
Where		
$y_i = 1 + \frac{x_i - 1}{4}$		
Problem 8		
$f(x) = \frac{1}{400}\sum_{i=1}^{n} \lvert x_i\rvert - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	$x \in [-5,5]^n$	0

Table 2: Test functions

Test/benchmark functions	Search region	global minimum
Becker and Lago function (BL)		
$f(x) = \sum_{i=1}^{n} \lvert\lvert x_i\rvert - 5\rvert$	$x \in [-10,10]^n$	0
Bohachevsky 1 function($BF1$)		
$f(x) = \lvert x_1\rvert + 2\lvert x_2\rvert - 0.3\cos(3\pi x_1) - 0.4\cos(4\pi x_2) + 0.7$	$x \in [-50,50]^2$	0
Bohachevsky 2 function ($BF2$)		
$f(x) = \lvert x_1\rvert + 2\lvert x_2\rvert - 0.3\cos(3\pi x_1)\cos(4\pi x_2) + 3$	$x \in [-50,50]^2$	0
Levy and Montalvo 1 function($LM1$)		
$f(x) = (\pi/n)\left(10\lvert\sin(\pi y_1)\rvert + \sum_{i=1}^{n-1} \lvert y_i - 1\rvert[1 + 10\lvert\sin(\pi y_{i+1})\rvert] + \lvert y_n - 1\rvert\right)$	$x \in [-10,10]^n$	0
with $y_i = 1 + \frac{1}{4}(x_i + 1)$		
Levy and Montalvo 2 function($LM2$)		
Modified Rosenbrock function (MRP)		
$f(x) = 100\lvert x_2 - \lvert x_1\rvert\rvert + \lvert 6.4\,[\lvert x_2 - 0.5\rvert] - x_1 - 0.6\rvert$	$x \in [-5,5]^2$	0
Periodic function (PRD)		
$f(x) = 1 + \lvert\sin(x_1)\rvert + \lvert\sin(x_2)\rvert - 0.1\exp(-\lvert x_1\rvert - \lvert x_2\rvert)$	$x \in [-10,10]^2$	0.9
Schaffer 2 function($SF2$)		
$f(x) = \left(\lvert x_1\rvert + \lvert x_2\rvert\right)^{0.25}\left[\lvert\sin\left(50\left(\lvert x_1\rvert + \lvert x_2\rvert\right)^{0.1}\right)\rvert + 1\right]$	$x \in [-100,100]^2$	0
Variably dimensioned function(VAR)		
$f(x) = \sum_{i=1}^{n+2} \lvert f_i(x)\rvert$	$x \in [-5,5]^n$	0
with $f_i(x) = x_i - 1,\ i = 1,\ldots,n,\ f_{n+1}(x) = \sum_{j=1}^{n} j(x_j - 1),\ f_{n+2}(x) = \left(\sum_{j=1}^{n} j(x_j - 1)\right)^2$		
Trigonometric function (TRI)		
$f(x) = \sum_{i=1}^{n} \lvert n - \sum_{j=1}^{n} x_j + i(1 - \cos(x_i)) - \sin(x_i)\rvert$	$x \in [-5,5]^n$	0

Table 3: Optimal values obtained by RTESLS using the curve of Example 1.

$Pb-num$	n	x^*	$f(x^*)$	$feval$
1	2	$(1.0005, 0.9999)$	2.00026	1068
2	2	$(-0.00045, -2.9956)$	-2.9731	1683
3	2	$(5.1692, 2.0271)$	44.9660	15864
4	3	$(0.0001, -0.0001, 0.0009)$	-2.6418	2445
5	4	$(0.4951, 0.4995, 0.5045, 0.4981)$	$1.18e-002$	2473
6	5	$(-0.0023, 0.0056, 0.049, -0.0060, -0.0005)$	$1.9e-002$	1386
7	5	$(0.9992, 0.9999, 1.0007, 1.0010, 0.9999)$	$1.75e-002$	22368
8	7	$(-0.0756, -0.0548, -0.0042, 0.0021, 0.0231, 0.0102, -0.0812)$	$4.7e-003$	55124

report respectively the approximation of a global minimiser x^* and the global minimum $f(x^*)$ obtained by the algorithm. The fifth column denotes the number of function evaluations required to reach the global minimum $(feval)$.

The RTESLS algorithm was also compared with Tilecutter (Price et al. (2012)), CRS (Price (1977)), ARS (Appel (2003)) and G-CARTopt (Robertson et al. (2013)) on a set of nonsmooth problems. The average (over ten runs) number of function evaluations to obtain an approximate solution of (10^{-4}) and (10^{-6}) accuracies are reported in tables 4 and 5.

Table 4: Average number of function evaluations required to reduce f below 10^{-4}

function	n	RTESLS	Tilecutter	CRS	ARS	G-CARTopt	
						Random	Halton
Becker & Lago	2	210	3014	2455	400	723	767
Bohachevsky 1	2	1215	4908	2414	1015	831	825
Bohachevsky 2	2	12230	4523	2351	945	812	802
Levy & Montalvo 1	3	4325	4744	4097	6203	932	930
Levy & Montalvo 2	3	2984	10153	4523	8 fails	1094	1095
Mod. Rosenbrock	2	1998	6119	13437	10478	1143	1676
Periodic (Price)	2	2229	1825	2722	1712	1341	909
Schaffer 2	2	17	10 fails	4132	10 fails	2429	2590
Trigonometric	5	89762	17655	17617	4797	2747	3162
Variably dimensioned	4	30945	6 fails	6751	2592	1990	2003
Variably dimensioned	8	31948	10 fails	29268	10 fails	8825	8705

Table 5: Average number of function evaluations required to reduce f below 10^{-4}

function	n	RTESLS	Tilecutter	CRS	ARS	G-CARTopt	
						Random	Halton
Becker & Lago	2	315	8915	3628	968	1365	1111
Bohachevsky 1	2	1280	12020	3415	1645	2050	2082
Bohachevsky 2	2	1252	11099	3298	1242	1123	1116
Levy & Montalvo 1	3	4569	12524	5417	6586	1281	1299
Levy & Montalvo 2	3	3988	35508	6286	10 fails	1580	1730
Mod. Rosenbrock	2	2170	14815	16483	5 fails	1844	2354
Periodic (Price)	2	2447	5936	3692	3450	2394	1680
Schaffer 2	2	19	10 fails	5790	10 fails	4370	4301
Trigonometric	5	93994	45917	22319	14903	4505	4266
Variably dimensioned	4	33002	9 fails	10162	5368	2958	2960
Variably dimensioned	8	34058	10 fails	37049	10 fails	16820	21773

4.1. Comments

Note first that all the solutions of the problems in Table 3 have been obtained during the generation of the second and third parametrized curves. A comparison between of RTESLS, Tilecutter, CRS, ARS and G-CARTopt, the RTESLS algorithm is better performing in terms of the number of evaluations of f in some problems. However, for some other type of problems our algorithm is less efficient. On the whole, RTESLS proves to be good for problems having a large number of minimisers (both local and global) whereas Metaheuristic methods are more adapted to problems with large dimensions.

5. CONCLUSION

In this paper we develop a new approach for solving a large class of global optimization problems. The objective function is only continuous, non-smooth and non-Lipschitzian defined on a rectangle of $\mathbb{R}^n$. This approach is based on the generation, in the feasible set, of a family of parametrized curves satisfying certain properties combined with the one-dimensional Evtushenko algorithm. To accelerate the corresponding mixed algorithm we have incorporated a stochastic local optimization method inspired by the deterministic local optimization method of Hooke and Jeeves. Preliminary numerical experiments indicate that the algorithm is promising. In a future work, curves with simpler analytical expressions will be sought for as well as matching one-dimensional global optimization algorithms.

REFERENCES

Appel, M. J., LaBarre, R., and Radulovic, D. (2003). On accelerated random search, *SIAM Journal on Optimization*, 14:708–731.

Cherruault,Y. (1999). *Optimisation, Méthodes locales et globales*, Presses Universitaires de France.

Evtushenko,Y. G. (1985). *Numerical Optimization Techniques*, Springer, Berlin.

Hooke, R., Jeeves, T. A. (1961). "Direct search" solution of numerical and statistical problems, *Journal of the ACM*, 8: 212-229.

Pintér, J. D. (2006). *Global optimization: scientific and engineering case studies*, Springer Science & Business Media.

Price, C. J., Reale, M., and Robertson, B. L. (2012). A cover partitioning method for bound constrained global optimization, *Optimization Methods and Software*, 27:1059-1072.

Price,W. L. (1977). A controlled random search procedure for global optimisation, *The Computer Journal*, 4: 367–370.

Robertson, B. L., Price, C. J., and Reale, M. (2013). A CARTopt method for bound-constrained global optimization, *The ANZIAM Journal*, 55:109–128.

Rahal, M., Ziadi, A. (2008). A new extension of Piyavskii's method to Hölder functions of several variables, *Applied mathematics and Computation*, 197:478-488.

Zhigljavsky, A. A., Zilinskas, A. (2008). *Stochastic global optimization*, Springer-Verlag, New York.

Ziadi, A., Cherruault, Y. (2000). Generation of α-dense curves and application to global optimization, *Kybernetes*, 29:71-82.

Ziadi, A., Cherruault, Y., Mora, G. (2001). Global optimization: A new variant of the Alienor method, *Computers & Mathematics with Applications*, 41:63-71.

Ziadi, R., Ellaia, R., and Bencherif-Madani, A. (2017). Global optimization through a stochastic perturbation of the Polak–Ribière conjugate gradient method, *Journal of Computational and Applied Mathematics*, 317:672-684.

PRIMAL-DUAL PARTITIONS IN LINEAR SEMI-INFINITE OPTIMIZATION

ABRAHAM BARRAGAN[1], LIDIA REBOLLAR[2], MAXIM TODOROV[3]

[1] UDLAP, Dep. of Actuarial Sciences, Physics and Mathematics, ab_beni211@hotmail.com
[2] BUAP, Faculty of Sciences Physics and Mathematics, lhernan@fcfm.buap.mx
[3] UDLAP, Dep. of Actuarial Sciences, Physics and Mathematics and on leave from IMI-BAS, Sofia, BG, maxim.todorov@udlap.mx

---

Abstract: *We consider two partitions over the space of linear semi-infinite programming parameters with a fixed index set and bounded coefficients (the functions of the constraints are bounded). The first one is the primal-dual partition inspired by consistency and boundedness of the optimal value of the linear semi-infinite optimization problems. The second one is a refinement of the primal-dual partition that arises considering the boundedness of the optimal set. These two partitions have been studied in the continuous case, this is, the set of indices is a compact infinite compact Hausdorff topological space and the functions defining the constraints are continuous.*

Keywords: *linear semi-infinite programming, bounded linear semi-infinite optimization problems, primal-dual partition, stability*

1. INTRODUCTION

We associate with each triplet $\pi \in \Pi = \mathbb{B}^n \times \mathbb{B} \times \mathbb{R}$ a primal problem

$$P: \quad \begin{array}{l} \inf \boldsymbol{c}'\boldsymbol{x} \\ s.a\ \ \boldsymbol{a}_t'\boldsymbol{x} \geq b_t,\ t \in T, \end{array}$$

and a dual problem in the sense of Haar

$$D: \quad \begin{array}{l} \sup \sum\limits_{t\in T} \lambda_t \boldsymbol{b}_t \\ s.a.\ \ \sum\limits_{t\in T} \lambda_t \boldsymbol{a}_t = \boldsymbol{c} \\ \lambda \in \mathbb{R}_+^{(T)}. \end{array}$$

Above and henceforth, $\mathbb{R}_+^{(T)}$ denotes the set of nonnegative general finite sequences, that is, functions $\lambda : T \to \mathbb{R}_+$ satisfying that $\lambda_t = 0$ for all $t \in T$ except maybe for a finite number of indices. In $\mathbb{R}_+^{(T)}$ we consider the norms l_∞ and l_1.

As both primal and dual problems are defined with the same data $\boldsymbol{a}$, b and $\boldsymbol{c}$, these are represented by the triplet $\pi := (\boldsymbol{a}, b, \boldsymbol{c})$.

The **parameters space** Π is defined as the set of all triplets $\pi = (\boldsymbol{a}, b, \boldsymbol{c})$ with n and T fixed, equipped with the pseudometrics $d : \Pi \times \Pi \to [0, +\infty]$, defined by

$$d(\pi^1, \pi^2) := \max\left\{ \|\boldsymbol{c}^1 - \boldsymbol{c}^2\|_\infty,\ \sup_{t\in T} \left\| \begin{pmatrix} \boldsymbol{a}_t^1 \\ b_t^1 \end{pmatrix} - \begin{pmatrix} \boldsymbol{a}_t^2 \\ b_t^2 \end{pmatrix} \right\|_\infty \right\},$$

where, $\pi^i = (\boldsymbol{a}^i, b^i, \boldsymbol{c}^i) \in \Pi$, $i = 1, 2$ and $\|\cdot\|_\infty$ represents the uniform norm.

We continue with the notations to follow in the rest of the work. We denote by $\mathbb{R}_+$ the set of positive real numbers, and by $\mathbb{R}_{++}$ the set of positive real numbers where the zero is not included. In the n-dimensional space $\mathbb{R}^n$ endowed with the Euclidean norm, $\boldsymbol{x}'$ stands for the transpose of the vector column $\boldsymbol{x}$, the null vector will be denoted by $\boldsymbol{0}_n$. If X is a set of any topological space, *int* X and *cl* X will denote the interior and closure, respectively. Given a nonempty set $X \subset \mathbb{R}^n$, *conv* X and *cone* X will denote its convex and conical hull, respectively. If C is a nonempty convex set, its recession cone $O^+(C)$ is:

$$cone\ \{\boldsymbol{y} \in \mathbb{R}^n\ :\ \boldsymbol{x} + \alpha\boldsymbol{y} \in C \text{ for all } \boldsymbol{x} \in C \text{ and for all } \alpha > 0\}.$$

The feasible set (optimal) of P and D will be denoted by F (F^*) and Λ (Λ^*), respectively. The optimal value of the primal (dual) problem P (D) will be denoted by $v^P(\pi)$ ($v^D(\pi)$) where, as usual, $v^P(\pi) = +\infty$ and $v^D(\pi) = -\infty$ when the corresponding problems become inconsistent.

With each parameter π we associate the first and second moment cones $M := cone\ \{\boldsymbol{a}_t,\ t \in T\}$ and $N := cone\ \{(\boldsymbol{a}_t, b_t)',\ t \in T\}$, and its characteristic cone $K := cone\ \{(\boldsymbol{a}_t, b_t)',\ t \in T; (\boldsymbol{0}_n, -1)'\}$. Remember that π satisfies the Slater condition if there exist $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ such that, $\boldsymbol{a}_t'\bar{\boldsymbol{x}} > b_t$ for all $t \in T$. Also, π satisfies the strong Slater condition if there are $\varepsilon > 0$ and $\bar{\boldsymbol{x}} \in \mathbb{R}^n$, such that, $\boldsymbol{a}_t'\bar{\boldsymbol{x}} \geq b_t + \varepsilon$ for all $t \in T$. From the definition we have that every parameter that satisfies the strong Slater condition is consistent and satisfies the Slater condition. However, the opposite is not true in general (see Example 4). In [Goberna and López (1998), Theorem 6.1] it is shown that a parameter satisfies the strong Slater condition if and only if $\boldsymbol{0}_{n+1} := (\boldsymbol{0}_n, 0)' \notin cl\ G$, where $G := conv\ \{(\boldsymbol{a}_t, b_t)',\ t \in T\}$. It is worth mentioning that in the continuous case strong Slater and Slater conditions coincide.

We will denote by Π_C^P, Π_{IC}^P, Π_B^P and Π_{UB}^P (Π_C^D, Π_{IC}^D, Π_B^D and Π_{UB}^D) the sets of parameters that have primal (dual) problem consistent, inconsistent, bounded (consistent with finite optimal value) and unbounded, respectively. Also, Π_S^P (Π_S^D) will denote the set of parameters with solvable primal (dual) problem which have bounded optimal set, while Π_N^P (Π_N^D) will denote the set of parameters with primal (dual) problem which is not solvable or has unbounded optimal set.

In the continuous case, the sets Π_S^P and Π_S^D are characterized in [Goberna and Todorov (2008); Barragán et al. (2016)]. These characterizations are presented in the next lemma.

► **Lemma 1.**
(i) $\pi \in \Pi_S^P$ *if and only if* $\binom{\boldsymbol{0}_n}{1} \notin cl\ K$ *and* $\boldsymbol{c} \in int\ M$.
(ii) $\pi \in \Pi_S^D$ *if and only if* $\boldsymbol{c} \in M$ *and* π *satisfies the Slater condition.*

In the first primal-dual partition, presented in [Goberna and Todorov (2009)], the primal and dual problems are classified in inconsistent (IC), bounded (B) and unbounded (UB) classes. This partition is showed in the Table 1.

Table 1:

$(D) \setminus (P)$	IC	B	UB
IC	Π_4	Π_5	Π_2
B	Π_6	Π_1	
UB	Π_3		

where,

$$\Pi_1 := \Pi_B^P \cap \Pi_B^D,\ \ \Pi_2 := \Pi_{UB}^P \cap \Pi_{IC}^D,\ \ \Pi_3 := \Pi_{IC}^P \cap \Pi_{UB}^D,$$
$$\Pi_4 := \Pi_{IC}^P \cap \Pi_{IC}^D\ \ \Pi_5 := \Pi_B^P \cap \Pi_{IC}^D,\ \textit{and}\ \Pi_6 := \Pi_{IC}^P \cap \Pi_B^D.$$

We conclude this section with the characterization of the sets Π_i, $i = 1, ..., 6$ where M, N and K play a crucial role, look at [Goberna and Todorov (2009)]. The next theorem, proved in [Ochoa and Vera de Serio (2012)], holds for the general linear semi-infinite optimization, hence in the particular case when $\boldsymbol{a}$ and b are bounded, as well.

▶ **Theorem 2.**
(i) $\pi \in \Pi_1$ *if and only if* $(\boldsymbol{0}_n, 1)' \notin cl\ N$ *and* $\boldsymbol{c} \in M$.
(ii) $\pi \in \Pi_2$ *if and only if* $(\boldsymbol{0}_n, 1)' \notin cl\ N$ *and* $(\{\boldsymbol{c}\} \times \mathbb{R}) \cap cl\ N = \varnothing$.
(iii) $\pi \in \Pi_3$ *if and only if* $\{\boldsymbol{c}\} \times \mathbb{R} \subseteq K$.
(iv) $\pi \in \Pi_4$ *if and only if* $(\boldsymbol{0}_n, 1)' \in cl\ N$ *and* $\boldsymbol{c} \notin M$.
(v) $\pi \in \Pi_5$ *if and only if* $\boldsymbol{c} \notin M$, $(\boldsymbol{0}_n, 1)' \notin cl\ N$ *and* $(\{\boldsymbol{c}\} \times \mathbb{R}) \cap cl\ N \neq \varnothing$.
(vi) $\pi \in \Pi_6$ *if and only if* $(\boldsymbol{0}_n, 1)' \in cl\ N$, $\boldsymbol{c} \in M$ *and* $\{\boldsymbol{c}\} \times \mathbb{R} \nsubseteq K$.

2. FIRST REFINED PRIMAL-DUAL PARTITION

The following theorem shows that the characterization of the interior of the sets that are generated with the primal-dual partition, in the case of bounded coefficients, is similar to the continuous case [Goberna and Todorov (2009)]. However, in the new case Slater condition is replaced by strong Slater condition, because in the case of bounded coefficient there are parameters, like in the general case [Ochoa and Vera de Serio (2012)], satisfying the Slater condition, but not the strong Slater condition (see Example 4).

▶ **Theorem 3.** *Let* $\pi = (\boldsymbol{a}, b, \boldsymbol{c})$ *a parameter with bounded coefficients. Then*

(i) $\pi \in int\ \Pi_1$ *if and only if* π *satisfies the strong Slater condition and* $\boldsymbol{c} \in int\ M$.
(ii) $\pi \in int\ \Pi_2$ *if and only if there exists* $\boldsymbol{y} \in \mathbb{R}^n$ *such that,*

$$\boldsymbol{c}'\boldsymbol{y} < 0 \text{ and } \boldsymbol{a}_t'\boldsymbol{y} > 0 \text{ for all } t \in T.$$

(iii) $\pi \in int\ \Pi_3$ *if and only if* $(\boldsymbol{0}_n, 1)' \in int\ N$.
(iv) $int\ \Pi_i = \varnothing$ *for* $i = 4, 5, 6$.

Observation. If $\boldsymbol{a}$ and b are bounded, then $O^+\ (cl\ G) = \{\boldsymbol{0}_{n+1}\}$.

A refinement of the primal-dual partition follows from classifying the bounded primal and dual problems in two categories. The first one, is formed by solvable problems with bounded optimal set (S). The second one, includes unsolvable problems and those that have unbounded optimal set (N). The refinement is called *refined primal-dual partition* and it is shown in Table 2.

Table 2:

$D \setminus P$		IC	B		UB
			S	N	
IC		Π_4		Π_5	Π_2
B	S		Π_1^1	Π_1^3	
B	N	Π_6	Π_1^2	Π_1^4	
UB		Π_3			

In the refinement,

$$\Pi_1^1 := \Pi_S^P \cap \Pi_S^D,\ \ \Pi_1^2 := \Pi_S^P \cap \Pi_N^D,$$
$$\Pi_1^3 := \Pi_N^P \cap \Pi_S^D \ \text{ and }\ \Pi_1^4 := \Pi_N^P \cap \Pi_N^D.$$

The other sets are the same as in the primal-dual partition.

According to the Duality Theorem ([Goberna et al. (2004), Theorem 4.2]), in ordinary linear programming ([Bazaraa (2005)]) we have $\Pi_1^i = \varnothing$ for $i = 2,3,4$. However, in the case of bounded coefficients, the mentioned sets are nonempty, which comes from the theorem in [Goberna and Todorov (2009)].

► **Theorem 4.** $\Pi_1^i \neq \varnothing$, $i = 1,...,4$.

The conditions characterizing the sets generated by the refined primal-dual partition in the continuous case, are as follows:

► **Theorem 5.** *[Goberna and Todorov (2008), Theorem 3.3] The following statements are true:*

(i) $\pi \in \Pi_1^1$ *if and only if* $\boldsymbol{c} \in int\ M$ *and* π *satisfies the Slater condition;*
(ii) $\pi \in \Pi_1^2$ *if and only if* $\binom{\boldsymbol{0}_n}{1} \notin cl\ K$, $\boldsymbol{c} \in int\ M$ *and* π *does not satisfy the Salter condition;*
(iii) $\pi \in \Pi_1^3$ *if and only if* $\boldsymbol{c} \in M \setminus int\ M$ *and* π *satisfies the Salter condition;*
(iv) $\pi \in \Pi_1^4$ *if and only if* $\binom{\boldsymbol{0}_n}{1} \notin cl\ K$, $\boldsymbol{c} \in M \setminus int\ M$ *and* π *does not satisfy the Slater condition.*

The condition that characterizes the set Π_S^P and which is presented in Lemma 1 is true in the case of bounded coefficients. In the following example, $c^1 \in M_1$ and π^1 satisfies the Slater condition. However, the dual problem associated with the parameter π^1 is unsolvable. Hence, we show that the condition that characterizes the set Π_S^D fails in the case of bounded coefficients.

Example 1. Let $T = [0,1]$ and $n = 2$. We define $\pi^1 := (\boldsymbol{a}^1, b^1, \boldsymbol{c}^1)$ such that, $\boldsymbol{a}_t^1 := (t,1)$ for all $t \in T$,

$$b_t^1 := \begin{cases} 1, & \text{if } t = 0, \\ 0, & \text{if, } 0 < t < 1, \\ -1 & \text{if, } t = 1, \end{cases}$$

and $\boldsymbol{c}^1 := (1/3, 1)$. In [Hernández (2004)] it is shown that $\boldsymbol{c}^1 \in M_1$ and π^1 satisfies the Salter condition, but the dual problem is not solvable.

In the above example, the parameter has unsolvable dual problem, which means that the optimal set of the dual problem is empty, and in particular it is bounded. In addition, the parameter satisfies the strong Slater condition. This leads to the following conjecture.

Conjecture. If $\boldsymbol{c} \in M$ and π satisfies the strong Salter condition, then Λ^* is bounded.

Resolving the conjecture above requires a result that is analogous to Carathéodory's Theorem for positive linear combinations [Rockafellar (1970), Corollary 17.1.2].

► **Lemma 6.** *If* $\sum_{t \in T} \lambda_t \boldsymbol{a}_t = \boldsymbol{c}$ *with* $\lambda_t \geq 0$ *for all* $t \in T$*, then*

$$\sum_{i=1}^{n+1} \gamma_i \boldsymbol{a}_{t_i} = \boldsymbol{c} \ \text{ and } \ \sum_{t \in T} \lambda_t = \sum_{i=1}^{n+1} \gamma_i.$$

► **Theorem 7.** *Let* π *a parameter, with* $|T| \geq n+2$*. If* $\boldsymbol{c} \in M$ *and* π *satisfies the strong Slater condition, then* Λ^* *is bounded with respect to the norm* l_1.

► **Corollary 8.** *If* Λ^* *is not bounded with the norm* l_1 *and* $|T| \geq n+2$*, then* π *does not satisfy the strong Slater condition.*

Observation. If Λ^* is bounded with the norm l_1, then it is bounded with the norm l_∞ too.

Observation. If $|T| < n+2$, we have a parameter in finite case.

► **Theorem 9.** *Let* $\pi = (\boldsymbol{a}, b, \boldsymbol{c})$ *a parameter with* $\boldsymbol{c} = \boldsymbol{0}_n$ *and* $|T| \geq n+2$*. If* $\boldsymbol{c} \in int\ M$ *and* π *satisfies the strong Slater condition, then* $\Lambda^* = \{\lambda \equiv 0\}$.

► **Theorem 10.** *Let* $\pi = (\boldsymbol{a}, b, \boldsymbol{c})$ *a parameter with* $\boldsymbol{c} \neq \boldsymbol{0}_n$ *such that,* $\boldsymbol{c} \in M$ *and* π *satisfies the strong Slater condition, then*

$$\inf \{||\lambda||_1 \ : \ \lambda \in \Lambda^*\} > 0.$$

With the examples below we will show that the conditions that characterize the sets generated by the refined primal-dual partition, in the continuous case, do not hold in the case of bounded coefficients.

Example 2. The primal problem is:

$$P_2 : \qquad \text{Min } \tfrac{1}{3}x_1 + x_2$$

$$s.a \quad \begin{array}{l} x_2 \geq 1 \\ tx_1 + x_2 \geq 0, \quad t \in (0,1), \\ x_1 + x_2 \geq -1 \end{array}$$

$\boldsymbol{c}^2 \in int\ M_2$ and π^2 satisfies the strong Slater condition, but the dual problem is not solvable. Therefore, $\pi^2 \notin \Pi_1^1$ in the case of bounded coefficients.

Example 3. Let $\alpha > 0$ and consider the following problem in $\mathbb{R}$:

$$P_3 : \qquad \inf\ \alpha x_1$$

$$s.a \quad tx_1 \geq t^2, \quad t \in (0,1].$$

The problem is solvable. Also,

$$\boldsymbol{c}^3 = \alpha \in int\ M_3 = int\ (cone\ \{t : t \in (0,1]\}) = \mathbb{R}^0_+.$$

On the other hand, π^3 does not satisfy the strong Slater condition. The dual problem is solvable with bounded optimal set. This way, we have a parameter $\pi^3 \in \Pi_1^1$, where $\boldsymbol{c}^3 \in int\ M_3$, but π^3 does not satisfy the strong Slater condition.

► **Corollary 11.** *Let* $\pi = (\boldsymbol{a}, b, \boldsymbol{c})$ *be a parameter with* $\boldsymbol{c} = \boldsymbol{0}_n$ *and* $|T| \geq n+2$*. If* $\boldsymbol{c} \in int\ M$ *and* π *satisfies the strong Slater condition, then* $\pi \in \Pi_1^1$.

As a consequence of the above corollary, we have that the feasible set of the system

$$\{\boldsymbol{a}_t \boldsymbol{x} \geq b_t,\ t \in T\},$$

(when $|T| \geq n+2$) is nonempty and bounded, if $\boldsymbol{0}_n \in int\ cone\ \{\boldsymbol{a}_t,\ t \in T\}$ and there are $\varepsilon > 0$ and $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ such that, $\boldsymbol{a}_t \bar{\boldsymbol{x}} > b_t + \varepsilon$ for all $t \in T$.

► **Corollary 12.** *Let* $\pi = (\boldsymbol{a}, b, \boldsymbol{c})$ *be a parameter with* $b \equiv 0$ *and* $|T| \geq n+2$*. If* $\boldsymbol{c} \in int\ M$ *and* π *satisfies the strong Slater condition, then* $\pi \in \Pi_1^1$.

The parameter π^3 presented in Example 3 shows us that $\binom{\boldsymbol{0}_n}{1} \notin cl\ N$, $\boldsymbol{c} \in int\ M$ and π without the strong Slater condition are not sufficient conditions for π to belong to Π_1^2.

The parameter π^2 of Example 2, also shows that $\binom{\boldsymbol{0}_n}{1} \notin cl\ N$, $\boldsymbol{c} \in int\ M$ and π without the strong Slater condition are not a necessary conditions for the belonging of π to Π_1^2.

Example 4. Consider the problem in $\mathbb{R}^2$ defined by:

$$P_4 : \qquad \inf\ x_1 + x_2$$

$$s.a \quad t^2 x_1 + t x_2 \geq t, \quad t \in (0,1].$$

The problem is solvable with unbounded optimal set. On the other hand, π^4 does not satisfy the strong Slater condition and $\boldsymbol{c}^4 = (1,1)' \in M_4 \setminus int\ M_4$. The dual problem is solvable and it has bounded optimal set. This way, $\pi^4 \in \Pi_1^3$ and although $\boldsymbol{c}^4 \in M_4 \setminus int\ M_4$, we have that π^4 does not satisfy the strong Slater condition.

Example 5. Consider the following problem in $\mathbb{R}^2$:

$$P_5: \qquad \inf x_2$$

$$s.a \quad \begin{array}{l} x_2 \geq t, \quad t \in [0,1), \\ x_1 \geq 0. \end{array}$$

The problem, is solvable with unbounded optimal set. On the other hand, π^5 satisfies the strong Slater and $\boldsymbol{c}^5 = (0,1)' \in M_5 \setminus int\ M_5$. The dual problem is not solvable. So, it has a parameter π^5 in which $\boldsymbol{c}^5 \in M_5 \setminus int\ M_5$ and π^5 satisfies the strong Slater condition, but $\pi^5 \notin \Pi_1^3$.

In the previous example

$$\pi^5 \in \Pi_1^4, \quad \begin{pmatrix} \boldsymbol{0}_n \\ 1 \end{pmatrix} \notin cl\ N_5 \ \text{ y } \ \boldsymbol{c}^5 \in M_5 \setminus int\ M_5,$$

but π^5 does not satisfy the strong Slater condition. This tells us that $\binom{\boldsymbol{0}_n}{1} \notin cl\ N$, $\boldsymbol{c} \in M \setminus int\ M$ and that π does not satisfy the strong Slater condition are not necessary conditions for $\pi \in \Pi_1^4$.

With the Example 4, we show that $\binom{\boldsymbol{0}_n}{1} \notin cl\ N$, $\boldsymbol{c} \in M \setminus int\ M$ and π without the strong Slater condition are not sufficient for π to belong to Π_1^4.

We have that $\Pi_5^1 = \varnothing$ holds in the case of bounded coefficients. We have thus that $\Pi_5 = \Pi_N^P \cap \Pi_{IC}^D$.

With the following example we see that $\Pi_6^1 = \Pi_{IC}^P \cap \Pi_S^D \neq \varnothing$ in the case of bounded coefficients.

Example 6. Consider the problem in $\mathbb{R}^2$ given by:

$$P_6: \qquad \inf 0$$

$$s.a \ \ tx_1 + tx_2 \geq 1, \qquad t \in (0,1].$$

The problem is inconsistent. The dual problem is bounded and solvable with bounded optimal set.

The set Π_6^2 is also nonempty. This could be seen if we look at the continuous case.

This section ends with the presentation of several necessary conditions.

The following two result are obtained from [Goberna and López (1998), Corollary 9.3.1] and Corollary 8.

- If $\pi \in \Pi_1^2$ and Λ^* is unbounded, then $\boldsymbol{c} \in int\ M$ and π does not satisfy the strong Slater condition.
- If $\pi \in \Pi_1^4$ and Λ^* is unbounded, then $\boldsymbol{c} \in M \backslash int\ M$ and π do not satisfy the strong Slater condition.

The next result is obtained from [Hernández (2004)].

- If $\pi \in \Pi_1^2$ and $\Lambda^* = \varnothing$, then there exist $\{\pi^r\}$ in Π such that, $\pi^r \to \pi$, $\boldsymbol{c}^r \in int\ M_r$ and π^r satisfies the strong Slater condition.

3. CONCLUSION

We conclude mentioning that we have obtained a sufficient condition for the boundedness of the optimal set, which might however be empty, of the dual problem. Conditions that guarantee the solvability of the dual problem turns out to be complicated task even in the continuous case. This could be a challenge problem for a future work.

REFERENCES

Barragán, A., Hernández, L., and Todorov, M. (2016). New primal-dual partition of the space of linear semi-infinite continuous optimization problems. *Comptes rendus*, 69:1263–1274.

Bazaraa, M. (2005). *Programación lineal y flujo en redes*. Limusa.

Goberna, M., Jornet, V., and Puente, R. (2004). *Optimización Lineal. Teoría, Métodos y Modelos*. Mc Graw Hill.

Goberna, M. and López, M. (1998). *Linear Semi-infinite Optimization*. Wiley.

Goberna, M. and Todorov, M. (2008). Generic Primal-dual solvability in continuous linear semi-infinite optimization. *Optimization*, 57:239–248.

Goberna, M. and Todorov, M. (2009). Primal-dual stability in continuous linear optimization. *Mathematical Programming*, 116B:129–147.

Hernández, L. (2004). *Representaciones del conjunto factible y estabilidad del problema dual en programación lineal semi-infinita*. PhD thesis, Benemérita Universidad Autónoma de Puebla.

Ochoa, P. and Vera de Serio, V. (2012). Stability of the primal-dual partition in linear semi-infinite programming. *Optimization*, 61:1449–1465.

Rockafellar, R. (1970). *Convex Analysis*. Princenton.

BALCOR 2018
XIII Balkan Conference on Operational Research

A LINEAR COST MODEL FOR 3D PRINTING NETWORKS

EVGENIA FRONIMAKI[1], MARIA MAVRI[2]

[1] University of the Aegean, Department of Business Administration, bad14009@ba.aegean.gr
[2] University of the Aegean, Department of Business Administration, m.mavri@ba.aegean.gr

Abstract: *3D printing describes the use of a collection of technologies capable of joining materials to manufacture complex products in a single process step. Flexible, rapid and cheap are the three main characteristics of this new technology. The cost of manufacturing a product using additive manufacturing processes exceeds that of traditional methods and may slow the adoption of this technology. Although of the high purchasing costs of AM machines, an AM industry can reduce its operational costs and increase its profitability, while providing goods in fewer competitive prices, if its network is being scheduled carefully. The goal of this work is: (a) to present a 3DP network, which consists of customers, suppliers, AM manufacturing plants, workforce, distribution centers, as well as the links among these entities, through which printed goods flow and (b) to propose a linear cost model that minimizes investment and operational costs and therefore maximizes profits of an AM industry in a 3DP network.*

Keywords: *3D printing, linear programming, production planning.*

1. INTRODUCTION

Additive manufacturing (AM), also known as 3D printing (3DP), is an automated production process of making three dimensional solid objects from a digital file. 3DP enables to the potential customer to produce complex (functional) shapes using less material than traditional manufacturing methods (3dprinting.com). Theoretically, AM technology is capable of producing any physically feasible product design compiled in a 3D model, because products are manufactured layer by layer (Weller et al. 2015). Currently, 3DP is primarily used for prototyping, for producing patterns, and for small production runs (Holmström et al. 2010; Weller et al. 2015). Actually, Gress and Kalafsky (2015) estimated that 36.5% of all 3DP globally is prototyping.

Using CAD software or 3D scanning, users and potential customers are able to create and share designs rapidly, enhancing their ability to adapt faster to marketplace's needs. As the market segment of personal 3D printers has achieved an average annually growth rate of approximately 170% from 2008 to 2013 (Li et al. 2017), Columbus (2015) stated that global spending on printers is forecasted to reach about $27B by 2019.

AM is a technology that has the potential to change the future of supply chains and how industries design, manufacture and/or repair goods in the future. 3DP will affect every area of the manufacturing value chain: Design; Manufacturing; Supply chain; and Distribution. 3DP may become a general purpose technology with wide-ranging implications for: supply chains; market structure, sustainability and production. 3DP is optimistically a new "industrial revolution" that will dramatically change supply chains, firm strategies, competition, and industrial geographies (Sasson and Johnson 2015). The barriers of 3DP applicability are the technological limitations, high material costs, lack of safety and quality standards and high energy costs (Berman 2012). Demand for specialized products and demand from specialized geographies will reduce 3D printing's raw materials costs (Ruffo et al. 2007), and advance 3DP technology.

In the manufacturing sector, forms of 3DP have been used for more than a decade. With the cost of technology inevitably dropping over time, and processes becoming more effective and efficient, it has already reached a point where a 3D printer is affordable for library use, and can produce anything from models of machine parts to jewelry or clothing (Massis 2013). AM technology can also bring significant benefits to the construction industry in terms of increased customization, reduced construction time, reduced manpower, and construction cost (Wu et al. 2016). 3DP can be applied to various manufacturing markets. The decision to invest in AM technologies must be linked to the market and product characteristics; products with a degree of customization; products with increased functionality through design optimization; products of low volume (Murmura and Bravi 2018; Mellor et al. 2014).

Because of the high purchasing and processing costs of AM machines, the planning and scheduling a 3DP network play a crucial role in reducing operational costs, providing services to customers in fewer prices and increasing the profitability of AM industries. Although cost functions have been developed in the literature, 3DP networks have not studied yet. In this work, we propose a linear cost model that minimizes investment and operational costs and therefore maximizes profits of an AM industry in a 3DP network.

The remainder of this paper is organised as follows. Section 2 presents a brief review of important literatures related to cost models for 3DP problems. Section 3 defines the problem and discusses the relevant assumptions. Section 4 presents the new multi-objective optimization model. Finally, Section 5 summarises the conclusions of this work and suggests further research.

2. LITERATURE REVIEW IN 3DP COST MODELS

3D Printer manufacturers are currently growing rapidly and the technology is used for production in several areas, including jewelry, dental implants, orthopedics and components for the automotive and aerospace industries (Hopkinson et al. 2006). As AM has only been used for manufacturing in more recent years, studies of how industries shift in 3DP using are currently scarce and there is a need for more empirical descriptions of how and why an industry adopts 3DP for manufacturing purposes (Sandström 2016).

The use of personal 3D printers could not only change the manufacturing methods but could also reshape business processes from the design of an object to its consumption. Consumer co-creation is increasingly used in the area of 3D product development. In AM process, the consumer is no longer passive; he/she takes an interactive role in the design of the good and becomes a prosumer. Consumers may provide ideas for new products that fill needs not yet met by the market, or might improve on existing products. Adopting a consumer co-creation concept enables producers to shift their business model from manufacturing-centric mass production to consumer-centric mass innovation or customization (Li et al., 2014).

According to our knowledge few econometric models at literature define the cost of a 3D object. Hopkinson and Dickens (2003) developed a model that can be used for different components and batch sizes to identify the specific unit costs in order to identify the break-even points. The approach of the authors provides a breakdown of costs into the following three elements: machine costs; personnel costs; and material costs. The authors suggest that industries with high capital investment, low volumes of production and complexity in design are more suited to AM and technical barriers become less significant.

Ruffo et al. (2006) used an activity based cost model, in order to calculate the cost of AM parts, in which each cost is associated with a particular activity. In their model, the total cost of a build is the sum of raw material costs and indirect costs. As an extension to the model of Hopkinson and Dickens, Ruffo et al. proposed a full cost model and observe more influences than the previously recognized material, personnel and machine costs. The cost model follows the allocation of cost-relevant activities to direct and indirect costs. The identified cost-related activities are material, software, hardware, personnel expenses, equipment purchase and maintenance, as well as production and management.

Gibson et al. (2010) separated the costs of AM in the following four categories: machine costs; production costs; material costs; and labor costs. The sum of these cost categories represents the total costs. The production costs are mainly depending on the time of printing.

Cost analysis in Atzeni and Salmi’s work (2012) presents how machine cost per part is the major term of cost. As the complexity of parts and products increases, costs remain relatively flat for 3DP while costs for conventional manufacturing tend to rise sharply. Marginal production costs of AM remain higher than with conventional technology, owing largely to high material costs and energy intensity. It costs the same to produce two different variants as two identical thereof and can reproduce shapes that are not feasible with the traditional manufacturing methods.

Baumers et al. (2016) present a cost model that draws on two direct costs of manufacturing, raw material cost and energy consumption, which are combined with the total indirect (time-dependent) costs incurred during the build.

3. PROBLEM DESCRIPTION

A 3DP process network comprises multiple entities and connections among entities. More specifically, a simple 3DP network consists of customers, suppliers, AM manufacturing plants, designers, factory workers, distribution centers and business partners i.e. subcontractors for outsourcing, as well as the links among these entities, through which printed goods and procured items flow.

As the 3DP technology grows, supply chains will become more flexible. By making objects in layers, no assembly is required. Less assembly will shorten supply chains, saving money on labor and transportation and shorter supply chains will be less polluting (Lipson and Kurman, 2013).

There are five entrepreneurial opportunities in 3DP along the industry's value chain: (1) the need for error-free technology and software; (2) a need to connect designers with consumers via intermediaries, to access the largest possible number of paying customers; (3) the need for online retailers of 3D printers and consumables for users, although renting or collaborative use of printers is also an option; (4) the ability to work with a commission-based model; and (5) the opportunity to work as consultants for any users at any point along the value chain (Printing the way to success, 2017). User entrepreneurs focus primarily on the combination of low opportunity exploitation cost and a large number of potential customers (Holzmann et al. 2017). In the world economy there is the emergence of advanced manufacturing technologies that are enabling more cost and resource-efficient small-scale production. Among them, 3DP is leading companies to rethink where and how they conduct their manufacturing activities.

Our goal is to develop a cost model based on selected costs and crurial related factors in order to minimize the overall costs associated with procurement, production and delivery over the design horizon. We propose a model for linear 3DP network design problem that minimizes investment and operational costs. The proposed model covers the simple 3DP network decisions (i.e. supplier, manufacturing plant, procurement, production and distribution) trying to minimize the cost of production, procurement and transportation.

The problem is formulated by a MILP model. The total cost in AM takes account of investment costs and operating costs (costs of supply, production, energy consumption, labour, manufacturing plant and staff idle cost, outsourcing and delivery) but not the costs of inventory and warehousing as AM industries store only digital files, reducing the initial investment for warehousing buildings. With 3DP, no molds or other tooling equipment need to be inventoried for each part, and manufacturers will eliminate expensive minimum restocking order requirements.

The objective function deals with the scope of the model on economic aspects, such as minimization of cost and maximization of profit. The entities that are considered to the model are suppliers, plants, customers, as well as to multiplicity of transportation types and time periods.

The problem aims to determine the 3DP structure satisfying customer demand and guaranteeing efficient and profitable performance. Various decisions have to be made such as:

- What is the appropriate number of owned plants and of what capacity? Should existing production facilities be expanded in new plants and in which location?
- Which product should be produced during each time period?
- Which suppliers should be selected?
- Which plant should be supplied by each supplier?
- What is the appropriate workforce at each owned plant?
- Which are the appropriate manufacturing subcontractors to collaborate?

4. MATHEMATICAL FORMULATION

AM industries could increase their sales by selling new innovative product schemes as they are flexible to operate without pressure to decrease manufacturing cost in order to increase outputs. 3DP is emerging as a viable alternative for low volume and customized manufacturing. The degree to which designers can create products that meet individual customers' demand is likely to become a key factor for firms that want to profit from AM (Weller et al. 2015).

The 3DP network design problem addressed here concerns manufacturing firms that transform their production process in 3DP process over a, typically long, planning horizon in order to satisfy the forecasted product demand, while minimizing associated costs. We have structured the description of this 3DP network design problem in the following components: (a) customer demand, (b) products and suppliers, (c) manufacturing plants, (d) distribution channels and (e) transportation.

(a) <u>Customer demand:</u> Customers may include retail outlets or individual customers. Also, in AM industries, goods are paid for before being manufactured. Customers' locations and their demand per period of the design time horizon are considered to be known. No late deliveries are allowed. In fact, in AM technology, once the part design is released, the production begins immediately.

(b) <u>Products and suppliers</u>: Commonly used materials in AM are plastics, metals, ceramics, nickel chromium, cobalt chromium, stainless steel, titanium and polymers (Mavri 2015). The variety of raw materials is less in AM industries and we can see a better use of materials and less waste material; no scrap, milling, or sanding as the 95-98% of waste material can be recycled in 3DP. The items of the BoM of each product family may be procured from one or more suppliers at the specified quantities. In 3DP production, the list of suppliers is shorter than in traditional production chain because fewer materials are used by 3D printers (Mavri 2015). We consider that each supplier may provide up to a maximum quantity per raw material family and per time period. The ability to integrate just-in-time manufacturing in AM industries will change the nature of the relationship between suppliers and assemblers.

(c) Manufacturing plants and production routings: A final good may be produced either by an existing plant (insourcing), or by subcontracting the full or part of production in external partners (outsourcing). Each plant may produce certain product families depending on the manufacturing procedures available in this plant. Each product family is manufactured following a certain routing. 3D printers require no up-front tooling and relatively little setup time, so manufacturers can move more quickly from initial design to prototype to finished product.

The capacity of each owned manufacturing plant for each procedure is given, and may be expanded at an investment cost with new machines, peripherals etc. In the proposed framework, the investment cost is accounted by considering the depreciation cost corresponding to each period of the planning horizon.

Furthermore, each plant employs direct labor staff for each procedure, the number of which is fitted according to production needs. As most 3DP process is highly automated, manpower required in the construction process can be significantly reduced. Except for some finishing operations, no manual labor by skilled technicians is needed. In AM fewer hours are needed to build a prototype. The production process becomes simpler in AM technology. The additive process enables the manufacture of highly complex parts that can't be built using traditional techniques and reduces component counts by allowing complex assemblies to be manufactured as a single part, speeding assembly times and reducing labor costs.

The plant's labor capacity concerns the availability of staff in labor time per period for each procedure. If a procedure is not utilized up to its capacity (equipment or labor), then the idle costs are taking account.

(d) Distribution channels: Products can be distributed through direct channel that contains three stages between the shipper and the client; the supplier, the manufacturer and the customer. Product can be distributed also through indirect channels including intermediaries between manufacturer and end customer, such as distribution channels. In this framework, we assume that the industry uses subcontracted distribution channels or uses direct channel for delivery the final goods.

(e) Transportation: Transportation demand is quantified by units of a certain product or raw material within a certain time period. Since the number of units of each product or raw material per unit is given, the transportation cost between two nodes can be defined.

According to raw materials' transportation, we must consider that in AM industry, by making objects in layers, no assembly is required. Less assembly will shorten supply chains, saving money on labor and transportation. Supply chains are expected to become more flexible as 3D printing technology develops.

3DP is set to facilitate digital object storage and digital object transportation. For example, if hobbyists have a 3D printer, will buy only the digital files from the store and the store (physical or online) will send the digital file to the customer via e-mail.

Prior to providing the formulation of the objective function and the related constraints, we present firstly the definition of the necessary sets, parameters, and decision variables.

Suppose a set of nodes ***N***. Nodes are consided by possible suppliers s_q, where q is an index with cardinality to n ($S=\left\{s_1,s_2,...,s_n\right\}$), owned plants p_j, where j takes values between 1 and w, ($P=\left\{p_1,p_2,...,p_w\right\}$) and customers c_k ($C=\left\{c_1,c_2,...,c_k\right\}$).

Each plant p_j consists of a set of machines AM_l^j (l is an index for the number of machines of the *j*-th plant). Each machine AM_l^j has different operation cost, deprecation cost, set-up cost, production efficiency and capacity.

Time periods $\left\{t_1,t_2,...,t_t\right\}$ are introduced to our model, by the set *T*

Three more set are necessary, the set of transportation types $R=\left\{r_1,r_2,...,r_p\right\}$, the set of product types $F=\left\{f_1,f_2,...f_h\right\}$ and the set of raw materials $M=\left\{m_1,m_2,...,m_y\right\}$. We also introduce M^f the set of raw materials of the f–th product type ($M^f \subseteq \boldsymbol{M}$) and the set of transportation types R_g available at the *i*-th node

Before continuing we have to determine related costs: suppliers' cost, manufacturing costs and transportations costs.

Supplier's cost:

$\boldsymbol{a_q^m}$: Procurement cost per unit of raw material $m \in M$ related to supplier s_q

Manufacturing costs:

$\boldsymbol{\beta_j}$: Depreciation cost per unit of capacity per period related to owned p_j th plant

$\boldsymbol{\gamma_j^f}$: Set up cost for the *f*-th product type when it is manufactured in owned p_j - th plant. The set up cost is the product of the direct labor cost rate multiplied by the total set up time of a product type $f \in F$.

$\boldsymbol{\delta_j^f}$: Production cost (associated with run time) for product type $f \in F$ when it is manufactured in p_j-th plant. The production cost is the product of the direct labor cost rate by the total production/run time of a product type $f \in F$.

$\boldsymbol{\varepsilon_j}$: Idle equipment cost per unit of capacity per period of owned p_j-th plant

$\boldsymbol{\zeta_j}$: Idle labor cost per unit of capacity per period in owned p_j-th plant

Transportation costs:

$\boldsymbol{\lambda_{ibr}^f}$: Transportation cost per unit of product type $f \in F$ from i-th node to node b by transportation type r, $r \in Ri \cap Rb$

$\boldsymbol{\mu_{ibr}^m}$: Transportation cost per unit of raw material family $m \in M$ from node i to node b by transportation type r, and $r \in Ri \cap Rb$

$\boldsymbol{\eta_j}$: Cost per dismissed employee due to new kind of procedure in the *l*-th machine AM_l^j, associated with owned AM_l^j

$\boldsymbol{\theta_j}$: Recruitment cost per new hire due to new kind of procedure in the *l*-th machine AM_l^j, associated with owned AM_l^j

Concerning the capacities and the time periods, we introduce to our model

Supplier's capacity:

$\boldsymbol{\nu_q^{mt}}$: Maximum number of raw material units $m \in M$ provided by supplier s_q in time period $t \in T$

Manufacturing capacity:

$\boldsymbol{H}$: Number of working time units an employee is available to work per time period

Manufacturing time:

$\boldsymbol{\xi_f}$: Production time in working time units required to process a unit of product type $f \in F$

$\boldsymbol{\pi_f}$: Set up time in working time units per time period required to process product type $f \in F$

The number of units of product type $f \in F$ required by customer c_k, at time period $t \in T$ *is represented by* $\boldsymbol{\rho_k^{ft}}$ (*Customer demand)and the* quantity of raw material $m \in M$ required to produce one unit of product type $f \in F$ is represented by $\boldsymbol{Q_f^m}$:

<u>Decision variables</u>

$\boldsymbol{\sigma_j^t}$: Total production capacity in working time units of owned plant p_j at time period $t \in T$

$\boldsymbol{\tau_{ibr}^{mt}}$: Quantity of raw material family $m \in M$ being transferred by transportation type $r \in R$ through arc (i,b), related to supplier s_q at time period $t \in T$

$\boldsymbol{\upsilon_j^{ft}}$: Binary variable for owned plant p_j manufactures product family $f \in F$ at time period $t \in T$

$\boldsymbol{\varphi_j^{ft}}$: Quantity of product family $f \in F$ that each plant p_j manufactures at time period $t \in T$

$\boldsymbol{\chi_j^t}$: Number of labor staff related to owned plant p_j at time period $t \in T$

$\boldsymbol{\psi_j^t}$: Number of dismissed employees related to owned plant p_j at the beginning of time period $t \in T$

$\boldsymbol{\omega_j^t}$: Number of hired employees related to owned plant p_j at the beginning of time period $t \in T$

$\boldsymbol{\iota\alpha_j^t}$: Labor idle capacity in working time units of owned plant p_j at time period $t \in T$

$\boldsymbol{\iota\beta_j^t}$: Equipment idle capacity in working time units of owned plant p_j at time period $t \in T$

$\boldsymbol{\iota\gamma_{ibr}^{ft}}$: Quantity of products of family $f \in F$ being transferred by transportation type $r \in R$ through arc (i,b), p_j at time period $t \in T$

The mathematical formulation of the proposed 3DP process network model is presented below. Objective function minimizes total investment and operational costs over the entire time horizon. Total cost comprises the following parts:

(a) cost of depreciation related to owned plants ($\sum_{j \in P} \sum_{t \in T} \beta_j * \sigma_j^t$)
(b) procurement cost of raw materials ($\sum_{i,b,r \in S} \sum_{m \in M} \sum_{t \in T} a_q^m * \tau_{ibr}^{mt}$),
(c) set up costs ($\sum_{j \in P} \sum_{f \in F} \sum_{t \in T} \gamma_j^f * \upsilon_j^{ft}$)
(d) production (run time) costs ($\sum_{j \in P} \sum_{f \in F} \sum_{t \in T} \delta_j^f * \varphi_j^{ft}$)
(e) dismissal ($\sum_{j \in P} \sum_{t \in T} \eta_j * \psi_j^t$) and
(f) recruitment costs related to employees of plants ($\sum_{j \in P} \sum_{t \in T} \theta_j * \omega_j^t$)
(g) labor ($\sum_{j \in P, q \in S} \sum_{t \in T} \zeta_j * \iota\alpha_q^t$) and
(h) equipment idle capacity costs ($\sum_{j \in P} \sum_{t \in T} \varepsilon_j * \iota\beta_j^t$)
(i) transportation cost of final products ($\sum_{i,b,r \in S} \sum_{f \in F} \sum_{t \in T} \lambda_{ibr}^f * \iota\gamma_{ibr}^{ft}$)
(j) transportation cost of raw materials ($\sum_{i,b,r \in S} \sum_{m \in M} \sum_{t \in T} \mu_{\iota br}^m * \tau_{ibr}^{mt}$).

$$\min \sum_{j \in P} \sum_{t \in T} \beta_j * \sigma_j^t + \sum_{i,b,r \in S} \sum_{m \in M} \sum_{t \in T} a_q^m * \tau_{ibr}^{mt} + \sum_{j \in P} \sum_{f \in F} \sum_{t \in T} \gamma_j^f * \upsilon_j^{ft}$$
$$+ \sum_{j \in P} \sum_{f \in F} \sum_{t \in T} \delta_\iota^f * \varphi_i^{ft} + \sum_{j \in P} \sum_{t \in T} \eta_j * \psi_j^t + \sum_{j \in P} \sum_{t \in T} \theta_j * \omega_j^t + \sum_{j \in P} \sum_{t \in T, q \in S} \zeta_j * \iota\alpha_q^t$$
$$+ \sum_{j \in P} \sum_{t \in T} \varepsilon_j * \iota\beta_j^t + \sum_{i,b,r \in S} \sum_{f \in F} \sum_{t \in T} \lambda_{\iota br}^f * \iota\gamma_{ibr}^{ft} + \sum_{i,b,r \in S} \sum_{m \in M} \sum_{t \in T} \mu_{\iota br}^m * \tau_{ibr}^{mt} \quad (1)$$

The constraints of the proposed model are presented in groups as follows:

<u>Supplier capacity</u>

The procurement of raw material family $m \in M$ from supplier s_q at time period $t \in T$ cannot exceed the supplier's capacity during the same period.

$$\sum_{i,b,r \in S} \boldsymbol{\tau_{ibr}^{mt}} \le \boldsymbol{\nu_q^{mt}} \quad (2)$$

<u>Material requirements</u>

The quantity t_{ibr}^{mt} of raw material family $m \in M^f$ shipped from all suppliers to owned plant p_j by transportation type $r \in R$ equals the quantity required to manufacture the final products of product family $f \in F$ in plant ***i***, during time period $t \in T$.

$$\sum_{i,b,r \in S} \tau_{ibr}^{mt} = \sum_{f \in F | m \in M_f} Q_f^m * \varphi_j^{ft} \quad (3)$$

<u>Demand</u>

The total production of product family $f \in F$ by all plants p_j until time period $t \in T$ must be equal or greater than the total demand for this product family by all customer nodes c_k until the same period.

$$\sum_{j \in P} \sum_{t=1}^{T} \boldsymbol{\varphi_j^{ft}} \ge \sum_{k \in C} \sum_{t=1}^{T} \boldsymbol{\rho_k^{ft}} \quad (4)$$

<u>Production and production time</u>

Any production requires set up, assuming that M ≫ 1.

$$\boldsymbol{\sigma_j^t} \le \text{M} * \boldsymbol{\upsilon_j^{ft}} \quad (5)$$

The required set up and production time in owned plant p_j at time period $t \in T$ cannot exceed this procedure's capacity at the same period.

$$\sum_{f \in F} (\boldsymbol{\xi_f} * \boldsymbol{\varphi_j^{ft}} + \boldsymbol{\pi_f} * \boldsymbol{\upsilon_j^{ft}}) \leq \boldsymbol{\sigma_j^t} \tag{6}$$

Department's idle capacity:

$$\boldsymbol{\sigma_j^t} - \sum_{f \in F} (\boldsymbol{\xi_f} * \boldsymbol{\varphi_j^{ft}} + \boldsymbol{\pi_f} * \boldsymbol{\upsilon_j^{ft}}) = \boldsymbol{\iota\beta_j^t} \tag{7}$$

Workforce

The variation of the workforce level in each procedure of an owned plant must be defined.

$$\boldsymbol{\chi_j^t} = \boldsymbol{\chi_j^{t-1}} - \boldsymbol{\psi_j^t} + \boldsymbol{\omega_j^t} \tag{8}$$

Ranges of Decision Variables

$$\boldsymbol{\upsilon_j^{ft}} \in \{\mathbf{0}, \mathbf{1}\} \tag{9}$$

$$\boldsymbol{\chi_j^t}, \boldsymbol{\psi_j^t}, \boldsymbol{\omega_j^t} \in \aleph \tag{10}$$

$$\boldsymbol{\sigma_j^t}, \boldsymbol{\tau_{ibr}^{mt}}, \boldsymbol{\varphi_j^{ft}}, \boldsymbol{\iota\alpha_j^t}, \boldsymbol{\iota\beta_i^t}, \boldsymbol{\iota\gamma_{ibr}^{ft}} \geq 0 \tag{11}$$

5. CONCLUSION-FURTHER RESEARCH

3DP enable customers to order their choices online, allow firms to profitably serve small market segments and enable industries to operate with little or no unsold finished goods inventory. Berman (2012) stated that 3DP will significantly reduce the advantages of producing small lot sizes in low-wage countries via reduced need for factory workers.

The research has focused on the most important aspect of 3DP processes, i.e. the manufacturing cost. This paper shows that AM technologies can add flexibility to production networks and when deciding whether or not to use AM technologies, production cost is a sufficient criterion. The redefinition of production cost seems to be necessary. We propose a cost model that provides the flexibility to manufacture with the minimum operational costs. Supplier capacity, material requirements, demand, production and production time, workforce are the constraints of the proposed model.

Early findings showed that the implementation of AM caused a shift in 3DP networks as AM industries redesign their production process, taking account the related factorial variables that affect the minimum cost of manufacturing.

In order to test the value of the proposed model in tackling the very significant complexities of current business reality, we apply the proposed model to a case study of an AM industry.

REFERENCES

Atzeni, E., & Salmi, A. (2012). Economics of additive manufacturing for end-usable metal parts. *International Journal of Advanced Manufacturing Technology*, 62, 1147-1155.

Baumers, M., Dickens, P., Tuck, C., & Hague, R. (2016). The cost of additive manufacturing: machine productivity, economies of scale and technology-push. *Technological Forecasting and Social Change*, 102, 193-201.

Berman, B. (2012). 3D printing: the new industrial revolution. *Business Horizons*, 55 (2), 155-162.

Columbus, L. (2016). Seven ways 3D-printing is making manufacturing more competitive, Forbes, available at: http://www.forbes.com/sites/louiscolumbus/2016/04/24/7-ways-3d-printing-is-making-manufacturing-more-competitive/#6a0824c62938 (accessed 17 February 2018).

Gibson, I., Rosen, D., & Stucker, B. (2010). *Additive manufacturing technologies. Rapid Prototyping to Direct Digital Manufacturing*. New York: Springer.

Gress, D., & Kalafsky, R. (2015). Geographies of production in 3D: theoretical and research implications stemming from additive manufacturing. *Geoforum*, 60, 43-52.

Holmström, J., Partanen, J., Tuomi, J., & Walter, M. (2010). Rapid manufacturing in the spare parts supply chain: alternative approaches to capacity deployment. *Journal of Manufacturing Technology Management*, 21 (6), 687-697.

Holzmann, P., Breitenecker, R., Soomro, A., & Schwarz, E. (2017). User entrepreneur business models in 3D printing. *Journal of Manufacturing Technology Management*, 28 (1), 75-94.

Hopkinson, N., & Dickens P. (2003). Analysis of Rapid Manufacturing – Using Layer Manufacturing Processes for Production. In *Proceedings of the Institution of Mechanical Engineers, Part C : Journal of Mechanical Engineering Science*, 217 (C1), 31-39.

Hopkinson, N., Hague, R.J.M., & Dickens, P, (2006). *Rapid Manufacturing: An Industrial Revolution for the Digital Age*. Wiley.

Li, P., Mellor, S., Griffin, J., Waelde, C., Hao, L., & Everson, R. (2014). Intellectual property and 3D printing: a case study on 3D chocolate printing. *Journal of Intellectual Property Law & Practice*, 9 (4), 322-332.

Lipson, H., & Kurman, M. (2013). *Fabricated: The New World of 3D Printing*. Hoboken: John Wiley & Sons.

Massis, B. (2013). 3D printing and the library. *3D printing and the library*, 114 (7/8), 351-354.

Mellor, S., Hao, L., & Zhang, D. (2014), Additive manufacturing: a framework for implementation. *International Journal of Production Economics*, 149, 194-201.

Mavri, M. (2015). Redesigning a Production Chain Based on 3D Printing Technology. *Knowledge and Process Management*, 22 (3), 141-147.

Murmura, F., & Bravi, L. (2018). Additive manufacturing in the wood-furniture sector. Sustainability of the technology, benefits and limitations of adoption. *Journal of Manufacturing Technology Management*, 29 (2), 350-371.

Ruffo, M., Tuck, C., & Hague, R. (2006). Cost Estimation for Rapid Manufacturing – Laser Sintering Production for Low to Medium Volumes. In *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 1417-1427.

Ruffo, M., Tuck, C., & Hague, R. (2007). Make or buy analysis for rapid manufacturing. *Rapid Prototyping Journal*, 13 (1), 23-29.

Sandström, C. (2016). The non-disruptive emergence of an ecosystem for 3D Printing-Insights from the hearing aid industry's transition 1989-2008. *Technological Forecasting & Social Change*, 102, 160-168.

Sasson, A., & Johnson, J. (2015). The 3D printing order: variability, supercenters and supply chain reconfigurations. *International Journal of Physical Distribution & Logistics Management*, 46 (1), 82-94.

Weller, C., Kleer, R., & Piller, F. (2015). Economic implications of 3D printing: Market structure models in light of additive manufacturing revisited. *Int. J. Production Economics*, 164, 43-56.

Wu, P., Wang, J., & Wang, X. (2016). A critical review of the use of 3-D printing in the construction industry. *Automation in Construction*, 68, 21-31.

3Dprinting.com, available at https://3dprinting.com/what-is-3d-printing/ (accessed 5 March 2018).

(2017). Printing the way to success. *Strategic Direction*, 33 (6), 4-6.

B13
Multiple Criteria Decision Analysis & Multi-Objective Optimization

BALCOR 2018
XIII Balkan Conference on Operational Research

MULTI-CRITERIA DECISION MAKING METHODS: DOES APPROACH MATTER?

NENAD MEDIĆ[1], ZORAN ANIŠIĆ[2]

[1] University of Novi Sad, Faculty of Technical Sciences, medic.nenad@uns.ac.rs
[2] Subotica Tech – College of Applied Sciences, Subotica, Serbia, azoran@vts.su.ac.rs

---

Abstract: *Multi-criteria decision making (MCDM) problems can be solved using numerous different approaches. For example, Preference Ranking Organization Method for Enrichment Evaluations (PROMETHEE) is an outranking method, while Technique for Order Performance by Similarity to Ideal Solution (TOPSIS) and VlseKriterijumska Optimizacija I Kompromisno Resenje (VIKOR) are compromise ranking methods. One of the issues in this field of research is related to differences in results based on the selected approach for problem solving. A comparative analysis of three above mentioned methods is presented in this paper. These methods were applied for selection of energy supply system for space heating of the residential building. All three methods yield different ranking of alternatives in this particular case. This implies that based on the approach and method used, results may vary. Decision makers should take this into consideration when choosing a method for solving MCDM problems.*

Keywords: *PROMETHEE, TOPSIS, VIKOR, Comparative analysis.*

1. INTRODUCTION

Multi-criteria decision making (MCDM) is the process of finding best option from all of the feasible alternatives in the presence of multiple, usually conflicting criteria (Zanakis et al., 1998). Since criteria are often conflicting there may be no solution satisfying all criteria simultaneously. Consequently, result of this process is often a compromise solution, which is obtained according to decision makers' preferences. Main steps of MCDM are defined as follows (Opricovic & Tzeng, 2004):

- Establishing system evaluation criteria that relate system capabilities to goals
- Developing alternative systems for attaining the goals
- Evaluating alternatives in terms of criteria
- Applying a normative multi-criteria analysis method
- Accepting one alternative as ''optimal'' (preferred)
- If the final solution is not accepted, gather new information and go into the next iteration of multi-criteria optimization.

Wide range of methods can be applied for solving MCDM problems. The problem of selecting appropriate MCDM method has been addressed in literature (Guitouni & Martel, 1998; MacCrimmon, 1972; Ozernoy, 1992). Nevertheless, there is no perfect solution for this problem, due to diversified approach of different methods. Often, different methods can yield different ranking of alternatives. The inconsistency in results occurs because (Zanakis et al., 1998):

- Algorithm for selection of preferred alternative differs
- Techniques use criteria weights differently in their calculations
- Algorithms attempt to scale the objectives, thus affecting the weights already chosen
- Algorithms introduce additional parameters that affect selection of preferred alternative.

For this reason there were some attempts to find similarities and differences among MCDM methods in literature. It was determined that change in number of alternatives or criteria, and change of criteria weights distribution are affecting the results (Zanakis et al., 1998). Comparing VIKOR, TOPSIS and SAW (Simple Additive Weighting) authors revealed that traditional method such as SAW cannot clarify decision-making as well as TOPSIS or VIKOR (Chu et al., 2007). Comparison of VIKOR and TOPSIS showed that different approaches in normalization of attributes and aggregation functions affect the order of alternatives (Opricovic & Tzeng, 2004). On the basis of a case study, outranking methods are compared to techniques based on the ideas of multi attribute utility theory (Kangas et al., 2001). In (Opricovic & Tzeng, 2007), VIKOR was compared with TOPSIS, ELECTRE (ELimination Et Choix Traduisant la REalite) and PROMETHEE. Differences between methods were notified and explained.

In this paper, PROMETHEE, TOPSIS and VIKOR will be compared on an empirical example. These MCDM methods were applied for selection of energy supply system for space heating of the residential building. It was found interesting to compare these three methods because of the difference in their approach to problem solving situation. PROMETHEE is an outranking method, while TOPSIS and VIKOR are compromise ranking methods. Also, these three methods are highly used for solving MCDM problems (Mardani et al., 2015). This paper contributes to the existing literature by confirming on an empirical example that there are differences in the order of preferred alternatives based on the selected approach for problem solving.

The rest of the paper is structured as follows. In Section 2 procedures of PROMETHEE, TOPSIS and VIKOR are explained. In Section 3 these three methods are applied on an empirical example and results are presented. Finally, Section 4 presents discussion of these results and conclusion of the paper.

2. MCDM METHODS

In this section we described PROMETHEE, TOPSIS and VIKOR methods as a theoretical basis for the following application.

2.1. PROMETHEE

The PROMETHEE family of outranking methods, including the PROMETHEE I for partial ranking of the alternatives and the PROMETHEE II for complete ranking of the alternatives, were developed by Brans (Brans, 1982). PROMETHEE II is described in this part of the paper, since majority of researchers have referred to this version of the method (Behzadian et al., 2010). This method is based on a pairwise comparison of alternatives in respect to each defined criterion. The implementation of PROMETHEE II requires two types of information. Decision maker needs to define weight and preference function for each criterion. Weight determines the importance of each criterion, while preference function serves to translate difference between the evaluations obtained by alternatives into a preference degree ranging from zero to one. There are six types of preference functions proposed: (1) usual criterion, (2) U-shape criterion, (3) V-shape criterion, (4) level criterion, (5) V-shape with indifference criterion and (6) Gaussian criterion (Vincke, 1985). The procedure of PROMETHEE II method is as follows (Brans, 1982):

Step 1: Determination of deviations based on pairwise comparisons

$$d_j(a,b) = g_j(a) - g_j(b) \quad (1)$$

Where $\mathrm{d_j(a,b)}$ denotes the difference between the evaluations of a and b on each criterion.

Step 2: Application of the preference function

$$P_j(a,b) = F_j\left[d_j(a,b)\right] \qquad j = 1, \ldots, k \quad (2)$$

Where $P_j(a,b)$ denotes the preference of alternative a with regard the alternative b on each criterion, as a function of $d_j(a,b)$.

Step 3: Calculation of an overall or global preference index

$$\forall\, a, b \in A, \qquad \pi(a,b) = \sum_{j=1}^{k} P_j(a,b)\, w_j \quad (3)$$

Where $\pi(a,b)$ of a over b (from 0 to 1) is defined as a weighted sum $p(a,b)$ of each criterion, and $\mathrm{w_j}$ is the weight associated with the expressing the decision maker's preference as the relative importance of the j-th criterion.

Step 4: Calculation of outranking flows

$$\phi^+(a) = \frac{1}{n-1}\sum_{x\in A} \pi(a,x) \quad (4)$$

$$\phi^-(a) = \frac{1}{n-1}\sum_{x\in A} \pi(x,a) \quad (5)$$

Where $\phi^+(a)$ and $\phi^-(a)$ denote the positive outranking flow and negative outranking flow for each alternative, respectively.

Step 5: Calculation of net outranking flow

$$\phi(a) = \phi^+(a) - \phi^-(a) \quad (6)$$

Step 6: Determine the ranking of all the considered alternatives depending on the values of $\phi(a)$. Higher value of $\phi(a)$, means better ranking of the alternative. Thus, the best alternative is the one having the highest $\phi(a)$ value.

2.2. TOPSIS

TOPSIS is a compromise ranking method developed to determine best alternative nearest to the positive ideal solution and farthest from the negative ideal solution (Hwang & Yoon, 1981). It has been successfully applied in many different research areas (e.g. Supply Chain Management, Manufacturing Systems, and Energy Management) (Behzadian et al., 2012). The stepwise procedure of TOPSIS method is as follows (Hwang & Yoon, 1981):

Step 1: Construct normalized decision matrix:

$$r_{ij}(x) = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}}, \ i = 1, \dots, n; \ j = 1, \dots, m \tag{7}$$

Where x_{ij} and r_{ij} are original and normalized score of decision matrix, respectively.

Step 2: Construct the weighted normalized decision matrix:

$$v_{ij}(x) = w_j * r_{ij}(x), \ i = 1, \dots, n; \ j = 1, \dots, m \tag{8}$$

Where w_j is the weight of the j criterion.

Step 3: Determine the positive ideal solution (PIS) and negative ideal solution (NIS):

$$PIS = A^+ = \{v_1^+(x), v_2^+(x), \dots, v_j^+(x), \dots, v_m^+(x)\}$$

$$= \left\{\left(\max_i v_{ij}(x) \mid j \in J_1\right), \left(\min_i v_{ij}(x) \mid j \in J_2\right) \mid i = 1, \dots, n\right\} \tag{9}$$

$$NIS = A^- = \{v_1^-(x), v_2^-(x), \dots, v_j^-(x), \dots, v_m^-(x)\}$$

$$= \left\{\left(\min_i v_{ij}(x) \mid j \in J_1\right), \left(\max_i v_{ij}(x) \mid j \in J_2\right) \mid i = 1, \dots, n\right\} \tag{10}$$

Where J_1 and J_2 are the benefit and the cost attributes, respectively.

Step 4: Calculate the separation values from PIS and NIS for each alternative:

$$D_i^* = \sqrt{\sum_{j=1}^{m} \left[v_{ij}(x) - v_j^+(x)\right]^2}, \ i = 1, \dots, n \tag{11}$$

and

$$D_i^- = \sqrt{\sum_{j=1}^{m} \left[v_{ij}(x) - v_j^-(x)\right]^2}, \ i = 1, \dots, n \tag{12}$$

Step 5: Calculate the relative closeness to the ideal solution:

$$C_i^* = D_i^- / (D_i^* + D_i^-), \ i = 1, \dots, n \tag{13}$$

Where $C_i^* \in [0,1] \ \forall \ i = 1, \dots, n$

Finally, the alternatives can be ranked by respective C_i^* value in descending order.

2.3. VIKOR

The VIKOR method was introduced as a technique for multi-criteria optimization of complex systems (Opricovic, 1998). It determines the compromise ranking list and the compromise solution. Assuming that each alternative is evaluated according to each criterion function, the compromise ranking could be performed by comparing the measure of closeness to the ideal alternative. The compromise ranking algorithm VIKOR has the following steps (Opricovic & Tzeng, 2004):

Step 1: Determine the best f_i^* and the worst f_i^- values of all criterion functions, i = 1, 2, … , n.

If the i-th function represents a benefit, then:

$$f_i^* = \max_j f_{ij}, \qquad f_i^- = \min_j f_{ij}$$

If the ith function represents a cost, then:

$$f_i^* = \min_j f_{ij}, \qquad f_i^- = \max_j f_{ij}$$

Step 2: Compute the values S_j and R_j, j = 1, 2, … , J, by the relations

$$S_j = \sum_{i=1}^{n} w_i\left(f_i^* - f_{ij}\right)/(f_i^* - f_i^-) \tag{14}$$

$$R_j = \max_i \left[w_i\left(f_i^* - f_{ij}\right)/(f_i^* - f_i^-)\right] \tag{15}$$

where w_i are the weights of criteria, expressing their relative importance for the decision maker.

Step 3: Compute the values Q_j, j = 1, 2, … , J, by the relation

$$Q_j = \frac{v(S_j - S^*)}{(S^- - S^*)} + \frac{(1-v)(R_j - R^*)}{(R^- - R^*)} \tag{16}$$

where

$$S^* = \min_j S_j, \qquad S^- = \max_j S_j$$

$$R^* = \min_j R_j, \qquad R^- = \max_j R_j$$

and v is introduced as weight of the strategy of the maximum group utility, whereas $1 - v$ is the weight of the individual regret.

Step 4: Rank the alternatives, sorting by the values S, R and Q, in decreasing order. The results are three ranking lists.

Step 5: Propose as a compromise solution the alternative (a') which is ranked the best by the measure Q (minimum) if the following two conditions are satisfied:

C1: "Acceptable advantage"

$Q(a'') - Q(a') \geq DQ$

where a'' is the alternative in the second position in the ranking list by Q; $DQ = 1/(J-1)$; J is the number of alternatives.

C2: "Acceptable stability in decision making":

Alternative (a') must also be the best ranked by S or/and R.

If one of the conditions is not satisfied, then a set of compromise solutions is proposed, which consists of:

- Alternatives a' and a'' if only condition C2 is not satisfied, or
- Alternatives a', a'', … , $a^{(M)}$ if condition C1 is not satisfied; and $a^{(M)}$ is determined by the relation $\mathrm{Q}\left(\mathrm{a}^{(\mathrm{M})}\right) - \mathrm{Q}(\mathrm{a}') < \mathrm{DQ}$ for maximum M (the positions of these alternatives are "in closeness").

3. COMPARING PROMETHEE II, TOPSIS, AND VIKOR ON AN ILLUSTRATIVE EXAMPLE

In this section, an empirical example on which PROMETHEE II, TOPSIS, and VIKOR will be applied and analyzed is introduced. Consequently, the results obtained by these three methods will be presented.

3.1. Illustrative example

For the purpose of this paper, selection of energy supply system for space heating of the residential building will be performed. There are several different energy supply systems for space heating of the residential building that could be used. Natural gas boiler, pellet boiler, coal boiler and heat pump are considered in this case. Heat pump and pellet boiler belong to renewable energy sources, while gas and coal boilers belong to non-renewable energy sources. Those alternatives are compared on several different and conflicting criteria. Literature review on the application of the MCDM techniques to the energy issues shows that evaluation criteria for alternative energy sources can be grouped into four main categories: technical, economic, environmental, and social (Wang et al., 2009). The most used criteria from previous research in this area are selected: investment costs, annual expenses, CO_2 emissions, system efficiency, and comfort of the end users (Kaya & Kahraman, 2011). All relevant information, necessary for the use of MCDM methods, is presented in Table 1. Attributes describing investment costs, annual expenses and CO_2 emissions are based on calculations, system efficiency is taken from the catalog, while comfort of the end user is estimated by energy experts.

Table 1: Parameters for multi-criteria decision analysis

	Criteria	Investment costs	Annual expenses	CO_2 emissions	System efficiency	Comfort of the end user
	Unit	€	€	kg/a	/	5-point
	Min/Max	Min	Min	Min	Max	Max
	Weight	0,30	0,25	0,20	0,10	0,15
Alternatives	Gas boiler	3.500,00	3.850,00	20143	0,90	Very good
	Pellet boiler	3.800,00	3.000,00	500	0,85	Good
	Coal boiler	2.900,00	2.900,00	48644	0,75	Average
	Heat pump	7.400,00	1.950,00	10957	4	Very good

3.2. Results

In this section results obtained by PROMETHEE II, TOPSIS and VIKOR are presented.

3.2.1. PROMETHEE II Results

Complete ranking of alternatives when PROMETHEE II is used for evaluation is presented in Table 2. Preference functions which were used in this case are: V-shape with indifference criterion was used for comparison of alternatives in regards to investment costs and annual expenses; V-shape criterion was used for comparison of alternatives in regards to CO_2 emissions and efficiency; Level criterion was used for comparison of alternatives in regards to comfort of the end user.

Table 2: PROMETHEE II complete ranking of alternatives

	ϕ^+	ϕ^-	ϕ	Rank
Heat pump	0.4182	0.2933	0.1250	1
Pellet boiler	0.2527	0.1291	0.1237	2
Gas boiler	0.2193	0.2293	-0.0099	3
Coal boiler	0.1368	0.3755	-0.2387	4

Ranking of alternatives is as follows: heat pump is the most preferred energy supply system (0.1250), followed by pellet boiler (0.1237), gas boiler (-0.0099), and coal boiler (-0.2387). Renewable energy sources such as heat pump and pellet boiler are better ranked then non-renewable energy sources such as gas and coal boilers. It is also worth noticing that ϕ values of heat pump and pellet boiler are very close to each other and far away from gas and coal boiler values. Therefore, it can be concluded that there is a clear distinction between renewable and non-renewable energy sources.

3.2.2. TOPSIS Results

Ranking of the alternatives when TOPSIS is used for evaluation is presented in Table 3.

Table 3: TOPSIS complete ranking of alternatives

	D_i^*	D_i^-	C_i^*	Rank
Pellet boiler	0.0919	0.2159	0.7014	1
Gas boiler	0.1305	0.1663	0.5603	2
Heat pump	0.1476	0.1812	0.5511	3
Coal boiler	0.2014	0.1478	0.4233	4

Ranking of alternatives is as follows: pellet boiler is the most preferred energy supply system (0.7014), followed by gas boiler (0.5603), heat pump (0.5511), and coal boiler (0.4233). In this case, renewable energy sources such as pellet boiler and heat pump are the first and the third ranked, while non-renewable energy sources such as gas and coal boilers are the second and the fourth ranked. Here, we do not have clear line between renewable and non-renewable energy sources. This is especially evident for gas boiler and heat pump which have very similar C_i^* values.

3.2.3. VIKOR Results

Ranking of the alternatives when VIKOR is used for evaluation is presented in Table 4.

Table 4: VIKOR complete ranking of alternatives

	S_j	R_j	Q_j	Rank
Pellet boiler	0.3701	0.1382	0.0576	1
Heat pump	0.3434	0.3000	0.5000	2
Gas boiler	0.4669	0.2500	0.6121	3
Coal boiler	0.5750	0.2000	0.6910	4

Ranking of alternatives is as follows: pellet boiler is the most preferred energy supply system (0.0576), followed by heat pump (0.500), gas boiler (0.6121), and coal boiler (0.6910). In this case, renewable energy sources such as pellet boiler and heat pump are better ranked then non-renewable energy sources such as gas and coal boilers. Even so, there is no clear difference between renewable and non-renewable energy sources since pellet boiler has far better Q_j value than the other alternatives.

4. DISCUSSION AND CONCLUSION

It is evident from the results presented in Tables 2, 3, and 4 that the order of alternatives varies based on the method used. PROMETHEE II (Table 2) and VIKOR (Table 4) results are the same in regards of renewable and non-renewable energy sources. The difference is in the order of renewable energy sources. PROMETHEE II ranks heat pump as the first and pellet boiler as the second, while with VIKOR rank of these two alternatives is reversed. The difference between TOPSIS (Table 3) and VIKOR is in the reversed order of the second and the third alternative (i.e. gas boiler and heat pump). The greatest difference in ranking of alternatives is between PROMETHEE II and TOPSIS. PROMETHEE II ranks heat pump, pellet boiler, and gas boiler as the first, the second, and the third respectively, while the ranking with TOPSIS is pellet boiler, gas boiler, and heat pump, respectively. The only thing in common for all three methods is the ranking of coal boiler as the most unfavorable choice.

There were some attempts to explain similarities and differences of MCDM methods in the literature. There is an argument that PROMETHEE II gives the same ranking as the ranking based on S_j value in VIKOR, which presents the maximum group utility (Opricovic & Tzeng, 2007). This is proven in the case where PROMETHEE II uses V-shape criterion (linear function) for all evaluations. Analyses in this paper are confirming this fact, with the addition that it is not necessary to use only V-shape criterion for evaluation in PROMETHEE II. In this paper three different types of preference function are used (i.e. V-shape with indifference criterion, V-shape criterion, and Level criterion) and the ranking still complies with the ranking based on S_j value in VIKOR. The difference in the final ranking lies in the fact that VIKOR additionally uses R_j values, which represent minimum individual regret, to evaluate alternatives. This is the main reason for different results of these methods in this particular case. Based on R_j values, pellet boiler is the best alternative while heat pump is the worst. Considering this information, it is expected that pellet boiler is the most preferred option in VIKOR method.

It seems that one of the biggest issues regarding MCDM methods, particularly TOPSIS method, is related to normalization procedure (Çelen, 2014). Comparative analysis of VIKOR and TOPSIS showed that linear normalization of attributes in both methods yields the same results. However, this is not the case if vector normalization is applied in TOPSIS method (Opricovic & Tzeng, 2004). This is particularly important for problems with qualitative attributes that should be transformed into quantitative ones using a Likert scale (Pavlicic, 2001). In the illustrative example presented in this paper, one criterion is qualitative (i.e. comfort of the end users). This could be the possible reason for differences in results of TOPSIS method compared to the results of PROMETHEE II and VIKOR.

Considering results of this study, it can be concluded that decisions are highly dependable on the selected approach for problem solving. This implies that completely different conclusions may be drawn based on the selected method. It should be stressed that differences in results in this particular example are only based on the differences in algorithms of proposed MCDM methods. It is important to emphasize, that there is no intension of authors to favor one method over another. Differences in methods do not necessarily mean that one approach is better than the other. It simply implies that it is important for decision makers to get familiar with different MCDM methods and their specificities so they can apply appropriate method for problem solving based on their preferences. Subjectivity is a part of the decision making process and better understanding of different MCDM methods could help decision makers to make more reliable decisions.

This paper is focused on different approaches of MCDM methods for problem solving. It is shown on an empirical example that results of the decision making process may vary based on the selected approach. All three methods presented in this paper (i.e. PROMETHEE II, TOPSIS, and VIKOR) yielded different ranking of alternatives solely because of the differences in algorithms. This paper contributes to the existing literature by confirming the fact that different MCDM methods can give different results when applied to the same problem.

This research is limited only to differences in algorithms of three highly used MCDM methods. Therefore, we propose that further research should include other MCDM methods (e.g. ELECTRE, AHP, ANP) and other issues in this research field (e.g. changes in criteria weights, introduction of fuzzy logic in MCDM).

REFERENCES

Behzadian, M., Kazemzadeh, R. B., Albadvi, A., & Aghdasi, M. (2010). PROMETHEE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research*, *200*(1), 198–215.

Behzadian, M., Khanmohammadi Otaghsara, S., Yazdani, M., & Ignatius, J. (2012). A state-of the-art survey of TOPSIS applications. *Expert Systems with Applications*, *39*(17), 13051–13069.

Brans, J. P. (1982). Lingenierie de la decision. Elaboration dinstruments daide a la decision. Methode PROMETHEE. In M. Nadeau, R., Landry (Ed.), *Laide a la Decision: Nature, Instrument set Perspectives* (pp. 183–214). Quebec, Canada: Presses de Universite Laval.

Çelen, A. (2014). Comparative analysis of normalization procedures in TOPSIS method: with an application to Turkish deposit banking market. *Informatica*, *25*(2), 185–208.

Chu, M. T., Shyu, J., Tzeng, G. H., & Khosla, R. (2007). Comparison among three analytical methods for knowledge communities group-decision analysis. *Expert Systems with Applications*, *33*(4), 1011–1024.

Guitouni, A., & Martel, J.-M. (1998). Tentative guidelines to help choosing an appropriate MCDA method. *European Journal of Operational Research*, *109*(2), 501–521.

Hwang, C.-L., & Yoon, K. (1981). *Multiple Attribute Decision Making*. *Sprinnger-Verlag, New York, NY*.

Kangas, A., Kangas, J., & Pykäläinen, J. (2001). Outranking Methods As Tools in Strategic Natural Resources Planning. *Silva Fennica*, *35*(2), 215–227.

Kaya, T., & Kahraman, C. (2011). Multicriteria decision making in energy planning using a modified fuzzy TOPSIS methodology. *Expert Systems with Applications*, *38*(6), 6577–6585.

MacCrimmon, K. R. (1972). An overview of Multiple Objective Decision MAking. In *Multiple Criteria Decision Making* (pp. 18–44).

Mardani, A., Jusoh, A., MD Nor, K., Khalifah, Z., Zakwan, N., & Valipour, A. (2015). Multiple criteria decision-making techniques and their applications – a review of the literature from 2000 to 2014. *Economic Research-Ekonomska Istraživanja*, *28*(1), 516–571.

Opricovic, S. (1998). Multi-criteria Optimization of Civil Engineering Systems.

Opricovic, S., & Tzeng, G. H. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *European Journal of Operational Research*, *156*(2), 445–455.

Opricovic, S., & Tzeng, G. H. (2007). Extended VIKOR method in comparison with outranking methods. *European Journal of Operational Research*, *178*(2), 514–529.

Ozernoy, V. M. (1992). Choosing the best multiple criteria decision-making method. *Infor*, *30*(2), 159–171.

Pavlicic, D. (2001). Normalisation affects the results of MADM methods. *Yugoslav Journal of Operations Research*, *11*(2), 251–265.

Vincke, J. P. B. and P. (1985). A Preference Ranking Organisation Method: (The PROMETHEE Method for Multiple Criteria Decision-Making). *Management Science*, *31*(6), 647–656.

Wang, J., Jing, Y., Zhang, C., & Zhao, J. (2009). Review on multi-criteria decision analysis aid in sustainable energy decision-making. *Renewable and Sustainable Energy Reviews*, *13*(January), 2263–2278.

Zanakis, S. H. S. H., Solomon, A., Wishart, N., & Dublish, S. (1998). Multi-attribute decision making: A simulation comparison of select methods. *European Journal of Operational Research*, *107*(3), 507–529.

BALCOR 2018
XIII Balkan Conference on Operational Research

MULTI-CRITERIA DECISION MAKING FOR SMART SPECIALISATION IN SERBIA

ĐURO KUTLAČA[1], LAZAR ŽIVKOVIĆ[2], DIJANA ŠTRBAC[3]

[1] University of Belgrade, Institute "Mihajlo Pupin", Belgrade, djuro.kutlaca@pupin.rs
[2] University of Belgrade, Institute "Mihajlo Pupin", Belgrade, lazar.zivkovic@pupin.rs
[3] University of Belgrade, Institute "Mihajlo Pupin", Belgrade, dijana.strbac@pupin.rs

---

Abstract: *Smart Specialization Strategy (S3) is in process of creation in Serbia. S3 methodology is based on rigorous concept for selection of priority sectors in one region or country, relying on high values of location quotient (LQ>1.5) calculated for (at least) single criterion representing available indicators of economic, scientific and innovation performance. In addition to this concept, valuable for identification of limited number of priority sectors, analysis with multiple criteria which allows creation of ranking list of all active sectors in country based on aggregation using Multiple Criteria Decision Making (MCDM) methods is conducted as well. This paper presents both methodological approach and first empirical results of S3 process in Serbia using Compromise Programming as one among numerous quantitative analytical tools of MCDM.*

Keywords: *Smart Specialization Strategy, MCDM, Serbia.*

1. INTRODUCTION

The Government of the Republic of Serbia has created an Interministerial Working Body (IWB) in early 2017 to develop a Research and Innovation Strategy for Smart Specialization (RIS3). The European Commission's Joint Research Centre (JRC) has supported these processes in partnership with the government of Serbia in developing RIS3.

The first action of the IWB was establishment of the Analytical and Operational Team. The most important task of the operational team is to prepare policy activities and to communicate with government institutions, research and innovation stakeholders, private sector, civil society representatives and JRC. The key activities of the analytical team are data collection, quantitative and qualitative analyses necessary for building a strong evidence-base for RIS3. The JRC has engaged Fraunhofer (FhG) ISI Institute from Karlsruhe as an expert support for the analytical team.

The paper presents some of the methodological aspects as well as the first results of realization of the RIS3 in Serbia achieved in 2017.

2. METHODOLOGICAL NOTES

2.1. The S3 methodology

Methodology for creation of smart specialization strategy is well defined procedure published as a guide for experts and practitioners in EU member countries, following the obligatory status of S3 in EU (S3 Platform, 2012). The concept of "smart specialisation" was developed within the EU research and innovation policy framework by the expert group *Knowledge for Growth* (Foray, D., David, P.A. & Hall, B., 2009) and has rapidly been implemented in EU policy. Elaborated by a group of academics in 2008, it very quickly made a significant impact on the policy audience, particularly in EU, as part of the preparation of the new Cohesion Policy for 2014–2020. Smart specialisation strategies should be used in regions to concentrate resources to a few key priority areas and business sectors. They should also be a key element in developing multi-level governance for integrated innovation policies and have to be closely linked with other policy domains and require an understanding of regional strengths relative to other regions and possible gains for inter-regional and transnational cooperation (European Commission, 2010).

To support national and regional actors in the process of developing RIS3, the European Commission has established the S3 Platform in Seville as support for information, seminars, peer reviews and guidelines providing six practical steps for designing national or regional RIS3 (S3 Platform, 2012):

- Step 1 – Analysis of the national or regional context and potential, in relation to other nations and regions;

- Step 2 – Governance: ensuring participation and ownership – set up an inclusive structure and incentives for securing broad stakeholder involvement;
- Step 3 – Vision – produce a shared vision among stakeholders – elaboration of an overall vision for the region;
- Step 4 – Prioritisation – Identification of priorities – selection of a limited number of priorities for regional development;
- Step 5 – Policy mix – Definition of a coherent policy mix, road maps and action plans – combination of a mixture of policy measures and support them with road maps or action plans to secure implementation;
- Step 6 – Evaluation and monitoring – Integration of monitoring and evaluation mechanisms – developing of systems for continuous and evidence-based monitoring of the process and follow up on results and effects, in order to learn and revise the policy mix.

Implementation of the RIS3 concept in Serbia in year 2017, among organisational issues, practically was realisation of the Step 1 – quantitative analysis of all available statistical data in order to identified priority sectors in economy of Serbia. JRC and FhG experts have developed more instrumentalised procedures for analysis of economic, scientific, and innovative potentials of four out of five statistical regions in Serbia:

- Region RS11: Belgrade
- Region RS12: Vojvodina
- Region RS21: Šumadija and Western Serbia
- Region RS22: Southern and Eastern Serbia
- Region RS23: Kosovo – data are not available.

Detailed roadmap for pursuing smart specialisation in 2017-2018 was adopted jointly by IWB and JRC with five major phases which are above mentioned steps re-arranged for implementation in Serbia (JRC and IWB, 2017):

1. Preparatory stage: preparing organisational resources necessary for effective cooperation.
2. Mapping economic, innovative and scientific potential: multi-dimensional quantitative analysis showing strongest sectors and areas of science at regional level.
3. Entrepreneurial discovery process: qualitative analysis and organized dialogue with business sector, representatives of the research community and public authorities – organized separately for each of the smart specialisation areas.
4. Establish monitoring and evaluation system: a system of input, output, context and result indicators (monitoring) and ex-post impact assessment (evaluation) that will be used to assess progress and results of the implementation of the RIS3; this must be based on a clear intervention logic stemming from the strategy.
5. Develop an implementation system: definition of organizational structure able to deliver the implementation of S3 strategy and sound financing system to enable the implementation.

The second phase which is realisation of the step 1 in original guide has become key activity of the analytical team, JRC and FhG experts. For the purpose of mapping, external experts have defined three sets of indicators and procedures for quantitative analysis. Three sets of indicators are data with detailed subdivision by economic/scientific field of activity, for most indicators according to NACE 3-digit categories (NACE is the acronym for “Nomenclature statistique des activités économiques dans la Communauté européenne”; EC, 2009), for scientific activity according to Web of Science/Frascati classifications (Kroll at all, 2017):

1. Economic Potential
    - employment, according to 2011-2016 labour force survey data
    - exports, according to 2012-2016 national export statistics.
2. Innovative Potential
    - innovating firms, according to the 2010-2014 national innovation survey
    - patents, according to indicators developed by the Mihajlo Pupin Institute, based on data provided by the Intellectual Property Office.
3. Scientific Potential
    - publications, according to indicators developed by the Faculty of Physics and Mihajlo Pupin Institute based on data collected by the Faculty of Physics

Priority sectors in one region are identified according to their *specialisation proper*, i.e. an in relative terms higher importance of the sector in the regional economy than is standard for the economy. Typical measure to determine this specialisation in literature is the *Location Quotient* (*LQ*) that contrasts the share of a sector in the local economy with the share of a sector in the national economy. It can reveal what makes a particular region “unique” in comparison to the national average. Example of calculation of LQ is given in formula 1 for economic indicator of employment:

$$LQ = \frac{\frac{e_{NACE\,X}}{e_{total}}}{\frac{E_{NACE\,X}}{E_{total}}} \tag{1}$$

$e_{NACE\,X}$ = (sectoral) employment in particular region
e_{total} – total employment in particular region
$E_{NACE\,X}$ = (sectoral) employment in the country
E_{total} – total employment in the country

Following the notion of identifying potential priority domains for smart specialisation by focusing on those in which a region is specialised, included are only NACE sectors with a location quotient of more than 1.5 i.e. the ones that hold at least 1.5 times of the share in the regional economy than in the national economy.

Statistical Office of the Republic of Serbia provided data for all requested indicators of economic and innovative potentials which indicate a high level of compliance with international standards. Nevertheless, calculation of the *Location Quotient* for four statistical regions with condition $LQ \geq 1.5$ have produced only few priority sectors performing at least 1.5 times the share in a regional economy than they do in the national economy but only for one or, and in very rear situations, for two indicators out of set of 6 indicators in all three predefined potentials (economic, innovative, scientific). Although situation with small number of identified priority sectors makes process of decision making rather simplified, use of single criterion – one indicator instead of all six indicators – criteria provide rationale for extension of basic S3 approach with implementation of multiple criteria aggregation in mapping exercise.

2.2. The MCDM for S3 methodology

Since decision-making problem involves greater number of criteria that must be respected, it is necessary to use multi-criteria decision-making methods. According to its characteristics, decision-making belongs to multi-criteria decision making that is used in circumstances where there are greater numbers of criteria, in order to get an optimal solution. Multi-criteria analysis will not get the best solution since the ideal solution to the problem of multi-criteria decision making does not exist. It focuses on the so-called "compromise" solutions, which take into account the preferences of all decision-makers, making concessions between the criteria.

MCDM methods are frequently used to solve real-world problems with multiple, conflicting, and incommensurate criteria. Each method provides a different approach for selecting the best among several preselected alternatives (Janic and Reggiani, 2002).

MCDM has grown as a part of operations research, concerned with designing computational and mathematical tools for supporting the subjective evaluation of performance criteria by decision-makers (Zavadskas, Turskis, & Kildienė, 2014). In the decision making approach, the selection is made from amongst the decision alternatives that are described by their attributes. Over time, a large number of MCDM techniques and approaches have been proposed, which are different in their theoretical background and the type of obtained results.

Introducing multiple criteria analysis as extension of the basic S3 quantitative analysis, the Ideal Point Method (IPM) Compromise Programming is applied (Zeleny, 1976). In the following model, preferences are fully cardinal in character, though in a fuzzy and adaptive way. The model is based on the ideal alternative, providing the highest score with respect to all individual attributes considered. Adaptive information gathering and an evaluation process are thus initiated. Partial decisions are made, inferior alternatives removed and post-decision dissonance reduction ensues. We allow for a re-adjustment of attribute weights, displacement of the ideal and new pre-decision conflict formation. Such man-machine interactive procedure leads to a decision in a finite number of iterations.

The model is implemented as follows: a set A of n objects is compared with respect to m criteria. All objects are compared with an object that has ideal values for all m criteria, a so-called ideal (a reference object). A point in m-dimensional space represents each object from the set A. The point representing the ideal object is referred to as the ideal point. The distance d (usually geometrical, as recommended by the author) of each point from the ideal one is calculated (see formula 2). The object that is the nearest to the ideal, i.e., that whose distance from ideal point is the shortest, is the best object. Calculated distances may be corrected by specifying different weights of criteria. The calculated distance could be used for forming a ranking list of objects.

In this method, single indicators (economic, innovation, etc.) for sectors in observed regions are analysed in a coordinate system which axes are just these indicators. The values of a single indicator for sector are the coordinates of the observed sector in the space of available indicators.

The ideal point in this coordinate system is a sector with specially defined values of a single indicator, so it can be referred to as an '*ideal sector'*, or, more appropriate for this analysis, a '*reference sector'*. The value of a single indicator for a reference sector may be defined in several ways, for example: (a) an unachievable, practically unrealisable value, (b) an imagined target value that is hard to realize, (c) a desired, realizable value for a particular indicator, etc.

$$d_i = \sqrt[1/L_p]{\sum_j k_j \times \left(\frac{IC_j - C_{ij}}{(C_{ij})_{max}} \right)^{L_p}} \quad ; \quad j = 1,...,m \ ; \quad i = 1,...,n \qquad (2)$$

where: ***ICj*** - a ***j***-th single indicator for '*reference sector'*, ***Cij*** - a ***j***-th single indicator of an ***i***-th observed sector, k_j - a weighting factor of a ***j***-th single indicator; ***m***– number of single indicators; ***n*** – number of observed sectors; ***Lp*** - used metrics, d_i – calculated distance for ***i***-th sector from *reference sector*. For the case *Lp*=2, the formula becomes a calculation of *Euclidean distance* between the observed and '*reference sector*', and this case is used by analytical team. It is important to note that weighting factors (*k*) are not under the *Lp*-degree (*Lp*=2). This is to avoid dependence of the aggregate distance on the number of indicators.

3. PRELIMINARY RESULTS

Selection of priority sectors using *Location Quotient* for four statistical regions with condition $LQ \geq 1.5$ have produced the following priorities (Kroll at all, 2017):

- Region RS11: Belgrade
  - Computer Programming and ICT
  - R&D and Technical Consultancy
  - Creative Economy
  - Monetary Intermediation
  - Potentially emerging innovative sectors: Beverages, Pharmaceuticals, Electrical Components, Transport Equipment
  - Science based sectors: various

- Region RS12: Vojvodina
  - Automotive
  - Agricultural Economy (including processing industries)
  - Petrochemical Industry
  - Plastics Industry
  - Potentially emerging innovative sectors: Agricultural Machinery, Measurement Instruments
  - Science based sectors: Computer Science, Telecommunications

- Region RS21: Šumadija and Western Serbia
  - Agri-/Horti-/Silvicultural Economy (including processing industries)
  - Automotive
  - Textile Industry
  - Plastics Industry
  - Metal Industry
  - Potentially emerging innovative sectors: Special Purpose Machinery
  - Science based sectors: mechanical engineering, pharmacy

- Region RS22: Southern and Eastern Serbia
  - Agri-/Horticultural Economy (including processing industries)
  - Textile Industry
  - Rubber Industry

- Electrical Engineering
- Potentially emerging innovative sectors: Food Products, Medical and Dental
- Science based sectors: electrical engineering

In addition to above listed priority NACE sectors, use of Compromise Programming have produced five ranking lists of sectors: four ranking lists for regions and fifth ranking list for Serbia (Tables 1, 2, 3, 4 and 5; data for year 2016, excerption – only first ten NACE sectors-groups are listed).

Table 1: Ranking list of NACE sectors-groups in Serbia total – excerption, year 2016

Rank	NACE Sectors – Groups
1.	J62.0 - Computer programming
2.	M73.1 - Advertising
3.	M71.1 - Architectural and engineering activities
4.	A1.1 - Growing of non-perennial crops
5.	G46.9 - Non-spec. wholesale trade
6.	M71.2 - Technical testing and analysis
7.	C28.2 - Manufacture of other general-purpose machinery
8.	C26.5 - Manufacture of instruments and appliances for measuring, testing and navigation; watches and clocks
9.	C10.8 - Manufacture of other food products
10.	M72.1 - R & D - Natural Science

Source: Analytical team, internal working documents, 2017

Ranking list of NACE groups in Serbia based on aggregation (Table 1) shows that the first three positions have the following groups: J62.0 - Computer programming, M73.1 – Advertising and M71.1 - Architectural and engineering activities. Among 10 top-ranked NACE groups the following sectors are prevailing: J - Information and communication (1 group identified ranked as the 1st), M - Professional, scientific and technical activities (4 groups, ranked as the 2nd, 3rd, 6th and 10th), A - Agriculture, forestry and fishing (1 group, ranked as the 4th), G - Wholesale and retail trade; repair of motor vehicles and motorcycles (1 group, the 5th position) and C – Manufacturing (3 groups ranked as the 7th, 8th and 9th).

Table 2: Ranking list of NACE sectors-groups in Region RS11: Belgrade – excerption, year 2016

Rank	NACE Sectors – Groups
1.	C26.3 - Manufacture of communication equipment
2.	J58.1 - Publishing of books etc.
3.	C26.2 - Manufacture of computers and peripheral equipment
4.	C27.4 - Manufacture of electric lighting equipment
5.	D35.1 - Electric power generation, transmission and distribution
6.	C20.2 - Manufacture of pesticides and other agrochemical products
7.	C20.4 - Manufacture of soap and detergents, cleaning and polishing preparations, perfumes and toilet preparations
8.	J62.0 - Computer programming
9.	C26.5 - Manufacture of instruments and appliances for measuring, testing and navigation; watches and clocks
10.	C28.2 - Manufacture of other general-purpose machinery

Source: Analytical team, internal working documents, 2017

In Belgrade region, ranking list of NACE groups based on aggregation indicates that dominant groups are from manufacturing sector (C) and information and communication (J). These results can be considered as partially in line with identified priority sectors using Location Quotient , since NACE groups identified according to Compromise Programming method are also included in two priorities identified using LQ: Computer Programming and ICT and Potentially emerging innovative sectors: Beverages, Pharmaceuticals, Electrical Components, Transport Equipment.

Table 3: Ranking list of NACE sectors-groups in Region RS12: Vojvodina – excerption, year 2016

Rank	NACE Sectors – Groups
1.	C28.1 - Manufacture of general-purpose machinery
2.	C28.3 - Manufacture of agricultural and forestry machinery
3.	C22.2 - Manufacture of plastics products
4.	C25.2 - Manufacture of tanks, reservoirs and containers of metal
5.	C10.4 - Manufacture of vegetable and animal oils and fats
6.	C10.9 - Manufacture of prepared animal feeds
7.	C15.1 - Tanning and dressing of leather; manufacture of luggage, handbags, saddlery and harness; dressing and dyeing of fur
8.	C32.9 - Manufacturing n.e.c.
9.	C23.1 - Manufacture of glass and glass products
10.	C30.9 - Manufacture of transport equipment n.e.c.

Source: Analytical team, internal working documents, 2017

According to the rank list of NACE groups, Region of Vojvodina is characterised with dominant presence of groups exclusively from the manufacturing section. Manufacturing section is also dominant in identified priority sectors using LQ, but they also include Science based sectors: Computer Science, Telecommunications.

Table 4: Ranking list of NACE sectors-groups in Region RS21: Šumadija and Western Serbia – excerption, year 2016

Rank	NACE Sectors – Groups
1.	C20.5 - Manufacture of other chemical products
2.	C25.4 - Manufacture of weapons and ammunition
3.	C26.1 - Manufacture of electronic components and boards
4.	C27.3 - Manufacture of wiring and wiring devices
5.	C28.9 - Manufacture of other special-purpose machinery
6.	C10.5 - Manufacture of dairy products
7.	C25.9 - Manufacture of other fabricated metal products
8.	E38.1 - Waste collection
9.	C14.1 - Manufacture of wearing apparel, except fur apparel
10.	C20.3 - Manufacture of paints, varnishes and similar coatings, printing ink and mastics

Source: Analytical team, internal working documents, 2017

Table 5: Ranking list of NACE sectors-groups in Region RS22: Southern and Eastern Serbia – excerption, year 2016

Rank	NACE Sectors – Groups
1.	C12.0 - Manufacture of tobacco products
2.	C13.2 - Weaving of textiles
3.	C15.2 - Manufacture of footwear
4.	C26.7 - Manufacture of optical instruments and photographic equipment
5.	M75.0 - Veterinary activities
6.	C23.3 - Manufacture of clay building materials
7.	C29.3 - Manufacture of parts and accessories for motor vehicles
8.	J58.2 - Software publishing
9.	C27.3 - Manufacture of wiring and wiring devices
10.	C21.1 - Manufacture of basic pharmaceutical products

Source: Analytical team, internal working documents, 2017

In terms of comparing Compromise Programming method results and priority sectors using LQ, similar conclusions can be made for region of Šumadija and Western Serbia and region of Southern and Eastern Serbia. Namely, ranks of NACE groups in these two regions show dominant presence of manufacturing section, while priority sectors identified using LQ include also other economic sectors: Agri-/Horti-/Silvicultural Economy; Science based sectors: mechanical engineering, pharmacy; Electrical engineering.

4. CONCLUDING REMARKS

Realisation of smart specialisation strategy for the Republic of Serbia is in the second year of activities, but only with completed either first step out of six steps proposed by the S3 Guide, or second phase out of five proposed by the roadmap for Serbia. Third phase is crucial for success of the S3 process; third phase is so called *Entrepreneurial discovery process* (EDP), or dialogue between all stakeholders and qualitative analysis of proposed priority NACE sectors-groups and ranking lists of NACE sectors in regions and for Serbia in total. Dialogue between business sector, research community and public authorities should be organised under the situation of well-structured and fact based decision making process were the “wishes” must be faced with reality presented in figures and facts. The smart specialisation is, in fact, development of sectors of economy based on knowledge, i.e. new industrial policy relying on integration of R&D and innovation with business (European Commission, 2017). Therefore, only NACE sectors already integrated with adequate R&D community which is big and competitive enough to solve all development challenges, could be selected as priority sectors of economy in final S3 document.

Results of applying Compromise Programming method on NACE groups of Serbian economy in 2016 have shown that the first three positions have the following groups: J62.0 - Computer programming, M73.1 – Advertising and M71.1 - Architectural and engineering activities. Among 10 top-ranked NACE groups the following sectors are prevailing: J - Information and communication (1 group identified ranked as the 1st), M - Professional, scientific and technical activities (4 groups, ranked as the 2nd, 3rd, 6th and 10th), A - Agriculture, forestry and fishing (1 group, ranked as the 4th), G - Wholesale and retail trade; repair of motor vehicles and motorcycles (1 group, the 5th position) and C – Manufacturing (3 groups ranked as the 7th, 8th and 9th).

Extension of the single-criterion based selection of NACE sectors (basic S3 process) with ranking list of all NACE sectors-groups created with all available criteria (Compromise Programming aggregation and ranking) which presents economic, scientific and innovative potentials, should be a basis for meaningful EDP dialogue.

Acknowledgement

Research presented in this paper was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, under the project: "Research and Development of the Platform for Science Based Management of the Scientific and Technological Development of the Republic of Serbia", reg. no. III 47005.

REFERENCES

European Commission (2010). Regional Policy Contributing to Smart Growth in Europe 2020. COM (2010) 553 final.

European Commission (2011). Europe 2020 Flagship Initiative Innovation Union SEC(2010) 1161. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, European Union.

European Commission (2012). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions ‘A European strategy for Key Enabling Technologies – A bridge to growth and jobs’. Brussels, 26.6.2012, COM(2012) 341 final.

European Commission (2017). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee, the Committee of the Regions and the European Investment Bank ‘Investing in a smart, innovative and sustainable Industry - A renewed EU Industrial Policy Strategy’. Brussels, 13.9.2017, COM(2017) 479 final.

Freeman, C. (1982). Technological infrastructure and international competitiveness. Draft paper submitted to the OECD Ad hoc-group on Science, technology and competitiveness, August 1982, mimeo. Later published as Freeman, C. (2004). Technological infrastructure and international competitiveness. *Industrial and Corporate Change*, 13: 540-52.

Foray, D., David, P.A. & Hall, B. (2009). Smart Specialisation – The Concept. Knowledge Economists Policy Brief no. 9, Knowledge for Growth Expert Group.
Foray, D. & Ortega-Argilés R. (2011). Smart Specialisation. From Academic Idea to Political Instrument, the Surprising Career of a Concept and the Difficulties in its Implementation. MTEI WP 2011-001.
Janic, M. and Reggiani, A. (2002), An application of the multiple criteria decision making (MCDM) analysis to the selection of a new hub airport, *European Journal of Transport and Infrastructure Research*, Vol. 2 No. 2, pp. 113-41.
JRC and IWB (2017). Smart Specialisation & Organisational Development: Roadmap 2017-2018 for the pilot project with Serbia. Internal working document.
Kutlaca, Dj. (2014). Erawatch Country Reports 2013: Serbia. European Commission Joint Research Centre Institute for Prospective Technological Studies. JRC91226, EUR 26768 EN, ISBN 978-92-79-39503-1 (PDF), ISSN 1831-9424 (online), doi:10.2791/1012. Luxembourg: Publications Office of the European Union.
Kroll, H., Schnabl, E., Horvat, Dj. (2017). Mapping of economic, innovative and scientific potential in Serbia. Fraunhofer ISI, Report commissioned by the JRC, European Commission
Kutlaca, Dj., Semencenko, D. (2015). National Innovation System in Serbia: Past, Present, Future. Publisher: University of Belgrade, Institute "Mihajlo Pupin" - Science and technology Policy Research Centre. ISBN 979-86-82183-14-3, UDK 005.591.6(497.11), 001.896:62(497,11), COBISS.SR-ID 216533516, pages: 151. Belgrade (book is in Serbian language) http://www.pupin.rs/cirnt/wp-content/uploads/2012/12/NIS2-P1-color.pdf
Kutlaca, Dj. (2015). Southeast Europe: Albania, Bosnia and Herzegovina, Croatia, Former Yugoslav Republic of Macedonia, Montenegro, Serbia, Slovenia. Chapter 10 in: *UNESCO Science Report: Towards 2030*. Paris, UNESCO, p. 273-295. http://en.unesco.org/unesco_science_report
Meske, W. (2004). FROM SYSTEM TRANSFORMATION TO EUROPEAN INTEGRATION: Science and Technology in Central and Eastern Europe at the beginning of the 21st century. Werner Meske (Ed.), Lit Verlag Münster.
OECD (1997). National Innovation Systems. Organisation for Economic Co-operation and Development, OECD, Paris.
RI-SSTD (2016). The Strategy of Scientific and Technological Development of Republic of Serbia for the period from 2016 to 2020 - "Research for Innovation". Ministry of Education, Science and Technological Development. http://www.mpn.gov.rs/wp-content/uploads/2015/08/Strategija-engleski-jezik.pdf
S3 Platform (2012). Guide to Research and Innovation Strategies for Smart Specialisation (RIS3). http://ec.europa.eu/regional_policy/sources/docgener/presenta/smart_specialisation/smart_ris3_2012.pdf
Zavadskas, E. K., Turskis, Z., & Kildienė, S. (2014). State of art surveys of overviews on MCDM/MADM methods. Technological and Economic Development of Economy, 20, 165–179
Zeleny, M. (1976). The theory of the displaced ideal. In M. Zeleny (ed.): *Multiple Criteria Decision Making*, Kyoto, Springer-Verlag, Berlin, pp.153-206.

XIII Balkan Conference on Operational Research
BALCOR 2018

ON FINDING A PATTERN-EFFICIENT SET OF NON-DOMINATED VECTORS TO A MULTI-OBJECTIVE OPTIMIZATION PROBLEM

BOGDANA STANOJEVIĆ[1], FRED W. GLOVER[2]

[1] Mathematical Institute of the Serbian Academy of Sciences and Arts, Belgrade, bgdnpop@mi.sanu.ac.rs
[2] College of Engineering & Applied Science, University of Colorado, Boulder, glover@colorado.edu

Abstract: *Pareto optimality is the fundamental construct employed to determine whether a given solution to a multicriteria mathematical optimization model is preferred to another solution. In this paper we describe an approach to generating a pattern-efficient set of non-dominated vectors to a multi-objective optimization problem. We employ an optimization model that aims to yield certain non-dominated vectors that can fill gaps between already generated non-dominated vectors. We describe a way to deal with the adjacency of generated non-dominated vectors, and quantify the gaps between them. We perform some experiments, first on a small illustrative example, then on an instance from the ZDT benchmark, and finally on an instance with 5 objective functions also obtained from the literature; and present graphically the numerical results from applying our method.*

Keywords: *multiple objectives, efficient frontier, pattern efficiency, diversified representations*

1. INTRODUCTION

The efficient frontier associated with Pareto efficiency has found special application in multi-objective (multi-criteria) decision making, where it is used to classify trade-offs among decisions according to their effect on different organizational or business goals. Pareto optimality is the fundamental construct employed to determine whether a given solution to a mathematical optimization model is preferred to another solution.

Non-linear multiple objective optimization has been widely studied, and several surveys can be found in the literature (see for instance Miettinen (1999)).

Das and Dennis (1997) reported some drawbacks of minimizing weighted sums of objectives when generating a Pareto set in multiple objective optimization problems. They provided several examples of convex Pareto curves, and showed the uneven distribution of weights that correspond to an even spread of points on those Pareto curves.

Ray et al. (2001) developed an evolutionary algorithm for generic multiple objective design optimization problems. Founding their approach on non-dominance, they avoided the use of scaling and aggregation that generally negatively affect the conventional penalty function methods for constraint handling.

Asafuddoula et al. (2013) presented a decomposition-based evolutionary algorithm that generates uniformly distributed reference points using systematic sampling, maintaining a fair balance between convergence and diversity by using two independent distance measures, and using a simple preemptive distance comparison scheme for association. They used an adaptive epsilon formulation to deal with the constraints.

Zitzler et al. (2000) provided a systematic comparison of various evolutionary approaches to multiple objective optimization proposing six carefully chosen test problems. Each of their test problems involves a feature that is known to cause difficulty in the evolutionary optimization process. We use one of their instances to illustrate our approach, described subsequently.

Chand and Wagner (2015) presented an overview of the recent developments in many-objective optimization; and suggested some research directions in that field.

The widespread relevance of multi-objective optimization to practical decision making in business and government makes it a central focus of our current study. Within this setting we disclose that classical Pareto efficiency analysis harbors two key limitations that create a need for a better way to evaluate decision trade-offs. The first limitation involves a reliance on assumptions that are often unrealistic in practical settings, causing the approach to exclude solutions that are highly relevant to decision analysis. The second involves the presence of a "blind spot" that makes the approach oblivious to the issue of generating solutions that are diverse in certain critical ways. We aim to generate non-dominated vectors that populate the criterion space in a representative manner without wide gaps between them.

The paper is organized as follows. Section 2 briefly describes the multiple objective optimization problem and introduces basic terminology. This section additionally introduces a special single-objective optimization model able to generate non-supported non-dominated vectors, provides a description of parameter settings, and gives a way to compute the generalized volume of an j-simplex in R^n. Our solution algorithm is introduced in Section 3 and numerical results are presented in Section 4. Finally, Section 5 offers concluding remarks.

2. PROBLEM FORMULATION

We consider the system of m optimization problems given by

$$\max_{x\in X} \left(f_i\left(x\right)\right)_{i\in M}, \tag{1}$$

where $M = \{1,\ldots,m\}$. The vector x is n dimensional with components represented by x_j, $j \in N, N = \{1,\ldots,n\}$ and the solution space X can be convex or non-convex. In a variety of optimization problems of interest, X may be written in the form

$$X = \{x \in R^n | Ax \leq b,\, g(x) \leq d,\, x_j \,\text{discrete},\, j \in D\}, \tag{2}$$

respectively identifying linear constraints, nonlinear constraints and discrete (e.g. integer) variable conditions. When the nonlinear constraints and discrete variable conditions are absent, X is a feasible set of the type customarily encountered in linear programming, and when in addition the function $f_i\left(x\right)$ is linear, the corresponding $i-$th maximization problem is a linear program.

The multi-objective optimization problem consists in finding solutions that solve the system (2) in a special sense, accounting for the need to balance trade-offs between the objectives $f_i\left(x\right)$, given that it is highly unlikely to find a single x vector in X that optimizes each of them simultaneously.

2.1. Terminology

Let $F(x) = (f_1(x),\ldots,f_m(x))$ represent the $m-$vector of objective function values for x. A solution x' is called feasible if $x' \in X$, and is said to dominate another feasible solution x'' if $F(x') \geq F(x'')$ (hence if $f_i(x') \geq f_i(x'')$ for all $i \in M$). Under such conditions, x' strictly dominates x'' if in addition $f_i(x') > f_i(x'')$ for at least one $i \in M$. A feasible solution is **non-dominated** if there is no feasible solution that strictly dominates it. The condition of being non-dominated expresses the condition in which a feasible solution cannot be made better relative to one of the objectives of (2) without being made worse for another. We will use the term **efficient solution** with the same meaning as **non-dominated solution**. Let $(f_1(x'),\ldots,f_m(x'))$ be called a **non-dominated vector** if and only if x' is a non-dominated solution.

Under circumstances where the classical assumptions (e.g. convexity) regarding X and the functions $f_i(x)$ are satisfied, the objective function values of efficient solutions can be visualized in a convenient manner because these $f_i(x)$ values lie on a boundary of a convex set. These $f_i(x)$ values are called **Pareto optimal** and they define the **efficient frontier**.

The use of the efficient frontier as a way of representing objective function values of non-dominated solutions has proved valuable in many applications, and is highly useful in analyzing trade-offs among such solutions in multi-objective optimization. Consequently, it becomes important to consider how to derive an appropriate analog of this representation under the more general conditions encountered in real world settings where the classical assumptions typically fail to apply.

The relevance of this challenge is underscored by the fact that the classical structure of the efficient frontier changes radically in the absence of convexity, or more accurately, loses its meaning. Even in classical settings, a practical representation of an efficient frontier is often not easy to create, and under a variety of circumstances can be highly misleading. The curves constructed by economists are typically based on the supposition that all relevant solutions are known, at least in principle (which is to say, the distribution of their objective function values can be inferred and graphed) – a supposition that is generally not valid.

We can generate non-dominated solutions directly by selecting any given vector c satisfying $c_i > 0$, $i \in M$ and $\sum\limits_{i\in M} c_i = 1$; and solving the following optimization problem over the space of solutions X

$$\max\left\{\sum_{i\in M} c_i f_i\left(x\right) | x \in X\right\}. \tag{3}$$

The efficient solutions x obtained by solving (3) with $c_i > 0$, $i \in M$, are called **supported efficient solutions**. The vector $y = F(x)$ is then called a **supported non-dominated vector**. For more details concerning the terminology of multi-objective optimization we refer the reader to Ehrgott (2005).

The reliance on positive c vectors is not sufficient to identify all non-dominated solutions. As shown by the example in Section 4.1, it may even be that such vectors will only identify a small subset of the efficient solutions, in this case, two out of an infinite number (constituting the points (1,0) and (0,1) out of all points on the line $x_1 + x_2 = 1$ in X space). Consequently, it is desirable to identify formulations that preserve the ability to identify supported non-dominated solutions, but that also make it possible to identify a much larger range of non-dominated solutions. Formulations that can identify all these solutions were provided in Glover (2015).

2.2. Optimization model

We recall from Glover (2015) the formulation (4), the one that showed better results in the experiments, and propose a general algorithm that may identify solutions that provide a diversified coverage of the solution space.

$$\max \left\{ cF(x) + p_0 z_0 + q_0 \sum_{i \in M} c_i \left(f_i(x) - z_0 y_i^0 \right) | x \in X, f_i(x) \geq z_0 y_i^0, i \in M, z_0 \geq 0 \right\} \tag{4}$$

We assume that in (4) p_0, q_0 are scalar constants; z_0 is a scalar variable; c, y^0 are $m-$dimensional constants; and all functions f_i, $i \in M$ are normalized. The key to such a formulation consists of introducing a scheme that penalizes violations of hypothetical lower bounds on y vectors.

In the following we will refer to an objective function vector y^0 without specifying a corresponding solution x_0. We will be interested in a situation where y^0 is generated directly as a specific convex combination of certain y vectors in Y, without generating a point x^0 that yields $y^0 = F(x^0)$. In fact, y^0 may not belong to Y, and hence there may be no x^0 in X that produces y^0. By rules given in the next section, y^0 will be produced for the goal of filling in gaps in the current set of vectors y^0 to give a diversified coverage of the Y space.

2.3. Parameters settings

The c vector used in Model (4) is generated using the same weights that are used to create the target vector y^0. To begin, we assume we have generated the extreme vectors that maximize each function $f_i(x)$ separately, simultaneously using small weights to maximize the sum of the remaining functions given the maximization of a given $f_i(x)$. Note that the determination of weights is done more readily by considering the normalized functions $f_i(x)$, since these are more commensurate in value.

To determine p_0 and q_0, we make use of a term $U_0 = \sum_{i \in M} c_i$. We have chosen to set q_0=1+multq$\cdot U_0$, where multq is a multiple selected from the interval from 0 to 1. (Suggested values for initial testing are 0.1, 0.2 and 1.) Finally, we set p_0 = multp$\cdot U_0$ where multp is chosen to be one of the multiples 5, 10, 20, 50, 100. These of course are simply values suggested for experimentation. If the value multp = 5 or 10 works well, there is no need to look at other values.

The optimal solutions for (4) are assured to lie on the efficient frontier, whether or not the term $cF(x)$ is included in the formulation. Depending on the values of the c_i coefficients, the value of p_0 may need to be increased in relation to q_0 to assure the priority of maximizing z_0.

2.4. Generalized volume of a j-simplex in R^n

The generalized volume of a j-simplex in R^n will be used in the algorithm introduced in the next section. To generate a pattern efficient set of non-dominated vectors to Problem (2), we need to measure somehow the gaps between already generated non-dominated vectors. In fact, our algorithm keeps information about "places" (gaps) where new non-dominated vectors are desired to be generated measuring the generalized volumes of such "places" (gaps) and ordering them accordingly.

For the case $m = 2$, which is very particular, we intend to generate a new non-dominated vector between two adjacent already generated non-dominated vectors. The relation of adjacency for $m = 2$ is evident, since the non-dominated vectors can be lexicographically ordered. Then the parameter y^0 needed in (4) will be chosen as the midpoint of the segment defined by those two consecutive non-dominated points that are the farthest distance from each other.

For the general case, we define the adjacency of m non-dominated vectors in a recursive way: when a new non-dominated vector y^{new} is generated with the help of m non-dominated vectors $y^1, \ldots, y^m$, it becomes adjacent separately to each group of $m-1$ vectors out of those initial m.

To decide which gap to cover first, we need to measure the generalized volume of each gap and keep them ordered, according to their volumes.

We use the Cayley-Menger determinant (Sommerville (1958)) identified in Formula (5), to compute the generalized volume of a j-simplex S in R^n formed with $j+1$ vertexes $v_1, v_2, \dots, v_{j+1}$:

$$V_j^2(S) = \frac{(-1)^{j+1}}{2^j (j!)^2} \det\left(\widehat{B}\right), \tag{5}$$

where $\widehat{B}$ is a $(j+2)\times(j+2)$ matrix obtained from the matrix $B = (\beta_{ik})_{i=1,\dots,j+1}^{k=1,\dots,j+1}$, $\beta_{ik} = ||v_i - v_k||_2^2$, bordered with a top row $(0,1,1,\dots,1)$ and a left column $(0,1,1,\dots,1)^T$.

In the general case of non-degeneracy of Problem (2) (i.e. the number of the objective functions is $m \geq 3$ and the number of distinct marginal non-dominated vectors is also m) we must compute the generalized volume of a $(m-1)$-simplex in R^m.

3. THE SOLUTION ALGORITHM

In this section we present the solution algorithm that provides a pattern-efficient set of non-dominated vectors to a multi-objective optimization problem. In the initialization step, the algorithm finds the non-dominated marginal vectors. Then, it uses adjacency as described in the previous section to quantify and record the gaps between non-dominated vectors already generated. Finally, the method generates new non-dominated vectors by solving single-objective optimization problems until either no gap remains or the desired number of non-dominated vectors is reached.

Data describing an instance of the multiple objective problem (1) together with proper values for the additional parameters needed in Model (4) are inputs for the solution algorithm.

1. Initial stage – generate m marginal non-dominated vectors $y^1, \dots, y^m$ and save them in the list (G) of generated non-dominated points. Define their corresponding coefficients $c^1, \dots, c^m$. The coefficient c^i is a vector of size m, containing a value close to 1 in the $i-$th position, and values close to 0 in the rest. Using it in (3) produces the $i-$th extreme solution.
2. Save problem $P_0(V, y^1, \dots, y^m, c^1, \dots, c^m)$ in the list (L) of problems prepared to be solved further. The argument V is the value of the generalized volume of the m-simplex defined by the vectors $y^1, \dots, y^m$. Details about how to compute the generalized volume are given in Section 2.4.
3. Set the maximum number of iterations $maxit$.
4. For $it = \overline{1, maxit}$ do:
   If L is not empty, then

   a. Load P_0, the first problem from the list L, and compute the corresponding parameters p_0 and q_0.
   b. For $i = \overline{2, m}$ do:
      For each distinct combination of i out of m indexes, collected in the set M_i, do:
      - Set $y^0 = \left(\sum\limits_{k\in M_i} y^k\right) / (m-i)$ and $c^0 = \left(\sum\limits_{k\in M_i} c^k\right) / (m-i)$.
      - Use Model (4) to obtain a new non-dominated vector y^{new}. If $y^{new} \notin G$, then add y^{new} to G; and add m new problems to L, as it follows. The j-th new problem, $j \in M$, is obtained from Problem P_0 loaded from L, replacing the vector y^j by y^{new}, c^j by c^0, and computing the generalized volume V of the new m-simplex.
   c. Remove the first problem from the list L;
   d. Order the list L with respect to the generalized volume V associated with each problem.

   Otherwise, break.

The output of the algorithm is the list G of the non-dominated vectors generated. The algorithm stops either after the maximum number of iterations is reached or when the list of problems L becomes empty.

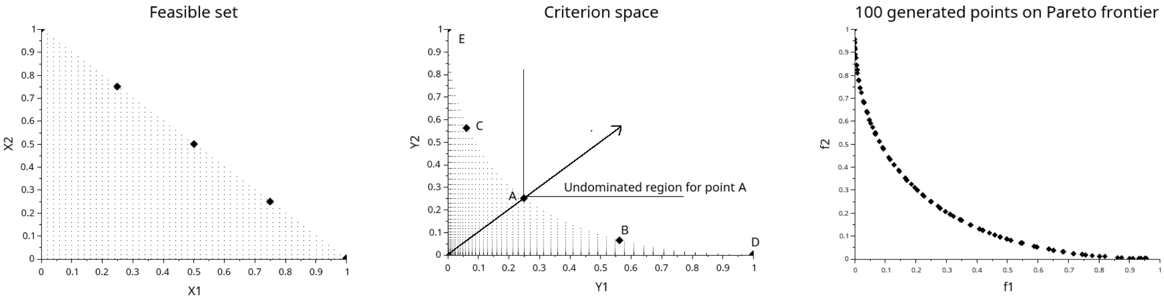


Figure 1 Graphic representation of the feasible set, the criterion space, and 100 non-dominated vectors on the Pareto frontier of Problem (6). The vectors D and E are supported non-dominated vectors, while A, B and C are non-supported non-dominated vectors.

4. NUMERICAL RESULTS

4.1. Small illustrative example

To emphasize the difficulties caused by the non-convexity of the Pareto front for weighted-sum based algorithms, we include the following example in the presentation of our numerical results. Consider the $2-$dimensional problem

$$\begin{array}{ll} \max & f_1(x) = x_1^2, \\ \max & f_2(x) = x_2^2, \\ \text{s.t.} & x_1 + x_2 \le 1, \\ & x_1, x_2 \ge 0. \end{array} \tag{6}$$

Although the feasible set is convex, the functions x_1^2 and x_2^2 are convex rather than concave (as is required in the case of maximization). As a result, the problem has only two supported efficient solutions: $x_1 = (1,0)$ and $x_2 = (0,1)$.

Let us consider now the relation of these two points to the feasible point $x = (1/2, 1/2)$. All three points yield non-dominated solutions, in that there exist no points in X that yield objective function values that dominate those produced by these solutions. From the fact that $F(x_1) = (1,0)$, $F(x_2) = (0,1)$ and $F(1/2,1/2) = (1/4,1/4)$ we see that $F(x)$ lies strictly below the line joining $F(x_1)$ and $F(x_2)$ and hence $y = F(x)$ cannot be obtained by solving (3). (Point A is under the line segment defined by D and E in Figure 1 (in the middle).)

Figure 1 also shows the region consisting of points that are non-dominated relative to the point A. It may be noted that this region ($y_1 \ge 0.25$, $y_2 \ge 0.25$) is the region that results by shifting the position of the point A to the origin. The line from the origin through the point A is also shown in Figure 1. This is the line that gives rise to the variable z_0 in formulation (4).

4.2. Zitzler–Deb–Thiele's problem ZDT2

We tested our algorithm on problem ZDT2 (introduced by Zitzler et al. (2000)) from the standard benchmark ZDT.

$$\begin{array}{ll} \min & f_1(x) = x_1 \\ \min & f_2(x) = g(x) \cdot h(f_1(x), g(x)) \\ \text{s.t.} & x_i \in [0,1],\ i \in \{1, \dots, 30\} \end{array} \tag{7}$$

where $g(x) = 1 + \frac{9}{29} \sum_{i=2}^{30} x_i$, and $h(f_1(x), g(x)) = 1 - \left(\frac{f_1(x)}{g(x)}\right)^2$. The ZDT2 has 30 decision variables and its bi-objective function has a non-convex Pareto frontier. Our numerical results are presented in Figure 2. The cover of the Pareto frontier defined by the generated non-dominated points is obviously pattern efficient, in all presented cases.

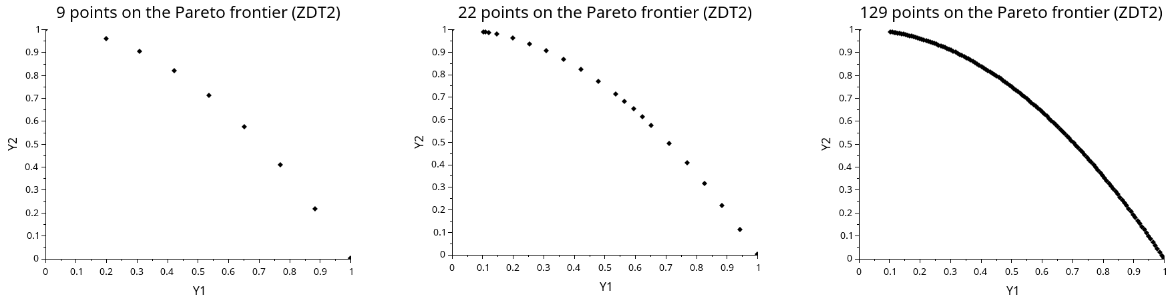


Figure 2 Generated Pareto frontiers for Problem (7)

Table 1: Parameters values used in the algorithm to obtain the 1510 non-dominated points to Problem (8)

Truncation on x, when memorizing	3 decimals
Truncation on y, when memorizing	2 decimals
Distance between two points, considered the same	0.6
Stop adding problems to the list	"generalized volume" < 0.5
No. of generated points	1510

4.3. Water Resource Management Problem

The Water Resource Management Problem (8) with five objective functions and three variables is introduced in (Ray et al. (2001)) and recalled in (Asafuddoula et al. (2013)).

$$\begin{array}{ll}
\min & f_1(x_1,x_2,x_3) = 106780.37(x_2+x_3)+61704.67 \\
\min & f_2(x_1,x_2,x_3) = 3000x_1 \\
\min & f_3(x_1,x_2,x_3) = \frac{(305700)2289x_2}{[(0.06)2289]^{0.65}} \\
\min & f_4(x_1,x_2,x_3) = (250)\,2289e^{-39.75x_2+9.9x_3+2.74} \\
\min & f_5(x_1,x_2,x_3) = 25\left(\frac{1.39}{x_1x_2}+4940x_3-80\right) \\
\text{s.t.} & g_1(x_1,x_2,x_3) = \frac{0.00139}{x_1x_2}+4.94x_3-0.08 \le 1 \\
& g_2(x_1,x_2,x_3) = \frac{0.000306}{x_1x_2}+1.082x_3-0.0986 \le 1 \\
& g_3(x_1,x_2,x_3) = \frac{12.307}{x_1x_2}+49408.24x_3+4051.02 \le 50000 \\
& g_4(x_1,x_2,x_3) = \frac{2.098}{x_1x_2}+8046.33x_3-696.71 \le 16000 \\
& g_5(x_1,x_2,x_3) = \frac{2.138}{x_1x_2}+7883.39x_3-705.04 \le 10000 \\
& g_6(x_1,x_2,x_3) = \frac{0.417}{x_1x_2}+1721.26x_3-136.54 \le 2000 \\
& g_7(x_1,x_2,x_3) = \frac{0.164}{x_1x_2}+631.13x_3-54.48 \le 550 \\
& 0.01 \le x_1 \le 0.45 \\
& 0.01 \le x_2 \le 0.10 \\
& 0.01 \le x_3 \le 0.10.
\end{array} \tag{8}$$

Observing some particularities, we may eliminate one objective function (due to its correlation with another one) and one variable (due to monotony). In this way, all constraints acquire the same form. Thus, we focus on solving the following reduced problem (9). We extend the resulting solutions to provide solutions to Problem (8) in order to compare our results to those from the literature.

$$\begin{array}{ll}
\min & f_1(x_1,x_2,x_3) = 3000x_1, \\
\min & f_2(x_1,x_2,x_3) = 28534690\cdot x_2, \\
\min & f_3(x_1,x_2,x_3) = 572250e^{-39.75x_2+2.839}, \\
\min & f_4(x_1,x_2,x_3) = \frac{34.75}{x_1x_2}-765, \\
\text{s.t.} & x_1x_2 \ge 0.0013487, \\
& 0.01 \le x_1 \le 0.45, \\
& 0.01 \le x_2 \le 0.10.
\end{array} \tag{9}$$

The parameters set in our algorithm for solving Problem (9) are shown in Table 1.

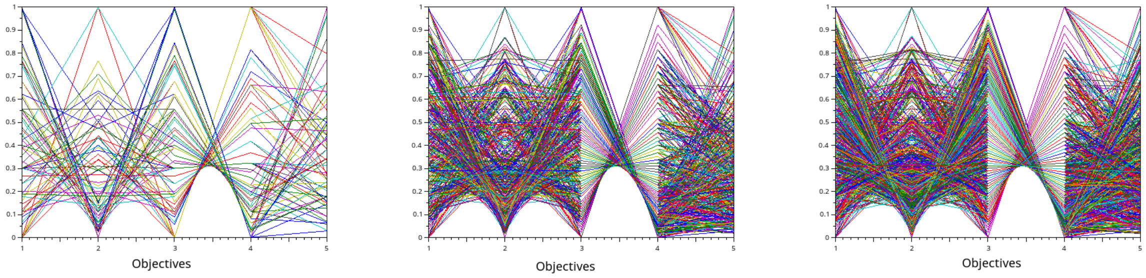


Figure 3 The non-dominated vectors generated for Problem (8) obtained by running 10, 100 and 200 iterations respectively

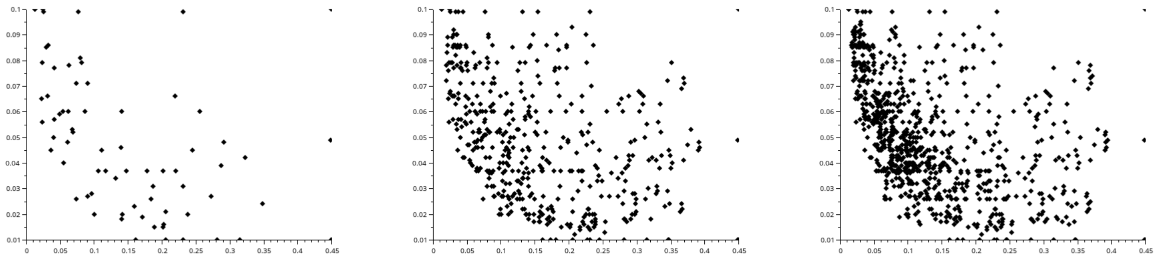


Figure 4 Efficient solutions generated for Problem (8), in the plane x_1Ox_2 with $x_3 = 0.01$, obtained by running 10, 100 and 200 iterations respectively

To analyze the extent to which the sets of generated points are "pattern efficient", we limited the number of iterations of the main loop of the algorithm to 10, 100, 200, and 1000 respectively.

The graphic representations of the generated sets in the cases 10, 100 and 200 are presented in Figures 3 (giving objective values) and 4 (giving efficient solutions). When 1000 iterations are allowed, then 1510 points are generated and the algorithm stops because the list of problems becomes empty. These points are presented in Figure 5 together with the reference points found in the literature. The non-dominated vectors belonging to the reference Pareto front were obtained using the JMetal framework for multi-objective optimization with meta-heuristics (Durillo et al. (2010)) and can be found at the web address "http://jmetal.sourceforge.net/problems.html".

5. CONCLUDING REMARKS

To provide an effective aid to multiple criteria decision analysis, it is essential to be able to populate the criterion space in a representative (pattern-efficient) manner, without producing knots or clumps of non-dominated vectors with wide gaps between them.

To find a pattern-efficient set of non-dominated vectors for a general multi-objective optimization problem, we employed a single-objective optimization model that derives certain non-dominated vectors that can fill

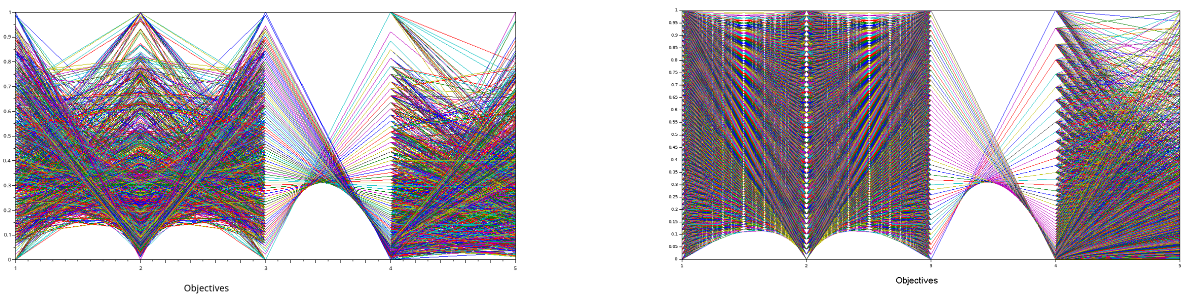


Figure 5 The representation of the 1510 generated non-dominated points (to the left), and 2429 reference points from the literature (to the right) for Problem (8)

gaps between already generated non-dominated vectors. We described a way to deal with the adjacency of non-dominated vectors, quantified the gaps between them, and proposed a solution algorithm for generating these vectors. We also performed some experiments to demonstrate the processes involved, first on a small illustrative example, then on an instance from the ZDT benchmark, and finally on an instance with 5 objective functions also obtained from the literature. The numerical results from applying our method were also presented graphically.

Our experiments showed that our solution algorithm succeeds in covering the Pareto front of several instances obtained from the literature in a pattern-efficient manner. A refined tuning of the algorithm's parameters may be expected to produce still better results.

Further research is in progress to study the application of our approach to instances where the Pareto sets have more complicated shapes. We are also investigating the use of pseudo-randomization strategies to improve the running time of our algorithm for large scale problems.

Acknowledgments

This research was partially supported by the Ministry of Education and Science, Republic of Serbia, Project number TR36006.

REFERENCES

Asafuddoula, M., Ray, T., and Sarker, R. (2013). A decomposition based evolutionary algorithm for many objective optimization with systematic sampling and adaptive epsilon control. In Purshouse, R. C., Fleming, P. J., Fonseca, C. M., Greco, S., and Shaw, J., editors, *Evolutionary Multi-Criterion Optimization*, pages 413–427, Berlin, Heidelberg. Springer.

Chand, S. and Wagner, M. (2015). Evolutionary many-objective optimization: A quick-start guide. *Surveys in Operations Research and Management Science*, 20:35–42.

Das, I. and Dennis, J. (1997). A closer look at drawbacks of minimizing weighted sums of objectives for pareto set generation in multicriteria optimization problems. *Structural optimization*, 14(1):63–69.

Durillo, J., Nebro, A., and Alba, E. (2010). The jmetal framework for multi-objective optimization: Design and architecture. In *CEC 2010*, pages 4138–4325, Barcelona, Spain.

Ehrgott, M. (2005). *Multicriteria Optimization*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Glover, F. (2015). Pattern-efficient frontiers for multi-objective optimization and implications for visualization. Research report, OptTek Systems.

Miettinen, K. (1999). *Nonlinear Multiobjective Optimization*. Kluwer Academic Publishers, Boston.

Ray, T., Tai, K., and Seow, K. (2001). Multiobjective design optimization by an evolutionary algorithm. *Engineering Optimization*, 33(4):399–424.

Sommerville, D. M. Y. (1958). *An Introduction to the Geometry of n Dimensions*. Dover, New York.

Zitzler, E., Deb, K., and Thiele, L. (2000). Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. *Evolutionary Computation*, 8(2):173–195.

BALCOR 2018
XIII Balkan Conference on Operational Research

FORECASTING THE ACCESSION OF THE REPUBLIC OF SERBIA TO THE EUROPEAN UNION BY USING THE ANALYTIC NETWORK PROCESS

ANA KRSTIĆ, JOVANA SAVIĆ[2], PREDRAG MIMOVIĆ[3]

[1] University of Kragujevac, Faculty of Economics, anakrstic@kg.ac.rs
[2] University of Kragujevac, Faculty of Economics, jokasavic92@gmail.com
[3] University of Kragujevac, Faculty of Economics, mimovicp@kg.ac.rs

---

Abstract: *The acquisition of full membership of Republic of Serbia in the European Union depends on a large number of factors, which is why, instead of accurate, the date of accession is mainly stated. The paper first describes the analytic network process (ANP), which is then applied in order to predict the most probable date of Republic of Serbia's entry into the European Union taking into account the economic, political and legal conditions that Republic of Serbia must fulfill in the stabilization and association process. The model is based on negotiating chapters that are considered crucial in the process of Eurointegration and their interactions and relationships, both internal and external. In addition to confirming the success of the ANP method for predicting in a highly complex and uncertain area, research results can help relevant institutions to make better strategic decisions in the process of Eurointegration.*

Keywords: *Eurointegration, prediction, accession date, analytic network process, Republic of Serbia, European Union.*

1. INTRODUCTION

Modern business conditions are characterized by high uncertainty and complexity, which is why the decision-making process is rather difficult and focused not on the best, but on the optimal solution. The choice of an optimal solution depends on a large number of heterogeneous criteria and their interdependencies. In the conditions of multicenter decision making, many methods can be used, and one of the most famous is the analytic hierarchical process (AHP). This method is based on a clear hierarchical structure and division of goals, criteria and decision-making alternatives, as well as neglecting the impacts and relationships that can exist between criteria, which can lead to decisions that are not optimal. In order to expand the scope of application of the analytic hierarchical process on the decision problems in conditions of interdependence and feedback between and within clusters, an analytic network process (ANP) has been developed. Many decision problems cannot be structured in the form of a hierarchy precisely because of the existence of interactions and dependencies between the elements at the higher and lower levels of the hierarchy, which makes the ANP structured in the form of a network (Saaty, 2001a). This approach allows us to more fully comprehend and encompass complex relationships, especially in situations characterized by risk and uncertainty, with a two-way comparison of pairs (Mimović, Talijan, 2017). In addition to solving decision problems, ANP is often used as an instrument for prediction in different areas (Niemira, Saaty, 2004; Ozorhon, Dikmen, Birgonul, 2007; Voulgaridou, Kirytopoulos, Leopoulos, 2009; Azis, 2010; Adamus, 2010; Saaty, Vargas, 2006, 2013).

In accordance with the all above, the subject of the research in this paper refers to the possibility of applying an analytic network process for predicting the date of acquiring full membership i.e. the accession of the Republic of Serbia to the European Union. The European Union, as one of the most powerful regional communities at the moment, is characterized by numerous political, economic and other complex problems stemming from the relations between member states and candidates which affect the survival of the European Union. Therefore, the acquisition of the status of a full member of the European Union depends not only on how quickly the reforms and harmonization of regulations in the Republic of Serbia are adjusted to the legal heritage of the European Union, but also how the other countries are satisfied with the progress of the Republic of Serbia as a candidate for the membership. Starting from the economic, political and legal conditions which the Republic of Serbia must fulfill in the process of stabilization and association, the aim of the research is to use the analytic network process to predict the most probable date of the Republic of Serbia entering the European Union. In accordance with the defined object and goal, the paper starts from the assumption that the implementation of the analytic network process, along with the respect of the economic, political and legal conditions included in the negotiating chapters that must be fulfilled, will enable a more accurate forecast of the date of the Republic of Serbia's entry into the European Union.

The paper is designed in the following way: the first part of the paper includes a review of the literature in which decision and prediction problems were presented using multi-criteria prediction methods or analytic hierarchical processes, and its extensions, analytic network process. In the second part of the paper, the research methodology is presented, first by describing the analytic network process, and then by describing the prediction problem and the construction of the model itself. The results of the model are presented in the third part, while the last part refers to concluding observations, work restrictions and guidelines for the future research.

2. LITERATURE REVIEW

To solve the problem of decision-making in the conditions of risk, uncertainty, dynamics and complexity, the ANP method has wide application. One of the important areas of application of the ANP method is to evaluate the performance and the choice of strategy. Domanović and Associates (2014) integrate the Balanced Scorecard model and ANP to measure performance and evaluate the strategy on the example of selected companies in Serbia. In this way, the authors showed that through the combination of these two approaches, an optimal business strategy can be chosen depending on business performance. A similar approach was adopted by Santos and Associates (2015) for evaluating the performance of the public health system in the city of mid-size in Brazil, who came to the conclusion that the impact of finance is most significant for achieving the best performance. The choice of an optimal strategy based on the ANP methodology was carried out by Wu and Lee (2007), in particular knowledge management strategies, and Jayant (2016) for choosing a strategy for managing green supply chains. Mimović and Talijan (2017), who have identified and evaluated the key factors for formulating the optimal business strategy and sustainable development of the Memorial Park, "Kragujevački oktobar", applied the combination of ANP methods and SWOT analysis. The ANP proved to be very suitable for selecting the best supplier (Bayazit, 2006; Tseng, Chiang, Lan, 2009), staff (Dağdeviren, Yüksel, 2007; Lin, 2010), optimal construction projects (Lin, Yang, 2016), development (Meade, Presley, 2002) and others.

In addition to the previously mentioned areas, ANP, as noted in the introductory part, is widely applied as a predictive instrument based on the distribution of the relative probabilities of future outcomes. Mimović (2012), on the example of predicting sales of the new car model, the Fiat 500L showed that the ANP can be used very successfully to structure the influence of various factors on the final outcome of the prediction process. Sales forecasting using ANP was carried out by Voulgaridou et al. (2009) who envisaged the volume of sales of the new product (the new edition of the book), as well as Shih and Associates (2012) who have been researching the scope of sales of printers in Taiwan. Ozorhon et al. (2007) applied the ANP to predict the performance of international construction joint ventures. Using the ANP model, the authors examined the determinants of the performance of these joint ventures and came to the conclusion that the most important determinants of their success are the relationships between partners, structural factors and partner compliance, primarily cultural. The ANP proved to be a good model for predicting the probability of failure in the business of construction companies in Turkey, including internal and external factors that determine the current business situation of the company and the links between them (Dikmen, Birgonul, Ozorhon, Egilmezer Sapci, 2010). In the field of economic forecasting, the ANP method is very often used to predict the likelihood of a financial crisis (Niemira, Saaty, 2004; Saaty, Vargas, 2006, 2013) or predicting the recovery of the economy (Saaty, Vargas, 2006, 2013; Blair, Mandelker, Saaty, Whitaker, 2010; Azis, 2010). Interesting research by the authors (Adamus, 2010; Saaty, Vargas, 2013) was carried out in connection with the forecasting Poland's entry into the Euro zone, which gave the result that for Poland the optimal alternative is the late entry (after 2011). Political problems are also common in the application of the ANP method. Saaty and Vargas (2006, 2013) by applying this method dealt with forecasting possible ways to overcome the conflict between China and Taiwan and America's response to North Korea's nuclear threats. There are no many sources of literature related to use of these methods for political decisions making. Saaty, Vargas and Zoffer (2015) have deal with this topic.

3. METHODOLOGY

3.1. Analytic Network Process

Due to the limited use of the AHP method for decision-making in terms of risk and uncertainty due to its inability to encompass complex relationships, Thomas L. Saaty (2001c) developed the ANP method as an extension of the AHP method. The essential concept of the ANP method is the impacts i.e. the interactions that exist between elements and clusters, which can be inner (between cluster elements) and outer (between clusters). In fact, this indicates that in this method it is not necessary to clearly define the levels of hierarchy, but that there is a network consisting of elements and influences. Therefore, the ANP model consists of two parts: the first part consists of a control hierarchy or a network of criteria and sub-criteria, which control the interactions in the system being studied, while the other part represents a network of influences between elements and clusters (Mimović, 2012). It is important to note that one ANP model can have more than one network of influences. The use of the ANP method often involves reliance on the control hierarchy, i.e. the control system of profit, costs, chances and risks, where the synthesis of results obtained from control systems is first performed by calculating the profit margin and the chances and costs and risks for each individual alternative, and then normalize the result at the level of all alternatives in order to achieve the best outcome (Saaty, 1999). The procedure for applying the ANP method involves the following five steps (Domanović, Mimović, Jakšić, 2014):

1. the decomposition of the problem – A decision problem is decomposed into its main components.
2. the cluster formation for the evaluation – After defining the decision-making objectives, it is also necessary to generate clusters for the evaluation purpose by a criterion, sub-criterion (if it is possible) and cluster alternative.
3. the structuring of the ANP model – The ANP is applied to different decision-making problems in the field of marketing, health, politics, military issues, society, predictions, etc. Their accuracy of forecasting proved in impressive applications in the field of economic trends, sports events and other events, whose outcome became known later.
4. a paired comparison and prioritization – In this step, it is necessary to compare the pairs of elements of decision-making as well as the synthesis of priorities for all the alternatives. The estimations are made by a fundamental scale 1-9 (Table 1), which the comparative study showed simulate human thinking most adequately.
5. the sensitivity analysis of the solution – It is finally possible to make a decision and carry out a sensitivity analysis in terms of the impact which, according to the importance of some criteria or sub-criteria, a final outcome has on a given solution; it is also possible to determine how big or small these indicators are through an analysis.

Table 1. Fundamental scale

1	Equal importance
3	Moderate importance
5	Strong importance
7	Very strong or demonstrated importance
9	Extreme importance
2, 4, 6, 8	Mean values
	Use reciprocal values for inverse comparison

Source: L Saaty, T. (2008). The analytic network process. *Iranian Journal of Operations Research*, *1(1)*, p.3

In relation to AHP, the ANP method allows each decision to appear in the form of a network rather than a strict hierarchy. Hierarchical decisions can be subjective and predetermined due to the imposed structure, while matrix decisions, whose structure includes dependencies and feedback, represent the real world more realistically and realistically, which makes the ANP method a more effective tool in relation to AHP for deciding in practice (Saaty, 2001b). Additionally, ANP is successfully applied in many areas of forecasting, since allows quick incorporation of feedback and simple comparison with actual results (Mimović, 2012). However, it is necessary to look at the deficiencies of this model, such as high complexity, generally and in comparison with the AHP method as well, comparison of clusters can often be unclear and confusing, the decision-making process is longer than in the AHP method, which is why it is more often used in relation to ANP methods, imprecise results can be obtained, etc.

3.2. Problem description and the construction of ANP model

In paper we use ANP, in order to predict the date of the Republic of Serbia's entry into the European Union, whose model is based on negotiating chapters that are grouped into three clusters: economic, political and legal conditions. Fulfilling these conditions is considered crucial for the European integration process for many reasons. According to the National Convention (2016) on the Accession of the Republic of Serbia to the European Union, the harmonization of the legislation of the candidate country with the regulations of the Union implies fulfillment of the first, political criteria and other economic criteria, i.e. the democratic organization of the state, stability of institutions, rule of law, protection of human and minority rights and existence functional market economy able to withstand the pressure of competition in the unified market of the European Union. The political issue of special importance refers to the normalization of relations with Kosovo, which, in the opinion of many analysts, including the Council of the European Union, together with the chapter on the rule of law, is considered fundamental to the negotiation process. Also, this chapter is one of the key aspects of foreign, security and defense policy. As a candidate, the Republic of Serbia is obliged to adopt the legal acquis of the European Union, which is why the most important issue is the harmonization of domestic legislation with this law and the construction of institutions that will ensure its implementation. Consequently, the most complex chapters of accession are chapters 23 and 24 dedicated to justice, human rights, justice, freedom and security (EU Info Center, 2014) which are called "the rule of law" in one name. Regarding legal requirements, the competition chapter is one of the most demanding and complex in the negotiation process, and this is one of the chapters for which transitional deadlines are often sought, mostly due to state aid policy (EU Info Center, 2015). Related to this chapter is the chapter on the right of companies that defines the conditions that will allow economic entities equal treatment in the European Union market. From the economic perspective, the Republic of Serbia has to take certain measures regarding the improvement of the investment climate and liberalization in foreign trade policy. According to the European Commission, the economic problems of the Republic of Serbia represent a high budget deficit and the absence of more stringent financial controls, as well as poor conditions in the labor market and high unemployment, while trade integration with the European Union has remained high, thus smooth implementation of the Transitional Trade Agreement continued (Rapaić, Dabić, 2013). An economic chapter of vital importance for the integration process relates to the principle of the free movement of goods, since it is the principle on which the European Union is based (EU Info Center, 2014). In the context of accession to the Eurozone, an important chapter relates to the regulation of economic and monetary policy and the adoption of the euro as the national currency. Another important economic aspect is the functioning of a customs union without which the EU's Common Trade Policy and Development, its common agricultural market and the effective coordination of economic and monetary policies would not be possible (EU Info Center, 2015). As for the definition of the accession date, which was widely spoken in 2017, according to Jean-Claude Juncker, President of the European Commission, in his mandate ending in 2019 there will be no enlargement of the European Union, which in turn indicates there is no room for entry Republic of Serbia before that period. An important indicator of entry is the European Union budget that is being prepared for the period 2021-2026, based on which it will be known whether the European Union plans to be expanden in that period. According to Jadranka Joksimović, the Minister for Eurointegration, the definition of a more precise time of entry is important for the planning of activities and budgets related to certain areas such as environmental protection, agriculture or food safety. Bearing in mind this statement, the fourth cluster is defined, which includes the alternative periods of the Republic of Serbia's entry into the European Union.

Starting from the description of the problem of forecasting and the current situation in which the Republic of Serbia is in the process of Eurointegration, an ANP model can be formed that includes the following clusters and their elements:

- Cluster *economic conditions* including negotiating chapters relating to freedom of movement of goods, economic and monetary policy, customs union and social policy and employment;
- Cluster *legal conditions* covering chapters related to competition policy and company law;
- Cluster *political conditions* including chapters of the judiciary and basic rights, justice, freedom and security, foreign, security and defense policy and relations with Kosovo;
- Cluster *alternative* that refers to potential time intervals of Serbia's entry into the European Union: between 2025 and 2030, after 2030, and the possibility that Serbia will not become a member of the European Union.

In accordance with the publication published at the time of the completion of the first phase of the negotiations, of the screening process in 2015, the conditions defined in the negotiation chapters taken in this paper are (see EU Info Center, 2015).

Interactions and relationships that exist between and within clusters, which are important for the comparison of pairs, can be determined based on the insight into the description of the problem. These are the following links:

- Clusters of *economic*, *political* and *legal conditions* affect the cluster of alternatives;
- Within the cluster, *economic*, *political* and *legal conditions* have internal dependence;

- Between cluster *economic conditions* and *political conditions* there is a two-way interdependence;
- Between the cluster *economic conditions* and *legal conditions* there is a two-way interdependence;
- Between the cluster *legal conditions* and *political conditions* there is a two-way interdependence;

Starting from the all stated above, the ANP model is shown in Figure 1.

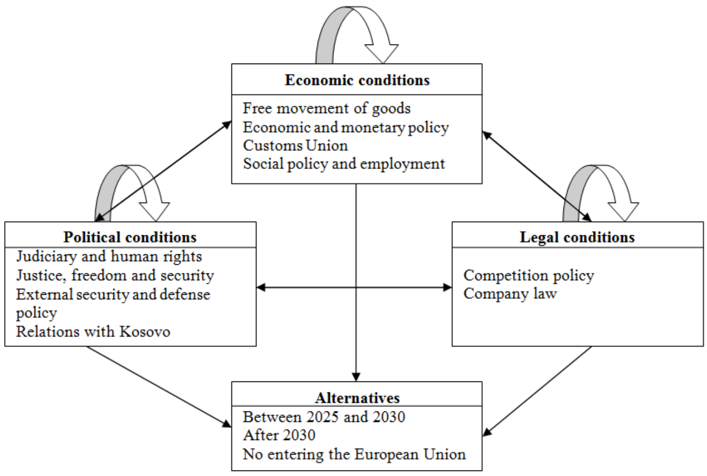


Figure 1. ANP model of the prediction of the indicative date of Serbia's entry into the EU
Source: Authors

3.3. Model Results

In Figure 1, the ANP model of the prediction of the indicative date of Serbia's entry into the EU is presented, whose structure is made up of clusters, elements and connections between them. These links, represented by arrows, indicate the direction of influence between the model elements established when structuring and modeling the problem. Comparison of pairs of model elements, clusters, or elements within the same cluster, or between different clusters, is done normally using the scale 1-9 (Table 1). Pairs' comparisons are basic to the AHP / ANP methodology. When comparing a few factors, one can determine the rationale of the relative importance, preference, or probability of these factors, depending on the need. This ratio is the ratio of two factors to be compared. Comparison of factors, or sub-criterion, within certain groups of conditions, is carried out according to scale 1-9 in relation to the factor that influences them either from the observed group of conditions (inner dependence) or from another group of conditions (outer dependence). This determines the relative importance of these factors. Comparison of alternatives is also done on scale 1-9, but in terms of the question of what is more likely to be achieved than the terms of observed factor, which determines the priority of the alternative, which in this case can be interpreted as its likelihood of achievement. Once inputs are made for each segment of the model, the information is synthesized in order to achieve the general preference of alternative outcomes. This synthesis gives a report that ranks alternatives (outcomes) with respect to the general goal. The report may include a detailed ranking showing how each alternative is evaluated against each criterion. After all the necessary comparisons have been made in accordance with theoretically based principles, taking into account the assumed inner and outer interdependencies within the cluster and between them, after executing calculations using the Superdecisions software package, developed as software support, ANP application, the results are shown in the tables 2-5. The assessment have been made by authors. Tables 2-4 show the relative importance of the sub-criterion within the criteria to which they belong. Thus, among the economic conditions, it was estimated that the greatest relative importance, in the context of the European Integration, has a sub-criterion of Free Movement of Goods (0.42595), among the political conditions, the sub-criterion External Security and Defense Policy (0.53294), and among the legal conditions, the sub-criterion of Competition Policy (0.56853). The obtained priorities of the alternative, shown in Table 5, are interpreted in terms of the probability of achieving the chosen alternatives. We can see that the highest estimated probability of achievement is the accession of Serbia to the EU between 2025 and 2030 (57.21%), then in the period after 2030 (34.3%), while the likelihood that Serbia will not become a member of the EU is negligible (0.09%). The Ideal column shows the results divided by the highest value, so that the highest ranking has priority 1 while remaining in the same proportion as in the Normalized value column.

Table 2. Relative importance of criteria *economic conditions*

Name	Normalized By Cluster	Limiting
Free movement of goods	0.42595	0.105976
Economic and monetary policy	0.34270	0.085262
Customs Union	0.15095	0.037556
Social policy and employment	0.08040	0.020003

Source: Authors

Table 3. Relative importance of criteria *political conditions*

Name	Normalized By Cluster	Limiting
Judiciary and human rights	0.20251	0.040801
Justice, freedom and security	0.11911	0.023997
External, security and defense policy	0.53294	0.107372
Relations with Kosovo	0.14544	0.029302

Source: Authors

Table 4. Relative importance of criteria *legal conditions*

Name	Normalized By Cluster	Limiting
Competition policy	0.56853	0.092653
Company law	0.43147	0.070317

Source: Authors

Table 5. Probability of achievement and ranking of alternatives

Alternatives	Total	Normal	Ideal	Ranking
After 2030.	0.1271	0.3287	0.5745	2
Between 2025. and 2030.	0.2213	0.5721	1.0000	1
Not entering the European Union	0.0384	0.0992	0.1734	3

Source: Authors

The objective of the forecast and the results obtained indicate the high level of uncertainty presented in the forecasting process, which only confirms the need for additional analysis of the sensitivity of the results to changes in the values of the key parameters of the model, in order to obtain a more complete and comprehensive estimation of the target value. Analysis of the sensitivity of the results by changing the level of significance of the higher level elements, i.e. may more or less significantly affect the order of importance and the assessment of the observed options, showing the performance of the alternative in relation to each criterion, and how these alternatives are sensitive to the changes in the importance of the criteria. Thus, Figure 2 shows that the growth of the relative importance of the Kosovo Relations, from 0.0001 to 0.9999, affects to a certain extent the likelihood of alternative outcomes: the probability of an alternative entry after 2030 is increased (from 28.4% to 34.5%), and the probability of expected entry between 2025 and 2030 is reduced (from 60.4% to 56%). At the same time, the likelihood that Serbia will not become a member of the EU remains virtually unchanged and negligible (from 11.3% to 0.09%), which means that the Kosovo Relations sub-criterion essentially influences Serbia's entry into the EU only in terms of deadlines, but will not stop this process, if the other presumed interdependencies of the factors are unchanged.

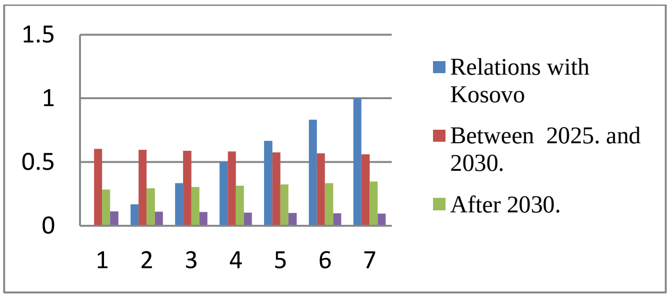


Figure 2. The Impact of Changing the Relative Importance of the Criteria Relations with Kosovo on Moving the Probability of Realizing Alternatives

If we consider the impact of changing the relative importance of the subcriteria External Security and Defense Policy on the likelihood of achieving alternative outcomes, one can notice that the growth of the relative importance of this sub-criterion leads to a somewhat significant increase in the probability of entering Serbia the EU after 2030 (from 26.6% 38%), decline in the probability of entry between 2025 and 2030 (from 60.1% to 54.8%) and somewhat lowering the likelihood that Serbia will not become a member of the EU (from 13.2% to 7.2%). Similar sensitivity analysis of the solution can be done for other sub-criteria in the model.

4. CONCLUSION

The paper presents the possibility of using the Analytic Network Process Methodology in the process of predicting the probability of alternative future outcomes, on the examle of Serbia and its entry into the European Union. Although work is somewhat simplified, which is quite understandable given that a large number of factors and their networks of influence have to be taken into account for a more serious analysis, its theoretical as well as methodological and practical implications can be seen. Theoretical implications are reflected in the fact that the effective ability of an analytic network process to conceptually encompass all relevant factors from the prediction / decision-making context is confirmed and that, despite the usual limitations encountered when applying qualitative predictive methods (precision, lack of information, bias, price, etc.) can serve as a satisfactory basis for the creative solution of decision-making issues in situations of growing complexity and uncertainty when a decision needs to be made quickly. Presenting the usage of ANP predictive model in the field of Eurointegration, which is extremely complex and uncertain, confirms its ability to be good support in the prediction process. In situations characterized by the pronounced dynamics and unpredictability of the environment and many factors that determine it, as is the case with the European Union and Serbia, prediction methods based on econometric models do not yield satisfactory results. Social justification is reflected in the possibility of using the results of the research by the competent institutions of the Republic of Serbia for redefining or making new, better strategic decisions in the process of stabilization and association. The research limitations relate primarily to the complexity of the ANP model, not only in terms of application, but also in determining the interactions and connections between elements and clusters, which is why certain relationships have been established, starting from the subjective evaluation of the author. Significant objectivity could be achieved by including a group of experts in the areas relevant to the EU integration process in the comparing process, and after obtaining the individual ranking lists of alternatives and their probability of achievement, using a geometric mean, obtain a unique and objective ranking list of alternative outcomes and a better assessment of their probability. Also, clusters cover only some negotiating chapters which may have an impact on the results obtained.

REFERENCES

Adamus, W. (2010). When shall Poland enter the euro zone? Decision making with the ANP. *The School of Business*, *1*, 31-52.

Azis, I. J. (2010). Predicting a recovery date from the economic crisis of 2008. *Socio-Economic Planning Sciences*, *44*(3), 122-129.

Bayazit, O. (2006). Use of analytic network process in vendor selection decisions. *Benchmarking: An International Journal*, *13*(5), 566-579.

Blair, A. R., Mandelker, G. N., Saaty, T. L., & Whitaker, R. (2010). Forecasting the resurgence of the US economy in 2010: An expert judgment approach. *Socio-Economic Planning Sciences*, *44*(3), 114-121.

Dağdeviren, M., Yüksel, İ. (2007). Personnel selection using analytic network process. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, *6*(11), 99-118.

Dikmen, I., Talat Birgonul, M., Ozorhon, B., & Egilmezer Sapci, N. (2010). Using analytic network process to assess business failure risks of construction firms. *Engineering, Construction and Architectural Management*, *17*(4), 369-386.

Domanović, V., Jakšić, M., Mimović, P. (2014). Balanced scorecard and analytic network process in performance measurement and strategy evaluation: a case study. *Teme: Casopis Za Društvene Nauke*, *38*(4). 1457-1474.

EU Info Centar. (2014). *Pregovaračka poglavlja – 35 koraka ka Evropskoj Uniji*, dostupno na: http://euinfo.rs/srbija-i-eu/

EU Info Centar. (2015). *Pregovaračka poglavlja – 35 koraka ka Evropskoj Uniji*. Drugo, dopunjeno izdanje, dostupno na: http://euinfo.rs/srbija-i-eu/

European Council. (2015). Normalisation of relations between Serbia and Kosovo, Chapter 35: Other issues. *Conference on accession to the European Union–Serbia*– Brussels, 30 November 2015, 1-8.

Jayant, A. (2016). An Application Of Analytic Network Process (ANP) To Evaluate Green Supply Chain Management Strategies: A Case Study. In *MATEC Web of Conferences*, *vol. 57*, 1-6.

Lin, H. T. (2010). Personnel selection using analytic network process and fuzzy data envelopment analysis approaches. *Computers & Industrial Engineering*, *59(4)*, 937-944.

Lin, J. H., Yang, C. J. (2016). Applying Analytic Network Process to the Selection of Construction Projects. *Open Journal of Social Sciences*, *4*(03), 41-47.

Meade, L. M., & Presley, A. (2002). R&D project selection using the analytic network process. *IEEE transactions on engineering management*, *49*(1), 59-66.

Mimović, P. (2012). Primena analitičkog mrežnog procesa u predviđanju prodaje automobila Fiat 500 L. *Ekonomski horizonti*, 165-176.

Mimović, P., Talijan, D. (2017). Višekriterijumski pristup vrednovanju strategije poslovanja muzejskih ustanova. *Megatrend Review*, *14*(1). 149-168.

Niemira, M. P., & Saaty, T. L. (2004). An analytic network process model for financial-crisis forecasting. *International Journal of Forecasting*, *20*(4), 573-587.

Ozorhon, B., Dikmen, I., Birgonul, M. T. (2007). Using analytic network process to predict the performance of international construction joint ventures. *Journal of management in engineering*, *23*(3), 156-163.

Rapaić, S., Dabić, D. (2013). Spoljnotrgovinski aspekt pristupanja Republike Srbije EU. *Međunarodni problemi*, *65(3)*, 341–364.

Saaty T.L. (1983). Conflict resolution and the Falkland Island invasions, *Interfaces 13 (6)* 68–83.

Saaty, T. L. (1999). Fundamentals of the analytic network process. In *Proceedings of the 5th international symposium on the analytic hierarchy process*, 12-14.

Saaty, T. L. (2001a). Analytic network process. In *Encyclopedia of Operations Research and Management Science*, 28-35.

Saaty, T. L. (2001b). Decision making with the analytic network process (ANP) and its "super-decisions" software the national missile defense (NMD) exANPle. *ISAHP 2001,* Berne, Switzerland. 365-382.

Saaty, T. L. (2001c). *Decision Making with Dependence and Feedback: The ANP*. Pittsburgh: RWS Publications.

Saaty, T. L. (2008). The analytic network process. *Iranian Journal of Operations Research*, *1*(1), 1-27.

Saaty, T. L., & Vargas, L. G. (2006). *Decision Making with the Analytic Network Process*. *Economic, Political, Social and Technological Applications with Benefits, Opportunities, Costs and Risks*. Springer US.

Saaty, T. L., & Vargas, L. G. (2013). *Decision Making with the Analytic Network Process*. *Economic, Political, Social and Technological Applications with Benefits, Opportunities, Costs and Risks*. 2nd edition, Springer US.

Saaty, T.L., Vargas, L., Zoffer, H. J. (2015): A structured scientific solution to the Israeli-Palestinian conflict: The analytic hierarchy process approach, *Decision Analytics, ISSN 2193-8636, Vol. 2*, Iss. 1, 1-53

Santos, M. A. R. D., Salomon, V. A. P., Marins, F. A. S. (2015). Analytic Network Process and Balanced Scorecard applied to the performance evaluation of Public Health Systems. *Pesquisa Operacional*, *35*(2), 353-361.

Shih, H. S., Lee, E. S., Chuang, S. H., & Chen, C. C. (2012). A forecasting decision on the sales volume of printers in Taiwan: An exploitation of the Analytic Network Process. *Computers & Mathematics with Applications*, *64*(6), 1545-1556.

Tseng, M. L., Chiang, J. H., & Lan, L. W. (2009). Selection of optimal supplier in supply chain management strategy with analytic network process and choquet integral. *Computers & Industrial Engineering*, *57*(1), 330-340.

Vargas, L. G. (1990). An overview of the analytic hierarchy process and its applications. *European journal of operational research*, *48*(1), 2-8.

Voulgaridou, D., Kirytopoulos, K., & Leopoulos, V. (2009). An analytic network process approach for sales forecasting. *Operational Research*, *9*(1), 35-53.

Wu, W. W., & Lee, Y. T. (2007). Selecting knowledge management strategies by using the analytic network process. *Expert systems with Applications*, *32*(3), 841-847.

XIII Balkan Conference on Operational Research

THE MCDM APPROACH TO THE SELECTION OF A SUPPLIER FOR MEDIUM VOLTAGE DISTRIBUTION EQUIPMENT

SANJA PUZOVIĆ[1], JASMINA VESIĆ VASOVIĆ[2], MIROSLAV RADOJIČIĆ[3], VLADAN PAUNOVIĆ[4],

[1] University of Kragujevac, Faculty of Technical Sciences Čačak, sanja.puzovic@ftn.kg.ac.rs
[2] University of Kragujevac, Faculty of Technical Sciences Čačak, jasmina.vesic@ftn.kg.ac.rs
[3] University of Kragujevac, Faculty of Technical Sciences Čačak, miroslav.radojicic@ftn.kg.ac.rs
[4] University of Kragujevac, Faculty of Technical Sciences Čačak, vladan.paunovic@ftn.kg.ac.rs

Abstract: *Procurement is closely associated with the formulation of a business strategy at the highest level, so, the supplier selection is a strategic priority and a particular challenge for modern manufacturers. This paper is aimed at proposing a decision support model for the multi-criteria supplier selection process in the case of the purchase of the vital equipment for mini water power plants. The proposed model combines the three techniques: the Delphi method, the Analytic Hierarchical Process (AHP) and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS), integrated with the fuzzy set theory in order to deal with ambiguities. The Fuzzy Delphi Method (DM) is used to select the most significant criteria to be included in the decision-making process; the Fuzzy AHP method is used to prioritize the criteria and to estimate the criteria relative importance, whereas the selection of the most suitable supplier is carried out by the Fuzzy TOPSIS method. The proposed model is illustrated on the case of supplier selection for the Medium Voltage (MV) distribution equipment for the needs of the mini water power plant "Seoce" Prijepolje.*

Keywords: *Supplier selection, MV distribution equipment, Fuzzy DM, Fuzzy AHP, Fuzzy TOPSIS*

1. INTRODUCTION

Development of an effective supply chain as an integrated system of the key business processes that creates additional value for the end user is the strategic priority for companies; moreover, proper supplier selection is of critical importance for its successful implementation. The selection of the most suitable supplier is of special importance in purchasing the equipment that represents the crucial part of a mini water power plant and which fully determines the performance of these plants, as is the *MV Distribution Equipment*. The fluctuations of supplier performances, as well as the ambiguity of information, always exist in real-world decisions. All this requires the provision of the hybrid multi-criteria decision-making (MCDM) model that will successfully be supportive of this process and provide realistic results, which is the aim of this paper.

In the literature, the supplier selection problem is viewed as one of the essential business problems; in order to deal with the complexity of this problem, a large number of integrated MCDM approaches have been developed. In the paper (Awasthi et al., 2018), an integrated fuzzy AHP – Multi-criteria Optimization and Compromise Ranking (VIKOR)-based approach to a sustainable global supplier selection that takes sustainability risks from sub-suppliers into account is developed. The paper (Banaeian et al., 2018) discusses the incorporation of the fuzzy set theory into the TOPSIS, VIKOR and Grey Relational Analysis (GRA) methods in the process of a green supplier evaluation. The integrated MCDM approach to supplier selection was developed by combining the Rank Order Centroid (ROC) method, used for criteria prioritization, and TOPSIS, as the basis of the process (Sureeyatanapas et al., 2018). For the same purpose, Luthra et al. (2017) proposed the MCDM approach based on the AHP and VIKOR, whereas the authors Sarkara et al. (2018) use a combination of the Analytic Network Process (ANP), fuzzy TOPSIS and Multiple Segment Goal Programming (MSGP). In the paper (Krishankumar et al., 2017), the extension of PROMETHEE under a fuzzy environment for solving the supplier selection problem with linguistic preferences is proposed. In the paper (Chul Park & Lee, 2018), a hybrid approach to supplier evaluation and selection is presented by using an Expectation Maximization (EM) algorithm for clustering, the Data Envelopment Analysis (DEA) for efficiency, and the AHP for importance.

2. THE PROPOSED METHODOLOGY

The proposed approach is aimed at supporting the *Medium Voltage Distribution Equipment* supplier selection process by combining the three MCDM methods, namely the Delphi, the AHP and the TOPSIS, all integrated in the Fuzzy set theory. The concept of the proposed methodology is discussed as follows.

2.1 The Fuzzy Delphi Method

The Delphi method enables the extracting of the most important evaluation criteria from a list of potential ones, based on the consolidation of experts' opinions about the significance of the importance of considering each of the criteria. The integration of the fuzzy set theory and the traditional Delphi method was first introduced by Kaufmann & Gupta (1988), which enables successful dealing with the ambiguity and vagueness of an expert's judgements.

Depending on experts' experience, their qualifications and the designation, it is necessary to assign different relative weights to each one of them, in which way experts' competence in handling the considered problem is taken into consideration. The assessment of experts' competence can be carried out by using the fuzzified linguistic descriptions shown in Table 1.

Table 1: The fuzzified scale for weights assignation by experts (Sultana et al., 2015)

Experience	Qualification	Designation	Department	Linguistic Variables	TFN
< 5	Under graduate	Up to executive	Maintenance department	Low	(0, 0.2, 0,4)
5-10	Graduate	Executive to Specialist	Procurement department	Average	(0.2, 0,4, 0.6)
10-15	Master graduation	Specialist to Manager	Management department	High	(0,4, 0.6, 0.8)
15 - above	Post graduate	Manager to GM	Power distribution department	Very high	(0.6, 0.8, 1.0)

In experts' opinion about the importance of considering each of the given criteria ($C_i, i = 1,2, \ldots, m$), a fuzzy decision matrix (1) is created. Experts are signed as E_j ($j = 1,2, \ldots, n$), their relative weights are signed as $\tilde{X}_j$, and experts' opinions about the importance of the criteria as $\tilde{a}_{ij}$; those opinions are expressed by the language expressions represented as triangular fuzzy numbers, according to the scale given in Table 2 and Figure 1.

$$\begin{array}{c} \\ \\ C_1 \\ C_2 \\ \vdots \\ C_m \end{array} \begin{array}{c} \begin{array}{cccc} \tilde{X}_1 & \tilde{X}_2 & \cdots & \tilde{X}_n \\ E_1 & E_2 & \cdots & E_n \end{array} \\ \begin{bmatrix} \tilde{a}_{11} & \tilde{a}_{12} & \cdots & \tilde{a}_{1n} \\ \tilde{a}_{21} & \tilde{a}_{22} & \cdots & \tilde{a}_{2n} \\ \vdots & \vdots & & \vdots \\ \tilde{a}_{m1} & \tilde{a}_{m2} & \cdots & \tilde{a}_{mn} \end{bmatrix} \end{array} \tag{1}$$

Table 2: The fuzzification of the linguistic scale for relative importance (Gupta et al, 2010)

Linguistic Variables	TFN
Very Low	(0, 0, 0.1)
Low	(0, 0.1, 0.3)
Moderate Low	(0.1, 0.3, 0.5)
Moderate	(0.3, 0.5, 0.7)
Moderate High	(0.5, 0.7, 0.9)
High	(0.7, 0.9, 1)
Very High	(0.9, 1, 1)

According to the constructed decision matrix, and assignated weights to the experts the average weight for each single criterion is determined (2).

$$\tilde{w}_i = \frac{\sum_{j=1}^{n} X_j \otimes a_{ij}}{n} \tag{2}$$

The next step represents the computing of the minimum acceptable weight for each criterion (3), where R_j stands for the percentage expressed minimum acceptable criterion weight multiplied by the criterion weight ($\tilde{w}_i$) defined according to the opinion of the j-th expert.

$$\tilde{R}_\delta = \frac{\sum_{j=1}^{n} X_j \otimes R_j}{n} \tag{3}$$

The criteria whose weight ($\tilde{w}_i$) is lower than the estimated minimum acceptable weight will be omitted from further consideration. The defuzzification of those values can be computed by applying the average method.

2.2 The Fuzzy Analytic Hierarchical Process

The Analytic Hierarchical Process (AHP) (Saaty, 1980) is an MCDM method which allows us to solve problems by structuring problems into a hierarchy of decision-making elements (the goal, the criteria and the alternatives), and by systematically evaluating them by pairwise comparison. The integration of the Fuzzy set theory within the AHP process is used in order to overcome ambiguity in handling linguistic variables. The criteria prioritization in this paper will be carried out by using the extent analysis method developed by Chang (1992). The first step in this process is the construction of a fuzzy pairwise comparison matrix $A = \{a_{ij}\}_{n\times m}$, which is presented in Equation (4), where a_{ij} represents the preference of the criterion i over the criterion j.

$$A_{ij} = \begin{vmatrix} 1 & a_{12} & \cdots & a_{1m} \\ a_{21} & 1 & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{vmatrix} \tag{4}$$

Those preferences relations are expressed as language statements: “Equally important”, “Weakly more important”, “Strongly more important”... In order to deal with the imprecise or vague nature of linguistic assessments successfully, they will be quantified by triangular fuzzy numbers (l, m, u) according to the preferences scale demonstrated in Table 3.

Table 3: The fuzzified criteria evaluation scale

Fuzzy number	Linguistic term	Scale of fuzzy number
‘1	Equally important	(1, 1, 1)
‘3	Weakly important	(2, 3, 4)
‘5	Essentially important	(4, 5, 6)
‘7	Very strongly important	(6, 7, 8)
‘9	Absolutely important	(7, 8, 9)
‘2,’4,’6,’8	Intermediate values (‘x)	(x-1, x, x+1)
1/’x	Between two adjacent judgments	(1/x + 1, 1/x, 1/x – 1)

The Fuzzy set theory introduced by Zadeh (1965) is oriented towards the rationality of uncertainty due to imprecision or vagueness. The TFN (Figure 1) is defined by three real numbers expressed as a triple (l, m, u), where $l \leq m \leq u$ for describing a fuzzy event.

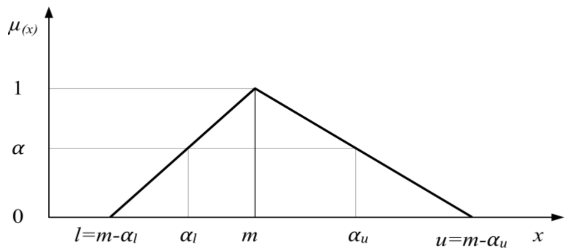


Figure 1: A triangular fuzzy number

Each TFN (l, m, u) has a membership function defined by Equation (5) which enables mapping from a real number to a closed interval $[0,1]$.

$$\mu_M(x) = \begin{cases} \frac{x-l}{m-l}, & x \in [l, m], \\ \frac{x-u}{m-u}, & x \in [m, u], \\ 0, & otherwise \end{cases} \tag{5}$$

In case several experts are involved in the decision-making process, their individual judgements might be aggregated by using the fuzzy geometric mean method, Equation (6), where a_{ijk} is the fuzzy relative importance according to the k^{th} decision maker’s opinion, and n is the number of decision makers.

$$M = \left(\prod_{i=1}^{n} a_{ijk}\right)^{\frac{1}{n}} = \{m_{ij}\} \tag{6}$$

After the construction of the aggregate fuzzy pairwise comparison matrix, the synthetic extent value S_i in the form of a fuzzy number can be computed by applying Equations (7) and (8).

$$S_i = \sum_{j=1}^{m} m_{ij} \oplus \left[\sum_{i=1}^{n}\sum_{j=1}^{m} m_{ij}\right]^{-1} \tag{7}$$

$$S_i = (l_i', m_i', u_i') \oplus \left(\frac{1}{\sum_{i=1}^{n} u_i'}, \frac{1}{\sum_{i=1}^{n} m_i'}, \frac{1}{\sum_{i=1}^{n} l_i'}\right) = \left(\frac{l_i'}{\sum_{i=1}^{n} u_i'}, \frac{m_i'}{\sum_{i=1}^{n} m_i'}, \frac{u_i'}{\sum_{i=1}^{n} l_i'}\right) \tag{8}$$

It is further necessary to determine the degree of the possibility that $S_2 = \left(l_2, m_{2,} u_2\right) \geq S_1 = (l_1, m_{1,} u_1)$ as follow:

$$V(S_2 \geq S_1) = sup\left[min(\mu_{S_1}(x), (\mu_{S_2}(y))\right] \tag{9}$$

which can be extracted in the following manner:

$$V(S_2 \geq S_1) = \mu_{S_2}(d) = \begin{cases} 1, & if\ m_2 \geq m_1 \\ 0, & if\ l_1 \geq u_2 \\ \frac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)} & oterwise \end{cases} \tag{10}$$

$$V(S_2 \geq S_1) = hgt(S_1 \cap S_2) = \mu_{S_1}(d) \tag{11}$$

By funding the preference of S_i and S_k, the degree of a possibility of obtaining a convex fuzzy number can be calculated as follows:

$$V\left(S \geq S_{1,} S_2, \dots, S_n\right) = [V(S \geq S_1)\ and \dots\ and\ V(S \geq S_n)] = min\ V(S \geq S_i)\ (i = 1,2, \dots, n) \tag{12}$$

As $\mathrm{d'(A_i)} = \min \mathrm{V(S_i \geq S_k)}$, $(\mathrm{k} = 1,2,3, \dots, \mathrm{n};\ \mathrm{k} \neq \mathrm{i})$, than the vector weight is given by:

$$\mathrm{W'} = (\mathrm{d'(A_1), d'(A_2) \dots d'(A_n)}) \tag{13}$$

By defuzzifying this value, a normalized vector can be obtained (14).

$$W = (d(A_1), d(A_2) \dots d(A_n)) \tag{14}$$

2.3 The Fuzzy Technique for Order Preference by Similarity to Ideal Solution

The TOPSIS method was first introduced by (Hwang & Yoon, 1981). It is a widely used MCDM technique for ranking alternatives according to their distance from the ideal solution. Let $\tilde{X} = \left[\tilde{x}_{ij}\right]_{m \times n}$ be the decision matrix (15) of the considered problem with n alternatives $(A_1, A_2 \dots, A_n)$, and m criteria $(C_1, C_2, \dots, C_m)$.

$$\tilde{X} = \begin{matrix} \\ A_1 \\ A_2 \\ \vdots \\ A_n \end{matrix} \begin{matrix} \begin{matrix} C_1 & C_2 & \cdots & C_m \end{matrix} \\ \begin{pmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \cdots & \tilde{x}_{1m} \\ \tilde{x}_{21} & \tilde{x}_{22} & \cdots & \tilde{x}_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ \tilde{x}_{n1} & \tilde{x}_{n2} & \cdots & \tilde{x}_{nm} \end{pmatrix} \end{matrix} \tag{15}$$

Where $\tilde{x}_{ij}$ represents the preferences of the alternative A_i over the alternative A_j, estimated by the experts involved in the decision-making process. In order to handle the vagueness of the decision-makers' subjective preferences, the expansion of the TOPSIS method by Fuzzy logic, or the application of the Fuzzy TOPSIS method (Chen & Hwang, 1992), is proposed. Preferential relationships will be expressed by using the linguistic expressions (Very Poor, Poor, Fair…) that will further be represented by triangular fuzzy numbers $(x_{ij}^l, x_{ij}^m, x_{ij}^u)$. The fuzzification of the linguistic expressions will be performed by applying the most frequently used scale in the literature, which is accounted for in Table 4.

Table 4: The fuzzified scales

Linguistic Scale	Triangular Fuzzy Scale
Very poor (VP)	(1,1,3)
Poor (P)	(1,3,5)
Fair (F)	(3,5,7)
Good (G)	(5,7,9)
Very good (VG)	(7,9,9)

vThe normalization of the decision matrix ($\tilde{R} = \left[\tilde{r}_{ij}\right]_{m \times n}$) will be executed by a transformation as follows:

$$\tilde{r}_{ij} = \left(\frac{x_{ij}^l}{x_{ij}^{u+}}, \frac{x_{ij}^m}{x_{ij}^{u+}}, \frac{x_{ij}^u}{x_{ij}^{u+}}\right) = (r_{ij}^l, r_{ij}^m, r_{ij}^u) \tag{16}$$

where x_{ij}^{u+} is $max\ \ (x_{ij}^u)$.

The next step is obtaining a weighted and normalized matrix ($\tilde{V} = \left[\tilde{v}_{ij}\right]_{m \times n}$) by multiplying the elements of the normalized matrix ($\tilde{r}_{ij}$) by the criteria weight (w_j).

$$\tilde{v}_{ij} = \tilde{r}_{ij} \otimes w_j = (v_{ij}^l, v_{ij}^m, v_{ij}^u) \tag{17}$$

The positive ideal solution (PIS) and the negative ideal solution (NIS) are determined as (18) and (19) respectively

$$A^{+} = ({v_{ij}^{l}}^{+}, v_{ij}^{m+}, {v_{ij}^{u}}^{+}) \quad (18)$$

$$A^{-} = ({v_{ij}^{l}}^{-}, v_{ij}^{m-}, {v_{ij}^{u}}^{-}) \quad (19)$$

The final ranking of the alternatives is determined on the basis of relative closeness to the ideal solution:

$$CC_i = \frac{d_i^-}{d_i^+ + d_i^-} \quad (20)$$

where d_i^+ is the distance of the alternative from PIS, and d_i^- the distance of the alternative from NIS.

$$d_i^+ = \sum_{j=1}^{n} d(\tilde{v}_{ij}, \tilde{v}_j^+), i = 1,2,\dots,n; j = 1,2,\dots,m \quad (21)$$

$$d_i^- = \sum_{j=1}^{n} d(\tilde{v}_{ij}, \tilde{v}_j^-), i = 1,2,\dots,n; j = 1,2,\dots,m \quad (22)$$

3. A CASE STUDY

In this section, the implementation of the model proposed as a systematic decision support for the selection of suppliers for *Medium Voltage Distribution Equipment* which includes: *Grid Automation and SCADA Software, Medium Voltage Switchgear, Medium-Voltage Transformers, Outdoor Equipment, Power Metering and Control* is illustrated. This process has been implemented for the needs of the mini water power plant "Seoce" Prijepolje.

Firstly, the most significant criteria for evaluating suppliers will be identified and prioritized by the Fuzzy Delphi and the Fuzzy AHP methods, which is then followed by the TOPSIS phase, in which the evaluation and ranking of the supplier according to the selected criteria will be performed. For that purpose, the five alternative suppliers are generated: Siemens, ABB, Schneider Electric, Loznicaelektro and TSN Maribor. These suppliers represent the most receptive MV Distribution Equipment producers available on our market. The phases of the proposed process are shown in Figure 2.

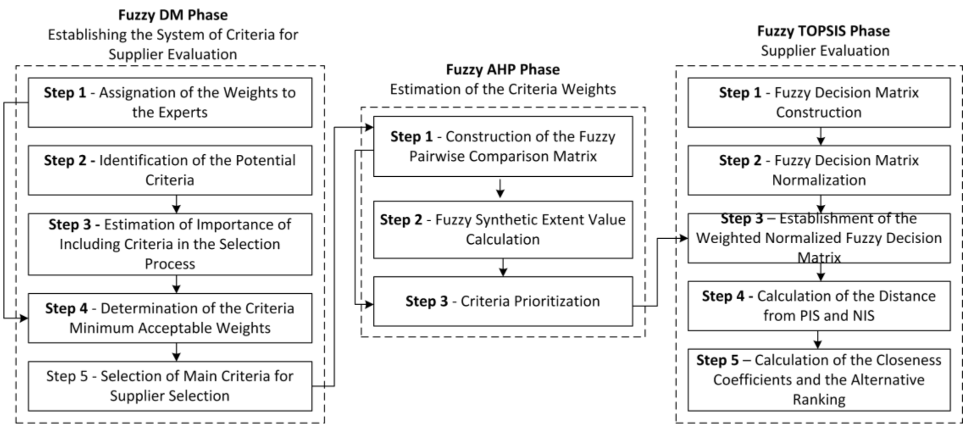


Figure 2: The proposed model

3.1 Establishing the System of Criteria for Supplier Evaluation by the Fuzzy Delphi Method

Step 1: The assignation of the Weights to the Experts - In this process, the four experts are involved, and the weights that represent their competences are assessed according to the following criteria: Experience, the Qualification Level, Designation and the Department to Belong to. The assessment is carried out by using the fuzzified linguistic descriptions shown in Table 1, the result being shown in Table 5.

Table 5: The experts' weights assessed according to the respective criteria for the competence evaluation

Experts	Experience	Qualification	Department	Designation	TFN
Expert 1	< 5	Master graduate	Power distribution department	Specialist to Manager	(0.35,0.55,075)
Expert 2	5-10	Graduate	Management department	Executive to Specialist	(0.25,0.45,0.65)
Expert 3	10-15	Graduate	Procurement department	Executive to Specialist	(0.3,0.5,0.7)
Expert 4	5-10	Under graduate	Procurement department	Up to executive	(0.15,0.35,0.55)

Step 2: The Identification of the Potential Criteria - In experts' opinions a list of the twelve potential supplier selection criteria are is identified. Those criteria are listed in Table 6. In order to shorten the list the following steps of the Delphi process will be conducted.

Step 3: The Estimation of the Importance of Including Criteria in Selection Process - Each expert expressed his/her opinion about the criteria importance by the language descriptions represented as triangular fuzzy numbers, according to the scale given in Table 2. Based on their weights and judgments, the average weight ($\widetilde{w}_i$) for each of the criteria is determined by applying Equation (2). The results are given in Table 6.

Step 4: The Determination of the Criteria Minimum Acceptable Weights - This step involves the determination of the minimum acceptable weight ($\tilde{R}_\delta$) for each criterion (3), as the basis for the decision to include the criterion in the further process of supplier selection. The assessed minimum accepted weight by each of the experts is given in Table 6 The defuzification of those values can be computed by using the average method.

Table 6: The weighted aggregate for criteria selection

Criteria	Criteria Weight				MAW %				Non-fuzzy Aggregated Criteria Weights	Non-fuzzy MAW Value	Selected Criteria
	E 1	E 2	E 3	E 4	E 1	E 2	E 3	E 4			
Product Quality Level	(0.9,1,1)	(0.5,0.7,0.9)	(0.7,0.9,1)	(0.9,1,1)	0.95	0.8	0.65	0.7	0.420	0.363	✓
Price and Purchase Cost	(0.7,0.9,1)	(0.7,0.9,1)	(0.9,1,1)	(0.3,0.5,0.7)	0.8	0.65	0.95	0.4	0.400	0.337	✓
Future Potential Purchase	(0.3,0.5,0.7)	(0,0.1,0.3)	(0.5,0.7,0.9)	(0.3,0.5,0.7)	0.6	0.45	0.7	0.65	0.240	0.278	
Discount	(0.3,0.5,0.7)	(0.7,0.9,1)	(0.3,0.5,0.7)	(0,0.1,0.3)	0.5	0.6	0.89	0.3	0.264	0.274	
After Sales Service	(0.7,0.9,1)	(0.5,0.7,0.9)	(0.1,0.3,0.5)	(0.3,0.5,0.7)	0.85	0.65	0.3	0.75	0.304	0.293	✓
Technical Expertise	(0.9,1,1)	(0.5,0.7,0.9)	(0.3,0.5,0.7)	(0.9,1,1)	0.9	0.7	0.4	0.5	0.375	0.296	✓
R&D Capability	(0.7,0.9,1)	(0.7,0.9,1)	(0.5,0.7,0.9)	(0.3,0.5,0.7)	0.75	0.9	0.6	0.53	0.371	0.326	✓
On Time Delivery Reliability Level	(0.5,0.7,0.9)	(0.5,0.7,0.9)	(0.7,0.9,1)	(0.1,0.3,0.5)	0.65	0.5	0.8	0.4	0.335	0.281	✓
Production Facilities and Capacity	(0.3,0.5,0.7)	(0.3,0.5,0.7)	(0.7,0.9,1)	(0,0.1,0.3)	0.6	0.9	0.85	0.6	0.268	0.343	
Procedural Compliance	(0,0.1,0.3)	(0.3,0.5,0.7)	(0.7,0.9,1)	(0,0.1,0.3)	0.65	0.4	0.8	0.3	0.216	0.261	
Reputation and Position in Industry	(0.5,0.7,0.9)	(0.5,0.7,0.9)	(0.3,0.5,0.7)	(0.1,0.3,0.5)	0.5	0.7	0.8	0.65	0.290	0.304	
Consistency	(0.3,0.5,0.7)	(0.3,0.5,0.7)	(0.5,0.7,0.9)	(0.3,0.5,0.7)	0.6	0.6	0.75	0.7	0.283	0.305	

Step 5: The Selection of Main Criteria for Supplier Evaluation - The criteria whose weight ($\widetilde{w}_i$) is lower than the estimated minimum acceptable weight ($\tilde{R}_\delta$) are omitted from further consideration. According to the results displayed in Table 6, the selected main supplier selection criteria are: *Equipment Quality Level, Price and Purchase Cost, After Sales Service, Technical Expertise, R&D Capability, On Time Delivery Reliability Level.*

3.2 The Estimation of the Criteria Weights by the Fuzzy AHP Method

Step 1: The Construction of the Fuzzy Pairwise Comparison Matrix - According to the experts' attitudes towards the importance of the considered criteria, the pairwise decision matrix (4) is formed by using the linguistic variables and triangular fuzzy numbers, according to the scale given in Table 2. This matrix (Table 7) was created by the aggregation of the judgements of all of the four experts involved in the decision-making process (Formula (6)).

Table 7: The aggregated fuzzy comparison matrix for the relative importance of the criteria

Criterion		C1	C2	C3	C4	C5	C6
C1	Equipment Quality Level	(1,1,1)	(1.32,1.73,2.15)	(1.07,1.32,1.57)	(1.68,2.28,2.83)	(0.84,1,1.19)	(2,2.59,3.13)
C2	Price and Purchase Cost	(0.47,0.58,0.76)	(1,1,1)	(2.14,2.51,2.91)	(1.52,1.97,2.45)	(0.5,0.61,0.76)	(0.9,1.16,1.41)
C3	After Sales Service	(0.64,0.76,0.93)	(0.34,0.4,0.47)	(1,1,1)	(0.84,1.04,1.32)	(0.54,0.67,0.84)	(1,1.32,1.68)
C4	Technical Expertise	(0.35,0.44,0.59)	(0.41,0.51,0.66)	(0.76,0.96,1.19)	(1,1,1)	(0.25,0.33,0.5)	(1.81,2.24,2.71)
C5	R&D Capability	(0.84,1,1.19)	(1.32,1.63,2)	(0.59,0.7,0.84)	(1.68,2.28,2.83)	(1,1,1)	(0.76,1,1.32)
C6	On Time Delivery Reliability Level	(0.32,0.39,0.5)	(0.71,0.86,1.11)	(0.59,0.76,1)	(0.37,0.45,0.55)	(0.76,1,1.32)	(1,1,1)

Step 2: The Fuzzy Synthetic Extent Value Calculation According to the constructed aggregate fuzzy pairwise comparison matrix, the fuzzy synthetic extent values (S_i) for each of the criteria are computed by applying Equations (7) and (8). The results are shown in Table 8.

Step 3: The Criteria Prioritization - The minimum degree of the possibility of the superiority of one criterion over another ($V(S_2 \geq S_1)$) is calculated (Equations (9)-(15)) on the basis of which the weight vector (W') and the normalized weight vector (W) are obtained by applying Equations (13) and (14). The results are listed in Table 8. According to the obtained results, the rank of the priority of one criterion over another is as follows: *Equipment Quality Level, Price and Purchase Cost, R&D Capability, Technical Expertise, After Sales Service, On Time Delivery Reliability Level.*

Table 8: The Synthetic Extent Values, the degrees of the possibility of superiority and the weight vectors for the criteria

Criterion	Synthetic Extent Value	Degree of the possibility of superiority for the selected criteria						Weight vector	Normalized Weight vector
		C1	C2	C3	C4	C5	C6		
Equipment Quality Level	(0.16,0.25,0.36)	-	1	1	1	1	1	1	0.357
Price and Purchase Cost	(0.13,0.19,0.28)	0.69	-	1	1	1	1	0.693	0.248
After Sales Service	(0.09,0.13,0.19)	0.17	1	-	0.93	0.5	1	0.174	0.062
Technical Expertise	(0.09,0.14,0.2)	0.25	0.53	1	-	0.58	1	0.253	0.09
R&D Capability	(0.13,0.19,0.28)	0.66	0.96	1	1	-	1	0.664	0.237
On Time Delivery Reliability Level	(0.08,0.11,0.16)	0.01	0.27	0.81	0.74	0.32	-	0.013	0.005

3.3 Supplier Evaluation by the Fuzzy TOPSIS method

Step 1: The Fuzzy Decision Matrix Construction **–** The aggregate decision matrix for the evaluation of the considered supplier alternatives is given in Table 9.

Table 9: The fuzzy decision matrix for the supplier evaluation

Alternative	Siemens	ABB	Electric Schneider	Loznicaelektro	TSN Maribor
Equipment Quality Level	(6.44,8.45,9)	(6.44,8.45,9)	(5,7,9)	(1.73,3.87,5.92)	(1.73,2.24,4.58)
Price and Purchase Cost	(1,1,3)	(1,3,5)	(2.28,4.4,6.44)	(6.44,8.45,9)	(5,7,9)
After Sales Service	(5.92,7.94,9)	(5,7,9)	(1.73,3.87,5.92)	(2.65,5.2,6.71)	(1,1.32,3.41)
Technical Expertise	(5,7,9)	(6.44,8.45,9)	(4.4,6.44,8.45)	(1.97,3.08,5.21)	(1.32,2.59,4.79)
R&D Capability	(5.44,7.45,9)	(7,9,9)	(5.21,7.3,8.45)	(1.32,2.59,4.79)	(1.73,3.87,5.92)
On Time Delivery Reliability Level	(7,9,9)	(5,7,9)	(3.41,5.44,7.45)	(1.32,3.41,5.44)	(1,1.32,3.41)

Step 2: The Fuzzy Decision Matrix Normalization **-** The normalized decision matrix (Table 10) is determined with the help of the transformation equation (16).

Table 10: The normalized fuzzy decision matrix

Alternative	Siemens	ABB	Electric Schneider	Loznicaelektro	TSN Maribor
Equipment Quality Level	(0.72,0.94,1)	(0.72,0.94,1)	(0.56,0.78,1)	(0.19,0.43,0.66)	(0.19,0.25,0.51)
Price and Purchase Cost	(0.11,0.11,0.33)	(0.11,0.33,0.56)	(0.25,0.49,0.72)	(0.72,0.94,1)	(0.56,0.78,1)
After Sales Service	(0.66,0.88,1)	(0.56,0.78,1)	(0.19,0.43,0.66)	(0.29,0.58,0.75)	(0.11,0.15,0.38)
Technical Expertise	(0.56,0.78,1)	(0.72,0.94,1)	(0.49,0.72,0.94)	(0.22,0.34,0.58)	(0.15,0.29,0.53)
R&D Capability	(0.6,0.83,1)	(0.78,1,1)	(0.58,0.81,0.94)	(0.15,0.29,0.53)	(0.19,0.43,0.66)
On Time Delivery Reliability Level	(0.78,1,1)	(0.56,0.78,1)	(0.38,0.6,0.83)	(0.15,0.38,0.6)	(0.11,0.15,0.38)

Step 3: The Establishment of the Weighted Normalized Fuzzy Decision Matrix **–** By applying Equation (17), the weighted matrix is obtained (Table 11) by multiplying the elements of the normalized matrix shown in Table 10 by the criteria relative weight given in Table 8.

Table 11: The Establishment of the Weighted Normalized Fuzzy Decision Matrix

Alternative	Siemens	ABB	Electric Schneider	Loznicaelektro	TSN Maribor
Equipment Quality Level	(0.26,0.34,0.36)	(0.26,0.34,0.36)	(0.2,0.28,0.36)	(0.07,0.15,0.23)	(0.07,0.09,0.18)
Price and Purchase Cost	(0.04,0.04,0.12)	(0.04,0.12,0.2)	(0.09,0.17,0.26)	(0.26,0.34,0.36)	(0.2,0.28,0.36)
After Sales Service	(0.23,0.32,0.36)	(0.2,0.28,0.36)	(0.07,0.15,0.23)	(0.11,0.21,0.27)	(0.04,0.05,0.14)
Technical Expertise	(0.2,0.28,0.36)	(0.26,0.34,0.36)	(0.17,0.26,0.34)	(0.08,0.12,0.21)	(0.05,0.1,0.19)
R&D Capability	(0.22,0.3,0.36)	(0.28,0.36,0.36)	(0.21,0.29,0.34)	(0.05,0.1,0.19)	(0.07,0.15,0.23)
On Time Delivery Reliability Level	(0.28,0.36,0.36)	(0.2,0.28,0.36)	(0.14,0.22,0.3)	(0.05,0.14,0.22)	(0.04,0.05,0.14)

Step 4: The Calculation of the Distance from PIS and NIS **–** As $\tilde{v}_{ij}$ is characterized by a triangular fuzzy number $(v_{ij}^{l}, v_{ij}^{m}, v_{ij}^{u}) \in [0,1]$, PIS and NIS are determined as: $A^{+} = (1,1,1)$ and $A^{-} = (0,0,0)$. The distances from PIS (d_i^{+}) and NIS (d_i^{-}) for each one of the alternatives are calculated by applying Formulas (21) and (22). The results are presented in Table 12.

Step 5: The Calculation of the Closeness Coefficients and the Alternative Ranking **–** Fuzzy relative closeness (CC_i) is also determined (shown in Table 12) with the help of Formula (20). According to the obtained non-fuzzy relative closeness coefficient, the final rank of the considered alternatives is determined.

According to the obtained results, the alternative ABB has the most powerful performance in terms of the quality of the equipment it offers, the price and the purchase cost, an after sales service, the technical expertise, an R&D capability and the on-time delivery reliability level.

Table 12: The final range of the alternatives

Alternative	d_i^+	d_i^-	Relative closeness	Alternative rank
Siemens	4.43	1.62	0.267	2
ABB	4.37	1.68	0.277	1
Electric Schneider	4.66	1.41	0.232	3
Loznicaelektro	4.97	1.11	0.183	4
TSN Maribor	5.2	0.88	0.144	5

4. CONCLUSION

In the paper, a realistic model for supplier selection that considers the problem from multiple aspects and provides the consolidation of the opinions of a group of experts is presented. The proposed approach represents the systematic supplier selection process formed by combining the three MCDM techniques, namely: the Delphi method – used for the selection of the most significant criteria to be included in the decision-making process; the AHP – used for criteria prioritization; and TOPSIS – which enables the ranking and selection of the most suitable supplier according to considered criteria. On the other hand, the incorporation of the fuzzy set theory into the proposed model makes it easier to handle with the ambiguities that this process is accompanied by. The model is fully illustrated through the process of the supplier selection for the *Medium Voltage Distribution Equipment* for the need of the mini water power plant “Seoce” Prijepolje. The obtained results point to the supplier with the most powerful performance in terms of the quality of the equipment it offers, the price and the purchase cost, an after sales service, technical expertise, an R&D capability and the on-time delivery reliability level.

REFERENCES

Awasthi, A., Govindan, K., & Gold, S. (2018). Multi-tier sustainable global supplier selection using a fuzzy AHP-VIKOR based approach. *International Journal of Production Economics*, 195, 106-117.

Banaeian, N., Mobli, H., Fahimnia, B., Nielsen, I. E., & Omid, M. (2018). Green supplier selection using fuzzy group decision making methods: A case study from the agri-food industry. *Computers & Operations Research*, 89, 337-347.

Chang, D. Y. (1992). Extent analysis and synthetic decision. *Optimization techniques and applications*, 1(1), 352-355.

Chen, S. J., & Hwang, C. L. (1992). Fuzzy multiple attribute decision making methods. *In Fuzzy multiple attribute decision making* (pp. 289-486). Springer, Berlin, Heidelberg.

Chul Park, S., & Lee, J. H. (2018). Supplier selection and stepwise benchmarking: a new hybrid model using DEA and AHP based on cluster analysis. *Journal of the Operational Research Society*, 1-18.

Gupta, R., Sachdeva, A., & Bhardwaj, A. (2010, October). Selection of 3PL service provider using integrated fuzzy Delphi and fuzzy TOPSIS. In *Proceedings of the world congress on engineering and computer* science (Vol. 2, No. 1, pp. 20-22).

Hwang, C. L., & Yoon, K. (1981). Methods for multiple attribute decision making. *In Multiple attribute decision making* (pp. 58-191). Springer, Berlin, Heidelberg.

Kaufmann, A., & Gupta, M. M. (1988). *Fuzzy mathematical models in engineering and management science*. Elsevier Science Inc..

Krishankumar, R., Ravichandran, K. S., & Saeid, A. B. (2017). A new extension to PROMETHEE under intuitionistic fuzzy environment for solving supplier selection problem with linguistic preferences. *Applied Soft Computing*, 60, 564-576.

Luthra, S., Govindan, K., Kannan, D., Mangla, S. K., & Garg, C. P. (2017). An integrated framework for sustainable supplier selection and evaluation in supply chains. *Journal of Cleaner Production*, 140, 1686-1698.

Saaty, T. L. (1980). The analytic hierarchy process: planning, priority setting, resources allocation. *New York: McGraw*, 281.

Sarkara, S., Pratiharb, D. K., & Sarkarc, B. (2018). An integrated fuzzy multiple criteria supplier selection approach and its application in a welding company. *Journal of Manufacturing Systems*, 46, 163-178.

Sultana, I., Ahmed, I., & Azeem, A. (2015). An integrated approach for multiple criteria supplier selection combining Fuzzy Delphi, Fuzzy AHP & Fuzzy TOPSIS. *Journal of Intelligent & Fuzzy Systems*, 29(4), 1273-1287.

Sureeyatanapas, P., Sriwattananusart, K., Niyamosothath, T., Setsomboon, W., & Arunyanart, S. (2018). Supplier selection towards uncertain and unavailable information: an extension of TOPSIS method. *Operations Research Perspectives*.

Zadeh, L. A. (1965). Information and control. *Fuzzy sets*, 8(3), 338-353.

XIII Balkan Conference on Operational Research
BALCOR 2018

PRIORITIZATION OF LOCATION ALTERNATIVES FOR ELECTRIC VEHICLE CHARGING STATIONS IN ISTANBUL

MELIKE ERDOĞAN[1], ALI KARAŞAN[2], İHSAN KAYA[1]

[1]Yildiz Technical University, Department of Industrial Engineering, 34349 Yıldız Beşiktaş, İstanbul
[2]Yildiz Technical University, Institute of Natural and Applied Sciences, 34220 Davutpaşa, Istanbul

Abstract: *With the rapid increase in air pollution and the acceleration of global warming, more serious steps have been taken in order to protect the environment and sustainability. One of the most important of them is to transform petroleum-based vehicles to electricity-driven vehicles. As these vehicles getting more involved in the traffic, one of the critical problems has arisen where the charging stations that feed these vehicles need to be installed. This problem has been considered as a multi criteria decision making (MCDM) problem. Taking this problem into consideration as a MCDM problem, an interval-valued intuitionistic fuzzy (IVIF) MCDM methodology has been suggested to solve this problem in this paper. The proposed methodology has been used to evaluate alternative locations for the electric vehicles (EVs) charging stations in Istanbul.*

Keywords: *Electric vehicles, charging stations, location selection, intuitionistic fuzzy sets, AHP*

1. INTRODUCTION

With the rapid increase in air pollution and the acceleration of global warming, some more serious steps have been taken in order to protect the environment and sustainability. By 2030, the global energy claim is expected to increase by 50% with a corresponding increase in CO_2 and greenhouse gas emissions, since most of this energy claim will be met by fossil fuels (Gonzalez et al., 2014). The burning of fossil fuels and greenhouse gas emissions drive an alternative road transport system to find a new way (Islam et al., 2015). Transportation electrification is one of the most essential elements in order to integrate intelligent city planning and electric vehicles into the transportation system in the future without problems (Lam et al., 2013). To increase environmental sustainability, many countries electrify their transportation systems, thereby significantly increasing the number of electric vehicles (EVs) in cities (Lam et al., 2014). EVs are considered to be one of the best solutions for road transport, because the usage of them can help to reduce CO_2 emissions, and dependence on fossil fuels (Islam et al., 2015). Today, many automobile companies are active in the development and production of EVs as a measure against global warming and depletion of fossil fuels (Kobayashi et al., 2011). Charging infrastructures are one of the most important structures of EVs. An EV must always be accessible to a charging station with a capacity of anywhere in the city (Lam et al., 2013). The charging stations should be constructed sufficiently to be able to be extended to all corners of the city (Lam et al., 2013). Increasing the number of centers to charge EVs is one of the most important factors for reducing the uncertainties associated with using a EVs (Eisel et al., 2014).

Turkey is one of the world's countries wishing to increase the number of EVs use the fast by following this trend. For this purpose, it promotes the use of EVs in crowded cities which are usually intensive about traffic. A lot of charging stations have been built for the spread of electric vehicle usage in Istanbul. Nevertheless it can be seen that this number is not sufficient when the urban population and traffic density are considered. For this reason, new charging stations should be built in Istanbul and the use of EVs should be further supported. In this paper, we are also dealing with this problem with suggesting new charging areas for EVs. For this purpose, we show in which regions the charging stations should be installed firstly by referring to the candidate charging station locations. We also use interval-valued intuitionistic fuzzy (IVIF) AHP method for finding the prioritization of alternative areas for charging stations. The rest of this paper has been organized as follows: Section 2 summarizes relevant literature about to determine the best location alternative for charging stations. Section 3 gives information about the proposed fuzzy based methodology. Section 4 includes a real case analysis for prioritizing of alternative areas for electric vehicle charging stations. Finally, Section 5 concludes the obtained results and give some suggestions for future research.

2. LITERATURE REVIEW

There are many approaches in the relevant literature to define the charging station location for electric vehicles. Some of them can be briefly summarized as follows: Wang and Lin (2009) presented a concept of set cover for proposing a refueling-station-location model by using a mixed integer programming method, based on vehicle-routing logics. The applied a case study focused on the siting of refueling stations for achieving multiple origin-destination intercity travel via electric vehicles on Taiwan. Pan et al. (2010) investigated how to best site these stations in terms of how they can support both the transportation system and the power grid. They used a two-stage stochastic program to optimally locate the stations prior to the realization of battery demands, loads, and generation capacity of renewable power sources. Schill (2011) presented usage of a game-theoretic model to analyze the impacts of a hypothetical fleet of plug-in electric vehicles on the imperfectly competitive German electricity market. Kley et al. (2011) presented a holistic approach to developing business models for electric mobility, which analyzes the system as a whole on the one hand and provides decision support for affected enterprises on the other. Wirges et al. (2012) suggested a dynamic spatial model of the development of a charging infrastructure for electric vehicles in the German metropolitan region of Stuttgart. Traut et al. (2012) presented an optimization model to determine optimal design of conventional vehicles, hybrid electric vehicles (HEVs), plug-in hybrid electric vehicles (PHEVs), and battery electric vehicles (BEVs) with optimal allocation of vehicle designs and dedicated workplace charging infrastructure in the fleet for minimum life cycle cost or greenhouse gas emissions over a range of scenarios. Jia et al. (2012) suggested an optimization process of the sizing and siting of electric vehicle charging stations, defines variables to represent the charging demand; abstracts the structure of road network to model and simulate with graph theory; solves the problem with Cplex. Liu et al. (2013) firstly identified the optimal sites of EV charging stations by a two-step screening method with environmental factors and service radius of EV charging stations considered and after develop a mathematical model for the optimal sizing of EV charging stations with the minimization of total cost associated with EV charging stations to be planned as the objective function and solved by a modified primal-dual interior point algorithm. Pazouki et al. (2013) presented optimal planning of Charging Stations (CSs) which consider loss and voltage. He et al. (2013) suggested an equilibrium modeling framework that captures the interactions among availability of public charging opportunities, prices of electricity, and destination and route choices of plug-in hybrid electric vehicles (PHEVs) at regional transportation and power transmission networks coupled by PHEVs. Lam et al. (2013) formulated the electric vehicle charging station placement problem, in which they minimize the total construction cost subject to the constraints for the charging station coverage and the convenience of the drivers for EV charging with the proposed method. Wagner et al. (2013) presented a point of interest based business intelligence system to determine the optimal locations for charging stations. Yi & Bauer (2014) developed a methodology than can easily incorporate additional constraints such as popular destinations, preferred parking, driver habits, available power infrastructure, etc. to initially reduce the search space for optimal charging station placement. Micari et al. (2014) aimed to analyze the Italian territory in order to provide an evaluation regarding the number of charging stations that satisfy the demand of the population. Barzani et al. (2014) presented a mixed-integer non-linear (MINLP) optimization approach for optimal placing and sizing of the fast charging stations. Eisel et al. (2014) suggested an approach for a location-planning problem of EV charging stations under special consideration of customer preferences. They used a semi-discrete planning model, as the charging stations can be located on nodes (junctions) as well as on edges (streets) within the considered road network. In this paper a MCDM methodology based on interval-valued intuitionistic fuzzy (IVIF) analytic hierarchy process (AHP) method has been suggested to determine the prioritization of alternative areas for charging stations.

3.1. PROPOSED METHODOLOGY

In this sub-section, we will give the preliminaries on interval-valued intuitionistic fuzzy numbers and steps of the proposed methodology (Wu et al., 2013).

3.1 Preliminaries on Interval Valued Intuitionistic Fuzzy Sets

Definition 1. Let X be a non-empty set. An interval-valued intuitionistic fuzzy set in X is an object $\tilde{A}$ given as in Eq. (1):

$$\tilde{A} = \{\langle x, [\mu_{\tilde{A}}^{-}, \mu_{\tilde{A}}^{+}], [v_{\tilde{A}}^{-}, v_{\tilde{A}}^{+}]\rangle; x \in X\} \quad (1)$$

where $0 \leq \mu_{\tilde{A}}^{+} + v_{\tilde{A}}^{+} \leq 1$ for every $x \in X$.

Definition 2. Let $\tilde{A} = ([\mu_{\tilde{A}}{}^{-}, \mu_{\tilde{A}}{}^{+}], [v_{\tilde{A}}{}^{-}, v_{\tilde{A}}{}^{+}])$ and $\tilde{B} = ([\mu_{\tilde{B}}{}^{-}, \mu_{\tilde{B}}{}^{+}], [v_{\tilde{B}}{}^{-}, v_{\tilde{B}}{}^{+}])$ be two interval-valued intuitionistic fuzzy numbers. Then,

$$\tilde{A} \oplus \tilde{B} = \left(\left[\mu_{\tilde{A}}^{-} + \mu_{\tilde{B}}^{-} - \mu_{\tilde{A}}^{-}\mu_{\tilde{B}}^{-}, \mu_{\tilde{A}}^{+} + \mu_{\tilde{B}}^{+} - \mu_{\tilde{A}}^{+}\mu_{\tilde{B}}^{+}\right], \left[v_{\tilde{A}}^{-} v_{\tilde{B}}^{-}, v_{\tilde{A}}^{+} v_{\tilde{B}}^{+}\right]\right) \quad (2)$$

$$\tilde{A} \otimes \tilde{B} = \left(\left[\mu_{\tilde{A}}^{-}\mu_{2}^{-}, \mu_{\tilde{A}}^{+}\mu_{2}^{+}\right], \left[v_{\tilde{A}}^{-} + v_{2}^{-} - v_{\tilde{A}}^{-}v_{2}^{-}, v_{\tilde{A}}^{+} + v_{2}^{+} - v_{\tilde{A}}^{+}v_{2}^{+}\right]\right) \quad (3)$$

Definition 3. Let $\tilde{r}_{ij}^{k} = ([\mu_{\tilde{r}}{}^{-}, \mu_{\tilde{r}}{}^{+}], [v_{\tilde{r}}{}^{-}, v_{\tilde{r}}{}^{+}])$ be the interval-valued intuitionistic fuzzy numbers where $k = 1,2, \ldots, n$. Then the aggregated IVIF number $(\tilde{r}_{ij}^{A})$ is obtained by using interval-valued intuitionistic fuzzy hybrid geometric (IIFHG) operator as in Eq. (4) (Wei & Wang, 2007):

$$\tilde{r}_{ij}^{A} = \langle [\textstyle\prod_{k=1}^{n}(\mu_k{}^{-})^{\omega_k}, \prod_{k=1}^{n}(\mu_k{}^{+})^{\omega_k}], \left[1 - \prod_{k=1}^{n}(1 - v_k{}^{-})^{\omega_k}, 1 - \prod_{k=1}^{n}\left(1 - v_k{}^{+}\right)^{\omega_k}\right] \rangle \quad (4)$$

where ω_k is the weight vector of expert k where $\sum_{k=1}^{n} \omega_k = 1$.

Definition 4. Let $\tilde{r}_1 = ([\mu_1{}^{-}, \mu_1{}^{+}], [v_1{}^{-}, v_1{}^{+}])$ and $\tilde{r}_2 = ([\mu_2{}^{-}, \mu_2{}^{+}], [v_2{}^{-}, v_2{}^{+}])$ be two IVIF numbers. The distance between these two IVIF numbers is obtained by Hamming distance as in Eq. (5) (Abdullah & Ismail, 2012):

$$HD = \frac{1}{4}\sum(|\mu_1{}^{-} - \mu_2{}^{-}| + |\mu_1{}^{+} - \mu_2{}^{+}| + |v_1{}^{-} - v_2{}^{-}| + |v_1{}^{+} - v_2{}^{+}| \quad (5)$$

3.2 The Proposed Methodology

Before presenting the steps of the IVIF AHP method, scale that is used for the comparison matrices is given in Table 1.

Table 1: Linguistic scale for weights of the criteria

Linguistic Term	IVIF Number
Certainly Low- CL	<[0.1, 0.25], [0.65, 0.75]>
Very Low- VL	<[0.15, 0.3], [0.6, 0.7]>
Low- L	<[0.2, 0.35], [0.55, 0.65]>
Below Medium- BM	<[0.25, 0.4], [0.5, 0.6]>
Exactly Equal- EE	<[0.5, 0.5], [0.5, 0.5]>
Above Medium- AM	<[0.5, 0.6], [0.25, 0.4]>
High- H	<[0.55, 0.65], [0.2, 0.35]>
Very High- VH	<[0.6, 0.7], [0.15, 0.3]>
Certainly High- CH	<[0.65, 0.75], [0.1, 0.25]>

The steps of the IVIF AHP method are given as below (Wu et al., 2013):

Step 1: Construct the pairwise comparison matrix $\left(\tilde{R} = (\tilde{r}_{ij})_{n\times n}\right)$ by using Table 1:

$$\tilde{R} = \begin{bmatrix} ([\mu_{11}^{-}, \mu_{11}^{+}], [v_{11}^{-}, v_{11}^{+}]) & \cdots & ([\mu_{1n}^{-}, \mu_{1n}^{+}], [v_{1n}^{-}, v_{1n}^{+}]) \\ \vdots & \ddots & \vdots \\ ([\mu_{n1}^{-}, \mu_{n1}^{+}], [v_{n1}^{-}, v_{n1}^{+}]) & \cdots & ([\mu_{nn}^{-}, \mu_{nn}^{+}], [v_{nn}^{-}, v_{nn}^{+}]) \end{bmatrix} \quad (6)$$

Step 2: Calculate the score judgment matrix $\left(\tilde{S} = (\tilde{s}_{ij})_{n\times n}\right)$:

$$\tilde{S} = \begin{bmatrix} [\mu_{11}^{-} - v_{11}^{+}, \mu_{11}^{+} - v_{11}^{-}] & \cdots & [\mu_{1n}^{-} - v_{1n}^{+}, \mu_{1n}^{+} - v_{1n}^{-}] \\ \vdots & \ddots & \vdots \\ [\mu_{n1}^{-} - v_{n1}^{+}, \mu_{n1}^{+} - v_{n1}^{-}] & \cdots & [\mu_{nn}^{-} - v_{nn}^{+}, \mu_{nn}^{+} - v_{nn}^{-}] \end{bmatrix} \quad (7)$$

where $\tilde{s}_{ij} = \left[\mu_{ij}^{-} - v_{ij}^{+}, \mu_{ij}^{+} - v_{ij}^{-}\right]$.

Step 3: Calculate the interval multiplicative matrix $\left(\tilde{A} = (\tilde{a}_{ij})_{n\times n}\right)$:

$$\tilde{A} = \begin{bmatrix} \left[10^{(\mu_{11}^{-} - v_{11}^{+})}, 10^{(\mu_{1n}^{+} - v_{1n}^{-})}\right] & \cdots & \left[10^{\left(\mu_{ij}^{-} - v_{ij}^{+}\right)}, 10^{\left(\mu_{ij}^{+} - v_{ij}^{-}\right)}\right] \\ \vdots & \ddots & \vdots \\ \left[10^{(\mu_{n1}^{-} - v_{n1}^{+})}, 10^{(\mu_{n1}^{+} - v_{n1}^{-})}\right] & \cdots & \left[10^{(\mu_{nn}^{-} - v_{nn}^{+})}, 10^{(\mu_{nn}^{+} - v_{nn}^{-})}\right] \end{bmatrix} \quad (8)$$

where $\tilde{a}_{ij} = \left[10^{\left(\mu_{ij}^{-} - v_{ij}^{+}\right)}, 10^{\left(\mu_{ij}^{+} - v_{ij}^{-}\right)}\right]$.

Step 4: Calculate the weights $(\widetilde{w}_i)$ interval for each criterion by using Eq. (9):

$$\widetilde{w}_i = \left[\frac{\sum_{j=1}^{n} \tilde{a}_{ij}^{-}}{\sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{a}_{ij}^{+}}, \frac{\sum_{j=1}^{n} \tilde{a}_{ij}^{+}}{\sum_{i=1}^{n}\sum_{j=1}^{n} \tilde{a}_{ij}^{-}}\right] \quad (9)$$

Step 5: Construct the possibility degree matrix $\left(P=\left(p_{ij}\right)_{m\times n}\right)$ by using Eq. (10):

$$P\left(w_i \geq w_j\right)=\frac{min\left\{L_{w_i}+L_{w_j},max\left(w_i{}^{+}-w_j{}^{-},0\right)\right\}}{L_{w_i}+L_{w_j}} \tag{10}$$

where $L_{w_i}=w_i^{+}-w_i^{-}$ and $L_{w_j}=w_j^{+}-w_j^{-}$ and $p_{ij}\geq 0.\, p_{ij}+p_{ji}=1,\, p_{ii}=1/2$.

Step 6: Prioritize the $P=\left(p_{ij}\right)_{m\times n}$ by Eq. (11):

$$w_i=\frac{1}{n}\left[\sum_{j=1}^{n}p_{ij}+\frac{n}{2}-1\right] \tag{11}$$

Step 7: Normalize the weights (w_i) and obtain the final weights.

4. REAL CASE ANALYSIS

In recent years, as a result of immigration from within and outside the country caused the rapid population growth in İstanbul and as well as the increase in the number of vehicles that are used. This issue also emerged problems such as pollution, increment of density of the population, and traffic jam. One of the solutions that is proposed by the local authorities is the popularize the usage of electric vehicles to avoid pollution. When the usage of electric vehicle in Turkey are evaluated, it can be seen that most of the cars is in Istanbul. Through this, Istanbul is decided as pilot zone, and new charging stations should be built in determined alternatives and the use of electric vehicles should be further supported increasing the number of electric charging stations. In this paper, we a methodology based on IVIF AHP to determine the prioritization of alternative areas for charging stations. For this aim 5 criteria which are C1 – Cost, C2 – Geography & Infrastructure, C3 – Reliability & Safety, C4 – Society, C5 – Others and 5 alternatives location are determined in Istanbul which are given as below with their brief characteristics:

AL1-Zorlu Center: Zorlu Center is full of showrooms and flats. There are 584 residences in the building. It is located in the intersection points of 4th Levent, Besiktas, Mecidiyekoy, and Etiler which are the most important business locations of the European Side. Therefore, it can be a suitable alternative for the location of charging station. **AL2-Gumussuyu Borusan Showroom:** The Borusan company is retailer of the BMW, Jaguar, Mini, and Land Rover. The company also sells BMW electric vehicles. The location is the intersection point of Taksim, Besiktas, and close to Mecidiyekoy. Also, lots of companies are located in the same tower. Therefore, it can be a suitable alternative for the location of charging station. **AL3-Istinye Park:** The area is composed of shopping centers, entertainment areas, and housing. There is a housing section of 190 thousand square meters in the area. Istinye Park has been hosted 75 million of visitors during last 5 years. Because of both its location and number of visitors, it can be a proper alternative for the location of charging station. **AL4-Ozdilek Levent:** The Ozdilek Levent is also composed of shopping centers, entertainment areas, and housing in the middle of 4th Levent area. The location is the center of business for the European Side of Istanbul. There are also 2 more campuses, Metro city and Kanyon, those are very close to the Ozdilek. Thus, we determined it as another alternative for the location of charging station. **AL5-Karakoy Dock:** Karakoy is one of the oldest commercial centers famous for its banks and business. The area is connection point between Eminonu and Besiktas. Because of both key location and number of bank and business centers, it is determined as another alternative for the location of charging station.

In the evaluation process, IVIF AHP method is applied to find the most appropriate location alternative for EVs charge station. The linguistic variables, results of the criteria weights and consistency ratio of the comparison matrix with respect to goal are given in Table 2.

Table 2: Pairwise comparison of criteria with respect to Goal

Goal	**C1**	**C2**	**C3**	**C4**	**C5**	**Weight**
C1	EE	EE	VH	VH	VH	0.176
C2	EE	EE	H	H	H	0.329
C3	VL	L	EE	AM	AM	0.185
C4	VL	L	BM	EE	EE	0.155
C5	VL	L	BM	EE	EE	0.155
CR	0.061					

It can be seen that, C2 is the most influential criterion with a value of 0.329 among the criteria with respect to goal. Also, the consistency ratio of the pairwise comparison matrix is equal to 0.061.

The linguistic variables, results of the alternative weights and consistency ratio of the comparison matrix with respect to C1 are given in Table 3.

Table 3: Pairwise comparison of alternatives with respect to C1

C1	AL1	AL2	AL3	AL4	AL5	Weight
AL1	EE	EE	AM	AM	VH	0.160
AL2	EE	EE	H	H	H	0.325
AL3	BM	L	EE	EE	H	0.167
AL4	BM	L	EE	EE	H	0.231
AL5	VL	L	L	L	EE	0.117
CR	0.082					

It can be seen that, AL2 is the most influential alternative with a value of 0.325 among the alternatives with respect to C1. Also, the consistency ratio of the pairwise comparison matrix is equal to 0.082.

The linguistic variables, results of the alternative weights and consistency ratio of the comparison matrix with respect to C2 are given in Table 4.

Table 4: Pairwise comparison of alternatives with respect to C2

C2	AL1	AL2	AL3	AL4	AL5	Weight
AL1	EE	EE	AM	AM	VH	0.159
AL2	EE	EE	AM	AM	VH	0.319
AL3	BM	BM	EE	EE	H	0.172
AL4	BM	BM	EE	EE	H	0.238
AL5	VL	VL	L	L	EE	0.112
CR	0.026					

It can be seen that, AL2 is the most influential alternative with a value of 0.319 among the alternatives with respect to C1. Also, the consistency ratio of the pairwise comparison matrix is equal to 0.026.

The linguistic variables, results of the alternative weights and consistency ratio of the comparison matrix with respect to C3 are given in Table 5.

Table 5: Pairwise comparison of alternatives with respect to C3

C3	AL1	AL2	AL3	AL4	AL5	Weight
AL1	EE	H	AM	AM	AM	0.161
AL2	L	EE	BM	BM	L	0.123
AL3	BM	AM	EE	EE	EE	0.212
AL4	BM	AM	EE	EE	EE	0.241
AL5	BM	H	EE	EE	EE	0.264
CR	0.028					

It can be seen that, AL5 is the most influential alternative with a value of 0.264 among the alternatives with respect to C3. Also, the consistency ratio of the pairwise comparison matrix is equal to 0.028.

The linguistic variables, results of the alternative weights and consistency ratio of the comparison matrix with respect to C4 are given in Table 6.

Table 6: Pairwise comparison of alternatives with respect to C4

C4	AL1	AL2	AL3	AL4	AL5	Weight
AL1	EE	EE	AM	AM	AM	0.168
AL2	EE	EE	AM	AM	AM	0.333
AL3	BM	BM	EE	EE	EE	0.143
AL4	BM	BM	EE	EE	EE	0.178
AL5	BM	BM	EE	EE	EE	0.178
CR	0					

It can be seen that, AL2 is the most influential alternative with a value of 0.333 among the alternatives with respect to C4. Also, the consistency ratio of the pairwise comparison matrix is equal to 0.

The linguistic variables, results of the alternative weights and consistency ratio of the comparison matrix with respect to C5 are given in Table 7.

Table 7: Pairwise comparison of alternatives with respect to C5

C5	AL1	AL2	AL3	AL4	AL5	Weight
AL1	EE	AM	AM	AM	AM	0.159
AL2	BM	EE	BM	BM	BM	0.139
AL3	BM	AM	EE	EE	EE	0.211
AL4	BM	AM	EE	EE	EE	0.245
AL5	BM	AM	EE	EE	EE	0.245
CR	0.042					

It can be seen that, AL4 and AL5 is the most influential alternative with the values of 0.245 among the alternatives with respect to C5. Also, the consistency ratio of the pairwise comparison matrix is equal to 0.042.

The final results of the alternatives and their ranks are given in Table 8.

Table 8: Final weights and ranks of the alternatives

Criteria & Weights		AL1	AL2	AL3	AL4	AL5
0.176	C1	0.160	0.325	0.167	0.231	0.117
0.329	C2	0.159	0.319	0.172	0.238	0.112
0.185	C3	0.161	0.123	0.212	0.241	0.264
0.155	C4	0.168	0.333	0.143	0.178	0.178
0.155	C5	0.159	0.139	0.211	0.245	0.245
Total Weight		0.160	0.260	0.180	0.230	0.170
Rank		5	1	3	2	4

As seen in Table 8, AL2- Gumussuyu Borusan Showroom is determined as best location for electric vehicle charging station. When we evaluate the total weights of the alternatives, AL4-Ozdilek Levent is the second alternatives in the rank, and AL3-Istinye Park, AL4-Ozdilek Levent, respectively. The least desirable location through the calculations is determined as AL1-Zorlu Center.

5. CONCLUSION AND FUTURE SUGGESTIONS

Interest in electric vehicles is increased with the increment in air pollution. Electric vehicle stations have been established in many points in Istanbul lately; however, works continue to establish more. In this paper, we have tried to choose the most suitable alternative for establishing a new electric charging station in Istanbul. We apply interval-valued intuitionistic fuzzy (IVIF) AHP to determine the prioritization of alternative areas for charging stations. As a result of our calculations, we determined at what point in Istanbul the electric vehicle charging station should be installed at firstly. For further research, new MCDM methods can be used, or different extensions of fuzzy sets can be used to evaluate the selected locations and the obtained results can be compared.

REFERENCES

Abdullah, L., & Ismail, W. K. (2012). Hamming distance in intuitionistic fuzzy sets and interval-valued intuitionistic fuzzy sets: a comparative analysis. *Advances in Computational Mathematics and its Applications*, 1(1), 7-11.

Atanassov, K. (1986). Intuitionistic fuzzy sets. Fuzzy sets and Systems, 20(1), 87-96.

Eisel, M., Schmidt, J., & Kolbe, L. M. (2014, December). Finding suitable locations for charging stations. In *Electric Vehicle Conference (IEVC), 2014 IEEE International* (pp. 1-8). IEEE.

Gonzalez, J., Alvaro, R., Gamallo, C., Fuentes, M., Fraile-Ardanuy, J., Knapen, L., & Janssens, D. (2014). Determining electric vehicle charging point locations considering drivers' daily activities. *Procedia Computer Science*, 32, 647-654.

He, F., Wu, D., Yin, Y., & Guan, Y. (2013). Optimal deployment of public charging stations for plug-in hybrid electric vehicles. *Transportation Research Part B: Methodological*, 47, 87-101.

Islam, M. M., Shareef, H., & Mohamed, A. (2015). A review of techniques for optimal placement and sizing of electric vehicle charging stations. *Elect. Review*, 91, 122-126.

Jia, L., Hu, Z., Song, Y., & Luo, Z. (2012, March). Optimal siting and sizing of electric vehicle charging stations. In *Electric Vehicle Conference (IEVC), 2012 IEEE Internationa*l (pp. 1-6). IEEE.

Kley, F., Lerch, C., & Dallinger, D. (2011). New business models for electric cars—A holistic approach. *Energy policy*, 39(6), 3392-3403.

Kobayashi, Y., Kiyama, N., Aoshima, H., & Kashiyama, M. (2011, June). A route search method for electric vehicles in consideration of range and locations of charging stations. *In Intelligent Vehicles Symposium (IV),* 2011 *IEEE*(pp. 920-925). IEEE.

Lam, A. Y., Leung, Y. W., & Chu, X. (2013, October). Electric vehicle charging station placement. In Smart Grid Communications (SmartGridComm), *2013 IEEE International Conference on* (pp. 510-515). IEEE.

Lam, A. Y., Leung, Y. W., & Chu, X. (2014). Electric vehicle charging station placement: Formulation, complexity, and solutions. *IEEE Transactions on Smart Grid*, 5(6), 2846-2856.

Liu, Z., Wen, F., & Ledwich, G. (2013). Optimal planning of electric-vehicle charging stations in distribution systems. *IEEE Transactions on Power Delivery*, 28(1), 102-110.

Micari, S., Napoli, G., Antonucci, V., & Andaloro, L. (2014, December). Electric vehicles charging stations network-A preliminary evaluation about Italian highways. In *Electric Vehicle Conference (IEVC), 2014 IEEE International* (pp. 1-5). IEEE.

Pan, F., Bent, R., Berscheid, A., & Izraelevitz, D. (2010, October). Locating PHEV exchange stations in V2G. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference* on (pp. 173-178). IEEE.

Pan, F., Bent, R., Berscheid, A., & Izraelevitz, D. (2010, October). Locating PHEV exchange stations in V2G. In *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on* (pp. 173-178). IEEE.

Park, J. H., Park, I. Y., Kwun, Y. C., & Tan, X. (2011). Extension of the TOPSIS method for decision making problems under interval-valued intuitionistic fuzzy environment. *Applied Mathematical Modelling*, 35(5), 2544-2556.

Pazouki, S., Mohsenzadeh, A., & Haghifam, M. R. (2013, December). Optimal planning of PEVs charging stations and demand response programs considering distribution and traffic networks. In *Smart Grid Conference (SGC)*, 2013 (pp. 90-95). IEEE.

Sadeghi-Barzani, P., Rajabi-Ghahnavieh, A., & Kazemi-Karegar, H. (2014). Optimal fast charging station placing and sizing. *Applied Energy*, 125, 289-299.

Schill, W. P. (2011). Electric vehicles in imperfect electricity markets: The case of Germany. *Energy Policy*, 39(10), 6178-6189.

Traut, E., Hendrickson, C., Klampfl, E., Liu, Y., & Michalek, J. J. (2012). Optimal design and allocation of electrified vehicles and dedicated charging infrastructure for minimum life cycle greenhouse gas emissions and cost. *Energy Policy*, 51, 524-534.

Wang, Y. W., & Lin, C. C. (2009). Locating road-vehicle refueling stations. *Transportation Research Part E: Logistics and Transportation Review*, 45(5), 821-829.

Wei, G., & Wang, X. (2007). Some geometric aggregation operators based on interval-valued intuitionistic fuzzy sets and their application to group decision making. In, 2007 (pp. *International Conference on Computational Intelligence and Security* (pp. 495-499). Harbin: IEEE.

Wirges, J., Linder, S., & Kessler, A. (2012). Modelling the development of a regional charging infrastructure for electric vehicles in time and space. *European Journal of Transport and Infrastructure Research*, 12(12), 391-416.

Wu, J., Huang, H. B., & Cao, Q. W. (2013). Research on AHP with interval-valued intuitionistic fuzzy sets and its application in multi-criteria decision making problems. *Applied Mathematical Modelling*, 37(24), 9898-9906.

Yi, Z., & Bauer, P. H. (2014, April). Energy Consumption Model and Charging Station Placement for Electric Vehicles. In *Smartgreens* (pp. 150-156).

BALCOR 2018
XIII Balkan Conference on Operational Research

AN ENTROPY BASED GROUP DECISION MAKING MODEL INTEGRATING ELECTRE AND VIKOR UNDER INTUITIONISTIC FUZZY ENVIRONMENT

SEDEF ÇALI[1], ŞEBNEM YILMAZ BALAMAN[2]

[1] Dokuz Eylul University, Department of Industrial Engineering, Izmir, Turkey, sedefcali@gmail.com
[2] Dokuz Eylul University, Department of Industrial Engineering, Izmir, Turkey, s.yilmaz@deu.edu.tr

Abstract: *This paper presents the integration of two well-known multiple criteria decision making methods, namely Elimination and Choice Translating Reality (ELECTRE) and Visekriterijumsko kompromisno rangiranje Resenje (VIKOR) under intuitionistic fuzzy environment to solve group decision making problems. This novel methodology employs entropy method to cope with the imprecise knowledge on weights of criteria and decision makers. An illustrative example regarding supplier selection is used for demonstrating the solution process of developed methodology. In addition, this model is implemented to four cases dealing with project manager selection, manufacturing system selection, financial evaluation of companies and university leader selection for the purpose of proving it's applicability in different fields.*

Keywords: *ELECTRE, VIKOR, Entropy, Intuitionistic Fuzzy Set Theory*.

1. INTRODUCTION

Multiple criteria decision making (MCDM) methods focus on finding the best option among a set of alternatives taking into account several conflicting criteria. One of the application field of these methods is group decision making (GDM) problems in which the assessment of alternatives are realized according to knowledge and experiences of two or more experts. The classical MCDM methods handling crisp data are insufficient for GDM problems in which the experts cannot precisely evaluate the alternatives due to the incomplete information. For this reason, intuitionistic fuzzy sets (IFSs) developed by Atanassov (1986) are employed to cope with the vagueness caused by incomplete information in GDM problems. IFSs differentiate from ordinary fuzzy sets proposed by Zadeh (1965) in terms of considering non-membership degree and hesitancy degree along with membership degree of an element to a set.

There are studies extending MCDM methods with IFSs for GDM problems. For instance, Roostaee et al. (2012) extended VIKOR method under intuitionistic fuzzy (IF) environment to solve supplier selection problem. Vahdani et al. (2013) proposed GDM methodology based on IF-ELECTRE and applied it to an example in flexible manufacturing systems. Yue (2014) presented TOPSIS-based methodology in which the group decision information expressed by crisp data were converted into intuitionistic fuzzy numbers and developed model was implemented to the problem of university leader selection. As for our research, we handled ELECTRE I and VIKOR methods.

The family of ELECTRE methods is outranking based methods. They benefit from pairwise comparisons of alternatives in order to obtain outranking relations. The first member of this family ELECTRE I was developed by Roy (1968) with the purpose of handling choice problematic which aims to find the best alternative among a small set of preferable alternatives. However, ELECTRE I has not capable of complete ranking of alternatives. On the other hand, VIKOR developed by Opricovic (1998) proposes compromise solution(s) which is the closest to ideal point and compromise rankings by employing an aggregation function handling maximum group utility of majority and minimum individual regret of opponent. VIKOR method was designed for ranking problematic which aims to rank the alternatives from best to the worst. The decision makers can obtain different rankings according to their preferences since VIKOR takes into account the weights of criteria and DM's strategies regarding group utility and individual regret in the process of ranking. Zandi and Rogharian (2013) extended fuzzy ELECTRE I based on VIKOR method to tackle the shortcoming of ELECTRE I in complete ranking. We combined IF-ELECTRE with VIKOR to overcome the partial ranking problem of IF-ELECTRE. In addition, we employed entropy method to cope with GDM problems in which the weights of criteria and weights of DMs are completely unknown.

The remaining of the paper involves four sections. Section 2 explains the basic definitions about IFS theory, Section 3 illustrates the developed methodology, Section 4 proposes experimental analysis and Section 5 presents conclusions.

2. PRELIMINARIES

Atanassov (1986) introduced IFSs as a generalization of fuzzy sets (Zadeh 1965).

Definition 2.1. Consider there exists a finite set X, the IFS A on X is defined as follows:

$$A = \{< x,\ \mu_A(x),\ v_A(x) > |\ x \in X\}$$

where the functions $\mu_A\colon \rightarrow [0,1]$ and $v_A\colon \rightarrow [0,1]$ define the membership degree and non-membership degree, respectively, of the element $x \in X$ to the set A with the condition $0 \le \mu_A(x) + v_A(x) \le 1$. In addition, $\pi_A\colon \rightarrow [0,1]$ is the hesitancy function which defines the hesitancy degree of x with $\pi_A(x) = 1 - \mu_A(x) - v_A(x)$ The IFS A can be reduced to ordinary fuzzy set when $\pi_A(x) = 0$.

Definition 2.2. The basic components of an IFS are the intuitionistic fuzzy numbers (IFNs) defined by membership degree (μ) and non-membership degree (v) of an object. Let $\alpha = (0.6, 0.2)$ be an IFN. The physical interpretation of this IFN is regarded as "the vote for resolution is 6 in favor, 2 against, and 2 abstentions" (Xu 2007). The score function and accuracy function are employed for comparison of IFNs. Let $\alpha = (\mu_\alpha, v_{\alpha)}$ be an IFN. The score value of this IFN can be calculated with the score function $s(\alpha) = \mu_\alpha - v_\alpha$. (Chen and Tan 1994). The higher score value indicates the bigger IFN. This function is insufficient in comparison when IFNs have same score values. Therefore, Hong and Choi (2000) introduced the accuracy function $h(\alpha) = \mu_\alpha + v_a$ to calculate accuracy values of IFNs. The higher accuracy value indicates the bigger IFN. That is to say, the bigger IFN has smaller hesitancy degree.

Definition 2.3. Distance measures indicate the differences between IFSs. Szmidt and Kacprzyk (2000) extended some basic distance measures such as Hamming, Euclidean and their normalized forms for fuzzy sets to propose distance measures for IFSs. In this paper, the Euclidean distance is employed to measure the differences between IFSs.

Let A and B be IFSs in $X = \{x_1, x_2 \dots x_n\}$, the Euclidean distance $d_E(A, B)$ is calculated with Eq. (1).

$$d_E(A,B) = \sqrt{\tfrac{1}{2}\textstyle\sum_{j=1}^{n}\left(\mu_A(x_j) - \mu_B(x_j)\right)^2 + \left(v_A(x_j) - v_B(x_j)\right)^2 + \left(\pi_A(x_j) - \pi_B(x_j)\right)^2} \tag{1}$$

3. IF-ELECTRE INTEGRATED WITH VIKOR

The developed MCGDM method whose main steps are illustrated in Figure 1 is explained in this section.

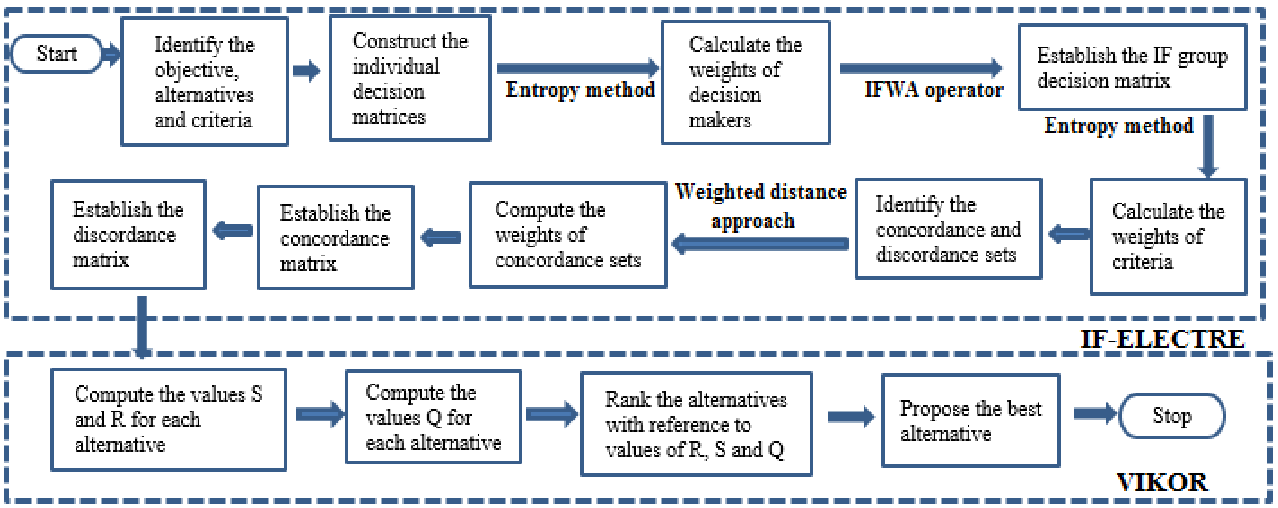


Figure 1. The developed methodology

The decision makers (DMs) denoted as $E = \{E_1, E_2, \dots, E_k\}$ evaluate each alternative defined as $A = \{A_1, A_2, \dots, A_m\}$ with reference to each criterion denoted as $X = \{x_1, x_2, \dots, x_n\}$ based on IFS theory. Alternative A_i is characterized by the IFS as $A_i = \{< x_j,\ \mu_{ij}, v_{ij} > | x_j \in X\},\ i = 1,2, \dots m$ where μ_{ij} indicates satisfaction degree of ith alternative with reference to jth criterion, v_{ij} indicates nonsatisfaction degree of ith alternative with reference to jth criterion. $\mu_{ij} \in [0,1]$, $v_{ij} \in [0,1]$, $\mu_{ij} + v_{ij} \in [1,0]$, $\pi_{ij} = 1 - \mu_{ij} - v_{ij}$ indicates hesitancy degree of ith alternative with reference to jth criterion. The weights of DMs are indicated by the set of $\delta = \{\delta_1, \delta_2, \dots\ \delta_k\}$. The weights of criteria are denoted as $W = \{w_1, w_2, \dots, w_n\}$.

Step 1. Identify objective, alternatives and criteria: The components of MCGDM problem are defined in this step. The scope of the problem, alternatives and the criteria used for evaluation of alternatives are defined.

Step 2. Construct the individual decision matrices: Each *k*th decision maker (DM) evaluates the alternatives regarding each criterion using linguistic variables defined by IFNs. Therefore, the individual decision matrices $R^{(k)}$ are established.

$$R^{(k)} = \begin{bmatrix} r_{11}^{(k)} & \cdots & r_{1n}^{(k)} \\ \vdots & \ddots & \vdots \\ r_{m1}^{(k)} & \cdots & r_{mn}^{(k)} \end{bmatrix} = \begin{bmatrix} (\mu_{11}^{(k)}, v_{11}^{(k)}, \pi_{11}^{(k)}) & \cdots & (\mu_{1n}^{(k)}, v_{1n}^{(k)}, \pi_{1n}^{(k)}) \\ \vdots & \ddots & \vdots \\ (\mu_{m1}^{(k)}, v_{m1}^{(k)}, \pi_{m1}^{(k)}) & \cdots & (\mu_{mn}^{(k)}, v_{mn}^{(k)}, \pi_{mn}^{(k)}) \end{bmatrix}$$

Step 3. Calculate the weights of decision makers: Each DM has different relative importance since each of them has different level of experience and knowledge about the alternatives. The weights of DMs are determined by using entropy measure. Each individual decision matrix $R^k = (r_{ij}^k)_{mxn}$ is regarded as an IFS and then the uncertain information of each R^k is calculated with Eq. (2).

$$E_{LT}^{IFS}\left(R^{(k)}\right) = -\frac{1}{mn\ln 2}\sum_{j=1}^{n}\sum_{i=1}^{m}[\mu_{ij}\ln\mu_{ij} + v_{ij}\ln v_{ij} - (1-\pi_{ij})\ln\left(1-\pi_{ij}\right) - \pi_{ij}\ln 2] \quad (2)$$

where $i = 1,2,\ldots m$ indicate alternatives and $j = 1,2,\ldots n$ indicate criteria.
After determination of divergence degrees $d_{R^{(k)}}$ with Eq. (3), the weights of DMs are computed with Eq. (4).

$$d_{R^{(k)}} = 1 - E_{LT}^{IFS}\left(R^{(k)}\right) \quad (3)$$

$$\delta_k = \frac{d_{R^{(k)}}}{\sum_{k=1}^{K} d_{R^{(k)}}} \qquad \text{where } \delta_k \geq 0,\ k = 1,2,\ldots K \text{ and } \sum_{k=1}^{K}\delta_k = 1. \quad (4)$$

Step 4. Establish IF group decision matrix: Each individual decision matrix is aggregated by IFWA operator (Xu 2007) to establish the group decision matrix *R* depicted as follows:

$$R = \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{bmatrix} = \begin{bmatrix} (\mu_{11}, v_{11}, \pi_{11}) & \cdots & (\mu_{1n}, v_{1n}, \pi_{1n}) \\ \vdots & \ddots & \vdots \\ (\mu_{m1}, v_{m1}, \pi_{m1}) & \cdots & (\mu_{mn}, v_{mn}, \pi_{mn}) \end{bmatrix}$$

The component r_{ij} indicating the overall judgment about *i*th alternative according *j*th criterion is determined with Eq. (5).

$$r_{ij} = IFWA_\delta(r_{ij}^{(1)}, r_{ij}^{(2)}, \ldots, r_{ij}^{(K)}) =$$
$$= [1 - \prod_{k=1}^{K}\left(1-\mu_{ij}^{(k)}\right)^{\delta_k},\ \prod_{k=1}^{K}\left(v_{ij}^{(k)}\right)^{\delta_k},\ \prod_{k=1}^{K}\left(1-\mu_{ij}^{(k)}\right)^{\delta_k} - \prod_{k=1}^{K}(v_{ij}^{(k)})^{\delta_k} \quad (5)$$

Step 5. Calculate the weights of criteria: Entropy measure can be used for criterion weighting. The information transmitted by criterion x_j can be measured with entropy. Lower entropy measure indicates lower uncertainty. Therefore, higher weight is assigned to criterion with lower entropy measure. The IF entropy measure (Vlochos and Sergiadis 2007) is employed to determine weights of criteria. The following steps are implemented (Hung and Cheng 2009). Firstly, the entropy measures are obtained via Eq. (6).

$$E_{LT}^{IFS}\left(x_j\right) = -\frac{1}{m\ln 2}\sum_{i=1}^{m}[\mu_{ij}\ln\mu_{ij} + v_{ij}\ln v_{ij} - (1-\pi_{ij})\ln(1-\pi_{ij}) - \pi_{ij}\ln 2] \quad (6)$$

After determination of divergence degrees d_j with Eq. (7), the weights of criteria are computed with Eq. (8).

$$d_j = 1 - E_{LT}^{IFS}\left(x_j\right) \quad (7)$$

$$w_j = \frac{d_j}{\sum_{j=1}^{n} d_j} \qquad \text{where } w_j \geq 0,\ j = 1,2,\ldots n \text{ and } \sum_{j=1}^{n} w_j = 1. \quad (8)$$

Step 6. Identify the concordance and discordance sets: Three different concordance and discordance sets are formed in this step. We inspired from Wu & Chen in which the formulations of sets are based on score function and accuracy function mentioned in Section 2. The strong, midrange and weak concordance sets for each pair of alternatives are formulated respectively as follow:

$$C'_{ab} = \{j | \mu_{aj} \geq \mu_{bj},\ v_{aj} < v_{bj}\ and\ \pi_{aj} < \pi_{bj}\} \quad (9)$$
$$C''_{ab} = \{j | \mu_{aj} \geq \mu_{bj},\ v_{aj} < v_{bj}\ and\ \pi_{aj} \geq \pi_{bj}\} \quad (10)$$
$$C'''_{ab} = \{j | \mu_{aj} \geq \mu_{bj},\ v_{aj} \geq v_{bj}\}. \quad (11)$$

The strong, midrange and weak discordance sets are formulated respectively as follow:

$$D'_{ab} = \{j | \mu_{aj} < \mu_{bj},\ v_{aj} \geq v_{bj}\ and\ \pi_{aj} \geq \pi_{bj}\} \quad (12)$$
$$D''_{ab} = \{j | \mu_{aj} < \mu_{bj},\ v_{aj} \geq v_{bj}\ and\ \pi_{aj} < \pi_{bj}\} \quad (13)$$
$$D'''_{ab} = \{j | \mu_{aj} < \mu_{bj},\ v_{aj} < v_{bj}\}. \quad (14)$$

Step 7. Compute the weights of concordance sets: We employed weighted distance approach adapted from (Zhang et al. 2017) for identifying the total dominance degrees of strong, midrange and weak concordance sets. The dominance degree of alternative A_a over alternative A_b in respect to criterion x_j, a member of concordance set, can be calculated with weighted distance measure $w_j * d(r_{aj}, r_{bj})$. The total dominance degrees of strong, midrange and weak concordance sets are defined as $dd' = \sum_{a,b=1, a\neq b}^{m} \sum_{j\in C'_{ab}} w_j * d(r_{aj}, r_{bj})$, $dd'' = \sum_{a,b=1, a\neq b}^{m} \sum_{j\in C''_{ab}} w_j * d(r_{aj}, r_{bj})$ and $dd'' = \sum_{a,b=1, a\neq b}^{m} \sum_{j\in C''_{ab}} w_j * d(r_{aj}, r_{bj})$,respectively.

Euclidean distance measure defined by Eq. (1) is used for calculations of distances between IFSs. The weights of strong, midrange and weak concordance sets are calculated based on total dominance degrees of these sets.

$$w_{C'} = dd'/(dd' + dd'' + dd''') \quad (15)$$
$$w_{C''} = dd''/(dd' + dd'' + dd''') \quad (16)$$
$$w_{C'''} = dd'''/(dd' + dd'' + dd''') \quad (17)$$

Step 8. Establish the concordance matrix: Concordance matrix is made up of concordance indices calculated on the basis of comprehensive concordance index (Figueira et al. 2010). It reflects the superiority of one alternative to another alternative. The concordance index g_{ab} between $A_a$and A_b is identified as:

$$g_{ab} = w_{C'} * \sum_{j\epsilon C'_{ab}} w_j + w_{C''} * \sum_{j\epsilon C''_{ab}} w_j + w_{C'''} * \sum_{j\epsilon C'''_{ab}} w_j \quad (18)$$

Step 9. Establish the discordance matrix: Discordance matrix is composed of discordance indices which indicate the inferiority degree of alternatives according to each other. The discordance index h_{ab} between $A_a$and A_b is identified with Eq. (19) (Zhang et al. 2017).

$$h_{ab} = \frac{\max_{j\in D'_{ab}\cup D''_{ab}\cup D'''_{ab}} w_j * d(r_{aj}, r_{bj})}{\max_{j\in J} d(r_{aj}, r_{bj})} \quad (19)$$

Step 10. Compute the values S (group utility) and R (individual regret) for each alternative: In this step, VIKOR method can be integrated to the system. Opricovic and Tzeng (2007) presented the similarities of discordance concept of ELECTRE and R_i in VIKOR as well as the similarities of concordance concept of ELECTRE and S_i in VIKOR. They developed aggregating discordance index and aggregating concordance index. With the help of these indices, the complete ranking is realized.

The aggregating discordance index of alternative A_a is calculated as follows:

$$h_a = R_a/C \quad (20)$$
$$C = \max_j w_j \text{ and } h_a = \max_i h_{ai} \quad i \neq a \quad i = \{1,2, \dots m\} \quad (21)$$

The aggregating concordance index of alternative A_a is calculated as follows:

$$g_a = 1 - S_a \quad (22)$$

$$g_a = \sum_{i\neq a}^{m} g_{ai} / (m-1) \quad (23)$$

where $g_{ai} = w_{C'} * \sum_{j\epsilon C'_{ai}} w_j + w_{C''} * \sum_{j\epsilon C''_{ai}} w_j + w_{C'''} * \sum_{j\epsilon C'''_{ai}} w_j$
where alternatives $i = \{1,2, \dots m\}$ and criteria $j = \{1,2, \dots n\}$

Therefore, the values R and the values S of each alternative i is calculated as follows:

$$R_i = h_i * C \quad (24)$$
$$S_i = 1 - g_i \quad (25)$$

Step 11. Compute the values Q (degree of closeness) for each alternative

$$Q_i = \gamma * \frac{S_i - S^*}{S^- - S^*} + (1-\gamma) * \left(\frac{R_i - R^*}{R^- - R^*}\right) \qquad i = 1,2, \dots m \quad (26)$$

where $S^- = max\, S_i$; $S^* = min\, S_i$; $R^- = maxR_i$; $R^* = minR_i$; and the notation of γ indicates a weight for the strategy of maximum group utility, whereas $(1-\gamma)$ indicates a weight of the individual regret.

Step 12. Rank the alternatives, sorting by the values S, R and Q in decreasing order: Three different ranking lists are formed based on ascending order of values Q, S and R. The compromise solution is the alternative with minimum Q value only if the following conditions are satisfied:

Condition 1: Consider A_1 is the first and A_2 is the second according the ranking of Q. Alternative A_1 has a meaningful advantage over the alternative A_2 by verifying the following equation:

$$Q(A_2) - Q(A_1) \geq DQ \text{ where } DQ = 1 / (m-1) \quad (27)$$

Condition 2: The alternative A_1 has acceptable stability in decision making if it is in the first position according to both ranking of S values and ranking of R values.

If alternative A_1 satisfies the only one condition, then a set of compromise solutions is proposed. If Condition 2 is not satisfied, then A_1 and A_2 are the compromise solutions. If Condition 1 is not satisfied, then A_1, A_1,… A_m are the compromise solutions where $Q(A_m) - Q(A_1) < DQ$ and $DQ = 1 / (m-1)$.

4. EXPERIMENTAL ANALYSIS

The supplier selection problem adapted from Rooestaee et al. (2012) is firstly handled to illustrate the calculation process of the developed methodology. An automotive company will purchase a component used in production of their new model automobiles. Three experts in the company will evaluate possible suppliers by considering several criteria. IF-ELECTRE integrated with VIKOR method is applied to this problem.

Step 1. The objective is to propose best supplier. There are three alternative suppliers denoted by A1, A2 and A3. These suppliers are evaluated by three experts $E_k (k = 1,2,3)$ according to these criteria: X1: On time delivery, X2: Closeness of relationship with the supplier, X3: Supplier's product quality, X4: Supplier's technological capability and X5: Price and Cost.

Step 2. Each DM evaluates the alternatives with reference to criterion using linguistic variables converted into IFNs. Each individual decision matrix is depicted in **Table 1**.

Table 1. Individual decision matrices

		X1	X2	X3	X4	X5
	A1	(0.60,0.30,0.10)	(0.70,0.20,0.10)	(0.50,0.40,0.10)	(0.80,0.10,0.10)	(0.50,0.40,0.10)
$R^{(1)} =$	A2	(0.70,0.20,0.10)	(0.80,0.10,0.10)	(0.80,0.10,0.10)	(0.80,0.10,0.10)	(0.80,0.10,0.10)
	A3	(0.80,0.10,0.10)	(0.60,0.30,0.10)	(0.70,0.20,0.10)	(0.70,0.20,0.10)	(0.70,0.20,0.10)
	A1	(0.70,0.20,0.10)	(0.60,0.30,0.10)	(0.70,0.20,0.10)	(0.70,0.20,0.10)	(0.50,0.40,0.10)
$R^{(2)} =$	A2	(0.70,0.20,0.10)	(0.80,0.10,0.10)	(0.80,0.10,0.10)	(0.80,0.10,0.10)	(0.60,0.30,0.10)
	A3	(0.70,0.20,0.10)	(0.70,0.20,0.10)	(0.60,0.30,0.10)	(0.80,0.10,0.10)	(0.70,0.20,0.10)
	A1	(0.60,0.30,0.10)	(0.50,0.40,0.10)	(0.70,0.20,0.10)	(0.80,0.10,0.10)	(0.50,0.40,0.10)
$R^{(3)} =$	A2	(0.60,0.30,0.10)	(0.80,0.10,0.10)	(0.70,0.20,0.10)	(0.80,0.10,0.10)	(0.70,0.20,0.10)
	A3	(0.50,0.40,0.10)	(0.80,0.10,0.10)	(0.80,0.10,0.10)	(0.60,0.30,0.10)	(0.60,0.30,0.10)

Step 3. The weights of DMs are computed with Eq. (2)-(3)-(4). The entropy measures for DM1, DM2 and DM3, respectively, are: 0.739, 0.766 and 0.787. The divergence values are obtained as 0.261, 0.234 and 0.213. Therefore, the weights of DMs are calculated as 0.369, 0.330 and 0.301, respectively.

Step 4. The group decision matrix is constructed with Eq. (5). All individual preferences are aggregated to obtain group opinion. The group decision matrix is depicted in **Table 2**.

Table 2. IF-group decision matrix

R=	X1	X2	X3	X4	X5
A1	(0.636,0.262,0.102)	(0.615,0.282,0.103)	(0.638,0.258,0.104)	(0.771,0.126,0.103)	(0.500,0.400,0.100)
A2	(0.673,0.226,0.101)	(0.800,0.100,0.100)	(0.774,0.123,0.103)	(0.800,0.100,0.100)	(0.716,0.177,0.107)
A3	(0.699,0.191,0.110)	(0.705,0.188,0.107)	(0.708,0.186,0.106)	(0.714,0.180,0.106)	(0.673,0.226,0.101)

Step 5. The weights of criteria are computed with Eq. (6)-(7)-(8). The entropy measures for criteria X1, X2, X3, X4 and X5, respectively are: 0.831, 0.744, 0.755, 0.644 and 0.857. The weights of criteria are determined as 0.144, 0.219, 0.210, 0.305 and 0.122, respectively.

Step 6. The strong, midrange and weak concordance sets are determined with Eq. (9)-(10)-(11), the strong, midrange and weak discordance sets are specified with Eq. (12)-(13)-(14). They are available below.

$$[C'_{12} = \{\emptyset\},\ C''_{12} = \{\emptyset\},\ C'''_{12} = \{\emptyset\},\ D'_{12} = \{1,2,3,4\},\ D''_{12} = \{5\},\ D'''_{12} = \{\emptyset\}]$$

$$[C'_{13} = \{4\},\ C''_{13} = \{\emptyset\},\ C'''_{13} = \{\emptyset\},\ D'_{13} = \{\emptyset\},\ D''_{13} = \{1,2,3,5\},\ D'''_{13} = \{\emptyset\}]$$

$$[C'_{21} = \{1,2,3,4\},\ C''_{21} = \{5\},\ C'''_{21} = \{\emptyset\},\ D'_{21} = \{\emptyset\},\ D''_{21} = \{\emptyset\},\ D'''_{21} = \{\emptyset\}]$$

$$[C'_{23} = \{2,3,4\},\ C''_{23} = \{5\},\ C'''_{23} = \{\emptyset\},\ D'_{23} = \{\emptyset\},\ D''_{23} = \{1\},\ D'''_{23} = \{\emptyset\}]$$

$$[C'_{31} = \{\emptyset\},\ C''_{31} = \{1,2,3,5\},\ C'''_{31} = \{\emptyset\},\ D'_{31} = \{4\},\ D''_{31} = \{\emptyset\},\ D'''_{31} = \{\emptyset\}]$$

$$[C'_{32} = \{\emptyset\},\ C''_{32} = \{1\},\ C'''_{32} = \{\emptyset\},\ D'_{32} = \{2,3,4\},\ D''_{32} = \{5\},\ D'''_{32} = \{\emptyset\}]$$

Step 7. The weights of concordance sets are calculated with Eq. (15)-(16)-(17). The weights of strong, midrange and weak concordance sets are 0.606, 0.394 and 0, respectively.

Step 8. After the determination of all concordance indices by using Eq. (18), the concordance matrix G is established as following form.

$$G = \begin{bmatrix} - & 0.0 & 0.185 \\ 0.580 & - & 0.492 \\ 0.274 & 0.057 & - \end{bmatrix}$$

For instance, the calculation of concordance index g_{21} is as follows:

$$g_{21} = 0.606 * (0.144 + 0.219 + 0.210 + 0.305) + 0.394 * (0.122) = 0.580$$

Step 9. After the determination of all discordance indices by using Eq. (19), the discordance matrix H is formed as follows:

$$H = \begin{bmatrix} - & 0.183 & 0.122 \\ 0.0 & - & 0.049 \\ 0.098 & 0.275 & - \end{bmatrix}$$

For instance, the discordance index h_{12} is calculated by:

$$h_{12} = \frac{max[(0.144 * 0.036), (0.219 * 0.183), (0.210 * 0.135), (0.305 * 0.027), (0.122 * 0.219)\ ]}{max[0.036, 0.183, 0.135, 0.027, 0.219]}$$

$$h_{12} = \frac{0.040}{0.219} = 0.183$$

Step 10-12. The characteristics of VIKOR are obtained by means of concordance and discordance matrices in this step. The values of R_i are calculated using Eq. (20)-(21)-(24) and the values of S_i are determined using Eq. (22)-(23)-(25). For instance, S_1 and R_1 are obtained as follows:

$$g_1 = \frac{g_{12} + g_{13}}{2} = \frac{0.0 + 0.185}{2} = 0.092, \qquad S_1 = 1 - g_1 = 1 - 0.092 = 0.908$$
$$C = max[w_1, w_2, w_3, w_4, w_5\,] = max[0.144, 0.219, 0.210, 0.305, 0.122] = 0.305$$
$$h_1 = max[h_{12}, h_{13}] = max[0.183, 0.122] = 0.183, \;\; R_1 = h_1 * C = 0.183 * 0.305 = 0.056$$

The degrees of closeness Q_i are obtained by Eq. (26). The results of calculations are demonstrated in **Table 3.** The ranking is obtained as A2 > A1 > A3 according to strategy of minimum individual regret (min R), whereas the ranking is A2 > A3 > A1 according to the strategy of maximum group utility (min S). Considering the weight as ($\gamma = 0.5$), we obtained the ranking A2 > A1 > A3 for consensus case. The alternative A2 satisfies Eq. (27) and at the first position both in ranking of S and ranking of R. Therefore, A2 can be proposed as compromise solution.

Table 3. The rankings of suppliers for group utility, individual regret and closeness to ideal

	Suppliers			Rankings
i=1,2,3.	A1	A2	A3	
Q (A*i*) ($\gamma = 0.5$)	0.797	0.00	0.918	A2 > A1 > A3
S (A*i*)	0.908	0.464	0.835	A2 > A3 > A1
R (A*i*)	0.056	0.015	0.084	A2 > A1 > A3

We also applied our methodology to four examples in different fields for testing whether the methodology proposes coherent solutions and rankings. We also handled the examples in which GDM problem is not under consideration. For all of these examples, our methodology gave consistent solutions depicted in **Table 4**.

Table 4. The comparison of rankings in example articles with rankings given by our methodology

Article	Problem	GDM	Solution in article	IF-ELECTRE-VIKOR (γ =0.5)
Wu and Chen (2011)	Project manager selection	No	A6>A4>A5>A3>A2>A1	A6>A5>A3>A2>A4>A1
Vahdani et al. (2013)	Manufacturing system selection	Yes	A3>A1>A2>A4>A5	A3>A4>A2>A1>A5
Joshi and Kumar (2014)	Financial evaluation of companies	No	A3>A1>A4>A2	A3>A1>A4>A2
Yue (2014)	University leader selection	Yes	A2>A3>A4>A1	A2>A3>A4>A1

5. CONCLUSION

We developed an innovative approach by extending IF-ELECTRE I based on VIKOR with the purpose of solving GDM problems. This hybrid methodology combines the strengths of these methods. It can propose different ranking lists for each DM who follows different decision making strategies as well as optimal solution of the problem. This methodology has advantages in terms of obtaining unknown weights of DMs and criteria by means of entropy and the weights of concordance sets via weighted distance approach. The applicability of the method to different areas is proved by the coherent solutions obtained by the developed methodology.

Acknowledgement

This work was supported by Dokuz Eylül University Scientific Research Coordination Unit. Project Number: 2017.KB.FEN.026

REFERENCES

Atanassov, K.T. (1986). Intuitionistic fuzzy sets. *Fuzzy sets and Systems*, 20(1), 87-96.

Chen, S.M., & Tan, I.M. (1994). Handling multicriteria fuzzy decision-making problems based on vague set theory. *Fuzzy Sets and Systems*, 67, 163-172.

Figueira, J.R., Greco, S., Roy, B., & Słowiński, R. (2010). ELECTRE methods: main features and recent developments. In Zopounidis, C. & Pardalos, P.M. (Eds.), *Handbook of multicriteria analysis*, Berlin: Springer, 51-89.

Hong, D.H., & Choi, C.H. (2000). Multicriteria fuzzy decision-making problems based on vague set theory. *Fuzzy sets and Systems*, 114(1), 103-113.

Hung, C.C., & Chen, L.H. (2009). A multiple criteria group decision making model with entropy weight in an intuitionistic fuzzy environment. In Castillo, H. & Huang, H. (Eds.), *Intelligent automation and computer engineering.* Netherlands: Springer, 17-26.

Joshi, D., & Kumar, S. (2014). Intuitionistic fuzzy entropy and distance measure based TOPSIS method for multi-criteria decision making. *Egyptian Informatics Journal*, 15(2), 97–104.

Opricovic, S. (1998). Multicriteria optimization of civil engineering systems. Faculty of Civil Engineering, Belgrade, 2(1), 5-21.

Opricovic, S., & Tzeng, G.H. (2007). Extended VIKOR method in comparison with outranking methods. *European Journal of Operational Research*, 178(2), 514-529.

Roostaee, R., Izadikhah, M., Lotfi, F. H., & Rostamy-Malkhalifeh, M., A (2012). Multi-criteria intuitionistic fuzzy group decision making method for supplier selection with VIKOR method. *International Journal of Fuzzy System Applications*, 2(1), 1-17.

Roy, B. (1968). Classement et choix en presence de points de vue multiples (La method ELECTRE). *Revue Francaise d'Informatique et de Recherche Operationnelle*, 8, 57–75.

Szmidt, E., & Kacprzyk, J. (2000). Distances between intuitionistic fuzzy sets. *Fuzzy Sets and Systems*, 114(3), 505-518.

Vahdani, B., Mousavi, S.M., Moghaddam, R. T., & Hashemi, H. (2013). A new design of the elimination and choice translating reality method for multi-criteria group decision-making in an intuitionistic fuzzy environment. *Applied Mathematical Modelling*, 37(4), 1781-1799.

Vlachos, I.K., & Sergiadis, G.D. (2007). Intuitionistic fuzzy information – Applications to pattern recognition. *Pattern Recognition Letters*, 28(2), 197–206.

Wu, M.C., & Chen, T.Y. (2011). The ELECTRE multicriteria analysis approach based on Atanassov's intuitionistic fuzzy sets. *Expert Systems with Applications*, 38(10), 12318–12327.

Xu, Z.S. (2007). Intuitionistic fuzzy aggregation operators. *IEEE Transactions on Fuzzy Systems*, 15(6), 1179-1187.

Yue, Z. (2014). TOPSIS-based group decision-making methodology in intuitionistic fuzzy setting. *Information Sciences*, 277, 141–153.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.

Zandi, A., & Roghanian, E. (2013). Extension of fuzzy ELECTRE based on VIKOR method. *Computers & Industrial Engineering*, 66(2), 258–263.

Zhang, H.Y., Peng, H.G., Wang, J., & Wang, J.Q. (2017). An extended outranking approach for multi-criteria decision-making problems with linguistic intuitionistic fuzzy numbers. *Applied Soft Computing*, 59, 462–474.

BALCOR 2018
XIII Balkan Conference on Operational Research

ELECTRE III APPROACH AS A TOOL FOR MIGRATION PROBLEMATIC

MLADEN STAMENKOVIĆ[1], MIHAIL ARANDARENKO[2], SALVATORE CORRENTE[3], MAJA JANDRIĆ[4]

[1] University of Belgrade, Faculty of Economics, mladen@ekof.bg.ac.rs

[2] University of Belgrade, Faculty of Economics, arandarenko@ekof.bg.ac.rs

[3] University of Catania, Department of Economics and Business, salvatore.corrente@unict.it

[4] University of Belgrade, Faculty of Economics, maja@ekof.bg.ac.rs

Abstract: *Internal migration is one of the most fundamental responses to different development dynamics of regions. In the most advanced societies, inter-regional migration is an important mechanism through which labor resources are redistributed geographically in response to changing economic and demographic forces. In this paper, we present the possibility for the use of multiple criteria decision aiding (MCDA) techniques, namely, ELECTRE III, as a tool for the problems related to the internal migration issues. To show the effectiveness of MCDA methods in this type of problems we use Serbian districts data, define criteria based on academic literature regarding determinants of internal migrations and develop a ranking of districts according to their migration potential. In that way, we show regions with strong emigration and immigration potential in the next period. Most importantly, ranking procedure will yield districts with the worst potential and show where should be policymakers' focus.*

Keywords: *multiple criteria decision aiding, ELECTRE, internal migrations, SRF method*

1. INTRODUCTION

Economic globalization has contributed to manifold transformations across the globe over the past decades. It has changed the patterns of socio-economic development in many cities and localities and turned these into places of origin or destination of ever more intensive flows of both internal and international migration. This outcome is compounded by the opposite effects of technological development in terms of the increase of distant work, and the increase of labor mobility, including mass migration and brain drain. Lastly, this phenomenon gives rise to the paradox of regionalization within globalization and fourth industrial revolution, characterized by the creation of economically integrated regions.

Regional centers are nowadays seen as the destinations of migration, where job opportunities and social networks tend to concentrate (Fauser, 2015). He states that the same global transformations that have attracted migrants to cities have simultaneously displaced many people from poorer regions. The Rural-urban discourse of migration has been studied extensively among economists. A vast literature in economic development sees internal migration as the central feature of future growth (Chernina et al., 2014). In most advanced societies, interregional migration is a major mechanism through which labor resources are redistributed geographically in response to changing economic and demographic forces (Greenwood, 1997). Chen and Rosenthal (2008) argue that the willingness of households to migrate is a primary driver of regional shifts in the supply of labor and the local level of human capital. Milanovic (2015) claims "migration is one of the most efficient ways of resolving global inequality". It is important to note that only around three per cent of the world's population – about 210 million people – are international migrants (Geddes and Korneev, 2015) and that most migratory flow happens within the country (Bell et al., 2015). Therefore, the effects of internal migration come as a result of labor market polarization and different development dynamics of regions as one of the most fundamental responses to the effects of technological transformations on employment.

Specific reaction of the different local labor markets to technological change has been widely documented. Local labor markets that have more jobs specialized in routine tasks have differentially adopted information technology, reallocated low-skill labor into service occupations, experienced earnings growth at the tails of the distribution, and received inflows of skilled labor (David and Dorn, 2013). Beaudry et al. (2010) show that localities with different levels of educational attainment adjust differently to a technology revolution, with more educated metropolitan areas having adopted new technologies faster, and experiencing a greater increase in the return to skills.

One of the most important issues that largely influences different positions and perspectives of the regional labor markets is technology's potential to substitute work. Academic literature shows that recent technological change has been skill-, routine-, and capital-biased. Digitalization tends to substitute for workers engaged in routine tasks, which are carried out by following well-defined procedures. On the other side, tasks that require intuition, creativity, complex social interaction and higher levels of perception and manipulation are still difficult to automate. In many advanced economies, significant expansion of employment at both ends of the skill spectrum were registered, at the expense of employment in middle-skill occupations. A surge in low-skill service jobs can be explained by the fact that higher incomes increase the demand for some of the services requiring low-skilled workers, and the manual non-routine tasks that are prevalent in the service occupations are not easily substitutable by computers (David and Dorn, 2013). However, rapid technological advances (artificial intelligence, use of big data, sophisticated algorithms, robotics, etc.) will probably soon make possible to automate an even wider set of manual tasks, which will make low-skilled workers even more vulnerable. The impact of contemporary technological changes on the labor market of a particular region depends on a large number of factors, while among the most important ones are economic structure and qualifications and occupational structure of the labor force. The unfavorable structure can lead to rising regional inequalities, which can reinforce internal migration, and in that way, make these disparities even larger.

The aim of this paper is to investigate the level of internal migration potential in Serbia and determine zones with a potential for large emigration and immigration in the next period. The focus of this research will be Serbian districts (NUTS 3 regional level). Migration potential will be obtained with the help of the ELECTRE III method as we want to define ranking of districts according to the selection of criteria that will represent determinants of migration potential. Earlier efforts to assess potentials and perspectives of regional labor markets in Serbia have been conducted by Arandarenko (2006) and Arandarenko and Jovicic (2007).

The paper reads as follows. Next section will explain methodological concepts of ELECTRE methods, whereas in third section we define main drivers of migration potential and define criteria. Results are presented in the fourth section, while final remarks are provided in the last section.

2. METHODOLOGY

Migration perspectives of the Serbian districts will be observed using ELECTRE III methodology as a convenient setting for the creation of their ranking according to migration potential. We will briefly describe general concepts of ELECTRE methods, starting with notation and afterward develop outranking relation of the method. The problem of the weights' choice for the model is explained in Section 2.2.

2.1. Notation and basic ELECTRE concepts

Consider decision aiding context in which we are facing with following (Greco et al., 2016):

(i) $A = \{a_1, a_2, ..., a_m, ...\}$ – set of potential alternatives defined in the problem,
(ii) $G = \{g_1, g_2, ..., g_n\}$ – coherent family of n pseudo-criteria,
(iii) $w_1, w_2, ..., w_n$ – corresponding weights for each criterion,
(iv) $g_k(a_i)$ – performance of alternative a_i relative to criterion g_k,

Without loss of generality, we can assume that decision maker wants to maximize performance on each criterion. In order to compare two alternatives, ELECTRE methodology introduces *indifference* and *preference* thresholds q_k and p_k for each criterion such that $p_k \geq q_k \geq 0$. These thresholds are introduced to take into account the imperfect character of the data from the computation of the alternative performances $g_k(a_i)$ as well as the arbitrariness that affects the definition of the criteria (Almeida-Dias et al., 2010; Roy et al., 2014). Based on the definition of such thresholds, the ELECTRE methods are handling the following preference situations concerning the comparison of two alternatives (Almeida-Dias et al., 2010; Figueira et al., 2013):

(i) *Indifference* corresponds to a situation where there are clear and positive reasons that justify an equivalence between the two alternatives (it leads to a reflexive and symmetric but not necessarily transitive binary relation). We will claim that there are no significant differences between two alternatives related to criteria k if $|g_k(a) - g_k(a')| \leq q_k$. In such a case we will see these alternatives as indifferent and denote it as aI_ka'. The subset of all criteria for which aI_ka' will be denoted by $C(aIa')$.
(ii) *Strict preference* corresponds to a situation where there are clear and positive reasons in favor of one (identified) of the two actions (it leads to a nonreflexive and asymmetric and usually transitive binary relation). In a modeling sense, we will claim that alternative a is strictly preferred to alternative a' on criterion k if $g_k(a) - g_k(a') > p_k$ and we will denote it as aP_ka' while $C(aPa')$ will be the subset of all criteria for which aP_ka'.

(iii) *Weak preference* corresponds to a situation where there are clear and positive reasons that invalidate strict preference in favor of one (identified) of the two alternatives, but they are insufficient to deduce either the strict preference in favor of the other alternative or indifference between both actions, thereby not allowing either of the two preceding situations to be distinguished as appropriate (it leads to a nonreflexive and asymmetric but not usually transitive binary relation). Formally, this is considered to be true when $q_k < g_k(a) - g_k(a') \leq p_k$ and this difference represents an ambiguity zone. The advantage of a over a' is too large to conclude about an indifference between a and a', but it is not enough to conclude about a strict preference in favor of alternative a. This means that there is a hesitation between indifference and strict preference. We will denote it as aQ_ka' while $C(aQa')$ will be the subset of all criteria for which aQ_ka'.

The aforementioned binary relations can be grouped into one partial outranking relation S_k comprising the three corresponding situations, $S_k = P_k \cup Q_k \cup I_k$ where aS_ka' means that alternative a is *at least as good* as alternative a' on criterion k. What we want to measure is the level of the statement aSa' for the whole set of criteria. Clearly, coalition of criteria for that statement is the union of the subsets already defined, that is, $C(aSa') = C(aIa') \cup C(aQa') \cup C(aPa')$. However, when we want to define an overall measure in order to determine if relation aSa' is valid, we have to take into account even situations where $a'Q_ka$. *The Concordance index* measures the strength of coalition in favor of the assertion aSa'. Formally,

$$C(a,a') = \sum_{\{k:g_k \in C(aSa')\}} w_k + \sum_{\{k:g_k \in C(a'Qa)\}} \varphi_k w_k. \tag{1}$$

Function φ_k measures the potential of statement that alternative a is indifferent to alternative a'. This function, therefore, should converge to one, as values of a and a' approach to each other, while it should go to zero as the difference $g_k(a') - g_k(a)$ approaches p_k. Formally:

$$\varphi_k = \frac{p_k - [g_k(a') - g_k(a)]}{p_k - q_k}. \tag{2}$$

The important feature of the ELECTRE methods is the principle of rejection of hypothesis if a large resistance exists, no matter how large the Concordance index is (Stamenković et al., 2016). When there is such a criterion g_k which strongly opposes to the fact that a is at least as good as a', g_k puts veto to this assertion. This principle is incorporated in the *Discordance index*. It represents the amount of discordance of criterion k on the fact that a is at least as good as a'. Of course it should not be neglected even if there is enough evidence in favor of aSa'. Discordance index is defined as

$$d_k(a,a') = \begin{cases} 1, & \text{if } g_k(a') - g_k(a) > v_k \\ \dfrac{[g_k(a') - g_k(a)] - p_k}{v_k - p_k}, & \text{if } p_k < g_k(a') - g_k(a) \leq v_k \\ 0, & \text{if } g_k(a') - g_k(a) \leq p_k \end{cases}.$$

Finally, taking into account both concordance and discordance index we have to incorporate these measures into a final value that will denote final recommendation regarding the relation between each pair of alternatives. We want to derive a measure upon which we can decide whether *a outranks a'* i.e. aSa'. Fuzzy measure obtained through multiple criteria aggregation procedure named *credibility index* takes into account all the concordance and discordance values for each criterion and builds a final measure

$$\sigma(a,a') = C(a,a') \prod_{\{k:d_k(a,a') > C(a,a')\}} \frac{1 - d_k(a,a')}{1 - C(a,a')}. \tag{3}$$

Crisp relation can be easily created using the credibility index values. Let λ denote the threshold credibility level, the minimum degree of credibility, which is considered or judged necessary by the decision maker to validate or not the statement a outranks a'. In other words, we will consider that alternative a outranks alternative a' for credibility level λ if $\sigma(a,a') \geq \lambda$. Also, for the same level λ we can say that a is preferred to a' if $\sigma(a,a') \geq \lambda$ and $\sigma(a',a) < \lambda$ and also we can impose *incomparability* among alternatives if neither $\sigma(a,a') \geq \lambda$ nor $\sigma(a',a) \geq \lambda$ holds. ELECTRE III builds outranking relation upon such crisp relation and creates final ranking based on upward and downward distillation (Roy, 1978).

2.2. The SRF method for weights elicitation

The problem in MCDA methods application related to real-life problems is often elicitation of weights for the defined set of criteria. One of the potential ways for inferring weights is a revised Simos procedure (Simos, 1990) defined in Figueira and Roy (2002) and called SRF. The idea of the procedure is the following. The DM is asked to rank the cards representing criteria from the least important to the most important. Criteria that are considered as equally important are getting the same rank. Moreover, if the DM wishes, he can put one or more blank cards between two successive subsets of criteria. The greater the number of blank cards separating two sets of indifferent criteria, the greater is the difference of importance between these sets of criteria (Corrente et al., 2016). The main distinction from the Simos method is that now DM is asked to state how many times the best criterion is more important than the worst one in the ranking. This value will be denoted by z. Following Corrente et al. (2016), let us denote with $I = \{1,...,m\}$ the set of considered criteria and let L_1 be the set of least important criteria, while L_v is the set of most important criteria, $L_1, L_2, ..., L_v \subseteq I$, $L_i \cap L_j = \emptyset$ for all $i \neq j$, $i, j = 1,..,v$. Assume that the number of blank cards between sets L_k and L_{k+1} is e_k, $k = 1,...,v-1$. Using such information obtained from the DM, a non-normalized weight for each criterion j is obtained as

$$w'_j = 1 + \frac{(z-1)\left[l(j) - 1 + \sum\limits_{s=1}^{l(j)-1} e_s\right]}{v - 1 + \sum\limits_{s=1}^{v-1} e_s}, \tag{4}$$

where $l(j)$ represents the rank of importance to which criterion j belongs. The obtained weights are therefore normalized so that they sum up to 1. For new advances on SRF method see Corrente et al. (2017).

3. CRITERIA FOR SELECTION – DETERMINANTS OF MIGRATION

Academic literature devoted to drivers of migrations is extensive and works in a manifold, sometimes opposing, which is another proof related to the validity of multiple criteria decision aiding approach as we intend to do. MCDA offers a solution to the problem facing opposing criteria that is the case when it comes to internal migrations problem. Theoretical models stem from gravity models based on population size and distance to extended models that include economic and labor market characteristics, and further to the more sophisticated models that encompass individual characteristics that determine propensity to migrate, like age and educational level. Empirical studies are in case of our research even more important as econometric studies defined drivers of migration based on all possible scenarios affecting internal migrations and we can find effects of technological development, economic factors, urban-rural divide or the effects of amenities and housing market.

Beside the, already mentioned, rural-urban aspect of internal migration, an interesting interregional flow of migration concerns interurban internal migration. This aspect is predominantly important in advanced economies, where the share of urban population is already high and there is little scope for further urbanization. Evidence points that there are large disparities in population growth of cities in industrialized countries: while some cities suffer from ongoing population decline, others have experienced increasing numbers of inhabitants (Buch et al., 2014). Growth or decline of cities' populations is mainly driven by migration flows (Buch et al., 2014; Chen and Rosenthal, 2008). The attractiveness of cities and regions from the migration point of view is determined by economic and non-economic factors. Economic factors encompass employment opportunities, regional wage differentials, housing market characteristics, other forms of expected income, etc., while non-economic factors refer mostly to local-specific amenities like climate, natural attractiveness, theaters, universities, accessibility, etc. The same authors find that small cities in Germany are, ceteris paribus, marked by less net in-migration than large cities, which points to specific benefits of living in large cities. It is important that local governance and urban planners distinguish main factors that drive internal migrations, since the ability to attract residents plays a fundamental role for cities' and regional prospects (Rodríguez-Pose and Ketterer, 2012).

As expected, high earnings and vast employment opportunities encourage people to move whereas high prices of houses discourage individuals to move. Academic literature confirms that specific housing conditions (regarding house prices, rents, and home-ownership rates) are known to affect labor market rigidities (Lux and Sunega, 2012). Previous research in the field led to the following selection of criteria that will be used in our study. Such selection will include the importance of all selected determinants of migration such as economic and labor indicators, housing market and amenities or demographic factors that might have a prevalent role.

All the criteria are presented in Table 1. We can see that the economic aspects, maybe the most important ones, are observed through indicators such as average wage, employment, but also with a gross regional product that will count for overall economic activity in the district. Housing is the important part of all the studies regarding internal migration and, therefore, it is observed in our analysis through the level of construction in the district (constructed dwellings per 1000 inhabitants), but in parallel, we will observe the possibility of citizens to acquire these dwellings. This is done by calculating the number of square meters that can be bought by average age in each district. Quality of life within each district is observed by the quality of medical service and, as well, by observing the life expectancy at birth. Urban agglomeration measure is the percentage share of urban population in the district divided by the total urban population in all districts as our proxy to the urban-rural aspect we discussed earlier. Also, index of the modern road surface is measured as the length of modern road surface divided by the total area of the district. The second column in the table describes preference direction for each criterion.

Table 1: Criteria based on determinants of migration

Criterion	Direction
Gross regional product per capita (GRPpc)	[max]
Employment rate (ER)	[max]
Unemployment rate (UR)	[min]
Average wage per employee (AWpE)	[max]
Share of young population (YP)	[min]
Average age (AA)	[max]
Life expectancy at birth (LEaB)	[max]
Urban agglomeration measure (UAM)	[max]
Square meters that can be bought by average wage (SfAW)	[max]
Constructed dwellings per 1000 inhabitants (CD)	[max]
Number of medical doctors per 1000 inhabitants (MD)	[max]
Share of children in pre-primary education, aged 0-3 (CPE)	[max]
Index of modern road surface (IMRS)	[max]

4. RESULTS

Based on the criteria defined in Table 1 we evaluate performance for each district in the Republic of Serbia. All indicators are presented in Table 2. Threshold values are set taking into the account overall situation in Serbian economy and industry. When it comes to weight elicitation we applied SRF method. Average wage per employee is selected as the most important criterion and we set one blank card between average wage and employment rate as the second most important one. This is largely in line with theoretical assumptions - a person, when deciding whether to move (within a country, as we only consider internal migrations) will primarily focus on employment possibilities (which is shown by employment rate), but prospective earnings in the destination region will also have a significant impact on his/her decision. After another blank card we set GRPpc, UAM, and SfAW, and after that follows UR and CD. To conclude, the last level of importance is reserved for the rest of defined criteria. The last part of the selection is in line with the description done in Section 3.

Results of the SRF method and thresholds values are presented in Table 3.

Using all the defined values we can now apply ELECTRE III method and use both distillations to get the final ranking of the Serbian districts according to internal migration potential. The calculation was conducted using open source software for MCDA methods, Diviz. Rankings are defined as presented in Table 4 as well as in Figure 1.

5. DISCUSSION AND CONCLUSION

Large regional disparities in Serbia lead, among other things, to significant internal migration flows. Internal migrations, on the other hand, reinforce existing regional disparities by causing the lack of human capital in the underdeveloped regions, which makes this issue extremely important to policymakers and their efforts to achieve a more balanced regional growth and development. Although net migration rates show previous trends in internal migration flows, it is important to predict the direction of these flows in the near future. The main

Table 2: Table of indicators

	GRPpc	ER	UR	AWpe	YP	AA	LEaB	UAM	SfAW	CD	MD	CPE	IMRS
Belgrade	7844.19	42.70	18.80	55551	16.92	42.10	76.26	31.48	0.41	2.40	3.60	29.42	96.62
Bor	3970.48	37.80	16.90	45581	16.21	45.25	74.88	1.66	0.98	1.20	3.30	18.14	28.36
Braničevo	4031.51	46.40	11.00	43761	17.02	44.69	73.98	1.66	0.61	0.70	2.50	18.00	49.18
Central Banat	4095.35	42.80	14.10	39054	17.39	42.99	73.44	2.22	0.70	0.60	2.20	10.66	26.02
Jablanica	2427.07	45.00	18.90	33502	17.97	43.05	74.82	2.19	0.53	0.60	2.70	9.91	36.44
Kolubara	3805.23	51.50	13.20	39404	16.81	44.14	75.04	1.72	0.54	1.00	2.40	16.87	53.30
Mačva	3042.39	46.30	15.70	36233	17.08	42.93	74.21	2.04	0.55	1.20	2.20	13.01	46.12
Moravica	4097.67	47.70	13.60	38103	16.47	44.09	75.67	2.68	0.54	1.60	2.30	25.70	43.76
Niš	3466.09	38.00	24.70	37993	17.12	43.60	75.80	4.99	0.44	1.80	4.00	16.75	43.36
North Bačka	4241.66	45.60	10.70	39216	17.48	42.66	73.72	2.82	0.77	0.50	2.20	20.41	21.51
North Banat	3564.43	40.50	14.90	38630	17.31	43.29	72.54	2.18	1.23	0.30	2.40	12.73	29.57
Pčinja	2169.64	36.30	15.80	33054	21.71	39.15	74.25	2.14	0.50	1.20	2.70	11.22	29.91
Pirot	4573.39	39.40	20.90	39548	15.91	46.06	76.24	1.35	0.70	0.90	2.90	13.70	24.47
Podunavlje	2088.24	40.70	18.90	39183	17.94	42.72	74.57	2.43	0.60	0.90	2.30	16.06	65.78
Pomoravlje	3264.48	38.30	19.00	34767	16.93	44.53	74.76	2.28	0.52	1.70	3.00	16.43	44.11
Rasina	2756.29	46.70	15.20	35224	16.44	44.47	75.87	2.10	0.51	1.00	2.20	18.18	45.27
Raška	2490.64	40.60	21.60	35103	19.44	39.22	75.46	3.85	0.44	2.10	2.50	15.97	36.12
South Bačka	6111.22	44.50	15.90	47445	18.10	41.04	74.99	10.15	0.47	1.40	3.10	37.63	29.93
South Banat	3888.30	38.40	20.90	45928	17.27	42.61	74.02	3.94	0.67	0.70	2.50	14.03	19.21
Srem	4083.75	40.60	18.30	39195	17.40	42.78	74.66	3.18	0.63	1.10	1.90	20.65	25.43
Šumadija	4450.43	41.00	20.00	39026	17.10	43.16	75.60	4.46	0.43	1.20	3.20	20.48	53.30
Toplica	2688.59	43.50	17.50	33569	17.86	43.64	73.97	1.07	0.83	0.70	2.70	10.96	28.47
West Bačka	3559.20	38.20	20.30	37593	16.78	43.99	73.85	2.36	0.80	0.40	2.20	14.37	25.01
Zaječar	3264.52	36.30	15.50	36716	14.68	47.12	74.31	1.62	0.65	0.80	3.30	22.14	31.18
Zlatibor	3660.73	45.60	15.00	37875	17.31	43.31	76.34	3.43	0.46	1.70	2.50	19.91	41.94

Table 3: Thresholds and weights for each criterion.

	GRPpc	ER	UR	AWpe	YP	AA	LEaB	UAM	SfAW	CD	MD	CPE	IMRS
q	480	2	2	5000	1	2	1	2	0,2	0,4	0,5	4	10
p	1700	5	5	9000	3	4	2,5	4	0,3	1	1	10	15
v	4800	10	10	15000	5	none	none	none	none	none	none	none	none
w	1.67	2.33	1.33	3	1	1	1	1.67	1.67	1.33	1	1	1

Table 4: Final ranks of Serbian regions.

Region	Rank	Region	Rank
Belgrade	1	Podunavlje	11
Bor	5	Pomoravlje	10
Braničevo	3	Rasina	7
Central Banat	8	Raška	13
Jablanica	12	South Bačka	2
Kolubara	2	South Banat	11
Mačva	7	Srem	11
Moravica	4	Šumadija	6
Niš	8	Toplica	11
North Bačka	5	West Bačka	13
North Banat	10	Zaječar	9
Pčinja	14	Zlatibor	5
Pirot	10		

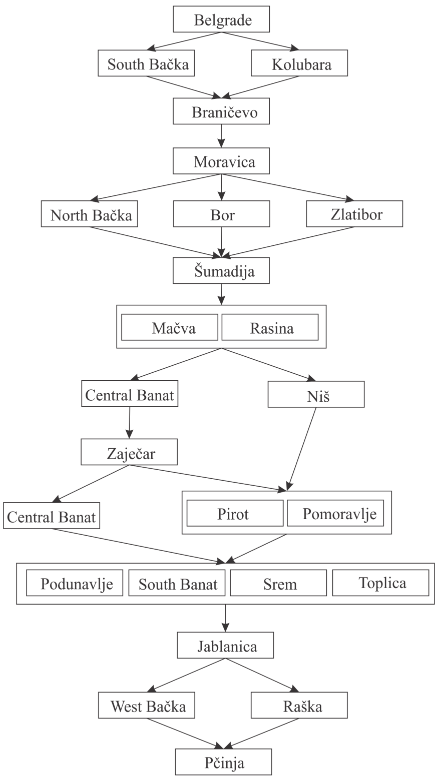


Figure 1 Ranking of Serbian districts.

aim of this paper was to present the potential of MCDA methods and show their effectiveness in this line of research. Using ranking technique, ELECTRE III, we measured the migration potential of each district on the basis of the main internal migrations determinants.

The results clearly indicate that the Belgrade will continue to attract most of the citizens, followed by Novi Sad region (South Bačka district). According to the obtained rankings, Pčinja district has the worst position (14th), while Raška and West Bačka district are in the 13th place. Some of the districts have unexpectedly good rankings (like Kolubara district). This is partly the result of the data on the employment rate, which is calculated according to Labour Force Survey (LFS) and, according to LFS methodology, it encompasses total employment in line with ILO (*International labor organization*) definition. Apart from the persons who have contracted employment and work with enterprises, institutions or other organizations or are active within private unincorporated enterprises, according to this definition, employed persons include several more categories. Included in the statistics are also individual farmers, unpaid family workers/supporting household members, as well as the persons who found and made an agreement (verbally or in written) on casual job conduct without contracting employment, and to whom the subject jobs were the only source of subsistence. Therefore, these data do not relate only to the formal employment status, while it is reasonable to assume that only higher quality jobs have stronger attractive power for potential internal migrants. One of the possible avenues to get more accurate rankings is to include some quality of employment aspects in the analysis. Maybe the most important findings are not the districts with strong immigration potential but the awareness which districts have the largest potential for emigration. These are the regions where there is the biggest expectation of the net outflow of citizens and policy measures need to be directed to those regions as the most critical ones.

Future research within this direction will opt to include other MCDA methods to defined migration potential of districts or even municipalities. Sorting approach might be a good direction for future research as the categories might be defined in line with migration potential showing either immigration or emigration status of districts. Also, additional work on criteria is also needed, and modern approach such as multiple criteria hierarchy process (Corrente et al., 2012) might be the right direction.

Acknowledgement

This paper is a part of the project entitled *Persistent unemployment on local labour markets and local development*, Dnr F16-1341:1, with funding from The Swedish Foundation for Humanities and Social Sciences. Authors from the University of Belgrade acknowledge financial support by the Ministry of Education and

Science of the Republic of Serbia, Project numbers: 179005 and 179065. Salvatore Corrente acknowledges funding by the “FIR of the University of Catania BCAEA3, New developments in Multiple Criteria Decision Aiding (MCDA) and their application to territorial competitiveness".

REFERENCES

Almeida-Dias, J., Figueira, J. R., and Roy, B. (2010). Electre Tri-C: A multiple criteria sorting method based on characteristic reference actions. *European Journal of Operational Research*, 204(3):565–580.

Arandarenko, M. (2006). Mapa tržišta rada srbije. *Beograd, Serbia: CEVES*.

Arandarenko, M. and Jovicic, M. (2007). Regional labour market differences in Serbia: assessment and policy recommendations. *The European Journal of Comparative Economics*, 4(2):299.

Beaudry, P., Doms, M., and Lewis, E. (2010). Should the personal computer be considered a technological revolution? Evidence from US metropolitan areas. *Journal of Political Economy*, 118(5):988–1036.

Bell, M., Charles-Edwards, E., Kupiszewska, D., Kupiszewski, M., Stillwell, J., and Zhu, Y. (2015). Internal migration data around the world: Assessing contemporary practice. *Population, Space and Place*, 21(1):1–17.

Buch, T., Hamann, S., Niebuhr, A., and Rossen, A. (2014). What makes cities attractive? The determinants of urban labour migration in Germany. *Urban Studies*, 51(9):1960–1978.

Chen, Y. and Rosenthal, S. S. (2008). Local amenities and life-cycle migration: Do people move for jobs or fun? *Journal of Urban Economics*, 64(3):519–537.

Chernina, E., Dower, P. C., and Markevich, A. (2014). Property rights, land liquidity, and internal migration. *Journal of Development Economics*, 110:191–215.

Corrente, S., Figueira, J., Greco, S., and Słowiński, R. (2017). A robust ranking method exending ELECTRE III to hierarchy of interacting criteria, imprecise weights and stochastic analysis. *Omega*, 73:1–17.

Corrente, S., Greco, S., and Słowiński, R. (2012). Multiple Criteria Hierarchy Process in Robust Ordinal Regression. *Decision Support Systems*, 53(3):660–674.

Corrente, S., Greco, S., and Słowiński, R. (2016). Multiple criteria hierarchy process for ELECTRE Tri methods. *European Journal of Operational Research*, 252(1):191–203.

David, H. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review*, 103(5):1553–97.

Fauser, M. (2015). Migration, transnationalization and urban transformations. In *Handbook of International Political Economy of Migration*.

Figueira, J. and Roy, B. (2002). Determining the weights of criteria in the ELECTRE type methods with a revised simos' procedure. *European Journal of Operational Research*, 139(2):317–326.

Figueira, J. R., Greco, S., Roy, B., and Słowiński, R. (2013). An overview of ELECTRE methods and their recent extensions. *Journal of Multi-Criteria Decision Analysis*, 20(1-2):61–85.

Geddes, A. and Korneev, O. (2015). The state and the regulation of migration. *Handbook of the international political economy of migration*.

Greco, A., Figueira, J., and Ehrgott, M. (2016). *Multiple Criteria Decision Analysis: State of the Art Surveys*. Springer, Berlin.

Greenwood, M. J. (1997). Internal migration in developed countries. *Handbook of population and family economics*, 1:647–720.

Lux, M. and Sunega, P. (2012). Labour mobility and housing: the impact of housing tenure and housing affordability on labour migration in the Czech Republic. *Urban Studies*, 49(3):489–504.

Milanovic, B. (2015). Global inequality of opportunity: How much of our income is determined by where we live? *Review of Economics and Statistics*, 97(2):452–460.

Rodríguez-Pose, A. and Ketterer, T. D. (2012). Do local amenities affect the appeal of regions in Europe for migrants? *Journal of Regional Science*, 52(4):535–561.

Roy, B. (1978). ELECTRE III: un algorithme de classement fondé sur une représentation floue des préférences en présence de critères multiples. *Cahiers du CERO*, 20:3–24.

Roy, B., Figueira, J., and Almeida-Dias, J. (2014). Discriminating thresholds as a tool to cope with imperfect knowledge in multiple criteria decision aiding: Theoretical results and practical issues. *Omega*, 43:9–20.

Simos, J. (1990). Evaluer l'impact sur l'environnement: Une approche originale par l'analyse multicritère et la négociation. In *Evaluer l'impact sur l'environnement: une approche originale par l'analyse multicritère et la négociation*. Presses polytechniques et universitaires romandes.

Stamenković, M., Anić, I., Petrović, M., and Bojković, N. (2016). An ELECTRE approach for evaluating secondary education profiles: evidence from PISA survey in Serbia. *Annals of Operations Research*, 245(1-2):337–358.

B15
O.R. in Agriculture and Forest Management

EFFECTIVE PRICING STRATEGIES FOR ORGANIC AGRICULTURE PRODUCTS UNDER THE COMMUNITY SUPPORTED AGRICULTURE (CSA) MODEL

NUR AYVAZ-CAVDAROGLU[1]

[1] Kadir Has University, Faculty of Business Administration, nur.cavdaroglu@khas.edu.tr

Abstract: *Although organic agriculture is gaining popularity around the world, organic producers need to be supported with better financial systems for the sustainability of this practice. This paper suggests that Community Supported Agriculture (CSA) models can be among the solution alternatives if applied properly. We develop and analyze a pricing model which aims to identify the conditions under which CSA can be feasible for small farmers in developing countries. We analyze the structure of the optimal solution and how it varies with problem parameters towards this end. Finally, we illustrate the proposed model on a realistic numerical case.*

Keywords: *Pricing, Organic Agriculture, Community Supported Agriculture, Mathematical Modeling.*

1. INTRODUCTION

Organic farming is gaining popularity across the world from year to year. This is mainly due to the increasing awareness of people towards the sustainability of resources and the increasing tendency of consumers towards healthier diets. However, performing organic cultivation is costly for producers in developing countries, mainly because of the lower yields, the pricey certification process, and the small size of the local organic markets. These producers have to sell the majority of their crop at relatively low prices to intermediary export firms, therefore not capturing a generous share of the profit in the organic food market.

This situation is frequently observed in Turkey, which is an important exporter for the imbalanced European organic market. In the absence of a strong domestic market, most of the Turkish organic producers work with intermediary firms on a contractual basis towards satisfying the demand of the Western European organic market (Demiryurek, 2011). This model, unfortunately, leads to several problems, especially regarding the producer welfare. When small farmers cannot perform distribution activities and instead rely on intermediaries to consolidate, process and distribute their products, the farmer only gets around 20% of the prices paid by the final consumer (Villalobos, et al., 2014). A potential problem in this situation would be farmers ceasing organic production.

One possible solution for this problem could be farmers and local consumers engaging in Community Supported Agriculture (CSA) practices. In this marketing model, consumers "buy" shares of the farm before the cultivation season and receive a portion of the produce throughout the year (Brown and Miller, 2008). This portion is usually sufficient to feed a family of four. This model is a win-win situation for both the farmers and the consumers. The consumers enjoy a healthier diet at much lower costs than the prices offered at the domestic organic market, while the producers can receive an all-year-long cash flow and a higher portion of the profits in the organic food industry.

Although this model is prevalent in countries like the U.S., it is still not common in Turkey. The organic producers selling to individual consumers usually do it by using a "pay as much as you buy" system, rather than engaging to a membership system with a fixed fee as in CSA. Although the former brings the advantage of dynamic pricing according to the changing supply and demand, it is clearly more difficult to manage and is susceptible to interruptions on the cash flows during the cultivation period. However, the CSA system must be handled carefully, too. The membership price must be set to cover operating costs and yield a fair return to the farmer's labor, bringing financial benefits for farmers. However, several case studies show otherwise, leading to the conclusion that the farmers are bad at determining the optimal pricing schedule for their produces (Lass et al. 2003, Brown and Miller, 2008).

In this paper, we aim to bring an operations management perspective to the pricing problem of the organic producers. By developing a mathematical model and analyzing the structure of the optimal solution, we hope to create insights towards the conditions for the feasibility of CSA systems in organic agriculture. The rest of the paper is organized as follows: In Section 2, we make a brief literature review of the CSA system and the OR models in agriculture. In Section 3, we discuss the details of our model and the solution.

In Section 4, we illustrate the solution on a realistic numerical problem. Finally, we summarize our findings by Section 5, Conclusion.

2. LITERATURE REVIEW

Although the production and supply chain management problem of agricultural producers has received much attention from the researchers in operations research field, the same thing cannot be said for the pricing problem. One reason for this could be the fact that the prices are usually imposed by outer parties or determined by the dynamics of the market in the agriculture sector. However, this is not the case in CSA model. Some innovative pricing or contracting models in agriculture are going to be mentioned below.

The work of Hovelaque et al. (2009) is one of the few papers that involve a mathematical model to investigate the effects of adding price contracts to the co-operative contract of the farmers. They focus on comparing the co-operatives with other buyer firms in the industry in the presence or absence of pricing contracts and illustrate their findings on a dairy farm case study by using Monte-Carlo simulation. Jang and Klein (2011) present mathematical models to answer some strategic questions of small farmers such as when to form a co-operative, the optimal size of the co-operative, what type of product to produce and how much, etc., and they find that small farmers must utilize a co-operative to increase their profits if local markets are relatively small or predictable. Agbo et al. (2013) build a theoretical model to study a market structure of a marketing cooperative with direct selling, in which many farmers are members of an agricultural marketing cooperative. The farmers can sell their products either to the cooperative or on a local market. However, the authors regard the local market as an oligopolistic setting, and they do not involve any pricing scheme which might induce higher profits to farmers when they sell to the co-op. Villalobos et al. (2014) depict a complete picture of the supply chain of fresh produce and state the reasons for the prices to be highly variable in this industry. In this environment, they state the recipe for small farmers' survival as "integration", i.e. associating with other farmers to be able to efficiently reach higher echelons in the value chain of fresh produce.

Another line of work involving mathematical models on agricultural supply chain management can be stated as follows. Kazaz (2004) investigates production planning decisions of an olive farmer under yield and demand uncertainty. He finds the optimal amount of farm space to be leased and the amount olive oil to be produced to maximize the profits of the grower, and illustrates the theoretical results on an empirical problem. Ahumada et al. (2012) present a stochastic tactical planning model for the production and distribution of fresh agricultural products using a two-stage stochastic program. Tan and Comden (2012) present a planning methodology for a firm who aims to match the random supply of annual fruits and vegetables from a number of contracted farms and the random demand from the retailers. Soto-Silva et al. (2016) present a review of the operational research models applied to the fresh food supply chains. To our knowledge, the majority of operational research papers in the field consider the planning problem of growers, and none of the annotated research involves pricing procedures and innovative practices such as CSA in particular.

Coming to the research on CSA practices, it is seen that the majority of this work involves case studies and survey-based qualitative research. Adam (2006) reports on the history of CSA in the U.S. and discusses various models that have emerged. There are basically two forms of CSA: "subscription CSA" which is farmer-driven where the farmer initiates the CSA practice, and "shareholder CSA", which is consumer-driven where a group of consumers come together, hire a farmer, and initiate the CSA practice. CSA practices have gained speed after computerization in the early 2000s. Cone and Mhyre (2000) find that motivations for membership in CSA organizations relate to the quality of food and environment, whereas the factors of "interest in community" and "price" come later. The authors also suggest first, that community-supported agriculture can offer solutions to the problems of modernity, and second, that those members who participated more extensively in their farms experienced greater rewards. Brown and Miller (2008) provide a review of the research regarding the impacts of CSA on farmers and consumers, and point out findings of several studies that indicate positive impacts on the local community. However, they also remark the significance of improving the financial situation of the farmers by better pricing policies to ensure sustainability of the CSA system.

3. THE MODEL

In this section, we will develop and analyze the pricing model of farmers in two settings. In the first, classical setting, the farmer can sell to the intermediary and in the local farmer's market. In the second setting, the farmer engages in a CSA model and sells to its member customers rather than in the farmer's market. Although in practice a farmer can produce and sell several different products, we base our analysis on a "generic" produce with the given demand function and an uncertain yield. Moreover, we consider the selling season as a single time period.

First, consider the classical setting for organic producers. In the absence of a CSA, the farmer has two sales channels: She can either sell to individual customers in the local market (which will be referred to as 'direct selling'), or she can prefer to sell the product in bulk to the intermediary. The unit price of the product sold directly to the end customer, p_D, is higher than the unit price of the product sold to the intermediary, p_m, which is in accordance with real life. The farmer's unit cost of production and handling when she sells to the intermediary is denoted by c, and the additional cost incurred in direct selling is denoted by c^0. This additional cost may accrue due to extra picking and preparing effort or transportation cost to the farmer's market. We assume that $p_D - c - c^0 > p_m - c$; that is, the farmer earns a higher profit in direct selling despite the extra effort. Finally, assume that the demand for the direct selling is given by $D(p_D) = K - ap_D$ for an appropriate price range, $p_D \in \left[\underline{p}, \overline{p}\right]$. This form of linear demand functions is frequently used to model the demand of agricultural products (e.g. see Kazaz (2004)). Hence, after the yield uncertainty is resolved, the farmer's profit in this problem is simply given by the following equation.

$$\pi_{f0}(V) = (p_D - c - c^0)min\{K - ap_D, V\} + (p_m - c)(V - K + ap_D)^+ \qquad (1)$$

where V is the yield of the product which has a probability distribution function $f(.)$ (with respective cdf $F(.)$) on the range $\left[\underline{V}, \overline{V}\right]$. The value of V is realized at the beginning of the sales period. Hence, the farmer's expected profit without a CSA before the harvest period is $E\left[\pi_{f0}(V)\right]$.

Next, suppose that the farmer engages in CSA. We assume that the customer base is the same as in the direct selling, therefore the demand function is the same. Moreover, due to the more efficient harvest process and reduced cargo costs due to bulk shipments (or, due to the improved cash flow throughout the year), we assume that the farmer does not incur the additional cost in direct selling, c^0, any more. The customers now pay a "lower" fixed price for the entire season, which we assume corresponds to a unit price of p_C for the generic product. This price is determined before the harvest time where the uncertainty in yield is resolved. Moreover, although the CSA philosophy states that whatever produce is available will be sent to the members each week; we will assume that the farmer guarantees to fulfill the demand of each member regardless of the yield amount. Therefore, whenever the product yield is realized at low levels, the producer has to satisfy the demand of the members from the direct selling market incurring a loss of $p_D - p_C$ per unit. On the other hand, if the amount of yield is larger than the amount of products to be sent to the CSA members, the rest of the produce can be sold to the intermediary firms as before.

Given the above assumptions, the expected profit function for the farmer under the CSA model is given by:

$$E\left[\pi_{fC}(p_C, V)\right] = \max_{p_C} \int_{\underline{V}}^{D(p_C)} \{y(p_C - c) - (D(p_C) - y)(p_D - p_C)\} f(y)dy + \int_{D(p_C)}^{\overline{V}} \{D(p_C)(p_C - c) + (y - D(p_C))(p_m - c)\} f(y)dy. \qquad (2)$$

The above function is in the stated form provided that $\underline{V} \le D(p_C) \le \overline{V}$. If $\underline{V} \ge D(p_C)$, the equation (2) will take the form: $E\left[\pi_{fC}(p_C, V)\right] = \max_{p_C} \int_{\underline{V}}^{\overline{V}} \{D(p_C)(p_C - c) + (y - D(p_C))(p_m - c)\} f(y)dy$. To avoid trivial solutions, we will also assume that $D(p_D) \le \overline{V}$ and $\underline{V} \le D(p_m)$.

We might also assume that the farmer is somewhat risk-averse. A risk-averse farmer is concerned about the possibility of losing money on his harvest, which we might model through a value-at-risk (VaR) constraint:

$$\Pr(\pi_{fC}(p_C, V) \le -\beta) \le \alpha \qquad (3)$$

Constraint (3) says that the probability of a loss as large as β or more must not be more than α. The value of α is small in practice (e.g., α~5%-10%). For a given α, the larger the value of β is, the less risk-averse is the farmer. Note that, as the demand function is decreasing in p_C, farmer's risk of losing money decreases if he sets a higher price. Hence, defining p_C^{var} as the minimum price to set within the acceptable price range in order to satisfy the VaR constraint (3), i.e.

$$p_C^{var} = \inf_{p_C\in[p_m,p_D]}\left\{p_C : Pr\left(\left[\int_{\underline{V}}^{D(p_C)}\{y(p_C-c)-(D(p_C)-y)(p_D-p_C)\}f(y)dy+\int_{D(p_C)}^{\overline{V}}\{D(p_C)(p_C-c)+(y-D(p_C))(p_m-c)\}f(y)dy\right]+\beta\le 0\right)\le\alpha\right\},$$

it is clear that the farmer should set the price p_C above p_C^{var}.

Finally, the CSA prices should be less than the usual direct-selling price but higher than the intermediary price, i.e.

$$p_m \le p_C \le p_D \quad (4)$$

Hence, the profit-maximizing problem of the farmer is stated as in the following line.

$$E\left[\pi_{fC}(V)\right] = \max_{p_C} E\left[\pi_{fC}(p_C,V)\right] \; s.to \; (3),(4). \quad (5)$$

Theorem: The optimal solution of the farmer's profit-maximizing problem is given by:

$$p_C^0 = max\{p_C^{var}, p_C^*\} \quad (6)$$

where p_C^* is the value that satisfies the following equation:

$$K-2ap_C^*+ap_DF\left(D(p_C^*)\right)+ap_m\left(1-F\left(D(p_C^*)\right)\right)=0 \quad (7)$$

Proof: The farmer's expected profit function stated in (2) can be shown to be concave in p_C. Hence, assuming an interior solution and ignoring the constraints, the optimal solution is found at the point which satisfies the first order condition (FOC), i.e. equation (7). This value is larger than p_m and smaller than p_D since trivial solutions are ignored. Taking into account the fact that the price should be as large as p_C^{var} in order to satisfy the risk constraint of the farmer, and that p_C^{var} is not larger than p_D by its definition, we obtain the form of the solution as in (6). □

The next Proposition establishes monotonicity results according to the problem parameters.

Proposition: The optimal solution of the unconstrained profit function, p_C^*, is decreasing in a; and increasing in K, p_D and p_m. The optimal expected profit function $E\left[\pi_{fC}(V)\right]$ is increasing in p_m and decreasing in p_D.

Proof: The first part of the proof can be found by using the interior function theorem, i.e.

$$\frac{\partial N}{\partial p_C^*}\Rightarrow\frac{\partial G}{\partial p_C^*}+\frac{\partial G}{\partial N}\frac{\partial N}{\partial p_C^*}=0\Rightarrow\frac{\partial N}{\partial p_C^*}=-\frac{\partial G/\partial p_C^*}{\partial G/\partial N}$$

for $N = K, p_D, a, p_m$, where $G(K,p_C^*,p_D,a,p_m) = K-2ap_C^*+ap_DF\left(D(p_C^*)\right)+ap_m\left(1-F\left(D(p_C^*)\right)\right)=0$. For instance,

$$\frac{\partial p_C^*}{\partial p_D}=-\frac{\partial G/\partial p_D}{\partial G/\partial p_C^*}=-\frac{aF\left(D\left(p_C^*\right)\right)}{-2a-a^2p_Df\left(D\left(p_C^*\right)\right)+a^2p_mf\left(D\left(p_C^*\right)\right)}>0$$ since $p_D > p_m$, $f(.)\ge 0, F(.)\ge 0$.

Similarly, it can be shown that $\frac{\partial p_C^*}{\partial p_m}\ge 0, \frac{\partial p_C^*}{\partial K}\ge 0, \frac{\partial p_C^*}{\partial a}\le 0.$

To see the second part, first consider the case where $p_C^0 \neq p_C^*$. In that case, $\frac{\partial}{\partial p_D} E\left[\pi_{fC}(V)\right] = \int_{\underline{V}}^{D(p_C^0)} \left(y - D\left(p_C^0\right)\right) f(y)dy \leq 0$ and $\frac{\partial}{\partial p_m} E\left[\pi_{fC}(V)\right] = \int_{D(p_C^0)}^{\overline{V}} \left(y - D\left(p_C^0\right)\right) f(y)dy \geq 0$.
Next, consider the case where $p_C^0 = p_C^*$. Note that the expected profit function can be written as:

$$E\left[\pi_{fC}(V)\right] = D(p_C^*)p_C^* - cE[V] + \int_{\underline{V}}^{D(p_C^*)} (y - D(p_C^*))p_D f(y)dy + \int_{D(p_C^*)}^{\overline{V}} (y - D(p_C^*))p_m f(y)dy$$

Noting that $\frac{\partial p_C^*}{\partial p_D} \geq 0$ and $\frac{\partial D(p_C^*)}{\partial p_D} = -a\frac{\partial p_C^*}{\partial p_D} \leq 0$:

$$\frac{\partial}{\partial p_D} E\left[\pi_{fC}(V)\right] = \frac{\partial p_C^*}{\partial p_D} D(p_C^*) - a\frac{\partial p_C^*}{\partial p_D} p_C^* + \int_{\underline{V}}^{D(p_C^*)} \left(y - D(p_C^*)\right) f(y)dy + \int_{\underline{V}}^{D(p_C^*)} a\frac{\partial p_C^*}{\partial p_D} p_D f(y)dy + \int_{D(p_C^*)}^{\overline{V}} a\frac{\partial p_C^*}{\partial p_D} p_m f(y)dy$$

$$= \frac{\partial p_C^*}{\partial p_D}\left\{D(p_C^*) - ap_C^* + ap_D \int_{\underline{V}}^{D(p_C^*)} f(y)dy + ap_m \int_{D(p_C^*)}^{\overline{V}} f(y)dy\right\} + \int_{\underline{V}}^{D(p_C^*)} \left(y - D(p_C^*)\right) f(y)dy \leq 0$$

which is because the term in the parenthesis is zero (check equation (7)) and $\int_{\underline{V}}^{D(p_C^*)} \left(y - D(p_C^*)\right) f(y)dy \leq 0$. Similarly, $\frac{\partial}{\partial p_m} E\left[\pi_{fC}(V)\right] \geq 0$.
□

The above Proposition establishes that unconstrained profit-maximizing CSA membership price would increase with the prices of the open market and the intermediary price. Moreover, the farmer would be able to charge higher values when the domestic market develops (i.e. K is large, a is small). However, if the profitability is concerned, higher open market prices reduce the profitability of the CSA scheme while intermediary payments create an opposite effect. That is, CSA would be the most effective way to improve the profits of a farmer if the prices of the direct selling market are not very profitable while the intermediary payments are better. The latter might seem counterintuitive, but remember that the extra volume of the product which is not sold in farmer's market or to the CSA members will still be sold to the intermediary firms.

The effect of the yield volatility on the optimal price and profitability of CSA models would also be interesting to analyze. Similarly, rather than the nominal increase in $E\left[\pi_{fC}(V)\right]$, it is interesting to see how the profit changes with respect to the profit under no CSA setting. However, it is difficult to develop closed form monotonicity results for a varying yield volatility, or the ratio of $E\left[\pi_{fC}(V)\right]/E\left[\pi_{f0}(V)\right]$, due to the complex mathematical equations. Therefore, we would like to analyze these issues on a numerical problem.

4. NUMERICAL PROBLEM

Consider the problem of a small farm owner who produces lentil. According to the price values taken from tazedirekt.com, an online selling platform of organic produces in Turkey, the unit price for 1 kg of organic lentil is assumed to be 12 TL. (Meanwhile, the unit price for non-organic lentil is around 6 TL per kg (source: sanalmarket.com).) According to TUIK data, the yield values for non-organic lentil is around 150 kg/decare, which we will adapt to our setting as $V \sim U[50, 150]$ kg/decare taking into account a 80% productivity rate of organic products and the best vs. the worst weather conditions (TUIK, 2018).

Assume that a small farmer has 200 decares of land on which she grows organic produces (this number is found by averaging the total agricultural land in Turkey with respect to the total number of growers). Assuming that all of the products that are grown by the farmer can be averaged by a single product, which is lentil, we denote the yield of this farmer to be $V \sim U[10{,}000, 30{,}000]$ kg. We will also assume a demand function in the following form:

$D(p) = 54{,}000 - 4000p$, for $p \in [6, 13.5]$.

That is, if the price of lentil is as low as 6, which is the price non-organic lentil is offered, the farmer can sell all of her produce in the consumer market. It is also safe to assume that in the current market price $p_D = 12$ of organic lentil, the farmer cannot sell the majority of her produce directly to the end customer.

Finally, we will set $p_m = 7, c = 5, c^0 = 3$. These numbers are purely assumptional, although it is safe to suppose that the intermediaries pay much lower prices than the direct selling option, and that unit cost of production can be as high as 80% of the prices offered by the intermediaries.

Under the given set of parameters, the expected profit of the farmer without a CSA model is:

$E\left[\pi_{f0}(V)\right] = \int_{V=10,000}^{30,000}\{(12-5-3)6{,}000 + (7-5)(V-6{,}000)\}\frac{1}{10{,}000}dV$=52,000TL

However, given a 5% VaR constraint with $\beta = 10{,}000$, the optimal price under CAS scheme and the resulting expected profit of the farmer becomes:

$p_C^* \approx 10.5$, $E\left[\pi_{fC}(V)\right]$ ≈81,500TL.

The effect of VaR constraint is insignificant for this set of parameters, hence $p_C^0 = p_C^*$. These figures mean that the consumers can enjoy lower (15% lower) organic product prices while the producer can obtain a much higher (57% higher) profit level.

To see how the price rate and expected profit levels change for a range of parameters, we will first start with the price and cost figures. Figure 1 shows how the optimal price p_C^0 changes as a percentage of the market price p_D (i.e. p_C^0/p_D) and how $E\left[\pi_{fC}(V)\right]/E\left[\pi_{f0}(V)\right]$ varies, as the market price for organic lentil, p_D, changes within the range [9, 13], and as the intermediary price, p_m, changes within the range [5, 9]:

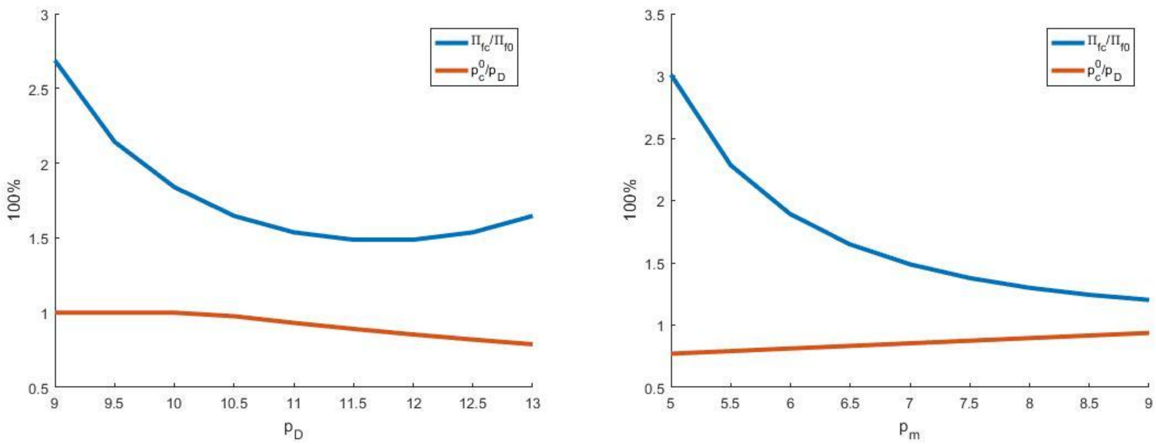


Figure 1. The ratios p_C^0/p_D and $E\left[\pi_{fC}(V)\right]/E\left[\pi_{f0}(V)\right]$ for different values of p_D and p_m

For lower values of market price, $E\left[\pi_{fC}(V)\right]/E\left[\pi_{f0}(V)\right]$ seems quite high, as the farmer sets the CSA price at the market price value and enjoys lower costs. This might not be too realistic with the current conditions of the domestic organic market in Turkey. Rather, the part where the market price is above 11 gives a better idea about how CSA scheme could be beneficial for the farmer and the consumers, still bringing an extra profit of above 50% to the farmer in all cases. $E\left[\pi_{fC}(V)\right]/E\left[\pi_{f0}(V)\right]$ decreases as the intermediary price p_m increases, in which case again the CSA price approaches the market price, which is to be expected. On a similar note, although we know from the Proposition that the optimal CSA price increases as the market price increases, this increase is less with respect to the increase in the p_D itself, bringing the end consumer an advantage under the CSA scheme for high values of the market price.

The next Figure shows the $E\left[\pi_{fC}(V)\right]/E\left[\pi_{f0}(V)\right]$ ratio for various values of c and c^0 (since p_C^0 is not affected by the cost parameters, we do not plot the p_C^0/p_D ratio).

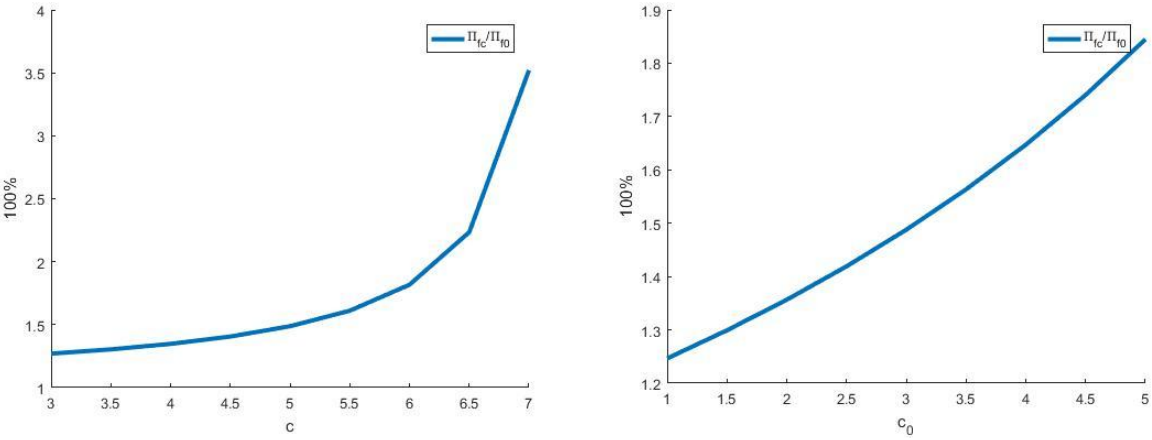


Figure 2. The ratios p_C^0/p_D and $E\left[\pi_{fC}(V)\right]/E\left[\pi_{f0}(V)\right]$ for different values of c and c_0

In both graphs, it is seen that the increase in the costs would mean that the CSA scheme would be much more profitable than the base case. The increase in profits is more for higher values of c, resulting from the fact that high values of c hurts $E\left[\pi_{f0}(V)\right]$ more than it hurts the profitability of the farmer with CSA. Moreover, the more the direct selling presents an extra cost, the higher the farmer's motivation would be to set up a CSA scheme, as expected.

Next, Figure 3 denotes the p_C^0/p_D and $E\left[\pi_{fC}(V)\right]/E\left[\pi_{f0}(V)\right]$ ratios for various values of the volume "range", namely the variability in the volume. In the base case, i.e. when $V \sim U[10{,}000, 30{,}000]$, the range is 30000-10000=20,0000. By keeping the expected volume constant at $E[V]$=20,000, we vary the range value from 5000 to 37,000 and obtain the following graph, indicating that although volatility adversely affects the CSA profits, the farmer still earns more under CSA even under high volatility values.

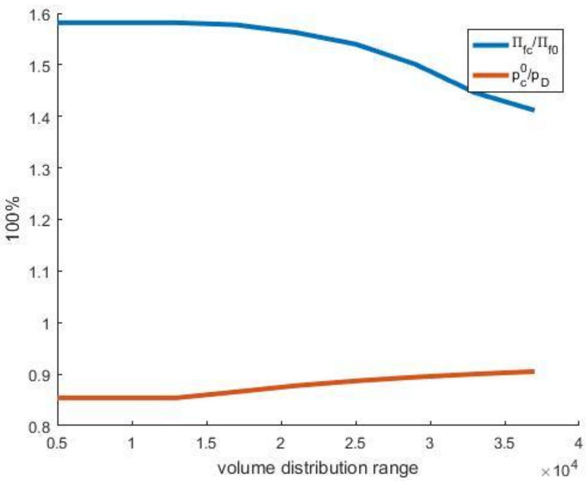


Figure 3. The ratios p_C^0/p_D and $E\left[\pi_{fC}(V)\right]/E\left[\pi_{f0}(V)\right]$ for different values of range

Finally, we would like to analyze the impact of price elasticity on CSA profitability and the optimal CSA price. To this end, we changed the parameter a in the demand function $D(p)$ within the range of [2500, 5000]. The following Figure shows the ratio of the optimal/market price, and the profit values.

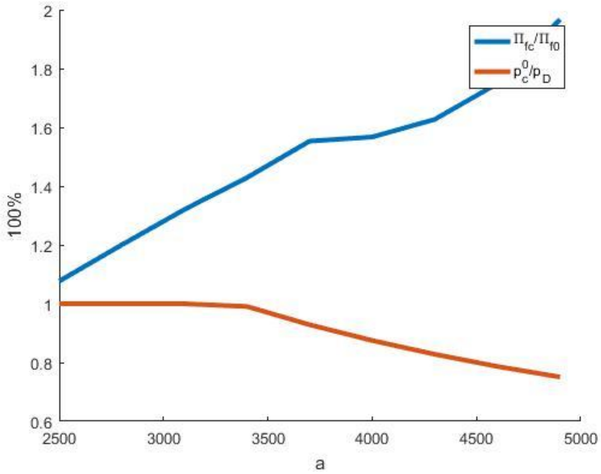


Figure 4. The ratios p_C^0/p_D and $E[\pi_{fC}(V)]/E[\pi_{f0}(V)]$ for different values of a

According to Figure 4, as the price elasticity of the demand increases, the profitability under CSA is improved with respect to the profit under the classical setting and the optimal price under the CSA scheme decreases. That is, if the domestic market is too sensitive to the price of the organic produces, the farmers will benefit more from the CSA scheme by offering lower prices and capturing more of the consumer demand.

5. CONCLUSION

In this paper, we developed a pricing model for CSA practice and analyzed the conditions for which this model would be more profitable for a small organic producer. According to the results of our analysis, the CSA practices seem to be most advantageous for the farmer when the market price is too high to let the domestic market grow substantially and the costs are high relative to the profits, which is mainly the case in Turkey and other developing countries. Moreover, although the volatility of the yield has an adverse effect on CSA profitability, the profits under the CSA scheme still seem to be much larger than the profits in the non-CSA setting up to very high volatility rates. Finally, if the demand function is more elastic meaning more sensitive to the changes in the organic market prices, the CSA scheme can be more profitable for the farmers with the correct pricing policy. In addition to the farmer benefits, individual consumers in the market would also benefit from the CSA practice by paying lower prices than the market premium. To our knowledge, our paper is the first attempt to analyze the impact of the CSA scheme on the farmer and consumer wealth from a mathematical modeling perspective.

However, our paper clearly has several limitations. First, we have assumed that the consumer demand is deterministic and known by the farmer with certainty, which might not be the case. Moreover, it is difficult to estimate the CSA demand without any past data; which led us to model it in the same manner as the regular market demand is modeled. We believe that these shortcomings of the paper can be remedied with further research when the researchers work in close connection with actual farmers and obtain first-hand data, which is in the scope of the future prospects of this project. Despite these shortcomings, we believe that the model we present here can be generalized to the pricing and planning problem of small growers especially in developing countries with small domestic organic markets. Hence, we hope that the insights we provide here can be used by other researchers and professionals from the field to improve the welfare and the living conditions of the farmers and the individual consumers all around the world.

REFERENCES

Adam, K. L. (2006). Community supported agriculture. Butte, MT: ATTRA-National Sustainable Agriculture Information Service.

Ahumada, O., Villalobos, J. R., & Mason, A. N. (2012). Tactical planning of the production and distribution of fresh agricultural products under uncertainty. Agricultural Systems, 112, 17-26.

Brown, C., & Miller, S. (2008). The impacts of local markets: a review of research on farmers markets and community supported agriculture (CSA). American Journal of Agricultural Economics, 90(5), 1298-1302.

Cone, C., & Myhre, A. (2000). Community-supported agriculture: A sustainable alternative to industrial agriculture?. Human organization, 59(2), 187-197.

Demiryürek, K. (2011). Organik tarım kavramı ve organik tarımın dünya ve Türkiye'deki durumu. GOÜ, Ziraat Fakültesi Dergisi, 28(1), 27-36.

Hovelaque, V., Duvaleix-Tréguer, S., & Cordier, J. (2009). Effects of constrained supply and price contracts on agricultural cooperatives. European Journal of Operational Research, 199(3), 769-780.

Jang, W., & Klein, C. M. (2011). Supply chain models for small agricultural enterprises. Annals of Operations Research, 190(1), 359-374.

Kazaz, B. (2004). Production planning under yield and demand uncertainty with yield-dependent cost and price. Manufacturing & Service Operations Management, 6(3), 209-224.

Lass, D., Bevis, A., Stevenson, G. W., Hendrickson, J., & Ruhf, K. (2003). Community supported agriculture entering the 21st century: Results from the 2001 national survey. Amherst: University of Massachusetts, Department of Resource Economics.

Soysal, M., Bloemhof-Ruwaard, J. M., Meuwissen, M. P., & van der Vorst, J. G. (2012). A review on quantitative models for sustainable food logistics management. International Journal on Food System Dynamics, 3(2), 136-155.

Soto-Silva, W. E., Nadal-Roig, E., González-Araya, M. C., & Pla-Aragones, L. M. (2016). Operational research models applied to the fresh fruit supply chain. European Journal of Operational Research, 251(2), 345-355.

TUIK, 2018. Statistical Tables about Turkey's Agricultural Production. http://www.tuik.gov.tr/PreTablo.do?alt_id=1001. Accessed **on** 10 March 2018.

Villalobos, J. R., Mason, A., Flores, H., Wishon, C., & Ahumada, O. (2014). Planning, consolidation and coordination the keys for small farmers to advance in the fresh produce value chain.

BALCOR 2018
XIII Balkan Conference on Operational Research

CONSUMERS' ATTITUTES TOWARDS PRODUCTS OF WOMEN'S COOPERATIVES

ANTONIOS TIGANIS[1], MANDO DIMITRIADOU[2], MARIA KARAISKOU[3] ANTONIOS AVGERIS[4]

[1]MSc Researcher, Dpt. of Agricultural Economics, Aristotle University of Thessaloniki, Thessaloniki, Greece, tiganisant@gmail.com
[2]MSc Researcher, Dpt. of Agricultural Economics, Aristotle University of Thessaloniki, Thessaloniki, Greece, mandod2189@gmail.com
[3]MSc Researcher, Dpt. of Agricultural Economics, Aristotle University of Thessaloniki, Thessaloniki, Greece, marykaraisk@gmail.com
[4]MSc Researcher, Dpt. of Agricultural Economics, Aristotle University of Thessaloniki, Thessaloniki, Greece, antoavge@agro.auth.gr

---

Abstract: *Women's Cooperatives are important for both women's and family income in rural areas. In order to position their products better they have to understand consumers' attitudes towards them. Thus, the central aim of the research is the identification of attitudes and beliefs by figuring out which attributes of women's' cooperatives products are considered with the greatest and which with the least importance. For the main objective we used the method of Best Worst Scaling in order to figure out which of the following eleven attributes consumers prefer most or least towards Women's Cooperatives products consumption. We found out that the three most preferred attributes are 1) Quality, 2) Handmade production, 3) Enhancement of Women's Cooperatives, while the three less favored are 1) Traceability, 2) Recipe, 3) Region. Two secondary aims were to define consumers' attitudes towards short chains and innovative products correlated with women's cooperatives.*

Keywords: *Best Worst Scale, Women's Cooperatives, Consumer Behavior.*

1. INTRODUCTION

Women of rural areas need pluriactivity in order to have their income raised up (Gidarakou et al., 2000). In Greece there are 141 active Women's Agro tourist Cooperatives (Ministry of Rural Development and Food, 2014) where women can both increase their income and upgrade their status into society. This pluriactivity need, expanded in Women's Cooperatives as well, mainly aimed at the fields of agro tourism, handicraft and home handicraft production, and producing and trading alternative agricultural products (Koutsou et al., 2003). The Women's Cooperatives hold an important role in these extra activities, where women can produce collectively rural and local products, that carry the dietary history and cultural or food heritage of the place they live.

Thus, we can propose that Women's Cooperatives are important not only for the financial contribution to woman and family income, but for the preservation of local culture as well. Moreover they enhance the collective spirit of rural areas. It is found, that farm-women prefer to work in terms of team or community, when involved in extra-agricultural activities (Koutsou et al., 2003).

This polyhedral significance of Women's Cooperatives demands a sufficient business strategy towards product's production and potential clients. Thus, defining the consumers' attitudes towards their products, Women's Cooperatives can segment their potential market and strengthen their strategy. The main objective, for accomplishing these aims, is to measure the preference or not towards eleven attributes of the Women's Cooperatives products by using Best Worst Scale. Furthermore, for the wider understanding of consumer perceptions we tried to investigate what they believe about short chains and innovative products. Short chains have a key role in reducing the cost of the product and thus its price, so it is crucial to define what consumers believe about them. Innovative products can play an important role on Women's Cooperatives strategy as they are differentiated products, thus they can both allure new clients and be sold in higher price.

A questionnaire was developed in November 2017 and administrated to 300 consumers during December 2017 and January 2017. At the beginning of the questionnaire we ask socio-demographic questions. Next, we are examining beliefs about Women's Cooperatives, and which of their products consumers buy, where, how frequently and how satisfied they are. Furthermore, we evaluate consumers' beliefs and preferences for short chains. We have used both Likert method and questions where consumers can choose which answer represents better their opinion. For examining their preference towards innovative products we have asked them to choose from a list two they prefer would most and two they would prefer less. Lastly, we used the Best Worst Scale method investigating which of the following eleven attributes consumers prefer most or least towards Women's Cooperatives products consumption. The attributes used were: 1) Quality 2) Availability 3) Price 4) Region of Origin 5) Tradition Preservation 6) Handmade production 7) Recipe 8) Traceability 9) Enhancement of Women's Cooperatives 10) PDO/PGI Certificate 11) Quality Certificates (ISO, HACCP ETC).

The final sample amounted to 300 consumers from three Greek cities (Thessaloniki, Naoussa and Karditsa), two of them located in the Greek regional unit of Central Macedonia (Thessaloniki, Naoussa) and one (Karditsa) in the Greek regional unit of Thessaly.

2. THEORETICAL FRAMEWORK

There are several factors that influence the consumer buying behavior towards food. The most important are the quality, the price, the place of origin, the taste, the production methods, the convenience to reach it and also factors as culture and religion. Nowadays consumers seem to value definitions like sustainability, locality and are also interested in environmental protection. The food choice is also based to personal preferences but may also be based in social and ethical factors (Koutroulou and Tsourgiannis 2011, Gracia 2013).

Other factors that influence the consumer behavior come from the negative aspects of food globalization. Food that travels through the conventional food supply chain is based on imports and is usually traveling a long distance to reach the final consumer, growing the distance between the place of production and consumption (Feldmann and Hamm, 2014). The globalization of food has raised the worries of the consumers since they feel uncertainty for the origin of the product they consume due to the difficultness of traceability which can be caused by the presence of many market intermediates whose presence makes the distribution channel more complex and the asymmetry of information severe (Aung and Chang 2014, Brunorei et al. 2016). In addition, the appearance of a number of food scandals such as the H1N1 flu, and the use of controversial methods and inputs such as the use of genetically modified organisms, have played a significant role in the loss of their trust towards the conventional food and making them question their safety (Banati, 2011).

The long distance between the food and the consumer is considered to have a negative effect on the quality, the nutrition level and taste of the product since the nature of agriculture products is quite sensitive and easily affected by the time and conditions of preservation, but also to the environment as it demands greater consumption of energy and fuels in order to transport to the final destination. That's why the consumers may appear more willing to pay a higher price for short distance products (Akkerman et al. 2010, Grebitus et al. 2013).

According to Eriksen and Sundbo, local food is the one that is produced and sold within a geographical area. This is recognized as the geographical proximity of these foods which cannot be measured with a certain number of miles even though some researchers have placed it between 100 to 25 miles or even at national level according to the urban or rural character of their living (Duram and Oberholtzer 2010, Carroll and Fahy 2014, Eriksen and Sundbo 2016).

Local and traditional food products are considered to fulfill the demands of the consumers according to freshness, higher quality, safety and superior taste, authenticity, but also appear additional attributes as the sustainable local development, the reduced environmental impact and the survival of tradition and cultural identity. Food is combined to tradition and in addition with the place of origin constructs the identity of the product which appears it unique and different from other similar ones (Denver and Jensen 2014, Luceri et al. 2016).

In order to protect traditional and local products from imitations of low quality and assure the consumers of their unique attributes and origin, a number of quality labels such as PDO (Protected Designation of Origin), PGI (Protected Geographical Indication) and TGS (Traditional Specialties Guaranteed) where developed within the EU. Those labels are considered to promote the buying intention as they may work as reassurance for quality standards (Grunert and Aachmann 2015, Luceri et al. 2016). According to the research of Tsakiridou et al. (2009) the presence of these labels may present an influence on the purchasing motives of the consumers who live in the specific region since they seem more informed for their significance.

According to Eriksen, local food has also a social proximity in the sense that it promotes the construction of relationships between the actors of the local food system but it also brings the consumer closer to the production area and the producer, especially when it is sold through direct distribution channels. This kind of relationships promotes the sense of trust but also the willingness of the consumers to support the local producers and generate also the local or national economy (Eriksen, 2013). Local food can promote the function of short food supply chains because it refers to specific boundaries where direct social interactions, the sense of trust and the sharing of information through direct communication between consumers and producers can overcome the need of multiple intermediates (Bareja-Wawryszuk and Golebiewski, 2014).

Both local food systems and short supply chains are considered as alternative food networks that consist an alternative solution towards the conventional food supply chain and the growth of globalized food. Local food systems refer primarily to the geographic distance between production and consumption and promotes the (re) territorialize of food. From the other hand, short food supply chains refer to the minimization of the number of intermediates which interfere between the consumers and the producers. That kind of chain may operate as direct sales channel where the number of market intermediates is zero or with the interference of a single intermediate who can be a cooperative shop, a supermarket, local shops, restaurants etc. (Kneafsey et al. 2013).

Without the increased number of market intermediates, producers can achieve a greater profit for their products, increasing in that way their income. Since price consists one of the factors that influence the purchasing intention, consumers are willing to pay for these type of food because they acknowledge their special features and also believe that they can reach them in reasonable price while the absence of intermediates closes the marketing margins (Sandika 2011, Bimbo et al. 2015).

There are several forms of SFSCs where local food can reach the consumers through farmers markets and shops, box schemes etc., which mostly refer to face-to-face interactions, but there is also the possibility for home deliveries and e-commerce in order to reach the consumers outside the specific region but inside the national boarders, and make them aware of the existence and quality of these products (Renting et al., 2003).

Women cooperatives seem to play an important role in the diffusion of local and traditional food, maintaining the cultural heritage of the region and at the same time offering high quality products. Even though women agriculture cooperatives are not very common in other EU countries, there is an important number of them in Greece. Most of them run their own shops within their facilities at local level since they were unable to expand due to financial difficulties and lack of education and entrepreneurship capacities (Vakoufaris et al. 2007, Sergaki et al. 2015).

The traditional character of these products makes them distinguish from the simple local definition and gain even more attention and meaning since they tie in with concepts such as cultural identity, ethnocentrism and nostalgia because of the reminisce to family made recipes (He and Wang 2014 , Renko and Bucar, 2014). Women cooperatives products are based on traditional recipes which survived and methods of production which survived through the years and they are the link with our history (Anthopoulou and Koutsou, 2010).

These foods are based to the loyalty of consumers which allows them to maintain their position in the market but due to the increased interest on them they can also play a significant role in the rise of tourism, which may occur to a spread on their demand in a wider level and even result to a flourish of exports (Dorota-Rudawska 2014, Madaleno et al. 2017).

Innovation is also an important key factor which may help to the greater expand of these products and that way it can help the weak women cooperatives. Women in general, are considered to have a tendency towards innovations and can successfully operate short distribution channels thanks to their greater sense of caring, the ability to multifunction and to their polite character (Zirham and Palomba, 2016).

Although innovation may attract the consumers, the unique nature of traditional food products must remain intact when it comes to the production methods and recipe in order to maintain the taste. Most of the consumers tend to accept innovations that have to do with the packaging of these products or with the size of them in order to become more convenient to use, and also the use of labels of origin. In other words, there is a positive correlation towards innovation and traditional products only if there are considerable health, safety and convenience advantages for the consumers while their character remains stable (Kuhne et al. 2010, Vanhonacker et al. 2013).

3. METHODOLOGY

The understanding of products' choice procedure by consumers is considered quite important for women's cooperatives in order to create the suitable conditions of increasing or ensuring their market share. The understanding of consumer's behavior requires data analysis which follows the collection of primary data that are compatible with research aims (Chrysochou, 2017).

A structured questionnaire was prepared and shared in a final number of 300 people, by random sampling method to three urban areas of Greece. The cities have been chosen according to existence of women's cooperatives. In the city of Thessaloniki (administrative region of Central Macedonia) there are three women's cooperatives producing traditional products, in the city of Naousa (administrative region of Central Macedonia) there is one women agro-tourist cooperative, in the city of Karditsa (administrative region of Thessaly) there are six women's cooperatives from which two are creating handicrafts and cottage products. These three different areas segmented the research to three study groups. The research lasted two months, from December 2017 to January 2018.

The questionnaire included six demography questions, nine research questions from which five were using Likert method and nine were using Best Worst Scale method. We used the Likert method in questions regarding consumer views on supply chain, on the level of satisfaction by the women's cooperatives products, and on the improvement of products, package and brands as well.

Furthermore, we have chosen to use Best Worst Scale method, as it requires less collected data and demonstrates a higher rate of validity compared with other formats (Hollis and Westbury, 2018). Particularly, requires from those answering the survey the suggestion of the ''best'' and the ''worst'' attribute between a total of attributes, and then it alternates the choices given in order to create a full list (Petrolia, Interis and Hwang, 2015). Those eleven attributes were chosen according to parameters and previous researches. By this way consumers were choosing in each question the attribute they were regarding as most important (best) or least important (worst) among five. There in total eleven Best Worst questions.

The attributes we used were:

1) Quality 2) Availability 3) Price 4) Region of Origin 5) Tradition Preservation 6) Handmade production 7) Recipe 8) Traceability 9) Enhancement of Women's Cooperatives 10)PDO/PGI Certificate 11) Quality Certificates (ISO, HACCP ETC).

4. RESEARCH QUESTIONS

i. Are consumers interested in short chains and are there any differences in their preferences among the different research areas?
ii. Which are the innovative products among the suggested, that consumers would prefer most or least?
iii. Which are the most and the least important attributes of women's cooperatives products?

5. RESULTS

The first results of the research were counted within the total sample (n=300) so that, we can have a first view of our market. Starting with the demographics we can define our consumers' profile. The majority of them are women (59.3%), their age is 20-29 (34%), and they are graduates from tertiary education. The average monthly income is 500-1000 euros (33.7%) while the average family consists of four people. There is a dispersion of the consumers' occupation. We have found 28.7% of them to be public servants, 24.7% private employee, 16.3% freelancers etc.

Concerning consumers' attitudes to women's cooperatives we observed that the majority believes that they respond efficiently to market's requirements and that they contribute into local society's traditions preservation. On the other hand, consumers were not very satisfied with the range of products offered by women's cooperatives.

The vast majority of consumers are heavily interested on short chains, without significant differences among sexes, different ages and different education's levels. By dividing our sample to the three research areas, we can observe that in the city of Thessaloniki there is a greater need for short chains than the other two areas.

Table1: Consumers' opinions on short chains

Research areas	It is better to purchase local products from short chains.					
	I strongly disagree	I disagree	I neither agree nor disagree	I agree	I strongly agree	Total
Naoussa	5 (5%)	3 (3%)	19 (18.8%)	44 (43.6%)	30 (29.7%)	101 (100%)
Thessaloniki	1 (1%)	1 (1%)	18 (18.6%)	38 (39.6%)	39 (40.2%)	97 (100%)
Karditsa	0 (0%)	3 (3.1%)	17 (17.3%)	44 (44.9%)	34 (34.7%)	98 (100%)
Total	6 (2%)	7 (2.4%)	54 (18.2%)	126 (42.6%)	103 (34.8%)	296 (100%)

Concerning the questions about innovative products we gave consumers the possibility of choosing up to two products they would prefer most or less from a list of nine products (gift package with products, cereal bars, stuffed olives, pasta with ready-made sauce, individual portion of sweet, sachet with tea, liqueurs with carbonate, handcrafted salves, yoghurt with jam). Surprisingly, the gift package with products figured in both categories. It would be preferred most by one segment of the market and less by another one. The first three preferred most innovative products were individual portion of sweet, sachet with tea and gift package with products. The first three preferred less innovative products preferred less are liqueurs with carbonate, gift package with products and cereal bars. Results are demonstrated both as a total and for each research area as well.

Table2: Innovative products consumer prefer most

Innovative Products	Research areas			
	Naousa	Thessaloniki	Karditsa	Total
Gift package with products	27 (13.4%)	30 (15.5%)	28 (13.7%)	85 (14.2%)
Cereal bars	29 (14.3%)	20 (10.3%)	29 (14.2%)	78 (12.9%)
Stuffed olives	9 (4.5%)	16 (8.2%)	18 (8.8%)	43 (7.2%)
Pasta with ready-made sauce	19 (9.4%)	24 (12.4%)	16 (7.8%)	59 (9.9%)
Individual portion of sweet	32 (15.8%)	39 (20%)	22 (10.8%)	93 (15.5%)
Sachet with tea	35 (17.3%)	24 (12.4%)	26 (12.8%)	85 (14.2%)
Liqueurs with carbonate	8 (4%)	4 (2%)	13 (6.4%)	25 (4.1%)
Handcrafted salves	29 (14.3%)	23 (11.9%)	31 (15.2%)	83 (13.8%)
Yoghurt with jam	14 (7%)	14 (7.3%)	21 (10.3%)	49 (8.2%)
Total	202(100%)	194 (100%)	204 (100%)	600 (100%)

Table3: Innovative products consumers prefer less

Innovative Products	Research areas			
	Naousa	Thessaloniki	Karditsa	Total
Gift package with products	27 (13.4%)	25 (12.9%)	40 (19.6%)	92 (15.3%)
Cereal bars	29 (14.3%)	22 (11.3%)	28 (13.7%)	79 (13.1%)
Stuffed olives	20 (9.9%)	28 (14.4%)	27 (13.2%)	75 (12.5%)
Pasta with ready-made sauce	18 (9%)	10 (5.2%)	19 (9.3%)	47 (7.8%)
Individual portion of sweet	11 (5.4%)	7 (3.6%)	19 (9.3%)	37 (6.1%)
Sachet with tea	11 (5.4%)	12 (6.2%)	14 (7%)	37 (6.1%)
Liqueurs with carbonate	39 (19.3%)	45 (23.2%)	19 (9.3%)	103 (17.3%)
Handcrafted salves	22 (10.9%)	18 (9.3%)	19 (9.3%)	59 (9.8%)
Yoghurt with jam	25 (12.4%)	27 (13.9%)	19 (9.3%)	71 (12%)
Total	202 (100%)	194 (100%)	204 (100%)	600 (100%)

Finally we are presenting the results of the Best Worst Scale. In the Table 4 are presented the BWS score each attribute gathered. The three attributes with the highest scores are Quality, Handmade and Enhancement, while the three attributes with the lowest scores are Trackability, Recipe and Region.

Table4: Attribute importance in Best Worst Scale

No. of Attributes	Attributes	Best	Worst	BWS Score	BWS Individual
ATT1	Quality	788	83	705	2,35
ATT6	Handmade	472	133	339	1,13
ATT9	Enhancement	345	167	178	0,59
ATT5	Tradition	264	180	84	0,28
ATT3	Price	366	324	42	0,14
ATT2	Availability	203	302	-99	-0,33
ATT10	PDO/ PGI Certificate	198	322	-24	-0,08
ATT11	Quality Certificate (ISO, HACCP)	227	427	-200	-0,67
ATT4	Region	131	411	-280	-0,93
ATT7	Recipe	149	440	-291	-0,97
ATT8	Trackability	115	510	-395	-1,32

5. CONCLUSION

First of all we observed a strong positive attitude about women's cooperatives. This is very important as the success of an entrepreneurship relies strongly on its fame. Here we have a whole market with a high reputation.

Regarding the short chains, we can suggest that it would be a very effective strategy for either women's cooperatives or other firms to start using them. We have found that consumers in each research area are highly interested in new ways of consuming, friendlier to environment and more cost effective. It is important that we didn't find differences according to sex, age, education level so we could propose that the women's cooperative or entrepreneurship that will launch first in the market a short chain, will gain a big market share.

The results we found from the questions about the innovative products can be used by women's cooperatives in order to produce products that consumers would like to purchase. Again we could mention the importance for taking the challenge to be the first who will produce those items. On the other hand, we found contradictory results regarding gift package with products. It could be a great idea for one segment of the market, and at the same time a very bad idea for another segment. Of course, a women's cooperation can take the initiative to produce such products as long as it offers a wide range of available choices.

Last but not least, the Best Worst Scale showed us the most and the least reputable attributes. The results showed us a balance between demanding consuming behavior and socially responsible consumer behavior. Consumers appreciate most quality as the most important attribute, followed by handmade production, which is possibly an indicator of authentic and less commercialized production. The third most important attribute is the enhancement of women's cooperative showing us that consumers value highly consumption choices that can be proven to help society as a whole. The least important attributes are region, recipe and trackability. This indicates that the place of origin in women's cooperatives products is not an important feature as obviously region and trackability are connected to it, but recipe can also be connected to a place of origin as well. Those findings are very crucial for the marketing or advertising campaigns that will be launched by women's cooperatives as it is crucial to demonstrate values and attributes consumers are attracted to.

REFERENCES

Aung, M., & Chang, Y. (2014). *Traceability in a food supply chain: Safety and quality perspectives. Food Control,* 39, 172-184. http://dx.doi.org/10.1016/j.foodcont.2013.11.007

Bánáti, D. (2011). *Consumer response to food scandals and scares. Trends In Food Science & Technology,* 22(2-3), 56-60. http://dx.doi.org/10.1016/j.tifs.2010.12.007

Bareja-Wawryszuk, O., & Golebiewski, J. (2014). *Economical, Environmental and Social Significance of Local Food Systems. Review Of Agricultural And Applied Economics,* 17(02), 74-77. http://dx.doi.org/10.15414/raae.2014.17.02.74-77

Bimbo, F., Bonanno, A., Viscecchia, R., & Nardone, G. (2015). *The Hidden Benefits of Short Food Supply Chains: Farmers' Markets Density and Body Mass Index in Italy. International Food and Agribusiness Management Review*, 18(1), 1-16.

Brunori, G., Galli, F., Barjolle, D., van Broekhuizen, R., Colombo, L., & Giampietro, M. et al. (2016). *Are Local Food Chains More Sustainable than Global Food Chains? Considerations for Assessment. Sustainability,* 8(5), 449. http://dx.doi.org/10.3390/su8050449

Carroll, B., & Fahy, F. (2014). *Locating the locale of local food: The importance of context, space and social relations. Renewable Agriculture And Food Systems,* 30(06), 563-576. http://dx.doi.org/10.1017/s1742170514000404

Chrysochou P. (2017) Consumer *Behavior Research Methods. In: Emilien G., Weitkunat R., Lüdicke F. (eds) Consumer Perception of Product Risks and Benefits. Springer, Cham*. https://doi.org/10.1007/978-3-319-50530-5_22

Denver, S., & Jensen, J. (2014). *Consumer preferences for organically and locally produced apples. Food Quality And Preference,* 31, 129-134. http://dx.doi.org/10.1016/j.foodqual.2013.08.014

Dorota-Rudawska, E. (2014). *Customer loyalty towards traditional products – Polish market experience. British Food Journal,* 116(11), 1710-1725. http://dx.doi.org/10.1108/bfj-10-2013-0299

Duram, L., & Oberholtzer, L. (2010). *A geographic approach to place and natural resource use in local food systems. Renewable Agriculture And Food Systems,* 25(02), 99-108. http://dx.doi.org/10.1017/s1742170510000104

Eriksen, S. (2013). *Defining local food: constructing a new taxonomy – three domains of proximity. Acta Agriculturae Scandinavica, Section B - Soil & Plant Science,* 63(sup1), 47-55. http://dx.doi.org/10.1080/09064710.2013.789123

Eriksen, S., & Sundbo, J. (2016). *Drivers and barriers to the development of local food networks in rural Denmark. European Urban And Regional Studies,* 23(4), 750-764. http://dx.doi.org/10.1177/0969776414567971

Feldmann, C., & Hamm, U. (2015). *Consumers' perceptions and preferences for local food: A review. Food Quality And Preference,* 40, 152-164. http://dx.doi.org/10.1016/j.foodqual.2014.09.014

Gidarakou, I., Xenou A., Theofilidou K. (2000). *Farm Women's New Vocational Activities: Prospects and Problems of Women's Cooperatives and Small On-Farm Businesses in Greece Journal of Rural Cooperation.* 28(1):19-37

Gracia, A. (2013). *Consumers' preferences for a local food product: a real choice experiment. Empirical Economics,* 47(1), 111-128. http://dx.doi.org/10.1007/s00181-013-0738-x

Grebitus, C., Lusk, J., & Nayga, R. (2013). *Effect of distance of transportation on willingness to pay for food. Ecological Economics,* 88, 67-75. http://dx.doi.org/10.1016/j.ecolecon.2013.01.006

Grunert, K., & Aachmann, K. (2016). *Consumer reactions to the use of EU quality labels on food products: A review of the literature. Food Control*, 59, 178-187. http://dx.doi.org/10.1016/j.foodcont.2015.05.021

He, J., & Wang, C. (2015). *Cultural identity and consumer ethnocentrism impacts on preference and purchase of domestic versus import brands: An empirical study in China. Journal Of Business Research,* 68(6), 1225-1233. http://dx.doi.org/10.1016/j.jbusres.2014.11.017

Hollis, G. and Westbury, C. (2018). *When is best-worst best? A comparison of best-worst scaling, numeric estimation, and rating scales for collection of semantic norms. Behavior Research Methods, [online] 50(1), pp.115-133. https://doi.org/10.3758/s13428-017-1009-0*

Kneafsey, M., Venn, L., Schmutz, U., Balázs, B., Trenchard, L., Eyden-Wood, T., ... & Blackett, M. (2013). *Short food supply chains and local food systems in the EU. A state of play of their socio-economic characteristics. JRC Scientific and Policy Reports. Joint Research Centre Institute for Prospective Technological Studies, European Commission.*

Koutroulou, A., & Tsourgiannis, L. (2011). *Consumers' attitudes towards GM Free products in a European Region. The case of the Prefecture of Drama–Kavala–Xanthi in Greece.* Appetite, 57(2), 448-458. http://dx.doi.org/10.1016/j.appet.2011.06.010

Koutsou, S., Iakovidou, O., Gotsinas N. (2003). *Women's Cooperatives in Greece: An On-going Story of Battles, Successes and Problems. Journal of Rural Cooperation.* 31(1): 47-57

Kühne, B., Vanhonacker, F., Gellynck, X., & Verbeke, W. (2010). *Innovation in traditional food products in Europe: Do sector innovation activities match consumers' acceptance?. Food Quality And Preference,* 21(6), 629-638. http://dx.doi.org/10.1016/j.foodqual.2010.03.013

Luceri, B., Latusi, S., & Zerbini, C. (2016). *Product versus region of origin: which wins in consumer persuasion?. British Food Journal,* 118(9), 2157-2170. http://dx.doi.org/10.1108/bfj-01-2016-0035

Madaleno, A., Eusébio, C., & Varum, C. (2017). *The promotion of local agro-food products through tourism: a segmentation analysis. Current Issues In Tourism,* 1-21. http://dx.doi.org/10.1080/13683500.2017.1296417

Ministry of Rural Development and Food. (2014). Women's Agrotourist Cooperatives . Retrieved at 10-02-2018 from, *http://www.minagric.gr/index.php/el/for-farmer-2/agrotourism/agricultural-association*

Petrolia, D., Interis, M. and Hwang, J. (2015). *Single-Choice, Repeated-Choice, and Best-Worst Elicitation Formats: Do Results Differ and by How Much?.* SSRN Electronic Journal, [online] 69(2), pp.365-393. https://doi.org/10.1007/s10640-016-0083-6

Renko, S., & Bucar, K. (2014). *Sensing nostalgia through traditional food: an insight from Croatia. British Food Journal,* 116(11), 1672-1691. http://dx.doi.org/10.1108/bfj-02-2014-0089

Renting, H., Marsden, T., & Banks, J. (2003). *Understanding Alternative Food Networks: Exploring the Role of Short Food Supply Chains in Rural Development. Environment And Planning A,* 35(3), 393-411. http://dx.doi.org/10.1068/a3510

Sandika, A. (2012). *Impact of Middlemen on Vegetable Marketing Channels in Sri Lanka. Tropical Agricultural Research And Extension,* 14(3), 58. http://dx.doi.org/10.4038/tare.v14i3.4843

Tsakiridou, E., Mattas, K., & Mpletsa, Z. (2009). *Consumers' Food Choices for Specific Quality Food Products. Journal Of Food Products Marketing,* 15(3), 200-212. http://dx.doi.org/10.1080/10454440902908217

Vanhonacker, F., Kühne, B., Gellynck, X., Guerrero, L., Hersleth, M., & Verbeke, W. (2013). *Innovations in traditional foods: Impact on perceived traditional character and consumer acceptance. Food Research International,* 54(2), 1828-1835. http://dx.doi.org/10.1016/j.foodres.2013.10.027

Zirham, M., & Palomba, R. (2016). *Female Agriculture in the Short Food Supply Chain: A New Path towards the Sustainability Empowerment. Agriculture And Agricultural Science Procedia,* 8, 372-377. http://dx.doi.org/10.1016/j.aaspro.2016.02.032

BALCOR 2018
XIII Balkan Conference on Operational Research

ANALYSIS OF RATIONALITY IN THE FOOD CONSUMPTION BY THE USING APRIORI ALGHORITM

RENATA BENDA PROKEINOVA[1], MARTINA HANOVA[2], JOHANA PALUCHOVA[3]

[1] Slovak University of Agriculture in Nitra, renata.prokeinova@uniag.sk
[2] Slovak University of Agriculture in Nitra, martina.hanova@uniag.sk
[3] Slovak University of Agriculture in Nitra, johana.paluchova@uniag.sk

Abstract: *The theory of the expected benefits produced very elegant and compelling framework in explaining of economic choices. Quite soon it became clear that he could not explain some empirical research results and economic experiments, such as St. Petersburg paradox, Allais paradox, Ellsberg paradox, but also a wide range situations of normal economic life. Psychologists and economists gathered during the 60s and 70s of the twentieth century a large body of evidence that people decide otherwise, as they are requested normative theory of expected good. In the practical part of the chapter the validity of selected examples from Behavioral Economics is tested. We decided to focus on the economic experiments and using Allais paradox in the context of consumer behavior. For the gaining data we used questionnaire with the specific questions related on the risk and uncertainty. Subsequently, collected data was statistically tested. Hypothesis that irrational people characterized by Allais paradox behave in the food consumption preferences differently compared to a rational behaving consumers. Examination of respondents in the preferences did not differ according to the data mining analytical tool – Apriori alhgoritm.*

Keywords: *Allais paradox, Association rule, consumer behavior, rationality*

1. INTRODUCTION

Every human decisions exclusively concern future events. But the future we can say with certainty only that it is uncertain. We never know how our decisions actually impact. It is logical that we try to describe the possible occurrence some future phenomena using by the probability theory. Basically, we can encounter two basic situations. If the probability distribution we know our decision is going on in terms of risk (Horská, 2013)

If the probability distribution is unknown, we make decisions under conditions of uncertainty. An example of such a decision is to predict trends in interest rates, weather and technological advances in society over the next two decades. We can only rely on past data, which, however, do not know whether they are relevant to estimating the likelihood of future events.

Decision-making in the conditions of a risk and uncertainty is the work that accompanies humanity constantly.

Regardless of the modeling approach and the underlying theory of choice and decision-making, these models have in common the assumption that decision-makers have perfect knowledge about the attributes of their choice alternatives. At the moment of choice, the values of these attributes are invariant, while individuals are assumed to hold perfect and complete knowledge about these attributes. All these models relate to choice and decision-making under conditions of certainty. Mura (2012)

Thus, the state of the consumer behavior and choosing of the foodstuff are inherently uncertain. Consequently, decision-makers always face conditions of uncertainty when choosing foodstuff on the base of price, taste, customs or recommendations. In that sense, it is surprising that applications of theories and models of decision making under conditions of uncertainty are relatively scarce in consumer behavior analysis. Moreover, the majority of studies, albeit also small in number, are concerned with uncertainty or variability in common view but do not address how individuals make decisions when facing such uncertainty and how it affects their consumer decisions. Considering the inherent uncertainty in the theory of the consumer behavior, the formulation and application of (improved) models of decision-making under conditions of uncertainty should be a field of research of high priority in behavior research. In the following contribution we would like to present analysis of the consumers, finding their rationality by using Allais paradox. On the base of results Allais paradox following we will analyze preferences of the foodstuff which are concluded and tracked in the consumer basket.

Aim of the paper is comparison in the food consumption preferences between rational and irrational consumers. Consumer are selected on the base of answers focused on Allais paradox. Comparing will be implemented through the data mining tool- association rules.

2. MATERIALS AND METHODS

As a method of research, we chose questionnaire. The survey was conducted online via Google Documents or through social networks and email but also at universities in printed version from September to December 2017.

When looking at the structure of respondents, we can say that the survey is mainly young people under the age of 34. Sometimes they live in their own homes, that is to say, because of the age they live with their parents. Hardly any young person under 24 own their own home. Household income also corresponds to the real statistics of the Statistical Office, but the sample is not homogeneous in all income groups. We are also interested in whether the respondents live in a house or apartment. This is also the source of many consumer habits. Surprisingly, 60% of respondents live in homes.

Table 1: Characteristics of the respondents

	Intuitive respondents	Logical respondents
Gender		
Male	20,00%	21,69%
Female	80,00%	78,31%
Age		
18 - 24 years	97,14%	97,87%
25 - 34 years	2,86%	2,13%
Way of living		
rent	24,29%	13,23%
own living	75,71%	86,77%
Income (netto)		
above 1200 EUR	32,86%	35,45%
1001 - 1200 EUR	14,29%	16,93%
801 - 1000 EUR	20,00%	21,16%
601 - 800 EUR	12,86%	8,47%
401 - 600 EUR	8,57%	8,47%
200 - 400 EUR	11,43%	9,52%
Kind of living		
flat	39,71%	41,94%
house	60,29%	58,06%

Source: own research

Chi-square test of goodness of fit, we verify the representativeness of the sample in terms of gender and income. The representativeness has been confirmed. The survey included 280 respondents, who were asked for a questionnaire on the topic of Consumer Behavioral Paradoxes. We sent the questionnaire to the students of our university. The aim was a homogeneous sample of young people. If the questionnaire was filled through social networks, we should have no control over the structure of respondents in terms of age.

Consequently, we came to analyze and compare respondents' opinions. Respondents were divided into two groups, based on their Allais paradox test response.

On the base of the presented goal, we need define dataset. Data was gained by questionnaire and respondents were asked by social networks. Question are consist of very simple questions focus on the preferences in consumer basket and last two part of the questionnaire identified rationality of the consumers by using Allais paradox –lottery possibilities. Allais (1979)

One of the earliest and best-known examples of systematic violation of linearity in the probabilities (or, equivalently, of the independence axiom) is the well-known Allais paradox (Allais, 1952). This problem involves obtaining the individual's preferred option from each of the following two pairs of gambles (readers who have never seen this problem may want to circle their own choices before proceeding):

a_1: {1.00 chance of \$1,000,000 versus a_2: { .10 chance of \$5,000,000; .89 chance of \$1,000,000; .01 chance of \$0 }

and

a_3: { .10 chance of \$5,000,000; .90 chance of \$0 } versus a_4: { .11 chance of \$1,000,000; .89 chance of \$0 }

Under the expected utility hypothesis, therefore, a preference for a1 in the first pair would indicate that the individual's indifference curves were relatively steep, which would imply a preference for a4 in the second pair. In the alternative case of relatively flat indifference curves, the gambles a2 and as would be preferred.

Apriori algorithm

Behavior patterns of customers using association rules and statistical methods. Association analysis is the process of discovering association rules, relationships and dependencies between attributes and their values. The analysis is performed on the incidence of these attributes and their values in the transactions. In the area of knowledge-based systems can be a recommendation using association rules and considered one of the possible methods of acquiring knowledge from a variety of data, or already known knowledge. To generate recommendations for the user may be using one of the at least three strategies. These vary in shape transactions, which are used for mining association rules, respectively, using different metrics in the final stages recommendations. To obtain rules from transaction data can be used Apriori algorithm.

Association rules: The concept of association rules was introduced in the paper of Agrawal (Agrawal, 1993,1994). From there the following definition where, L = I1, I2, Im is a set of binary attributes called items. Than T is a database transaction. Each transaction is represented as a binary vector, where t[k] = 1 if t buys item Ik and t[k] = 0 other. May X is a set of specific items L. We call it, a transaction t corresponds X if it is for all items Ik v X, t[k] = 1. Under Association rules are represented in the form of implication X=> Ij, where X is set of certain items in L and Ij is one item from L, which isn´t real in X. The rule is satisfactory in the set of transactions T with factor of trust $0 \leq c \leq 1$ if minimum c % from transactions in T, that are significant to X and Ij too. Trust is usually referred to as c.

Given the set of transactions T, we are interested in generating all rules that satisfy certain additional constraints of two different forms: Given the set of transactions T, we are interestedin generating all rules that satisfy certain additional constraints of two difierent forms:

1. Syntactic Constraints: These constraints involve restrictions on items that can appear in a rule. For example, we may be interested only in rules that have a specific item Ix appearing in the consequent, or rules that have a specific item Iy appearing in the antecedent. Combinations of the above constraints are also possible - we may request all rules that have items from some predefined itemset X appearing in the consequent, and items from some other itemset Y appearing in the antecedent.
2. Support Constraints: These constraints concern the number of transactions in T that support a rule. The support for a rule is defined to be the fraction of transactions in T that satisfy the union of items in the consequent and antecedent of the rule.

Support should not be confused with confidence. While confidence is a measure of the rule's strength, support corresponds to statistical significance. Besides statistical significance, another motivation for support constraints comes from the fact that we are usually interested only in rules with support above some minimum threshold for business reasons. If the support is not large enough, it means that the rule is not worth consideration or that it is simply less preferred (may be considered later).

In this formulation, the problem of rule mining can be decomposed into two subproblems:

1. Generate all combinations of items that have fractional transaction support above a certain threshold, called minsupport. Call those combinations large itemsets, and all other combinations that do not meet the threshold small itemsets. Syntactic constraints further constrain the admissible combinations.For example, if only rules involving an item Ix in the antecedent are of interest, then it is suficient to generate only those combinations that contain Ix.

2. For a given large itemset Y = I1,I2...Ik, k >2, generate all rules (at the most k rules) that use items from the set I1,I2...Ik,. The antecedent of each of these rules will be a subset X of Y such that X has k -1 items, and

the consequent will be the item Y - X. To generate a rule X → Ij │ c, where X = I1,I2...Ij+1,.Ij-1....Ik, take the support of Y and divide it by the support of X. If the ratio is greater than c then the rule is satised with the confidence factor c; otherwise it is not. Note that if the itemset Y is large, then every subset of Y will also be large, and we must have available their support counts as the result of the solution of the firrst subproblem. Also, all rules derived from Y must satisfy the support constraint because Y satisfies the support constraint and Y is the union of items in the consequent and antecedent of every such rule. Having determined the large itemsets,the solution to the second subproblem is rather straightforward. In the next section, we focus on the first subproblem. We develop analgorithm that generates all subsets of a given set of items that satisfy transactional support requirement.To do this task eficiently,we use some estimation tools and some pruning techniques.

Promotion rule is defined as the percentage of transactions in L, which contents XIj. It denotes as s. Support essentially represents the frequency of occurrence of a given set of items in the database. Support and confidence are measures (metrics) for association rules.

Trust (confidence) is the probability of the right hand side rule condition occurrence left side. It is therefore the percentage of rules whose left side is X and Y right of all whose left side is X.

Lift (interest): This rate determines how many times more often X and Y occur together than would be if they were statistically independent. In contrast to expectations is dependent on rules of thumb. The formula for calculating metrics lift:

$$\text{lift}\ (X \rightarrow Y) = \text{lift}\ (Y \rightarrow X) = \frac{p\ (x \text{ and } Y)}{p\ (X)\ p\ (Y)} = \frac{\text{trust}\ (X \rightarrow Y)}{\text{support}\ (Y)} = \frac{\text{trust}\ (Y \rightarrow X)}{\text{support}\ (X)} \qquad (1)$$

Apriori: To find frequently occurring sets of items can be used Apriori algorithm, which is stated in the paper Agrawal. Apriori sequentially generates sets of frequent items, the proceeds from the smallest (with the fewest elements) to largest. As far as possible, from the frequent sets with n elements generating sets with n +1 element. Set of frequent sets having n elements is called Ln. The procedures recommendation using association rules from the said general scheme differs in that instead of Neighbourhood Formation is the algorithm used data mining association rules. Its outputs are the rules containing some items on the left and right sides. In the third phase is recommended for all items that are listed in the consequences (on the right) obtained rules. Therefore it is possible to take a limited number of items (N best), or any that meet certain criteria, such as where a degree exceeds a defined threshold.

3. RESULTS AND DISCUSSION

Food consumption refers to the quantity and quality of food intake by households or individual family members. Though often measured in terms of food expenditures, it is conceptually closer to “food intake” as measured by calories or broken down into different nutrients (Migotto et al., 2005). Individual level consumption of food is an important indicator to examine the relationship between intra-household resource allocation and household food security. Reliable data concerning food consumption of individuals are needed for various reasons.

Association analysis of the output is sorted by the values Lift, therefore, in determining the rate of how many times more often X and Y occur together than would be if they were statistically independent.

A lift value greater than 1 indicates that X and Y appear more often together than expected; this means that the occurrence of X has a positive effect on the occurrence of Y or that X is positively correlated with Y.

A lift smaller than 1 indicates that X and Y appear less often together than expected, this means that the occurrence of X has a negative effect on the occurrence of Y or that X is negatively correlated with Y.

A lift value near 1 indicates that X and Y appear almost as often together as expected; this means that the occurrence of X has almost no effect on the occurrence of Y or that X and Y have Zero Correlation. Thus, lift is a value between 0 and infinity

Slovakia continues a long-term trend of high consumption of animal fats and proteins. Slovaks generally do not accept the recommended amounts of carbohydrates and crude fiber. As we can see in a detailed chart of consumption of butter and pork ointment, it is a long-term decreasing trend in the consumption of these vital fats.

The biggest deficiency in our diet is only about half greater intake of fat than recommended. Annual consumption in the SR is 23 kg. Although in previous years decreased bad fat structure is highlighted.

Consumption of pork ointment decreases only slightly, but we still consume it twice as much as the recommended consumption, but the consumption of butter does not reach even the recommended average values.

Table 2: Consumption in the fats and oils - rules

1. olive oil=consume rapeseed oil=not consume 157 ==> margarin=not consume Allais=rational 70 conf:(0.45) < lift:(1.18)> lev:(0.04) [10] conv:(1.11)
2. olive oil=consume rapeseed oil=not consume Allais= irrational 78 ==> margarin=not consume 70 conf:(0.9) < lift:(1.11)> lev:(0.03) [7] conv:(1.67)
3. olive oil=consume rapeseed oil=not consume Allais=irrational 78 ==> butter=consume margarin=not consume 65 conf:(0.83) < lift:(1.11)> lev:(0.02) [6] conv:(1.38)

The highest lift value of 1.18 has a 4-element rule. Support for this rule means that respondents who prefer olive oil consumption and do not favor rape oil consumption do not consume margarine at the same time and, according to the Allais test, are rational consumers. is in the file 85%. Confidence (CONF) for this rule is 45% and expresses the likelihood that if the consumer consumes olive oil and does not consume rape oil, it certainly does not favor margarine.

The other rule has a lift value of 1.11. The confidence value is 0.9 and expresses the likelihood that respondents consume olive oil, do not consume rape oil and are irrational and do not consume margarine.

In the posted rule, the lift value is greater than 1 (1.11). Confidence is 0.83 and points to irrational respondents who consume olive oil, do not consume rapeseed oil and consume butter and do not consume margarine. From the above results, irrational and rational respondents prefer similar foods and therefore there is no apparent difference in consumption of oils and fats in terms of rational and irrational behaviour

Over the past 20 years, the Slovaks' dining menu has changed. Compared to 1993, the average Slovaks are almost half of less potatoes. On the contrary, the most significant increases in consumption were attributed to the other two decades of pasta.

Slovaks eat pasta up to 70% more than 20 years ago. It has gained popularity in our latitudes, so-called Italian cuisine.

Annual consumption per person increased from 4.6 kilograms in 1993 to 7.8 kilograms in 2013.

Table 3: Consumption pattern in the side dish - rules

1. dumpling=consume 190 ==> rice=consume potatoes=consume pasta=consume health_side dish=not consume 121 conf:(0.64) < lift:(1.13)> lev:(0.05) [13] conv:(1.18)

In this case, the rule repeated several times. That's why we did not mention other forms. The lift of this rule is 1.13. From the point of view of rationality or irrationality, no relevant rule emerged. All we have found is that with probability 64% of respondents consume dumplings, rice, and potatoes and paste while not consuming healthy supplements. Healthy attachments were bulgur, couscous and others.

Table 4: Consumption pattern in the meat - rules

1. pork=consume wild_meat=not consume 145 ==> poultry=consume rabbit=not consume 120 conf:(0.83) < lift:(1.17)> lev:(0.07) [17] conv:(1.64)
15. poultry=consume pork=consume wild meat=not consume 145 ==> rabbit=not consume 120 conf:(0.83) < lift:(1.13)> lev:(0.05) [14] conv:(1.51)

Over the past period, the total consumption of meat per bones per capita stopped at 55.8 kilograms. We have reached the recommended meat consumption band, but the structure of consumption of individual types of meat is unfavorable.

Critically low is the consumption of beef and veal. Of the recommended dose of 17.4 kg per year, one inhabitant averages barely a quarter. With an annual thirty-kilogram consumption per pig head, Slovakia is 38% above the recommended healthy consumption.

In terms of meat consumption, it is worth pointing out the wrong structure of the individual meats consumed. The fact that the consumption of higher value added foods is reduced to the detriment of lower-value foods which are less suitable for the consumer's nutritional point of view.

Even in the case of meat, the rules on the rationality and irrationality of respondents did not appear in the rules. Obviously, both respondents responded similarly. The first rule reached the lift value of 1.17. With 83% probability, respondents consume pigs, poultry and do not consume wild meat and rabbit.

We can say that respondents prefer all available meats, except for the divine and the rabbit. Divina and rabbit is considered nutritionally valuable and dietary meat, yet people prefer either a cheaper alternative - poultry, or pork or beef.

4. CONCLUSION

The Allais paradox has not escaped counterarguments. When it was first proposed by Allais, many adherents of expected utility argued that it was simply a failure to understand the nature of the choice being made. The Independence Axiom that underlies expected utility theory was seen as a simple “Pareto improvement” argument: If one of the lotteries in a compound lottery is replaced by a better one, the new compound lottery should rank higher in the person’s preferences. The claim was that once this was properly explained, behaviour would conform to expected utility theory. However, this was found not to be the case, and it is now broadly accepted that behaviour conforming to the Allais paradox is perfectly consistent with the basic axioms of rationality, namely completeness and transitivity.

The next line of criticism says that paradoxical behaviour occurs only for extremes of probabilities. Many more people are likely to show behaviour conforming to expected utility theory. However, this counterargument does not quite rescue expected utility theory. For one thing, many actual choices do involve extreme probabilities; for example default risks on many bonds are quite small. And more basically, the counterargument only reinforces the basic Kahneman-Tversky argument that people’s perceptions are importantly affected by their reference point. The general idea that equal changes in probabilities are evaluated differently starting from different initial probabilities entails an overall utility function on random prospects that is not linear in probabilities, concludes Kahneman (2003)

In this article we have presented a example of the demonstration of analyze of the consumption behavior. The present work has been concerned primarily with the descriptive question focus on the finding of the consumption preferences, as well questionnaire consist of the psychology of choice is also relevant to the normative question which identify rationality of the respondents. In order to avoid the difficult problem of justifying values, the modern theory of rational choice has adopted the coherence of specific preferences as the sole criterion of rationality. This approach enjoins the decision-maker to resolve inconsistencies but offers no guidance on how to do so.

Applied to economics, rational choice theory is presumed to be ethically neutral, because it “does not question people’s preferences; it simply studies how they seek to maximize them.” However, McCumber (2011) argues that rational choice theory is not ethically neutral, because its parent philosophy is not ethically neutral. “Whatever my preferences are, I have a better chance of realizing them if I possess wealth and power.

Individual rationality is limited by their ability to conduct analysis and think through competing alternatives. The more complex a decision, the greater the limits are to making completely rational choice.

One of the principles of economics is that people respond to incentives. Rational choice theory attempts to provide an explanation why. Therefore, a criticism of the rational choice theory consists of either providing a better explanation of why people respond to incentives or of showing that people do not respond (only) to incentives. A legitimate objection to the argument of this paper is that it ignores uses of the rational choice theory outside the analysis of “generalized demand” such as in game theory. To what extent is the argument of this article applicable to these uses, remains an open question. But the general principle still holds: any evaluation of the theory needs to take into account what the theory professes to explain.

Acknowledgement

Supported by the Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences VEGA (Project No. 1/0890/16 – “Quantification of the sustainable consumption by using modelling of consumer behaviour in the food consumption”.

REFERENCES

Acrey, J. C., McCumber, W. R., & Nguyen, T. H. T. (2011). CEO incentives and bank risk. *Journal of Economics and Business*, *63*(5), 456-471.

Agrawal R., Imielinski T., Swami A. (1993). *Mining Association Rules between Sets of Items in Large Databases.* Proceedings of ACM SIGMOD-93, 1993. p. 207-216. [cit. 2014-06-28].

Agrawal R., Srikant R. (1994). *Fast algorithms for mining association rules.* Proceedings of the 20th VLDB Conference, (1994). p. 487-499, Santiago, Chile [cit. 2014-06-28].

Allais, M. (1979). The foundations of a positive theory of choice involving risk and a criticism of the postulates and axioms of the American school. Pp. 27-145 in M. Allais, editor; and O. Hagen, editor. , eds. *Expected Utility Hypotheses and the Allais Pardox*. Dordrecht, Holland: D. Reidel Publishing Company

Horská, E., Yespolov, T.I. (2013). *Sustainability in Business and Society: A Theoretical and Methodological Overview*. p. 9-23. In Horská, E., Yespolov, T.I. 2013. *Sustainability: in Business and Society: Global Challenges - Local Solutions.* Kraków: Wydawnictwo Episteme, 2013. p. 166. ISBN 978-83-7759-015-7.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. *American psychologist*, *58*(9), 697.

Kozelová, D., Fikselová, M., Mura, L., Mendelová, A., Vietoris, V. (2012). *Analysis of Consumer Preferences focusing on food Additives.* In: Acta Universitatis Agriculturae et Silviculturae Mendeleianae Brunensis. MZLU. Brno: Mendelova univerzita v Brně, No. 6/2012, p. 197-203. ISSN 1211–8516.

Lušňáková, Z., Kleinová, K. (2012). *The place of corporate social responsibility and its activities in the retail firm management.* In: Polityki europejskie. Vol. 56, no. 7 (2012), p. 53-61. ISSN 2081-3430.

Migotto, M., Davis, B., Carletto, C., & Beegle, K. (2007). Measuring food security using respondents' perception of food consumption adequacy. *Food Security: Indicators, Measurement, and the Impact of Trade Openness. B. Guha-Khasnobis, SS Archaya and B. Davis, Eds. UNU-WIDER and ICSSR*.

B16
Operations Management

BALCOR 2018
XIII Balkan Conference on Operational Research

A BRIEF LITERATURE ANALYSIS ABOUT USAGE OF THE FUZZY SETS AND MARKOV CHAIN ON MAINTENANCE PROCESSES

MELIKE ERDOGAN[1], IHSAN KAYA[2]

[1] Yıldız Technical University, Industrial Engineering Department, melike@yildiz.edu.tr
[2] Yıldız Technical University, Industrial Engineering Department, ihkaya@yildiz.edu.tr

Abstract: *In maintenance studies, the two most common approaches can be described as Markov Chain implementations and fuzzy logic. As these methods are used separately, it is possible to encounter hybrid versions of these approaches in the literature. In this paper, we have examined academic studies that have included Markov Chain and Fuzzy Logic methodologies in the area of maintenance. We also present a literature analysis about usage of the fuzzy sets and Markov Chain into maintenance applications. We aimed to guide by our literature study to those who want to work in this area by evaluating it in many different aspects such as publication year, methods, publication titles, and application area.*

Keywords: *Fuzzy Set Theory, Maintenance, Markov Chain, Review*

1. INTRODUCTION

Today, a significant part of the operating costs constitute the cost items spent for the continuity of the process. The most important part of these items belong to maintenance activities. Therefore, one of the most important problems affecting the cost of the enterprises is the establishment of the optimum maintenance model (Arı and Unal, 2003). Maintenance plays an important role by reducing the downtime and rework time to increase system reliability (Liao and Li, 2015). Traditional reliability analysis is an unsuitable method for complex systems without sufficient probabilistic knowledge (Kumar and Kumar, 2011). In fact, reliability data is either inadequate or have uncertainty (Ge and Asgarpoor, 2008). The concept of the fuzzy sets to overcome this problem has been successfully used to evaluate the reliability of a system (Kumar and Kumar, 2011). The fuzzy sets (FS) are often used in maintenance problems where the data is ambiguous. Markov chain is also one of the frequently used methods in maintenance studies such as determine the best maintenance policy for machines, aircrafts or pavements etc.

Fuzzy logic and Markov Chain methods are the most common approaches that we encounter in maintenance papers. The revealing of how much these methods play a part in the literature together, as well as the separately, will lead to the researchers working on this field. It is also often the case that usage of these two methods in the reliability calculations are highly applied. For this reason, we have analyzed the usage of these methods in the maintenance and have made a literature analysis about it. So we aimed to being a road map for future research on this area.

For this aim, the rest of this paper has been organized as follows: Section 2 briefly mentions about fuzzy logic. Section 3 gives summary information about Markov Chain. Section 4 presents a literature analysis about usage of the fuzzy sets and Markov Chain on maintenance processes. Finally, the obtained results and future suggestions have been discusses into Section 5.

2. FUZZY LOGIC

Uncertainty is often thought of as a random variable. However, it is also important to mathematically define another uncertainty derived from human cognitive behavior, utility, and subjectivity. In this case, the fuzzy logic is one of the most useful theories to deal with uncertainty (Hasuike, and Katagiri, 2016, August). Over the last decade, fuzzy systems have gained considerable acceptance in many areas such as control and automation, image recognition, medical diagnosis and prediction. This is mainly due to the need to consider a number of qualitative and quantitative variables that affect a value (Karimov, 2010). Fuzzy logic is used fuzzy sets, membership functions and fuzzy numbers to convert the uncertain data to the available information. The first time was introduced by Zadeh fuzzy logic in 1965 (Zadeh, 1965). Fuzzy sets define the data with the low, medium, and maximum limit values that reflect the uncertainty rather than defining them strictly (Lee and Chen, 2008). Shortly, fuzzy sets can be explained as follows:

If X is a collection of elements denoted by a, then a fuzzy set $\tilde{A}$ in X is a set of ordered pairs (Zadeh, 1972):

$$\tilde{A} = \{(a, \mu\_A\ (a) \mid a \epsilon X)\} \quad (1)$$

where $\tilde{A}$ in X satisfies the following conditions:

$\tilde{A}$ is normal,

$\tilde{A}$ is a closed interval for every a ∈ [0, 1],

The support of $\tilde{A}$ must be bounded.

$\mu\_\tilde{A}$ (a) is entitled as the membership function of element a which maps to X.

Fuzzy logic can be used in many different ways in the field of maintenance. For example, fuzzy analytic network process is applied for maintenance policy selection for an industrial unit by Kumar and Maiti (2012). Hennequin et al. (2009) presented an approach for the optimization of imperfect preventive maintenance and corrective actions performed on a single machine based on fuzzy logic. Fuzzy logic can also be applied for decision-making of maintenance to enhance production decisions (Lu and Sy, 2009). Shortly, fuzzy logic approach is referenced in the maintenance area can appear in many studies. By further customizing the research, we focus on the papers that uses the fuzzy logic and markov chain together in the maintenance area.

3. MARKOV CHAIN

Markov chain is known for its performance in modeling a randomly changing system, assuming that future system states depend only on the current state and nothing with previous states system (Liao and Li, 2015).
Consider the stochastic process $\{X_n, n = 0, 1, 2, ...\}$, which can take the final or countable possible values. Unless otherwise specified, the process is expressed as {0, 1, 2, ...} with a non-negative integer set of possible outcomes. If $X_n = i$, the process is called i at time n. Whenever the process is in state i, it is always assumed that the probability of the next state with probability p_{ij} is j. That is, for all $i_0, i_1, ..., i_{n-1}$, i, j states and for $n \geq 0$:

$$P\{X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots X_1 = i_1, X_0 = i_0\} = P_{ij}$$

Markov chain applications are often meet where the relevant data are not fully known. In most cases the data required in Markov decision processes must be estimated through different measures and naturally includes the uncertainty of the observed system (Kurano et al., 2006). Markov chain is also an approach used in many areas of work in the field of maintenance. For example, it is applied for maintenance planning considering repair time and periodic inspection (Lee et al., 2013), optimization on maintenance of wind turbines (Wu and Zhao, 2010), maintenance strategy optimization (Zhao et al., 2011), road maintenance optimization (Zhang and Gao, 2012) or aircraft maintenance (El Afia and Aoun, 2017).

Sometimes there may be situations where the exact data is not known in Markov Chain applications. Fuzzy logic can be applied to handle with this kind of uncertainty (Kurano et al., 2006). Fuzzy-Markov chain makes simpler the calculation process and achieves accurate prediction results with imprecise and incomplete information (Liao and Li, 2015). Fuzzy Markov is an effective tool for calculating reliability with uncertain information where failure rate, recovery rate and maintenance rate are sometimes best expressed at uncertain value (Binh and Khoa, 2006). In the light of this information, we try to review the Markov chain applications that are used with fuzzy logic in our paper for maintenance studies.

4. A LITERATURE ANALYSIS ON USAGE OF FUZZY LOGIC AND MARKOV CHAIN IN MAINTENANCE

In this paper, we examined the papers that include Markov chains and fuzzy logic approaches in the field of maintenance. For this aim the Scopus database has been used to search the related papers with keywords "maintenance", "Markov chain" and "fuzzy logic". As a result of this search, twenty-three studies conducted with these approaches in the field of maintenance have been examined. The papers that we encountered from this research are listed in Table 1. The methods fuzzy sets and Markov chain have been classified by independent or together methods named as “Together” or “Separately”.

Table 1: A classification for the papers

#	Authors	Methods (Together or Separately)	Application Area	Paper Type	Year	Publication Tittle
1	Kurano et al.	Together	Machine maintenance	Article	2006	Fuzzy Sets and Systems
2	Cannarile et al.	Separately	Nuclear Power Plant Maintenance	Article	2017	Annals of Nuclear Energy
3	Zeng et al.	Separately		Article	2017	Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability
4	Wang et al.	Together	Offshore Power Systems maintenance	Conference Paper	2009	Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference
5	Liao and Lee	Together	Aero-engine maintenance	Conference Paper	2015	2015 IEEE International Conference
6	Verma et al.	Separately	Machinery Health Monitoring	Conference Paper	2012	Fuzzy Information Processing Society (NAFIPS), 2012 Annual Meeting of the North American
7	Binh and Khoa	Together	Power Systems maintenance	Conference Paper	2006	Transmission & Distribution Conference and Exposition: Latin America
8	Bandara and Gunaratne	Together	Pavement maintenance	Article	2011	Journal of Transportation Engineering
9	Sal et al.	Together	Safety instrumented system	Article	2017	International Journal of System Assurance Engineering and Management
10	Min et al.	Together	System availability analysis	Conference Paper	2009	Intelligent Computation Technology and Automation, 2009. ICICTA'09
11	Kumar and Kumar	Separately	Availability modeling for biscuit manufacturing plant	Article	2011	International Journal of System Assurance Engineering and Management
12	Mohanta et al.	Together	Maintenance Scheduling	Conference Paper	2005	IEEE Transactions on Power Systems
13	Chen et al.	Separately	Machine prognosis	Conference Paper	2011	IEEE Transactions on Industrial Electronics
14	Ge and Asgarpoor	Together	Power system reliability	Conference Paper	2008	Probabilistic Methods Applied to Power Systems PMAPS'08
15	Kageyama	Together	Machine maintenance	Article	2008	Journal of Interdisciplinary Mathematics
16	Liu and Huang	Together	Non-repairable power generator	Article	2011	Multiple-Valued Logic and Soft Computing
17	Kurano et al.	Together	Machine maintenance	Article	2005	Lecture notes in computer science
18	Xing et al.	Together	Parameter Prediction for Weld Hidden Damage Status	Article	2017	Jixie Gongcheng Xuebao/Journal of Mechanical Engineering
19	Zhou et al.	Separately	Multi-state manufacturing systems	Article	2014	Jisuanji Jicheng Zhizao Xitong/computer Integrated Manufacturing Systems Cims
20	Ge and Asgarpoor	Together	Power Systems maintenance	Article	2010	IEEE Transactions on Power Systems
21	Tanrioven and Alam	Separately	PEM fuel cell power plants	Article	2006	Renewable Energy
22	Zhang et al.	Separately	Risk of electric power transformers	Article	2016	Power System Protection and Control
23	Peng et al.	Separately	Machine maintenance	Conference Paper	2011	IIE Annual Conference. Proceedings (p. 1). Institute of Industrial and Systems Engineers

While the table is being prepared, the article and the conference papers have been divided, the years of the studies and the authors have been specified and area which also papers are made is revealed. In some of these studies, fuzzy logic and Markov chain approaches are considered together, in which case they are regarded as separate approaches. Figure 1 shows the number of papers according to the years based on published.

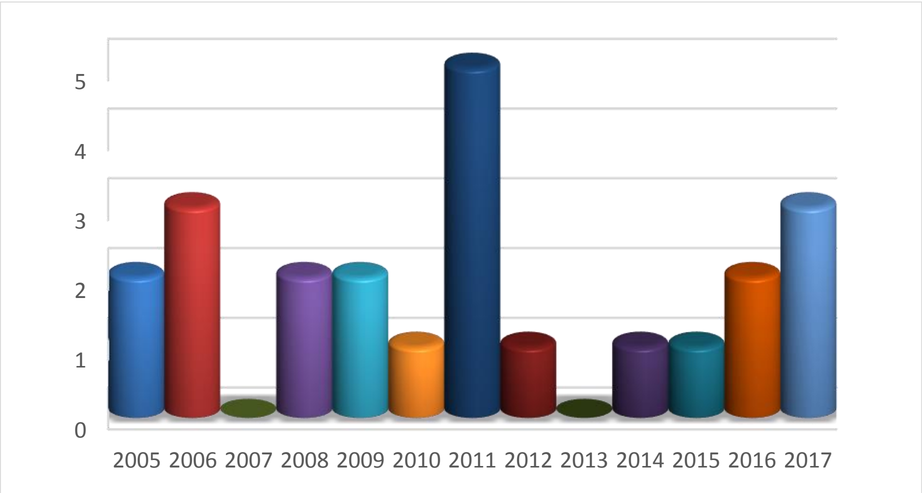


Figure 1. Number of the papers based on years

When we look at the Figure 1, we can see that most of the studies related to these areas have been published in 2011. In 2007 and 2013, we see that no work has been published that adopts these two approaches in maintenance area together or separately.

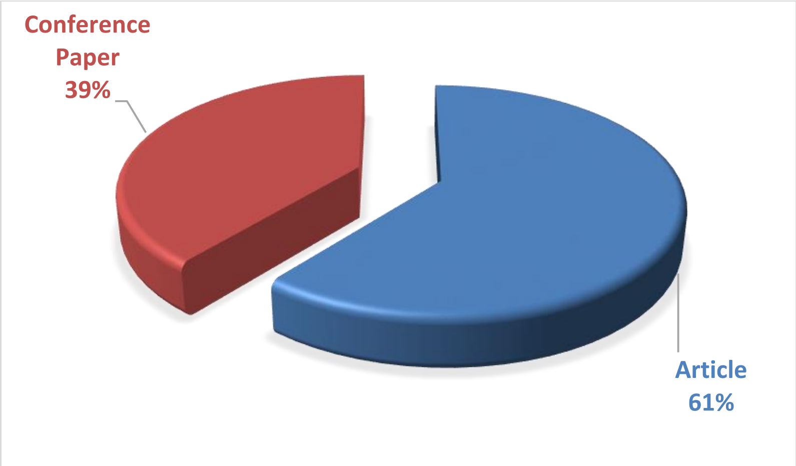


Figure 2: Type of papers

When we examine Figure 2, it is observed that the papers performed are generally published in the journals as articles with the percentages of 61%. By the way, 39% of the papers are as conference papers.

4.1. Usage of fuzzy logic and Markov chain as separate methods in maintenance

In this sub-section, the papers that use the fuzzy logic and Markov chain methods in the maintenance area are briefly summarized as follows. Tanrioven and Alam (2006) proposed a methodology for modeling and calculating Proton Exchange Membrane Fuel Cell Power Plants reliability with estimating failure and repair rates based on the fuzzy set theory and expert knowledge. Chen et al. (2011) investigated the trained adaptive neuro–fuzzy inference systems and its modelling noise constitute an mth-order hidden Markov model to describe the fault propagation process. Kumar and Kumar (2011) aimed to capture the effect of coverage factor, failure and repair on its fuzzy availability from a biscuit manufacturing plant consisting of six sub-systems. Peng et al. (2011) proposed a reliability estimation method based on production quantity which used a Markov process to identify the reliability degradation of the production machines, and integrates fuzzy theory to generate the estimation of machine reliability. Verma et al. (2012) offered a simple fuzzy logic based method for quantifing features of Machinery Health Monitoring systems such as detectability and prognostic ability and incorporated in a multi-objective maintenance optimisation model based on Markov process and genetic algorithm. Zhou et al. (2014) developed a multi-state reliability modeling method in a maintenance cycle based on the performance degradation by using a Markov model of machine performance degradation combined with fuzzy analytic hierarchy process for defining the relationship between the subsystems and system. Zhang et al. (2016) proposed a full condition model of transformer based on condition-based maintenance and Markov Process for building an effective assessment method for the risk of electric power transformers. They also used an entropy-weighted fuzzy method to quantify the serious degree. Zeng et al., (2016) used fuzzy similarity, feed-forward neural network and hidden semi-Markov model to assess the prediction capability of prognostic methods. Cannarile et al. (2017) offered a model based on the fuzzy expectation-maximization algorithm, which integrates the evidence of the field inspection outcomes with information taken from the maintenance operators about the transition times from one state to another.

4.2. Usage of fuzzy logic and Markov chain together in in maintenance

In this sub-section, the studies that are used together in the field of maintenance by hybridizing the applications of fuzzy logic and Markov chain are summarized. First paper that we met is the conference paper of Mohanta et al. (2005). They proposed a fuzzy Markov model to efficiently incorporate maintenance scheduling effects as well as age of production units on the failure-repair cycle. The proposed model was different from the classical models based on probability; the probabilistic Markov approach was used in combination with the fuzzy logic. In this model, mean time to failure and mean time to repair are calculated by using blurred numbers to transfer the uncertainty arising from expert assessments to the best way. In second paper, Kurano et al., (2005) benefited from fuzzy logic to determine benefited from the transition probabilities of the Markov chain. They tried to evaluate the optimal expected reward, which is named as fuzzy perceptive value, with a numerical example on a machine maintenance problem. Binh ve Khoa, (2006) developed a model for building reliability curve from the initial state using alpha-cut technique and defuzzification with fuzzy Markov chain. They discussed the usefulness of the proposed method calculating with large state space by using transition probability and initial values. Kurano et al. (2006) defined uncertain transition matrices of Markov decision processes which allow for fluctuating transition matrices at each step in time by the use of fuzzy sets. The also determined a Pareto optimal policy maximizing the infinite horizon fuzzy expected discounted reward over all stationary policies under some partial order relation. Ge and Asgarpoor (2008) extended maintenance models with fuzzy transition parameters in Markov Chain Decision Processes and compared results with traditional non-fuzzy method. Kageyama, (2008) presented the optimality gap between perfectly precise information and fuzzy information for uncertain Markov Decision Processes. He used a fuzzification operator $\mathcal{L}_{\sigma}$ with the deviation parameter σ ($0 \leq \sigma \leq \infty$). Min et al. (2009) proposed a fuzzy availability model for systems based on Markov model and affine arithmetic, and give a solution method to this model by recursive network. Wang et al. (2009) presented a type-2 fuzzy hidden Markov model to analyze the reliability indices of the offshore power system. Ge and Asgarpoor (2010) developed an approach to improve a fuzzy Markov model which incorporates parameter uncertainty and probability in aging equipment models and existing reliability models.

Bandara et al. (2011) proposed a mode for a future pavement condition prediction method based on the current evaluations and subjective probabilistic estimates of condition transition. Liu and Huang (2011) presented a fuzzy multi-state element model for overcoming the deficiencies of the conventional multi-state element theory. They determined fuzzy mean time between replacement and fuzzy performance reward via the proposed fuzzy Markov reward model, and the expected fuzzy average profit per unit time is also calculated. They used the parametric programming algorithm for the membership functions of the quantities of interest. Liao and Li (2015) developed a practical prognostics tool is given to better realize effective condition-based maintenance with fuzzy-Markov chain which simplifies the calculation process and achieves accurate prediction results with small sample and incomplete information. Sal et al. (2017) considered epistemic uncertainty sources encountered in risk assessment which results from lack of knowledge. They used a model based on fuzzy numbers for dealing with parameter uncertainty in Markov chain. Xing et al., (2017) proposed a quantitative metal magnetic memory prediction model which based on unbiased gray theory and fuzzy weighted Markov theory in order to solve the difficulty in quantitatively predicting the evolution of hidden damage for welded joints by applying metal magnetic memory technology. All of the papers dealt with in fuzzy logic and Markov chain in the field of maintenance are summarized in this sub-section.

4.3. Discussion

It is quite reasonable to use two methods that we mention together in the maintenance area. Because when the two are used as separately, there are quite a few numbers in the literature. While we examine the fuzzy logic and Markov chain methodologies in the field of maintenance, we see that the two schemes together do not draw enough attention, even though these are very powerful tools in that area. It can be easily seen that there is gap on the relevant literature in this hybridization. With the extension of the usual fuzzy logic applications (such as hesitant fuzzy sets, etc.), it is possible to make prominent papers in this area, and also new approaches can be developed.

5. CONCLUSION AND FUTURE DIRECTIONS

In this paper, we analyzed the papers that use fuzzy logic and Markov chain in the area of maintenance. First of all, the papers which use these two approaches in a same paper have been investigated. We have gathered papers on where to publish the conference or journals and briefly summarized. Then we have reviewed the publications on which these two methods are used together. We have again summarized these papers in the same way. Finally, we made a table for the papers that we reviewed which aimed to make it easy for the readers to search. It is observed that the papers performed are generally published in the journals as articles with the percentages of 61%. We recognized that most of the studies related to these areas have been published in 2011. In 2007 and 2013, we see that no work has been published that adopts these two approaches in maintenance area together or separately. We believe that this paper will lead to the researchers who want to work on this area.

REFERENCES

Bandara, N., & Gunaratne, M. (2001). Current and future pavement maintenance prioritization based on rapid visual condition evaluation. *Journal of Transportation Engineering*, 127(2), 116-123.

Cannarile, F., Compare, M., Rossi, E., & Zio, E. (2017). A fuzzy expectation maximization based method for estimating the parameters of a multi-state degradation model from imprecise maintenance outcomes. *Annals of Nuclear Energy*, 110, 739-752.

Chen, C., Zhang, B., Vachtsevanos, G., & Orchard, M. (2011). Machine condition prediction based on adaptive neuro–fuzzy and high-order particle filtering. *IEEE Transactions on Industrial Electronics*, 58(9), 4353-4364.

Ge, H., & Asgarpoor, S. (2010). Reliability evaluation of equipment and substations with fuzzy Markov processes. *IEEE Transactions on Power Systems*, 25(3), 1319-1328.

Kageyama, M. (2008). On optimality gaps for fuzzification in finite Markov decision processes. *Journal of Interdisciplinary Mathematics*, 11(1), 77-88.

Kumar, K., & Kumar, P. (2011). Fuzzy availability modeling and analysis of biscuit manufacturing plant: a case study. *International Journal of System Assurance Engineering and Management*, 2(3), 193-204.

Kurano, M., Yasuda, M., Nakagami, J. I., & Yoshida, Y. (2005). Perceptive evaluation for the optimal discounted reward in Markov decision processes. *Lecture notes in computer science*, 3558, 283.

Kurano, M., Yasuda, M., Nakagami, J. I., & Yoshida, Y. (2006). A fuzzy approach to Markov decision processes with uncertain transition probabilities. *Fuzzy Sets and Systems*, 157(19), 2674-2682.

Liu, Y., & Huang, H. Z. (2011). Optimal Replacement Policy for Fuzzy Multi-State Element. *Multiple-Valued Logic and Soft Computing*, 17(1), 69-92.

Mohanta, D. K., Sadhu, P. K., & Chakrabarti, R. (2005). Fuzzy Markov model for determination of fuzzy state probabilities of generating units including the effect of maintenance scheduling. *IEEE Transactions on Power Systems*, 20(4), 2117-2124.

Sal, R., Nait-Said, R., & Bourareche, M. (2017). Dealing with uncertainty in effect analysis of test strategies on safety instrumented system performance. *International Journal of System Assurance Engineering and Management*, 8(2), 1945-1958.

Tanrioven, M., & Alam, M. S. (2006). Reliability modeling and analysis of stand-alone PEM fuel cell power plants. *Renewable Energy*, 31(7), 915-933.

Xing, H., Sun, X., Wang, B., Ge, H., Dang, Y. & Yu, Z. (2017). Quantitative MMM Characteristic Parameter Prediction for Weld Hidden Damage Status Based on the Fuzzy Weighted Markov Chain, *Jixie Gongcheng Xuebao/Journal of Mechanical Engineering* 53(12), 70-77.

Zeng, Z., Di Maio, F., Zio, E., & Kang, R. (2017). A hierarchical decision-making framework for the assessment of the prediction capability of prognostic methods. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 231(1), 36-52.

Zhang, H., Bengang, W. E. I., & Kejun, L. I. (2016). Research on risk assessment technology based on Markov state evaluation model for power transformer and entropyweighted fuzzy comprehensive evaluation. *Power System Protection and Control*, 44(5), 134-140.

Zhou, F., Li, A., Xie, N., Xu, L. Y., & Shao, H. (2014). Reliability analysis of multi-state manufacturing systems based on performance degradation. *Jisuanji Jicheng Zhizao Xitong/computer Integrated Manufacturing Systems Cims*, 20(6), 1424-1431.

Karimov, A., Offıce Rent Varıatıon In Istanbul Cbd: An Applıcatıon Of Mamdanı And Tsk-Type Fuzzy Rule Based System, The Degree Of Master Of Scıence In Fınancıal Mathematıcs, Mıddle East Technıcal Unıversıty 2010.

Ari ,Y. & Unal R. (2003). Güvenilirlik Merkezli Bakim (Rcm) Yöntemi Ile Asansör Bakimi. Bakım Teknolojileri Kongresi ve Sergisi, 16-19 Ekim 2003-Denizli, Turkey.

Tosun, Ö. (2007), Tamir – Bakım Planlaması Ve Bir Uygulama, Akdeniz Üniversitesi, Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Antalya.

Aytemiz T., Şengönül A., "Markov Zincirlerinin Ekonomik Bir Probleme Uygulanması: Perakende Alışverişlerde Bireysel Olarak Kullanılan Madeni Para Stratejilerinin Karşılaştırmalı Analizi", Dokuz Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Cilt 6, Sayı 4, (2004), 29-43.

Binh, P. T. T., & Khoa, T. Q. D. (2006, August). Application of fuzzy markov in calculating reliability of power systems. In *Transmission & Distribution Conference and Exposition: Latin America, 2006. TDC'06. IEEE/PES* (pp. 1-4). IEEE.

Ge, H., & Asgarpoor, S. (2008, May). Markov processes with fuzzy parameters-A case study. In *Probabilistic Methods Applied to Power Systems, 2008. PMAPS'08. Proceedings of the 10th International Conference on* (pp. 1-6). IEEE.

Min, Y., Zhengyou, H., & Qingquan, Q. (2009, October). Fuzzy Availability Based on Affine Arithmetic and Markov Process. *In Intelligent Computation Technology and Automation, 2009. ICICTA'09. Second International Conference on* (Vol. 2, pp. 704-707). IEEE.

Wang, Z. X., Chang, C. S., Yang, F., & Tan, W. W. (2009, October). Adaptive type-2 fuzzy maintenance advisor for offshore power systems. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on* (pp. 4520-4526). IEEE.

Peng, Y., Shi, J., & Yadav, O. (2011, January). Study of Three Maintenance Policies Under Quantity Based Reliability Prediction. In *IIE Annual Conference. Proceedings (p. 1). Institute of Industrial and Systems Engineers (IISE)*.

Verma, A. K., Srividya, A., Goyal, A., & Ramesh, P. G. (2012, August). Application of fuzzy logic for selection of Machinery Health Monitoring strategies. In *Fuzzy Information Processing Society (NAFIPS), 2012 Annual Meeting of the North American* (pp. 1-6). IEEE.

Liao, W. Z., & Li, D. (2015, December). An improved prediction model for equipment performance degradation based on Fuzzy-Markov Chain. *In (IEEM), 2015 IEEE International Conference* on (pp. 6-10). IEEE.

Hasuike, T., & Katagiri, H. (2016, August). A Subjective and Objective Constructing Approach for Reasonable Membership Function Based on Mathematical Programming. *In Soft Computing and Intelligent Systems (SCIS) and 17th International Symposium on Advanced Intelligent Systems*, 2016 Joint 8th International Conference on (pp. 59-64). IEEE.

Lee, L.W., Chen, S.M., 2008. Fuzzy Multiple Attributes Group Decision-Making Based on The Extension of TOPSIS Method and Interval Type-2 Fuzzy Sets. *Proceedings of the Seventh International Conference on Machine Learning and Cybernetics*, IEEE. 3260-3265, Kunming.

Zadeh, L. (1972). Fuzzy-set-theoretic interpretation of the linguistic hedge. *Cybern*. 2(1972). 4-34.

Kumar, G., & Maiti, J. (2012). Modeling risk based maintenance using fuzzy analytic network process. *Expert Systems with Applications*, 39(11), 9946-9954.

Hennequin, S., Arango, G., & Rezg, N. (2009). Optimization of imperfect maintenance based on fuzzy logic for a single-stage single-product production system. *Journal of Quality in Maintenance Engineering*, 15(4), 412-429.

Lu, K. Y., & Sy, C. C. (2009). A real-time decision-making of maintenance using fuzzy agent. *Expert Systems with Applications*, 36(2), 2691-2698.

Lee, S., Li, L., & Ni, J. (2013). Markov-based maintenance planning considering repair time and periodic inspection. *Journal of Manufacturing Science and Engineering*, 135(3), 031013.

Wu, Y. R., & Zhao, H. S. (2010, October). Optimization maintenance of wind turbines using Markov decision processes. In *Power System Technology (POWERCON), 2010 International Conference* on (pp. 1-6). IEEE.

Zhou, Y., Ma, L., Mathew, J., Sun, Y., & Wolff, R. (2011). Maintenance strategy optimization using a continuous-state partially observable semi-Markov decision process. *Microelectronics Reliability*, 51(2), 300-309.

Zhang, X., & Gao, H. (2012). Road maintenance optimization through a discrete-time semi-Markov decision process. *Reliability Engineering & System Safety*, 103, 110-119.

El Afia, A., & Aoun, O. (2017, March). Data-driven based aircraft maintenance routing by markov decision process model. In *Proceedings of the 2nd international Conference on Big Data, Cloud and Applications* (p. 74). ACM.

BALCOR 2018
XIII Balkan Conference on Operational Research

KPI BASED MODEL FOR IMPROVEMENT OF PROCUREMENT PROCESS MANAGEMENT

TEODORA RAJKOVIĆ[1], DANICA LEČIĆ-CVETKOVIĆ[1], LENA ĐORĐEVIĆ[1], ZORANA MIJALKOV[1]

[1]University of Belgrade, Faculty of Organizational Sciences, teodora.rajkovic@fon.bg.ac.rs

Abstract: *Procurement is one of the key processes in every manufacturing company. The procurement process should ensure continuous production, high quality of products as well as reduction of inventory costs and good relationship management with suppliers. This paper presents the developed simulation model for the improvement of the procurement process management based on three key performance indicators. This model was applied to the real case study of a company from the automotive industry.*

Keywords: *Procurement Process Management, KPIs, Model, Simulation.*

1. INTRODUCTION

Procurement, as one of the key process in the manufacturing company, provides items needed for production. Beside raw and basic materials, companies procure high-value complex components needed for the realization of the main activity. The procurement process must ensure the continuous supply of items (materials, parts and components) in order to avoid the interruption of production, unplanned delays and high costs of the resources needed for production. Such problems increase operating costs and lead to inability to provide final products for customers. The success of the procurement process depends on the quality of relationship with suppliers, based on the analysis of their capabilities and cooperation. In order to avoid problems in procurement, this process must be thoroughly planned to balance the relationship between supply and demand. Managing the large amount of data is necessary to achieve efficiency in the procurement process, which requires software support. This paper examines the procurement process management on the case study of a company from the automotive industry which produces wipers and wiper systems. For the purpose of the procurement process management, this company uses the ERP software SAP ECC 6.0, as the solution for remote sites. However, this software does not have the ability to evaluate the performance of the supplier.

Consequently, the simulation model based on the Key Performance Indicators (KPIs) developed for the improvement of the procurement process management, implemented and simulated in *MS Excel* is presented in this paper. The paper is organized as follows. Basic terms related to the procurement process management are presented in the second chapter. The third chapter presents performance indicators, with an emphasis on the procurement process KPIs. This chapter highlights the importance of using and benefits of applying KPIs in procurement process management. In the fourth chapter is presented the implementation of the procurement process management model and simulation on the real case study of a company from the automotive industry. This model is based on KPIs with the aim to improve the procurement process management. The fifth chapter presents conclusions.

2. THE PROCUREMENT PROCESS MANAGEMENT

The modern procurement process is implemented through: providing a continuous supply, minimizing inventory, improving product quality, developing relationships with suppliers and the lowest total inventory costs (Bowersox et al., 2002). One of the main procurement objectives is to provide continuity in the procurement of materials, parts and components (Bowersox et al., 2002).

A determination of a quality procurement management method is necessary to define. The authors Naoum and Egbu in (Naoum et al., 2015) define the procurement method as "a mechanism for linking and coordinating members of the building team together throughout the building process in a unique systematic structure, both functionally and contractually. Functionally observed via roles, authority and power; contractually observed via responsibilities and risks. The main aim is to deliver a project that meets the objectives and fulfills the client criteria needs and expectations." Some of the possibilities for improvement of the procurement process management can be detected by monitoring and analyzing this process. These activities can be used for better detection of problems and their elimination, termination of cooperation with suppliers who operate irresponsibly and continuation of cooperation with the responsible and most important suppliers (Gebauer et al., 2000).

Many researchers have studied the problems of the procurement process management because a lot of different type of problems can appear in this process. The paper (Makajić-Nikolić et al., 2013) presented a tactical production planning model based on the application of Petri nets focused on production triggering decision on a weekly level, based on supply chain performance KPIs. For the model created to follow and to improve supplier delivery, the authors (Knowles et al., 2005) have used the Six Sigma methodology for performance and quality improvement. The paper (Kaur et al., 2017) presents a flexible dynamic sustainability procurement framework (FDSP) for global supply chains as a solution to procurement problems. The authors of this paper considered qualitative parameters: reliability, environmental and social factors and quantitative preferences (total cost and supplier capacity) for selection of suppliers. In (Diaz-Madronero et al. 2017), a mathematical programming model is used for integration of production and procurement transport planning decisions in manufacturing systems. The authors (Diaz-Madronero et al. 2017) suggest that MRP IV can be used as a model for defining indicators for procurement transport efficiency and total planning costs. In the paper (Xia et al., 2007), the authors were solving the problem of defining the best supplier and the quantity of material for procurement by using AHP analysis.

3. KEY PERFORMANCE INDICATORS OF THE PROCUREMENT PROCESS MANAGEMENT

Performance management is a continuous process of identifying, measuring and developing of performances, by linking the performance and stakeholder objectives with the overall mission and company objectives. The performance management model presents a performance management system that connects parts of the system and creates one unity (Golubović et al., 2013). Performance measurement is a process of determining how successful organizations or individuals are in achieving their objectives and strategies (Ofori, 2012). KPIs present a set of the most important performance indicators, focused on the most critical aspects of organizational performance for current and future success of the company.

The main objective of a manufacturing company is to ensure stable procurement process and new ways of procurement process improvement. The KPI presents a guide for continuous improvement to achieve more efficient and sustainable procurement process. Procurement performance indicators indicate the ability of the procurement process to achieve company objectives and tasks with minimal costs (Van Weele, 2010). Van Weele in (Van Weele, 2010) specifies two main objectives of the procurement process: efficiency and effectiveness. The effectiveness of procurement process indicates fulfilled quantity of previously defined objectives and tasks. It is considered as the relationship between realized and planned values of performance of certain activities in the procurement process. The same author points out that procurement efficiency is the relationship between the planned and actual needs for the resources necessary to achieve the set of objectives and tasks and planned and realized costs. In order to fulfill these two objectives, measurement and monitoring of certain performance indicators of supplier have the highest importance.

The company from the automotive industry, analyzed in this paper, has the main problem with supplier's delivery performance and the quality of delivered items. Suppliers deliver defective items and dispatch items in quantities different from a purchase order. This affects inventory and final production of a company. The authors (de Araújo et al., 2017) presented a systematic literature review of the criteria and the methods used in the phases of selecting and evaluating suppliers in projects. They defined following categories of the criteria for the selection and performance evaluation of the suppliers: quality, cost/price, staff features, financial issues, company management, experience, time, technical/technology, relationship with stakeholders, supplier performance (past and current), site capacity/facilities, health and safety, procurement process, flexibility and responsiveness and other categories (which include market, location, risk, etc.). Every category contains subcategories and criteria for their observation. Evaluation criteria for the selection of suppliers are defined according to the most frequently used criteria in the reviewed literature. According to (de Araújo et al., 2017) some of the most important KPIs of procurement process are: completion of defects in quality failures, time performance, technical capacity per supplier, business results, delivery ability and factors pertaining to the customer (de Araújo et al., 2017). This paper analyzes three KPIs, in the context of the mentioned company: *Rejection Ratio* (*RR*), *Purchase Demand Fulfillment Ratio* (*PDFR*) and *Total Coverage Time* (*TCT*). Terminology and equations are adopted from (SAP Community Wiki, 2018) in order to define KPIs.

Rejection Ratio (*RR*) is the percentage of items with unsatisfactory quality from the considered delivery. It is calculated as the ratio of rejected items and total ordered quantity, as presented in equation (1).

$$RR = \frac{RQ}{LQ} * 100\ [\%] \quad (1)$$

where:

- *RR*: rejection ratio per delivery;
- *RQ*: rejected quantity per delivery;
- *LQ*: total lot quantity per delivery.

Purchase Demand Fulfillment Ratio (*PDFR*) presents the percentage of fulfilled items from a purchase order. A supplier receives a purchase order created by a company from the automotive industry according to demand from the customers (purchase request). Order quantity in a purchase order and delivered quantity are compared in order to calculate *PDFR* (equation 2).

$$PDFR = \frac{QD}{TN} * 100\ [\%] \quad (2)$$

where:

- *PDFR*: purchase demand fulfillment ratio;
- *QD*: quantity of delivered items from purchase order;
- *TN*: total number of purchase order items.

Total Coverage Time (*TCT*) is the most important indicator in the procurement process. This indicator has the highest impact on decision making related to the procurement process in the company. This KPI takes into account stock level at the end of the previous day and average demand from all customers for the next 60 days. Parameter *D* presents total deterministic demand from all customers for next 60 days. Indicator *TCT* shows the number of days for which company can meet demand based on purchased items. Total coverage time is expressed in days. It is calculated for each item separately, in accordance with equation (3).

$$TCT = \frac{SLP}{\frac{D}{60}}\ [\text{day}] \quad (3)$$

where:

- *TCT*: total coverage time;
- *SLP*: stock level at the end of the previous day;
- *D*: total deterministic demand from all customers for 60 days;
- $\frac{D}{60}$: average demand from all customers for 60 days.

4. THE PROCUREMENT MANAGEMENT MODEL BASED ON KPI

This chapter presents the model based on KPIs, developed for the improvement of the procurement process management. The model is used for simulation of the procurement process in the company from the automotive industry. Procurement planner monitors daily level procurement process using ERP software SAP. In particular, the planner monitors the amounts of items needed to be provided for production in a defined time period, in order to avoid a lack of items. The SAP system provides the ability of monitoring supplier deliveries of requested quantities. Considering that delivered items may include defective items or less than ordered quantity, it is important to develop a model to control the order quantity of the item. This quantity presents the minimum quantity that supplier must dispatch in the next delivery in order to satisfy the production plan. A model based on the principle of a feedback loop and three KPIs that are presented in the previous part of this paper is developed, in order to facilitate monitoring and managing of procurement performances. The usage of this model is intended for procurement planners.

4.1. Introduction to the problem

Companies define objectives and their values on an annual basis. These values present desired values and companies must define activities to meet the objectives. A feedback model can be used to define activities for achieving desired values. The process is affected by inputs (applied signals) and outputs (produced signals of particular interest) (Stefani et al., 2002). The elementary feedback control system has three components: a system (the object to be controlled, called the system or the plant), a sensor (to measure the output of the plant) and a controller (to generate the plant's input) (Antić et al. 2012). Proper control actions are defined by comparing reference and results obtained through the model. When one or more output variables need to follow a certain reference over time, inputs are manipulated in order to obtain a planned effect on the output of the system, which affects increasing of results quality (Bubnicki, 2005).

The purchase demand fulfillment (*PDFR*) indicator presents the percent of items that were not sent with last delivery. Also, the quality of items is not 100 [%]. In the case of the automotive industry analyzed in the paper, it can be caused because of: bad cavity of a machine, inadequate internal control at supplier's company, damage during item packing and/or storage, damage during loading and/or unloading, damage during transport, inadequately packaged item, inadequate packaging of items, etc. Due to these reasons, the rejection ratio per delivery (*RR*) is considered as one of the key indicators. Also, one of the company's objectives is to reduce *TCT* per supplier, as the third KPI. The company's objective is *TCT* per item per supplier less than 10 days. The company defined three days as the lowest value of *TCT* indicator, to ensure availability of items in a case when supplier does not send ordered quantity. Therefore, the value of this indicator must be between 3 and 10 days, both included. When the coverage time is above 10 days, the company needs to implement some of the defined measures in order to decrease coverage time. Indicators *RR* and *PDFR* are used in the simulation model as input data at a defined time period *t* considering quantities of defective and not delivered items. The *TCT* affects the objective function that should be minimized. The developed model is used to define the minimum quantity of the order item. This quantity should ensure the item coverage per supplier, between 3 and 10 days.

The objective function of the model includes minimization of: the percentage of defective items per delivery, the difference between planned and delivered quantity and *TCT* (between 3 and 10 days). The output, as realized value, consists of inventory level and fulfillment of supplier delivery plan. The plan fulfillment considers the difference between the quantity of ordered item and the quantity of the item which supplier has delivered. Since the input and output values are calculated in the same units, the sensor transmits a signal from system output to the comparator for *RR*, *PDFR* and *TCT*. *RR* and *PDFR* should be converted from percentages to pieces. The comparator compares realized output values with desired values of all three KPIs. The controller is defined through control action or quantity to be ordered.

4.2. Mathematical modeling

The problem is modeled as a dynamic discrete time system control process (Kostić, 2009). The cooperation between the company and suppliers is based on the compensation principle. This principle implies shipping of all defective items from the previous delivery and all items ordered but not delivered, with the next delivery. The control variable presents order quantity that ensures item coverage per supplier within the allowed limits. The constraints are the storage space constraint for each supplier and the number of days of item coverage. Model input data are: inventory level per item for each supplier; total demand for the next 60 days; quantity of defective items received in the previous delivery; quantity of planned but not delivered items.

The assumptions of the model are:

- Demand from all customers is observed for the next 60 days from the day *t*;
- Single supply inventory model is developed for single item of single supplier;
- Lead time is predefined for each supplier;
- The main principle of item ordering is visibility of a day when the item needs to arrive at the company. Based on this, supplier plans deliveries in advance and shipping days (*t*-*LT* days). Shipped delivery arrives on the day *t* in the company.

In order to explain this model, we use the following notations:

- t: discrete time unit, t=0, 1, ..., T;
- D: total deterministic demand from all customers for 60 days;
- p_t: planned production for day t;
- m: total number of items per pallet; m=1, ..., 20;
- r_t: received quantity of items;
- X_t : stock level in period t;
- u_t: order quantity in period t;
- Y_t: number of defective items from the previous delivery that will be included in next order quantity;
- J_t: number of TCT_t days above allowed coverage time of 10 days;
- RR_t: rejection ratio;
- $PDFR_t$ quantity of non-delivered items;
- TCT_t: total coverage time;
- LT: lead time;
- G: maximum number of items in available storage space.

Because the values of parameters X_t, u_t and Y_t are presented in pieces, the indicator $PDFR_t$ is also presented in pieces. The indicator $PDFR_t$ presents the quantity of the item that was not sent. It is important to note that the control variable u_t presents order quantity, but procurement planner cannot be sure that the supplier will deliver this quantity. The company cannot rely on the supplier to comply with the requirement and to send items of the 100 [%] required quality, which may be influenced by other factors that do not originate from the supplier. These factors are mentioned in chapter 4.1. LT presents time necessary for the shipment from the supplier to the company, including time of loading and time for customs (export and import). This parameter depends on the location of the supplier, the number of drivers and length of the legal break that drivers must have during the drive. The developed model considers fixed LT.

The maximum allowed value of the indicator RR_t is 5 [%] of the ordered quantity. This is presented by equation (1). In the case when the number of defective items exceeds 5 [%] of delivery, it is necessary to ask the supplier for compensation with the next delivery (equation 4). Missing items for production are taken from stock. This missing quantity will be included in next order quantity u_t.

$$Y_t = \begin{cases} 0, & RR_t \leq 5\ [\%]\ of\ delivered\ parts\ shipped\ in\ t-LT \\ RR_t * u_{t-LT}, & otherwise \end{cases} \tag{4}$$

Demand is considered as the deterministic variable. The control variable (u_t) represents order quantity calculated in accordance with the average demand for next 60 days and missing quantities from the previous delivery (defective and non-delivered quantity). This is shown in inequality (5). The value of control variable must be non-negative and integer (presented as the constraint (6)). In this company items are ordered per pallet (one pallet contains m pieces of the item). Less than one pallet of items cannot be obtained. Quantity u_{t-LT} presents order quantity in time $t-LT$, which arrives in period t. Parameters RR_t and $PDFR_t$ are defined in time period t, for received delivery. According to these parameters and average demand for next 60 days, the order quantity u_t is defined. This quantity will arrive in the company for $t+LT$ days.

$$u_t \geq \frac{D}{60} + Y_t + PDFR_t \tag{5}$$

$$u_t/m \in Z_{+}^{N} \tag{6}$$

Variable X_t represents system state, i.e. stock level per item. The starting period for measurement of TCT_t indicator is $t=0$, where X_0 presents initial stock value. Items in stock have an inflow and increase only when a new quantity arrives. The new delivery from the supplier arrives for LT days from the date of ordering. When the delivery arrives ($\frac{t}{LT} \in Z^N$), a state of a system is calculated as the stock level at the end of the previous period reduced by the quantity defined by production plan p_t. In the opposite case, the state from the previous day is increased for delivered quantity and decreased by the quantity of defective items, the quantity of the item that was planned, but did not arrive with delivery, as well as the quantity defined by production plan p_t (equation 7).

$$X_t = \begin{cases} X_{t-1} + u_{t-LT} - (RR_t + PDFR_t) - p_t, & \frac{t}{LT} \in Z^N \\ X_{t-1} - p_t, & \frac{t}{LT} \notin Z^N \end{cases} \tag{7}$$

Storage space must be sufficient to store the existing stock level increased for quantity planned to be received (equation 8). In the meantime the item from the warehouse is consumed and frees the space for future deliveries. The total available storage space is defined in advance, as the maximum number of items per supplier.

$$X_{t-1} + u_{t-LT} \leq G \tag{8}$$

The TCT_t parameter is calculated as the ratio of the stock level at the end of the previous day and the average demand in next 60 days (equation 9). The company seeks to have minimal deviations in the planned number of days of total coverage time for a period of 10 days. Minimization of these deviations presents the objective function of the model. The assumption of the model is that the minimum stock level should cover demand for three days. The initial value of TCT_t in period t=0 is three days. Value of control variable u_t, i.e. order quantity should ensure that TCT_t is at least three days for every observed period, shown in inequality (10). The total coverage time per supplier should be between 3 days (minimum value) and 10 days (maximum value). The objective function in period t takes value 0 when TCT_t is equal or less to 10 days, because it is within the allowed number of days. In opposite case, when TCT_t is over maximum value, the objective function is calculated as TCT_t value reduced by maximum allowed value (10 days). This is shown in equation (11). The objective function should be minimized for all 10 days of the observation period. Performance criteria presents the cumulative value of the objective function for a discrete time period t= 1, ..., T (equation 12).

$$TCT_t = \frac{X_{t-1}}{\frac{D}{60}} \tag{9}$$

$$TCT_t \geq 3\ days \tag{10}$$

$$J_t = \begin{cases} TCT_t - 10, & TCTt > 10 \\ 0, & TCTt \leq 10 \end{cases}, \qquad \text{for } t=0, 1, ..., T \tag{11}$$

$$(min)\ J = \sum_{t=1}^{T} J_t \tag{12}$$

4.3. Model implementation and results of simulation

The single supplier inventory model is implemented in a spreadsheet (*MS Excel*) and the simulation is performed for single item from single supplier for the period of T=10 days. The spreadsheet model was developed in accordance with the previously described mathematical model. Values of input data used in this simulation present weighted values taken from the real case study of a company from the automotive industry. Input data, used for simulation, are initially exported from the SAP and include: r_t, RR_t, $PDFR_t$, p_t G, m, LT and D. Indicator $PDFR_t$ is presented as the quantity of non-delivered items. Simulation results are presented in Table 1.

Table 1: Input values and simulation results

G	*m*	*LT days*	*Data/time period*	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*(min) J*
			rt		100	0	0	295	0	0	185	0	0	176	
			RRt		8	0	0	10	0	0	9	0	0	8	
			PDFRt		30	0	0	25	0	0	15	0	0	4	
			pt		50	50	70	90	100	120	70	80	20	20	
		3	*D*		800	900	950	700	800	800	850	800	950	1000	
			ut		320	0	0	200	0	0	180	0	0	70	
			Xt	*100*	200	150	80	256	156	36	135	55	35	177	
			Yt		26	0	0	20	0	0	17	0	0	6	
			Jt	0	0	4	0	0	10	2	0	1	0	0	17
			rt		60	0	60	0	214	0	225	0	100	0	
			RRt		15	0	2	0	5	0	25	0	5	0	
			PDFRt		40	0	50	0	36	0	45	0	20	0	
			pt		50	50	70	90	100	120	70	80	20	20	
4000	10	2	*D*		800	900	950	700	800	800	850	800	950	1000	
			ut		110	0	250	0	270	0	120	0	60	0	
			Xt	*100*	200	150	139	49	153	33	132	52	127	227	
			Yt		17	0	0	0	0	0	30	0	0	0	
			Jt	0	0	4	0	2	0	2	0	0	0	0	8
			rt		47	53	35	88	73	80	78	80	38	80	
			RRt		5	3	7	11	5	5	2	1	1	1	
			PDFRt		3	7	15	2	7	10	12	20	2	0	
			pt		50	50	70	90	100	120	70	80	20	20	
		1	*D*		800	900	950	700	800	800	850	800	950	1000	
			ut		60	50	90	80	90	90	100	40	80	20	
			Xt	*100*	150	152	115	104	74	30	37	37	55	135	
			Yt		0	0	7	9	0	0	0	0	0	0	
			Jt	0	0	0	0	0	0	0	0	0	0	0	0

According to total daily demand from all customers for the period of 60 days, an average demand for next 60 days from the day t can be calculated using formula $\frac{D}{60}$. The objective function should achieve the minimum value. The simulation is conducted for three scenarios: *LT*=1 day; *LT*=2 days and *LT*=3 days in accordance with the real-life data related to the case study of a company from the automotive industry. For all three cases, the value of the objective function in period t=0 is 0 days. The control variable is defined in accordance with constraint (5). This variable must be divisible by the quantity of the item on the pallet and to be non-negative and the integer value (6). The item is ordered per pallet and it strives to fulfill the objective function. The control variable u_t defines order quantities in order to minimize the objective function (12).

Simulation results of defined scenarios are different. The objective function will achieve the minimum value when the item is ordered every day (*LT*=1 day). The increase of *LT* number of days, proportionally decrease the number of orders from a supplier on a weekly level, consequently manual ordering becomes more complex. This directly affects the value of TCT_t indicator and the objective function. As already mentioned in the paper, the simulation model was tested for single item of single supplier and observation period of 10 days. In the case of ordering from s suppliers, for s = 1, ..., S; q^s as the total number of items observed by the supplier; q = 1, ..., Q *and* m^s quantities of item per pallet per supplier, for m = 1, ..., M; simulation for described problem will become complex and it will not be possible to define control variables u_t manually. Complexity is reflected in increasing of model dimensions. The problem becomes NP-hard and some heuristic or metaheuristic methods should be applied in order to solve it.

5. CONCLUSION

This paper presents a model created for measuring of procurement process KPIs in order to improve the procurement process management and monitoring supplier deliveries. The model is based on three performance indicators: *Rejection Ratio, Purchase Demand Fulfillment Ratio* and *Total Coverage Time*. Based on reference values of performance indicators and calculated values, the model provides suggestions for further management actions. This model was applied to the real case study of a company from the automotive industry. Based on the current stock level and required deliveries, this model is used to suggest order quantities for the next delivery, in order to improve values of considered performance indicators and the procurement process management. The model can be expanded by increasing of model dimensions (number of items, number of suppliers, time horizon, etc.) and adding more constraints. If it becomes NP-hard problem, it should be solved by heuristic and metaheuristic methods. That would be the direction of further research of the authors of this paper.

REFERENCES

Antić, S., Kostić, K., & Đorđević, L. (2012). Spreadsheet Model of Inventory Control Based on Modern Control Theory. In *Proceedings* of the *XIII International Symposium SymOrg*,1365-1373.

Bowersox, D. J., Closs, D. J., & Cooper, M. B. (2002). *Supply Chain Logistics Management* (Vol. 2). New York, NY: McGraw-Hill.

Bubnicki, Z. (2005). *Modern Control Theory*. New York, NY: Springer Science & Business Media.

De Araújo, M. C. B., Alencar, L. H., & de Miranda Mota, C. M. (2017). Project Procurement Management: A Structured Literature Review. In *Proceedings* of *International Journal of Project Management, 35*(3), 353-377.

Díaz-Madroñero, M., Mula, J., & Peidro, D. (2017). A Mathematical Programming Model for Integrating Production and Procurement Transport Decisions. In *Proceedings* of *Applied Mathematical Modelling, 52*, 527-543.

Gebauer, J., & Segev, A. (2000). Emerging Technologies to Support Indirect Procurement: Two Case Studies from the Petroleum Industry. In *Proceedings* of *Information Technology and Management, 1*(1-2), 107-128.

Golubović, D., Rajković, D., Đurđević, S., Gojković, R., & Moljević, S. (2013). An Approach To Performance Measurement of Standardized Management System (in Serbian). In *Proceedings* of the *8th Scientific Conference with International Participation "Quality 2013"*, 71-76.

Kaur, H., & Singh, S. P. (2017). Flexible Dynamic Sustainable Procurement Model. In *Proceedings* of *Annals of Operations Research*, 1-41.

Knowles, G., Whicker, L., Femat, J. H., & Canales, F. D. C. (2005). A Conceptual Model for the Application of Six Sigma Methodologies to Supply Chain Improvement. In *Proceedings* of *International Journal of Logistics Research and Applications, 8*(1), 51-65.

Kostić, K. (2009). Inventory Control as a Discrete System Control for the Fixed-Order Quantity System. In *Proceedings* of *Applied Mathematical Modeling, 33*(11), 4201-4214.

Makajić-Nikolić, D., Lečić-Cvetković, D., Atanasov, N., & Babarogić, S. (2013). An Approach to Production Planning for Supply Chain Performance Improvements. In *Proceedings* of the *XI Balkan Conference on Operational Research*, 357-366.

Naoum, S., & Egbu, C. (2015). Critical Review of Procurement Method Research in Construction Journals. In *Proceedings* of *Procedia Economics and Finance*, *21*, 6-13.

Ofori, G. (Ed.). (2012). *Contemporary Issues in Construction in Developing Countries*. Routledge.

SAP Community Wiki, (2018). Retrived from: https://wiki.scn.sap.com/wiki/display/KPI/ Purchasing+ Demand+Fulfillment+Ratio

Stefani, R. T., Shahian, B., Savant, C. J., & Hostetter, G. H. (2002). *Design of Feedback Control Systems* (4th ed.). New York: Oxford University Press, Inc.

Van Weele, A. J., (2010). *Purchasing and Supply Chain Management*. *Analysis*, *Planning and Practice*. Cengage Learning EMEA.

Xia, W., & Wu, Z. (2007). Supplier Selection with Multiple Criteria in Volume Discount Environments. In *Proceedings* of *Omega*, *35*(5), 494–504.

B17
Optimization Software

BALCOR 2018
XIII Balkan Conference on Operational Research

ON POSSIBLE CRYPTOGRAPHIC OPTIMIZATION FOR SECURE MOBILE APPLICATION

MILAN MARKOVIĆ[1], GORAN ĐORĐEVIĆ[2]

[1] Paneuropean University Apeiron, Banja Luka, Republic Srpska, milan.z.markovic@apeiron-edu.eu
[2] AET Europe, Arnhem, The Netherlands, goran.djordjevic@aeteurope.com

Abstract: *In the paper, we focus on possible optimization of cryptographic algorithms implemented in the secure Android mobile client application. The presented experimental results justify that security operations related to X.509v3 digital certificate generation and XML/WSS digital signature creation/verification are feasible on some current smart phones and justify the use of the proposed optimization techniques for implemented cryptographic algorithms. A possible usage of this Android-based secure mobile client application in a secure m-healthcare model is presented. The presented model consists of SOA based m-healthcare online system with secure mobile communication between patients and/or medical professionals with medical and/or insurance organizations.*

Keywords: *Cryptographic algorithms, Digital signature, Android application, m-Healthcare system.*

1. INTRODUCTION

This paper is related to consideration of possible secure m-healthcare model and applying secure Android Web services based mobile client application in it.

Overviews of possible secure systems based on similar model, secure JAVA mobile Web service application and SOA-Based central platform are given in (Marković and Đorđević 2010), (Marković and Đorđević 2011), (Marković and Đorđević 2012), and (Marković and Đorđević 2013) where the model is conceptually and theoretically presented and evaluated in domains of m/e-government and m/e-banking.

In this paper, as an extension of the previous work, a possibility of applying the similar model in domain of m-healthcare systems is considered. Additionally, a possibility of using the secure Android based mobile client application in the proposed m-healthcare model is considered and experimentally evaluated.

First, we consider a possible model of secure SOA-based m-healthcare online systems, i.e. about secure mobile communication between patients and/or medical professionals with the medical and/or healthcare insurance organizations for different purposes. This model could be considered in both local and cross-border case. The latter means either crossing borders of municipalities/regions in the same country or crossing borders between countries (e.g. some medical organizations in different countries).

As a main goal of this paper, we consider a possible usage of the Android-based secure mobile Web service client application in the proposed secure m-healthcare model. A feasibility of using such Android based secure mobile client application is experimentally evaluated in the paper. An emphasis is given on possible optimization techniques of cryptographic algorithms implemented on the Android platform. In this sense, we give two approaches of possible optimization of RSA private key operations. The proposed optimization techniques are experimentally verified in the paper.

The paper is organized as follows. Security requirements in m-healthcare systems are elaborated in Section 2. The architecture of the proposed m-healthcare model is proposed in Section 3. Information about some related work in literature is given in Section 4, while some features of the secure mobile client applictions are presented in Section 5. Proposed optimization cryptographic techniques are described in the Section 6. Experimental results obtained by the secure Android-based mobile client application is given in Section 7 while conclusions are given in Section 8.

2. SECURITY REQUIREMENTS IN M-HEALTHCARE SYSTEMS

This Section deals with the basics of security mechanisms/requirements in m-healthcare systems. Key players in Healthcare systems are: medical organizations (hospitals, clinics, pharmaceutical organizations), insurance organizations, healthcare professionals (doctors, physicians, nurses, pharmacists, etc.), and patients – end users.

Most modern Healthcare systems are information systems based on TCP/IP computer networks and they work fast move toward the electronic business in Healthcare industry – electronic Healthcare (e-Healthcare). In this environment, security mechanisms for e-business must be implemented with necessary adaptation to the Healthcare environments. There are a lot of technical and security issues for these systems that include, between the others: electronic patient record or electronic health record (EHR) must be fully private, central database of patient electronic records must be enabled for use from all players (medical organizations, professionals, insurance, patients), privacy protection of the patient records, secure communications between all players in the system, electronic order entry, enabling mobile Healthcare, HIPAA compliance, etc.

Thus, security mechanisms that are necessary to be implemented in these e-healthcare systems are: strong user authentication procedure, digital signature technology, confidentiality protection of data in the system on the application, transport and network layers, privacy protection of the patient personal data, strong protection of the central healthcare database based on multiple firewall architecture, and PKI systems, which issues X.509 digital certificates for all users of the system (Healthcare professionals and patients) - digital identities (IDs) for the users.

However, with nowadays extreme penetration of mobile communications and usage of smart mobile phones/devices, earlier e-business models and systems move fast towards m-business models and systems. The same holds for e-healthcare systems and thus in this paper we considered, elaborated and experimentally evaluated a possible m-healthcare systems based on Secure Android based Web service mobile client application and SOA based Web service front end m-healthcare system. Some initial considerations of security requirements that need to be applied in the m-Healthcare systems are given in (Marković et al 2006).

3. POSSIBLE SECURE M-HEALTHCARE MODEL

The proposed secure m-Healthcare model, depicted in Figure 1, consists of:

- **Mobile users** (patients, medical professionals) who send some Web services requests to m-healthcare platform for different purposes (sending some patient data to the central system, asking for some medical advices, checking some information about patients, checking insurance data, etc.). These users use secure Android mobile Web service client application on their mobile devices (mobile phones, smart phones, tablets, etc.) for such purpose.
- **SOA based Web service endpoint implementation** on the Platform’s side that implements a complete set of server based security and business features. Well processed requests with all security features positively verified, the Web service platform’s application proceeds to other application parts (i.e. legacy subsystems) of the proposed SOA-Based platform of the medical or insurance organizations.
- **External entities** such as: PKI server with XKMS server as a front end, the Authentication server, and TSA (Time Stamping Authority).

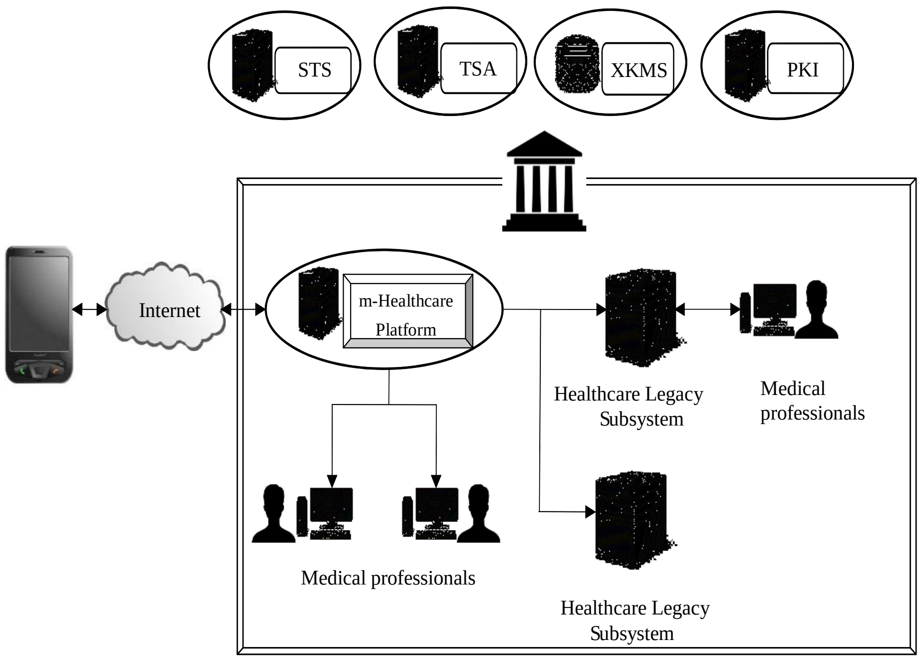


Figure 1: A proposed secure m-healthcare model

Functions of the proposed external entities are following:

- **PKI server** is responsible for issuing PKI X.509v3 electronic certificates for all users/entities in the proposed m-healthcare model (patients, medical professionals, administrators, servers, platforms, etc.). Since some certificate processing functions could be too heavy for mobile users, the PKI services (certificate location/validation) could be exposed by the XKMS server which could register users, as well as locate or validate certificates on behalf of the mobile user. This is of particular interests in all processes that request signature verification on mobile user side.
- **Authentication server (e.g. STS (Security Token Service))** is responsible for strong user authentication based on PKI X.509v3 electronic certificate issued to users and other entities in the proposed model. Possible communication between the authentication server and the user's mobile Web service application could be SOAP-based and secured by using WS-Security features. Possible scenario is that, after the successful user authentication, the STS server issues a SAML token to the user which will be subsequently used for the user authentication/authorization to the Web service of the proposed m-healthcare platform. The SAML token is digitally signed by the STS server and could consist of the user role for the Platform's user authorization. The alternative is that it could be a general-purpose Authentication server which will authenticate users by using any kind of authentication credentials, such as: username/password, OTP, PKI digital certificates, etc. In the latter case, there could be possible Web service based communication between the SOA-based central platform and the authentication server in order to authenticate users.
- **TSA server** is responsible for issuing time stamps for user's requests as well as for platform's responses (signed electronic documents). Time stamping of requests/documents could be requested from users, from the platform or from both entities.

Security operations in electronic business (e-government, e-healthcare, e-banking, e-commerce, e-payment, etc.) and mobile business (m-government, m-healthcare, m-banking, m-commerce, m-payment, etc.) systems are mostly based on two secure actions:

- Strong user authentication
- Transaction authorization

In the proposed model, the strong user authentication is based on the X.509v3 digital certificate as unique identifiers of users. Regarding the transaction authorization, it is based on digital signature of the electronic documents with additional usage of the timestamping. Since both choices represent techniques of the highest cryptographic level which are required in the Healthcare based systems, we believe that this model is the best suited for m-healthcare systems. Besides, in the proposed model, we use the encryption technique (WS-Encryption) in order to preserve confidentiality of information transmitted which represents an additional reason why this model is the best suited for m-healthcare systems.

4. RELATED WORK

There are no many similar works in the literature. One work worth mentioning is the session based Web application system presented in (Lee et al 2007). Compared to a session based Web/application platform, presented in (Lee et al 2007), in this paper we proposed a usage of the SOAP-based request-response technologies which is much better fitted to mobile environment.

The model proposed in this paper could have the following advantages compared to the model given in (Lee et al 2007):

- Web service based request-response system is much more efficient system in the mobile environment than the session based Web application system. Especially when some back office processing (Healthcare legacy systems) are needed to respond on the user requests.
- Web service based model provides much more flexibilities and an easier way to implement all security features (e.g. XML security, WS-Security, Time Stamping, XKMS, PKI) compared to the Web based solution.
- Web service based system provides much more flexibilities compared to the session based Web application system in cross-border scenarios when business process includes also some processing of the user request outside of the contacted government organization.

Also, there are some conceptual discussions about security issues in the m-government systems, given in (Kumar et al 2008). In this paper, we go further in experimental approving the usage of the secure Android mobile client application in the context of complex m-healthcare model presented in this paper.

Compared to the m-government system based on mobile qualified electronic signature in Austria (http://www.buergerkarte.at/langswitch.php?lang=en), where the mobile phone is used as a strong user authentication tool and where a server based signature is employed (user's private key is on the HSM on server side – generated and used), our proposed model is based on the „fat" client on the mobile user side where all cryptographic mechanisms are implemented in the Android based secure mobile client application. Thus, the system implemented in (http://www.buergerkarte.at/langswitch.php?lang=en) has emphasized on the authentication part of the security operations and for the transaction authorization it is implemented on the server side. In our model, both activities, strong user authentication and transaction authorization is done by using security mechanisms implemented in the mobile application.

Also, compared to some LSP (Large Scale Pilot) projects, e.g. STORK (https://www.eid-stork.eu/) and STORK 2.0 (https://www.eid-stork2.eu/), where some very complex interoperability authentication model is proposed, our proposed model could be more comprehensive and complete since the STORK models are mostly based only on user authentication mechanisms and their interoperabilities in cross-border usage. Unfortunately, there are no much discussions about possibilities of transaction authorization in the cross-border case.

Besides the above mentioned references, the authors of this paper could not find similar works in the literature related to m-healthcare systems based on Web services and Android clients. Thus, unfortunately, the presented experimental analysis does not contain a comparative experimental analysis to other achievements from the literature.

5. SECURE MOBILE WEB SERVICE CLIENT APPLICATION

The proposed secure mobile Web service client application could comprise of following functionalities:

- **Graphical User Interface (GUI)** for presenting business functionalities to the end user. The GUI object of the proposed mobile Web service application is responsible to show user interface that enable calling of function for authentication of the end user and presenting the core functionalities to the end user. According to this, the GUI object communicates with following modules:
    - User Authentication module for mobile client application of the Security module
    - User PKI Registration module (XKMS module) of the Security module
    - User Authentication and Authorization module for the m-government platform (SAML module) of the Security module
    - Business functionalities
- **Business (core) functionalities** of the application – m-healthcare functionalities. Business functionalities have links to Security and Communication modules of the secure mobile Web service application.
- **Security functionalities.** The Security module of the considered secure mobile Web service application is responsible for overall application-level security functionalities.
- **Communication.** The communication module is responsible for establishment of secure communication between patients and medical/insurance organizations.

The security functionalities of the proposed Secure Android Mobile Client application consist of the following modules:

- **Authentication module** of the secure mobile application. User authentication for the secure mobile application should be two-step process:
    - The first step would be a combination of username/password for accessing the application (password should be changeable by the user). This should be done immediately after the application starts. These credentials will be generated during the user registration process. During the initial phase of the registration application, the user will obtain the username and default password. The application has to force the user to change the initial password on the first application start.
    - The second step will be in presenting a corresponding PIN code for accessing the asymmetric private key just before digital signing different m-healthcare requests.

  The generation of user asymmetric public/private key pair and corresponding digital certificate should be done through user registration function of the XKMS protocol. The User Authentication module is called from the GUI object.
- **XKMS module.** XML Key Management Specification enables to simplify the use of PKI by mobile client systems.
- **STS module.** The STS module is responsible for the communication with the STS server in order to receive a SAML assertion (token) that will be used afterwards to enable access to the business functionalities by the client. The user first sends a RequestSecurityToken message to the STS

(Security Token Service) server by using a SAML protocol. A protection is done by using WS Security mechanisms. After successful authentication of the user based on the client's X.509v3 digital certificate, the STS server issues a SAML token to the user which is digitally signed by the STS server. This token is securely communicated to the end user by using the WS security mechanisms.

- **XML security module.** XML security module is responsible for implementation of standard XML signature and XML encryption components. XML security module consists of:
  - Implementation of the RSA private key operation for creating digital signature, as well as a function for signature verification.
  - Implementation of hash functions (SHA-1, SHA-224, SHA-256, SHA-384, SHA-512).
  - Implementation of different symmetrical cryptographic algorithms (3DES, AES).
  - Implementation of the RSA private key operation for decryption of encrypted symmetric message key in digital envelope.
  - Implementation of the RSA public key operation for encryption of symmetric message key in digital envelope.
- **WS-Security module.** Web Service (WS) Security module is implemented as standard security mechanisms for protection of SOAP messages. WS-Security module is very important module of the Security module since it is used for protection:
  - Communication with STS server.
  - Communication with the proposed SOA-Based m-healthcare platform.

  This way, the WS-Security module communicates with SAML module of the Security object as well as with Business functionalities object. The SAML module communicates with WS security module of the Security object as well as with the Communication object.
- **Time-Stamping module.** This module is responsible for communication with the TSA. A time-stamping service supports assertions of proof that a datum existed before a particular time. The user's application requests a time-stamp token by sending a request to the TSA. As the second message, the TSA responds by sending a response, i.e. actual timestamp, to the requesting entity.

The secure mobile Web service application could secure communicate with all mentioned external entities in Section 4, i.e. it has all security functions mentioned implemented:

- Secure mobile Web service application sends Request for Security Tokens to the STS server by using WS-Secured (WS-Signature and WS-Encryption) SOAP communication.
- Secure mobile Web service applications sends digitally signed (XML signature) m-healthcare request to the Web service of the proposed m-healthcare platform by using WS-Encrypted SOAP communication. The sent request includes the SAML token issued and signed by the STS server.
- The request is timestamped by sending a timestamp request and obtaining the corresponding timestamp response (digitally signed by the TSA).

The secure mobile Web service application also receives the signed and timestamped response from the m-healthcare platform through WS-Encrypted communication and performs all necessary signature verifications and certificate validations (by help of the XKMS server) actions.

6. OPTIMIZATION OF CRYPTOGRAPHIC ALGORITHMS IN SECURE MOBILE WEB SERVICE CLIENT APPLICATION

The Android platform ships with a cut-down version of Bouncy Castle - as well as being crippled. It also makes installing an updated version of the libraries difficult due to class loader conflicts. Different versions of Android operating system have implemented different versions of Bouncy Castle library releases. In order to avoid lack of interoperability between different devices that have implemented different operating systems and get more flexible code we used Spongy Castle functions (http://rtyley.github.com/spongycastle/).

The Spongy Castle package contains low-level lightweight API implementing all the underlying cryptographic algorithms and a provider for the Java Cryptography Extension (JCE) and the Java Cryptography Architecture. The basic package that supports the cryptographic algorithms and padding schemes is the org.spongycastle.crypto package. The org.spongycastle.asn1 package supports the parsing and writing ASN.1 objects, which is useful in processing X.509 certificates. The utility classes in org.spongycastle.util can be used for producting and reading Base64 and Hexadecimal strings. The utility is useful if the ciphertext is required to be displayed as a Base64 string.

In order to achieve smaller and faster implementation we have partly modified Spongy Castle functions. The modification of Spongy Castle funtions is achieved in org.spongycastle.jce package. We don't want to use JCE functionalities of genuine Spongy Castle implementation because that adds a significant memory overhead. In order to avoid using the heavyweight provider for the JCE that contains implementation of many unnecessary functions we cut off a lot of functions and implement only the necessary ones. We directly call necessary Spongy Castle functions without using java.security.Provider functionalities at all. Using this approach we got smaller and faster code.

Because mobile devices have limited resources, an application designed for mobile devices should be as compact as possible. An obfuscator is a useful tool for minimizing the size of an application. We used ProGuard obfuscator that shrinks, optimizes, and obfuscates code by removing unused code and renaming classes, fields, and methods with semantically obscure names. The result is a smaller sized .apk file that is more difficult for being reversely engineered.

In order to additionally improve performances we have considered possibility of implementation of private key RSA operations (creation of digital signature, open digital envelope) using native code in Android Native Development Kit (NDK). A basic reason for this decision was a fact that the native code is compiled to binary code and run directly on mobile phone OS. We implemented native code on a basis of usage of OpenSSL package. In this case, a lot of cryptographic functions are implemented using C programming language. In order to additionally speedup operations with RSA private key we implemented some functions directly in assembler code. One of implemented functions in assembler code is procedure of Montgomery modular multiplication. Montgomery multiplication is a method for computing $a*b$ mod m for positive integers a, b and m. It reduces execution time on a CPU when there are a large number of multiplications to be done with same modulus m, and with a small number of multipliers. In particular, it is useful for computing a^n mod m for a large value of n. The number of multiplications modulo m in such computation can be reduced to a number substantially less than n by successively squaring and multiplying according to the pattern of the bits in the binary expression for n ("binary decomposition").

7. EXPERIMENTAL ANALYSIS

This Section is dedicated to the experimental analysis of the cryptographic operations implemented on Android mobile phone, i.e. smart phones with Android operating system (Reto Meier 2012), as a possible example of the proposed secure mobile client application that could be used in the proposed m-healthcare model. Also, the proposed model and presented experimental results on Android mobile operating systems represent m-healthcare extension compared to the discussion presented in (Braga and Nascimento 2012). The presented experimental results are generated using devices (mobile phone, tablet, PC laptop and PC desktop) described in (Marković and Đorđević 2013).

Experimental results that are presented in this section are based on the modified version of Spongy Castle functions as well as partly modified version of OpenSSL native code. In Tables 1, 2, 3, 4, a row with title 'Native code' is shown with results where operations with RSA private key are done by using native code (C code + assembler). During testing phase we have measured average time by using some number of iterations. The actual number of iterations used are shown in each table. The same code and packages are used during testing procedure in all devices. Throughout this Section, all presented experimental results are given in miliseconds – ms.

In order to evaluate the possibility of using the mobile phone for secure mobile Android-based client application in m-healthcare systems based on Web service we measured times needed for creation X509 v3 self-signed certificate comprising a creation of PKCS#10 certificate request (Table 1). As a signature algorithm we used SHA-1 hash algorithm and RSA asymmetric cryptographic algorithm. Then we measured time for creation of XML-Signature and Web Service (WS) Signature (Table 2, Table 3), respectively. In all these experiments, we used a file of 1KB, RSA asymmetric algorithm and SHA-1 hash function. We also analyzed possibility of WS Decryption mechanisms (Table 4).

Table 1: Create X509 v3 self-signed certificate

Device		RSA private key length (bits), n=50000 iterations				
		512	1024	2048	3072	4096
Mobile Phone	SpongyCastle	18.11	27.41	84.41	216.89	467.95
	Native code	13.48	21.26	73.48	206.53	426.34
Tablet		34.49	46.27	116.05	283.22	548.40
PC Laptop		1.78	9.01	58.07	180.97	414.14
PC Desktop		1.33	6.78	43.92	137.36	312.90

Table 2: XML-Signature creation

Device		RSA private key length (bits), n=50000 iterations				
		512	1024	2048	3072	4096
Mobile Phone	SpongyCastle	29.64	38.15	95.87	228.20	479.10
	Native code	25.01	32.01	84.94	217.84	437.49
Tablet		59.73	73.78	144.08	319.65	586.54
PC Laptop		2.12	9.38	58.50	181.54	414.74
PC Desktop		1.57	7.05	43.85	137.87	312.79

Table 3: WS-Signature creation

Device		RSA private key length (bits), n=50000 iterations				
		512	1024	2048	3072	4096
Mobile Phone	SpongyCastle	63.76	74.51	131.00	266.29	507.47
	Native code	59.13	68.36	120.07	255.93	465.86
Tablet		126.99	147.68	216.81	384.48	663.93
PC Laptop		2.79	10.07	59.18	182.1	415.19
PC Desktop		2.02	7.50	44.57	138.03	311.66

Table 4: WS-Decryption mechanism

Device		RSA private key length (bits), n=50000 iterations				
		512	1024	2048	3072	4096
Mobile Phone	SpongyCastle	34.96	44.20	102.48	232.75	486.20
	Native code	30.33	38.05	91.55	222.39	444.59
Tablet		80.91	119.74	169.87	339.54	609.42
PC Laptop		2.41	9.67	58.88	181.79	415.98
PC Desktop		1.77	7.28	44.36	138.01	313.88

Some observations of the presented experimental analysis are:

- The creation of the self-signed X.509v3 digital certificate with 2048 bits key by using the mobile phone takes 84.61 ms and even 73.48 by using optimized native code which is similar to the results obtained by PC computers.
- The operation of digital signature of XML message (XML-Signature and WS-Signature mechanisms), using 2048-bit private RSA key, takes on mobile phone 95.87 and 131 ms, respectively, and in optimized native code version 84.94 and 120.07 ms, respectively, which are comparable to the results obtained by other devices..
- The operation of decryption of WS-Encrypted message using 2048-bit private RSA key, takes 102.48 ms and in the optimized native code version 91.55 ms. It means that in one second can be implemented about 10 operations of decryption WS-Encrypted message using 2048-bit RSA private key.

These observations could lead to the conclusion that mobile phone could be used in real time for implementation of RSA private key operations in times comparable to the ones obtained on PC computers, especially when the optimized native code is used.

8. CONCLUSION

In this Paper, we presented an overview of possible secure model of m-healthcare systems as well as an analysis of possibility and feasibility of using secure Android-based web service mobile client application in it, as well as some considerations of possible optimization of applied cryptographic algorithms.

First, this paper is related to the consideration of some possible SOA-based m-healthcare online systems, i.e. about secure mobile communication between patients and medical professionals with medical and insurance organizations.

Second, the paper presented a possible example of an Android-based secure mobile client application that could be used in the described m-healthcare model and which is experimentally evaluated. An emphasis is given on possible optimization techniques of cryptographic algorithms implemented on the Android platform. In this sense, we give two approaches of possible optimization of RSA private key operations. The proposed optimization techniques are experimentally verified in the paper

Presented experimental results justify that security operations related to RSA private key operations (creation of X.509v3 digital certificate, XML/WS digital signature, WS-Encryption) are feasible for usage on some current smart phones. Thus, we could conclude that this application could serve as a basis for implementing secure m-healthcare system based on the model described in this paper. Also, presented experimental analysis justifies the usage of the proposed optimization of cryptographic techniques implemented on a basis of C and assembler code.

REFERENCES

Braga, A. M., Nascimento, E. N. (2012). Portability Evaluation of Cryptographic Libraries on Android Smartphones, *Cyberspace Safety and Security, Lecture Notes in Computer Science*, Volume LNCS-7672, pp 459-469.

Lee, Y., Lee, J., Song, J. (2007). Design and implementation of wireless PKI technology suitable for mobile phone in mobile-commerce, Computer Communication, 30 (4): 893-903.

Kumar, M., Hanumanthappa, M., Reddy, B. L. (2008). Security Issues in mGovernment, H. Jahankhani, K. Revett, and D. Palmer-Brown (Eds.), In *Proceedings of ICGeS 2008, CCIS 12*, Springer-Verlag, Berlin Heidelberg, pp. 265-273.

Marković, M., Savić, Z, M., Kovačević, B. (2006). Secure Mobile Health Systems: Principles and Solutions, chapter in the book *M-Health, Emerging Mobile Health Systems*, Series: International Topics in Biomedical Engineering Istepanian, Robert; Laxminarayan, Swamy; Pattichis, Constantinos S. (Eds.), XXX, 624 p. 182 illus., Hardcover, ISBN: 0-387-26558-9, pp. 81-106.

Marković, M., Đorđević, G. (2010). On Possible Model of Secure e/m-Government System, *Information Systems Management*, Taylor & Francis Group, LLC, 27:320-333.

Marković, M., Đorđević, G. (2011). On Secure SOA-Based e/m-Government Online Services, in *Handbook "Service Delivery Platforms: Developing and Deploying Converged Multimedia Services"*, Taylor & Francis, pp. 251 – 278, Chapter 11.

Marković, M., Đorđević, G. (2012). On Secure m-government and m-banking model, in *Proc of 6th International Conference on Methodologies, Technologies and Tools Enabling e-Government,* Belgrade, Serbia, pp. 100-111.

Marković, M., Đorđević, G. (2013). Secure Android Application in SOA Based Mobile Government Systems, in *Proc of 7th Int. Conference on Methodologies*, *Technologies and Tools Enabling e-Government*, Luis Alvarez Sabucedo and Luis Anido Rifon (Eds.), Vigo, Spain, pp. 117-126.

Reto Meier, P. (2012). *Professional Android 4 Application Development*, John Wiley & Sons, Inc., Indianapolis, Indiana.

B19
Risk Analysis
& Management

BALCOR 2018
XIII Balkan Conference on Operational Research

A MODEL FOR ASSESSMENT OF RISKS WITH TRAPEZOIDAL INTERVAL TYPE-2 FUZZY SETS

DAJANA ŽIVKOVIĆ[1], ALEKSANDAR ALEKSIĆ[2], DANIJELA TADIĆ[3], ALEKSANDAR ĐORĐEVIĆ[4]

[1] University of Kragujevac, Faculty of Engineering Science, zivkovicdajana@yahoo.com
[2] University of Kragujevac, Faculty of Engineering Science, aaleksic@kg.ac.rs
[3] University of Kragujevac, Faculty of Engineering Science, galovic@kg.ac.rs
[4] Higher Technical School of Professional Studies Zvečan, adjordjevic@kg.ac.rs

Abstract: *The risk assessment problem for any organization system which existing in uncertain environment is a proactive approach for loss or elimination of consequence severities that may arise due to the materialization of risk factors. Uncertainties into severities of consequences, and the existence of adequate risk management procedures are described by pre-defined linguistic expressions modeled by trapezoidal interval type-2 fuzzy sets. The membership function of type-2 fuzzy sets have uncertainty associated with it's so that are more suitable than type-1 fuzzy sets. In this paper, a reliable and friendly used method for risk assessment of each risk factor is proposed. The proposed model is applied to environmental risk assessment problem for production supply chain.*

Keywords: *risk assessment, fuzzy data, trapezoidal interval type-2 fuzzy sets, fuzzy algebra.*

1. INTRODUCTION

In recent years, the risk assessment problem has become one of the most important operational management problems in any organization systems, especially in supply chain that exists in an environment that is rapidly and continuously changing. There are widely literature review of research papers in which is considered that the risk value depend on severities of consequences arising due to the materialization of Risk Factors (RFs), and frequency of occurrence of RFs (Aven et al., 2011). In recent time, many authors suggest that risk prevention procedure should be implemented in risk assessment. In the literature, there are many papers discussing risk assessment problem on the conceptual level and by applying the proposed models (Pinto 2014; Đapan et al. 2015). The identification of all possible RFs should be obtained by different ways, for instance by applying of risk checklists or literature sources (Waters, 2011, Song**,** Ming, and Liu, 2016). The checklist is a simple procedure and easy to understand. The effectiveness of this procedure depends on the knowledge, skills and expertise of decision makers it can be denoted as its main lack. In practice, the consequences at the level of each identified RF are determined by decision makers and depend on to their experience and knowledge. In this paper environmental risk assessment are considered, so that identification of RFs is performed according to (Giannakis and Papadopulos, 2016) and the set of consequences which are associated each RF are defined by decision makers.

In the literature, the frequencies of RFs occurrence are impacted by the current condition of organization, its robustness, preparedness and resilience. The most usually, the frequencies of RFs occurrence are determined according to the evidence data. Therefore, frequency values may be expressed by precise numbers. In general practice, uncertainties into risk assessment problem are estimated by decision makers that express their estimations by precise numbers (Eckert and Gatzert, 2017), stochastic variable with Poisson distribution (Hess, 2011) or linguistic expressions which are modeled by fuzzy sets theory (Pinto 2014; Đapan et al. 2015). The concept linguistic variable is introduced in Zadeh (1975), which may be defined as a variable whose values are expressed in linguistic terms. By applying fuzzy set theory (Dubois & Prade, 1980; Zimmermann, 2001), linguistic variable may be quantitatively presented. In the literature a great number papers can be found in which different uncertainties, vagueness and imprecise are modeled by type-1 fuzzy sets (Kaya and Kahraman, 2011; Aleksić et al, 2014).

A type-1 fuzzy set is represented by its membership function whose parameters are the shape and location in the universe of discourse. In other words, the membership functions of type-1 fuzzy sets are two-dimensional. Creating the membership function problem is considered in the literature (Pedrycy and Gomide, 1988). It can be said that creating of membership functions of type-1 fuzzy sets is based on one's experience, subjective belief of decision makers, intuition and contextual knowledge about the concept modeled (Zimmermann, 1978). However, subjectivity in determining the membership function has been considered as the weakest point in the fuzzy sets theory. Jointly used shapes of triangular and trapezoidal functions offer a good compromise between descriptive power and computational simplicity. The concept of a type-2 fuzzy set was introduced by Zadeh (1975) and presents an extension of the concept of a type-1 fuzzy set. It is known, the linguistic terms may be understood in different ways by other people (Karnik and Mendel, 2001). Hence, determining of membership function for a fuzzy set in exact manner is very difficult. In these cases, the third dimension of membership function of fuzzy set is added. It can be concluded that type-2 fuzzy sets are more suitable to represent uncertainties than type-1 fuzzy sets. Type-2 fuzzy sets require complex and enormous computational onerous operations so that have not wide spread application in modeling of real uncertainties. The interval type-2 fuzzy set is introduced by Mendel at al (2008) and presents a special case of type-2 fuzzy set. The computational effort with the interval type-2 fuzzy sets is reduced so that they are widely used to solve different decision making problems (Chen, and Wang, 2013). Zhang and Zhang (2013) have defined some operations on trapezoidal interval type-2 fuzzy sets which are used in (Kahraman et al, 2014). The correct versions of notions and operations of trapezoidal type-2 fuzzy sets with a rigorous proof are presented in (Zhang and Zhang, 2017). In this paper, modeling of existing uncertainties is based on trapezoidal interval type-2 fuzzy sets and its operations (by analogy Zhang and Zhang, 2017).

The main objectives of the paper is quantitatively describe imprecise dimensions of RFs, to determine a risk values in exact way, and to determine adequate management initiatives, which should lead to a reduction or eliminated of the identified RFs impact and thus to increase the effectiveness and sustainability of organizational system in long-term period.

In the literature, there are almost no papers in which interval trapezoidal type-2 fuzzy sets for modeling uncertain and imprecise data in risk research domain are used.

The paper is organized in the following manner. Section 2 presents basic definition of trapezoidal interval type-2 fuzzy sets. The proposed methodology is given in Section 3. In Section 4, the proposed model is illustrated with data, which come from SC automotive industry that exists in developing country. Conclusion is presented in Section 5.

2. PRELIMINARIES

In this Section, some basic definitions related to the current study are presented.

Definition 1. (Mendel et al., 2008). A type 2 fuzzy set, $\tilde{\tilde{A}}$ in the universe of discourse X can be represented by a type-2 membership function $\mu_{\tilde{\tilde{A}}}$ shown as follows:

$$\tilde{\tilde{A}} = \left\{ \left((x,u), \mu_{\tilde{\tilde{A}}}(x,u)\right) \middle| \forall x \in X, \forall u \in J_X \subseteq (0,1), 0 \le \mu_{\tilde{\tilde{A}}}(x,u) \le 1 \right\} \tag{1}$$

The type-2 fuzzy set $\tilde{\tilde{A}}$ also can be represented as:

$$\tilde{\tilde{A}} = \int_{x\in X} \int_{u\in J_x} \mu_{\tilde{\tilde{A}}}(x,u)/(x,u) = \int_{x\in X} \left(\int_{u\in J_x} \mu_{\tilde{\tilde{A}}}(x,u)/u \right) / x \tag{2}$$

where x is the primary variable, $J_X \subseteq [0,1]$ is the primary membership of x, u is the secondary variable, and $\int_{u\in J_x} \mu_{\tilde{\tilde{A}}}(x,u)/u$ is the secondary membership function at x. The union among all admissible x and u is denoted as $\int$.

Definition 2. (Mendel et al., 2008). Let $\tilde{\tilde{A}}$ be a type-2 fuzzy set in the universe of discourse X represented by the type-2 membership function $\mu_{\tilde{\tilde{A}}}(x,u)$. If all $\mu_{\tilde{\tilde{A}}}(x,u)=1$, then $\tilde{\tilde{A}}$ is called an interval type-2 fuzzy set which can be regarded as a special case of a type-2 fuzzy set, shown as follows:

$$\tilde{\tilde{A}} = \int_{x\in X} \int_{u\in J_x} 1 \ /(x,u) = \int_{x\in X} \left(\int_{u\in J_x} 1/u \right) / x \tag{3}$$

where x is the primary variable, $J_X \subseteq [0,1]$ is the primary membership of x, u is the secondary variable, and $\int_{u \in J_x} 1/u$ is the secondary membership function at x.

The upper membership function ($\tilde{A}^U$) and lower membership function ($\tilde{A}^L$) of $\tilde{\tilde{A}}$ are two type-1 membership functions that bound the footprint of uncertainty (FOU), so that $FOU = \bigcup_{x \in X} J_X$.

Definition 3. (Mendel et al., 2008). If X is a set of real numbers, then a type-2 fuzzy set and an interval type-2 fuzzy set in X are called a type-2 fuzzy number and an interval type-2 fuzzy number, respectively.

Definition 4. (Chen and Lee, 2008). If the upper membership function and lower membership function of $\tilde{\tilde{A}}$ are two trapezoidal type-1 fuzzy numbers, than $\tilde{\tilde{A}}$ is referred to as trapezoidal interval type-2 fuzzy number, $\tilde{\tilde{A}} = \left(\tilde{A}^U, \tilde{A}^L\right)$ so that:

$$\tilde{\tilde{A}} = \left(\tilde{A}^U, \tilde{A}^L\right) = \left(\left(a_1^U, a_2^U, a_3^U, a_4^U; H_1\left(\tilde{A}^U\right), H_2\left(\tilde{A}^U\right)\right), \left(a_1^L, a_2^L, a_3^L, a_4^L; H_1\left(\tilde{A}^L\right), H_2\left(\tilde{A}^L\right)\right)\right) \tag{4}$$

where:

$H_i = \left(\tilde{A}^U\right)$ denotes the membership value of the element d_{i+1}^U, in the upper trapezoidal membership function $\tilde{A}^U$, $1 \le i \le 2$

$H_i = \left(\tilde{A}^L\right)$ denotes the membership value of the element d_{i+1}^L, in the lower trapezoidal membership function $\tilde{A}^L$, $1 \le i \le 2$

$$H_i = \left(\tilde{A}^U\right) \in [0,1], \; H_i = \left(\tilde{A}^L\right) \in [0,1], \; 1 \le i \le 2 \tag{5}$$

Definition 5. Let us two trapezoidal interval type-2 fuzzy numbers, $\tilde{\tilde{A}}$, and $\tilde{\tilde{B}}$

$$\tilde{\tilde{A}} = \left(\left(a_1^U, a_2^U, a_3^U, a_4^U; H_1\left(\tilde{A}^U\right), H_2\left(\tilde{A}^U\right)\right), \left(a_1^L, a_2^L, a_3^L, a_4^L; H_1\left(\tilde{A}^L\right), H_2\left(\tilde{A}^L\right)\right)\right) \tag{6}$$

, and

$$\tilde{\tilde{B}} = \left(\left(b_1^U, b_2^U, b_3^U, b_4^U; H_1\left(\tilde{B}^U\right), H_2\left(\tilde{B}^U\right)\right), \left(b_1^L, b_2^L, b_3^L, b_4^L; H_1\left(\tilde{B}^L\right), H_2\left(\tilde{B}^L\right)\right)\right) \tag{7}$$

The union operation is introduced by (Zhang and Zhang, 2013) and the some arithmetic operations are introduced by (Mendel et al., 2008):

a)

$$\tilde{\tilde{A}}\cup\tilde{\tilde{B}}=\left(\begin{array}{l}\left(a_1^U\vee b_1^U,a_2^U\vee b_2^U,a_3^U\vee b_3^U,a_4^U\vee b_4^U;H_1\left(\tilde{A}^U\right)\wedge H_1\left(\tilde{B}^U\right),H_2\left(\tilde{A}^U\right)\wedge H_2\left(\tilde{B}^U\right)\right),\\ \left(a_1^L\vee b_1^L,a_2^L\vee b_2^L,a_3^L\vee b_3^L,a_4^L\vee b_4^L;H_1\left(\tilde{A}^L\right)\wedge H_1\left(\tilde{B}^L\right),H_2\left(\tilde{A}^L\right)\wedge H_2\left(\tilde{B}^L\right)\right)\end{array}\right) \tag{8}$$

b) The addition operation, which is denoted as, $\tilde{\tilde{A}}+\tilde{\tilde{B}}$ can be defined as:

$$\tilde{\tilde{A}}+\tilde{\tilde{B}}=\left(\begin{array}{l}\left(a_1^U+b_1^U,a_2^U+b_2^U,a_3^U+b_3^U,a_4^U+b_4^U;\min\left(H_1\left(\tilde{A}^U\right),H_1\left(\tilde{B}^U\right)\right),\min\left(H_2\left(\tilde{A}^U\right),H_{21}\left(\tilde{B}^U\right)\right)\right),\\ \left(a_1^L+b_1^L,a_2^L+b_2^L,a_3^L+b_3^L,a_4^L+b_4^L;\min\left(H_1\left(\tilde{A}^L\right),H_1\left(\tilde{B}^L\right)\right),\min\left(H_2\left(\tilde{A}^L\right),H_2\left(\tilde{B}^L\right)\right)\right)\end{array}\right) \tag{9}$$

c) The multiplication operation, which is denoted as, $\tilde{\tilde{A}}\cdot\tilde{\tilde{B}}$ can be defined as:

$$\tilde{\tilde{A}}\cdot\tilde{\tilde{B}}=\left(\begin{array}{l}\left(a_1^U\cdot b_1^U,a_2^U\cdot b_2^U,a_3^U\cdot b_3^U,a_4^U\cdot b_4^U;\min\left(H_1\left(\tilde{A}^U\right),H_1\left(\tilde{B}^U\right)\right),\min\left(H_2\left(\tilde{A}^U\right),H_{21}\left(\tilde{B}^U\right)\right)\right),\\ \left(a_1^L\cdot b_1^L,a_2^L\cdot b_2^L,a_3^L\cdot b_3^L,a_4^L\cdot b_4^L;\min\left(H_1\left(\tilde{A}^L\right),H_1\left(\tilde{B}^L\right)\right),\min\left(H_2\left(\tilde{A}^L\right),H_2\left(\tilde{B}^L\right)\right)\right)\end{array}\right) \tag{10}$$

Definition 6. Let us trapezoidal interval type-2 fuzzy number, $\tilde{\tilde{A}}$, and crisp value k. The inversion of the $\tilde{\tilde{A}}$, which is denoted as $\left(\tilde{\tilde{A}}\right)^{-1}$ and arithmetic operations between $\tilde{\tilde{A}}$, and crisp value k are defined (Chen and Lee, 2008):

a) $$\left(\tilde{\tilde{A}}\right)^{-1}=\left(\left(1/a_4^U,1/a_3^U,1/a_2^U,1/a_1^U/k;H_1\left(\tilde{A}^U\right),H_2\left(\tilde{A}^U\right)\right),\left(1/a_4^L,1/a_3^L,1/a_2^L,1/a_1^L;H_1\left(\tilde{A}^L\right),H_2\left(\tilde{A}^L\right)\right)\right) \tag{11}$$

b) $$k\cdot\tilde{\tilde{A}}=\left(\left(k\cdot a_1^U,k\cdot a_2^U,k\cdot a_3^U,k\cdot a_4^U;H_1\left(\tilde{A}^U\right),H_2\left(\tilde{A}^U\right)\right),\left(k\cdot a_1^L,k\cdot a_2^L,k\cdot a_3^L,k\cdot a_4^L;H_1\left(\tilde{A}^L\right),H_2\left(\tilde{A}^L\right)\right)\right) \tag{12}$$

Definition 7. The defuzzificated the interval trapezoidal type-2 fuzzy numbers approach (DTraT) is proposed (Kahraman et al, 2014):

$$DTraT=\frac{1}{2}\cdot\left\{\frac{\left(a_4^U-a_1^U\right)+\left(\beta^U\cdot a_2^U-a_1^U\right)+\left(\alpha^U\cdot a_3^U-a_1^U\right)}{4}+a_1^U+\left[\frac{\left(a_4^L-a_1^L\right)+\left(\beta^L\cdot a_2^L-a_1^L\right)+\left(\alpha^L\cdot a_3^L-a_1^L\right)}{4}+a_1^L\right]\right\} \tag{13}$$

3. THE PROPOSED APPROACH

The modelling of uncertainties with trapezoidal interval type-2 fuzzy numbers and the proposed Algorithm is explored.

3.1 Definition of finite set of environmental RFs

In general, the (environmental) RFs (in production SCs) can be defined by decision makers. Their assessments are based on experience, knowledge or literature sources. They should be presented by set of indices $I=\{1,...,i,...,I\}$. The total numbers of environmental RFs is denoted as I and i, i=1,..,I is index of environmental RF.

3.2 Definition of finite set of consequence

It may be comprehended that the realization of each environmental RF i, i=1,…, I, could lead to the occurrence of one or more consequences that may be generally presented with the set of indices $\boldsymbol{\Lambda_i} = \{\mathbf{1}, \dots, \mathbf{j}, \dots, \mathbf{J_i}\}$, j=1,…,$J_i$; i=1,…,I. The overall number of possible consequences related to the environmental RF i, i=1,…,I, is indicated by J_i and the index of each identified consequence is j_i. Generally, consequences for each environmental RF i, i=1,..,I are determined according to evidence data and the results of the best practice.

3.3 Modelling of the severity of consequences

It is assumed that the consequences that may appear due to realization of environmental RF i, i=1,…,I should be considered with respects to cost dimension. Severity of consequence j, $j=1,..,J_i; i=1,...,I$ are not measurable variables. Their values are given by assessment of decision makers which have used seven pre-defined linguistic variables that are modelled by trapezoidal interval type-2 fuzzy sets:

very low value (VL)- $((0,0.1,0.3,0.5;1,1),(0.05,0.1,0.3,0.45;0.75,0.75))$,

low value (L)- $((0.2,0.3,0.5,0.6;1,1),(0.25,0.3,0.5,0.55;0.75,0.75))$,

medium value (M)- $((0.3,0.4,0.6,0.7;1,1),(0.35,0.4,0.6,0.65;0.75,0.75))$,

high value (H)- $((0.4,0.5,0.7,0.8;1,1),(0.45,0.5,0.7,0.75;0.75,0.75))$,

very high value (VH)- $((0.5,0.9,1,1;1,1),(0.55,0.9,1,1;0.75,0.75))$.

The domains of these trapezoidal interval type-2 fuzzy sets should be defined into the interval [0-1]. The value 0 i.e. the value 1 denotes that values of considered uncertainties are the lowest, i.e. the highest, respectively.

3.4 Modelling of risk prevention procedure

In many companies, there are procedures with rules as to how to proceed when materializing the RF are defined. Levels of procedure appropriateness could be described by following linguistic expression

- *The unacceptable procedure (UP)*- $((1,1,2,3.5;1,1),(1,1,2,3;0.8,0.8))$,
- *The acceptable procedure (PP)*- $((1,2.5,3.5,5;1,1),(1.5,2.5,3.5,4.5;0.8,0.8))$,
- *The adequately procedure (AP)*- $((2.5,4,5,5;1,1),(3,4,5,5;0.8,0.8))$.

The domains of these interval type-2 fuzzy sets should be defined into the interval [1-5]. The value 1 i.e. the value 5 denotes that almost doesn't exists, i.e. there is adequate procedure, respectively.
Overlapping of the interval type-2 fuzzy sets indicates that there is not enough evidence data, nor results of good practice.

3.5 The proposed Algorithm

The proposed research Algorithm may be realized through a certain number of steps, which are further discussed.

Step 1. Calculate the fuzzy severity of all consequences that may appear due to materialization of RF i, $\tilde{\tilde{S}}_i$, i=1,..,I by using union operation (Zhang and Zhang, 2013):

$$\tilde{\tilde{S}}_i = \bigcup_{j=1,..,J_i} \tilde{\tilde{s}}_{ji}, \ i=1,..,I \tag{1}$$

where $\tilde{\tilde{s}}_{ji}$ is severity of consequence j, $j=1,..,J_i$ at the level of RF I, i=1,..,I

Step 2. It is known, if there are adequate risk procedures that arise from the realization of any RF, it is lower and vice versa. Determination of reciprocial interval trapezoidal type-2 fuzzy sets, $\left(\tilde{\tilde{p}}_i\right)^{-1}$ is based on fuzzy algebra rules (Kahraman, et al., 2014).

Step 3. Calculation of the risk value $\tilde{\tilde{R}}_i, i=1,..,I$ that may occur due to the materialization of each RF i, i=1,..,I:

$$\tilde{\tilde{R}}_i = \tilde{\tilde{S}}_i \cdot f_i \cdot \left(\tilde{\tilde{p}}_i \cdot\right)^{-1} \tag{2}$$

Step 4. The representative scalar of TFN $\tilde{\tilde{R}}_i$, R_i is calculated by extended procedure which is proposed in (Kahraman et al., 2014):

$$R_i = DTraT\,(\tilde{\tilde{R}}_i) \tag{3}$$

Step 5. Ranking of RFs according to the obtained risk values R_i, i=1,..,I.

Step 6. Propose management initiatives that may lead to acquisition of optimal requirements and RFs impact reduction on realization of business goals.

4. ILLUSTRATIVE EXAMPLE

The developed model for environmental risk assessment is tested on the data originated from automotive supply chain operating in the region of developing county. The environmental RFs in supply chain is defined according to (Giannakis and Papadopulos, 2016) and they are: Natural disasters (i=1) which is described as rare but severe disruptions caused by natural disasters (hurricanes, flood, storms, earthquakes); Inefficient use of resources (i=2), which is described as inefficient resource (e.g., energy, recyclable wastes) use for the production and delivery of goods and services, Environmental pollution (i=3), which is described as air, water, soil or other contamination due to facility operations or products. Hazardous waste generation, which is described as unusable or unwanted substance or material, produced during, or as a result of a process, such as manufacturing or transportation (i=4). Number and kind of consequences which may arise from materialization of considered environmental RFs, as well as their severities are determined by decision makers. They base their assessments on the knowledge, experience and results of good practice. The frequencies of occurrence of all environmental RFs are determined by using evidence data. The possibility of procedures application was assessed according to the available documentation. The input data are presented in Table 1.

Table 1: Severities and frequencies of identified consequences and procedure appropriateness

RFs	The consequences / cost severity/ safety severity	Relatively frequency	Procedure appropriateness
i=1	breakdown of production due to infrastructure damage/*H*; interruption of production due to damage of production equipment /*M*	*0.1*	*UP*
i=2	decrease production process effectiveness /*H*; decrease distribution process effectiveness /*L*	*0.6*	*PP*
i=3	air pollution /*M*; water pollution *H*, soil contamination V*L*	*0.95*	*AP*
i=4	hazardous waste contamination */VH*	*0.22*	*PP*

The procedure for calculating of the fuzzy severity of all consequences (Step 1 of the Algorithm) is illustrated for example RF (i=2):

$$\tilde{\tilde{S}}_2 = \bigcup_{j=1,2} \tilde{\tilde{s}}_{ji} = \left((0.2,0.3,0.7,0.8;1,1),(0.25,0.3,0.7,0.75;0.75,0.75)\right)$$

Similarly, we calculate the severity of the consequences arising from the materialization of the other environmental RFs considered.

The inverse value is given by procedure (Step 2 of the proposed Algorithm):

$$\left(\tilde{\tilde{p}}_2\right)^{-1} = \left((0.2,0.29,0.4,1;1,1),(0.22,0.29,0.4,0.67;0.8,0.8)\right)$$

The inverse values for other RFs are calculated similarly.

$$\left((0.75,0.9,1,1;1,1),(0.8,0.9,1,1;0.75,0.75)\right)$$

The risk value is assigned to a RF (i=2) and it is given according Step 3 of the proposed Algorithm:

$$\tilde{\tilde{R}}_2 = \tilde{\tilde{S}}_2 \cdot f_2 \cdot \left(\tilde{\tilde{p}}_2 \cdot\right)^{-1} = ((0.024,0.052,0.168,0.480;1,1),(0.033,0.052,0.168,0.302;0.75,0.75))$$

The representative scalar of the trapezoidal interval type-2 fuzzy number $\tilde{\tilde{R}}_2, R_2$ is calculated (Step 4 of the proposed Algorithm) is

R_2 =DTraT ($\tilde{\tilde{R}}_2$)=0.153

The fuzzy risk values and their scalar values are presented in Table 2. Also, rank of environmental RFs is presented in Table 2.

Table 2: The fuzzy risk values and corresponding representative scalars of environmental RFs and their rank

RFs	The fuzzy risk values	The representative scalars	Rank
i=1	$((0.006,0.015,0.028,0.080;1,1),(0.007,0.015,0.028,0.050;0.75,0.75))$	0.027	4
i=2	$((0.024,0.052,0.168,0.480;1,1),(0.033,0.052,0.168,0.302;0.75,0.75))$	0.153	1
i=3	$((0,0.019,0.166,0.304;1,1),(0.010,0.019,0.166,0.235;0.75,0.75))$	0.109	2
i=4	$((0.022,0.057,0.088,0.220;1,1),(0.027,0.057,0.088,0.147;0.75,0.75))$	0.084	3

Respecting the obtained rank, it can be clearly seen that following environmental RF has the highest influence on the SC existing SC is *inefficient use of resources* (i=2). Some of possible management initiatives are further analysed are to use resources that are obtained in the recycling processes.

5. CONCLUSION

In this paper, prioritization of management initiatives which should lead to a reduction in impact of RFs is based extended risk assessment model. Realization of management initatives according to priority leads to the less use of financial resources, which are further propagated on the increase of the business effectiveness over long-term period.

The existing uncertainties are modelled by the interval trapezoidal type-2 fuzzy numbers. Incorporating footprint of uncertainty into type-1, the different kinds of uncertainties can be described in an exact manner. Handling of uncertainties into the proposed model is based of fuzzy logic rules. There is a significant difference between results which are given by using risk assessment model with type-1 fuzzy sets and the extended risk assessment model with the interval type-2 fuzzy sets.

The main advantage of the proposed method is flexible to the changes in: (1) the numbers of RFs, (2) the number and severities of consequence, (3) the RFs frequency of occurrence, (4) the procedural adequacy, and (5) can be easily extended to the analysis of risk assessment problems in different research areas. In practical domain, the proposed model is easy to be comprehended and utilized, and as such, it may be very useful for decision makers in long-term period.

The general limitation of the proposed model is need for wide range of evidence data which sometimes cannot be easily found. Future research should include risk analysis that may occur due to different RFs and SCs that exist in many industrial branches.

Acknowledgement

Research presented in this paper was supported by Ministry of Science and Technological Development of Republic of Serbia, Grant TR 35033 and Grant III44010.

REFERENCES

Aleksić, A., Stefanović, M., Tadić, D., Arsovski, S. (2014). A fuzzy model for assessment of organizational vulnerability. *Measurement*, 51, 214-223.

Aven, T., Renn, O., Rosa, E.A. (2011)., The ontological status of the concept of risk. *Safety Science*, Vol. 49(8-9), pp. 1074–1079.

Chen, S.M., Wang, C.Y. (2013). Fuzzy decision making systems based on interval type-2 fuzzy sets, *Information Sciences*, 242, 1-21.

Chen, S-M., Lee, L-W. (2008). Fuzzy multiple attributes group decision-making based on the interval type-2 TOPSIS method. *Expert Systems with Applications* 37, 2750-2798.

Đapan, M. J., Tadić, D. P., Mačužić, I. D., & Dragojović, P. D. (2015). A new fuzzy model for determining risk level on the workplaces in manufacturing small and medium enterprises. Proceedings of the Institution of Mechanical Engineers, Part O: *Journal of Risk and Reliability*, 229(5), 456-468.

Dubois, D., & Prade, H. (1980). Systems of linear fuzzy constraints. *Fuzzy sets and systems*, 3(1), 37-48.

Eckert, C., & Gatzert, N. (2017). Modeling operational risk incorporating reputation risk: An integrated analysis for financial firms. *Insurance: Mathematics and Economics*, 72, 122-137.

Giannakis, M., & Papadopoulos, T. (2016). Supply chain sustainability: A risk management approach. *International Journal of Production Economics*, 171, 455-470.

Hess, C. (2011). The impact of the financial crisis on operational risk in the financial services industry: Empirical evidence. *The Journal of Operational Risk*, 6(1), 23.

Kahraman, C., Öztayşi, B., San, U.Î., Turanoğlu,E. (2014). Fuzzy analytic hierarchy process with interval type-2 fuzzy sets. *Knowledge-Based Systems*, 59, 48-57.

Karnik, N.N. and Mendel, J.M. (2001). Operations on type-2 fuzzy sets. *Fuzzy sets and Systems*, 122, 327-348.

Kaya, T., & Kahraman, C. (2011). An integrated fuzzy AHP–ELECTRE methodology for environmental impact assessment. *Expert Systems with Applications*, 38(7), 8553-8562.

Mendel, J.M., John, R.I., Liu, F. (2008). Interval type-2 fuzzy logic systems made simple. IEEE Trans. Fuzzy Systems 14 (6), 808-821.

Pedrycy, W., Gomide, F. (1988). *An introduction to fuzzy sets, Analysis and Design*. MIT-Press: Cambridge Massachusetts.

Pinto, A. (2014). QRAM a Qualitative Occupational Safety Risk Assessment Model for the construction industry that incorporate uncertainties by the use of fuzzy sets. *Safety Science*, 63, 57-76.

Song, W., Ming, X., & Liu, H. C. (2017). Identifying critical risk factors of sustainable supply chain management: A rough strength-relation analysis method. *Journal of Cleaner Production*, 143, 100-115.

Waters, D. (2011). *Supply chain risk management: vulnerability and resilience in logistics*. Kogan Page Publishers.

Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-1. *Information sciences*, *8*(3), 199-249.

Zhang, Z., Zhang, S. (2013). A novel approach to multi attribute group decision making based on trapezoidal interval type-2 fuzzy soft sets. *Applied Mathematical Modelling* 37, 4948-4971.

Zhang, Z., Zhang, S. (2017). Comments on A note on A novel approach to multi attribute group decision making based on trapezoidal interval type-2 fuzzy soft sets. *Applied Mathematical Modelling*, 41, 702-710.

Zimmermann, H. J. (2011). Fuzzy set theory—and its applications. *Springer Science & Business Media.*

Zimmermann, H.J. (1978). Results of empirical studies in fuzzy set theory (ed. G.J. Klir). *Applied General Systems Research*, Plenum Publishing Corporation, 303-311.

TERRORISM RISKS ASSESSMENT OF TOURISM DESTINATIONS

DRAGANA MAKAJIĆ-NIKOLIĆ[1], MARIJA KUZMANOVIĆ[2], BILJANA PANIĆ[3]

[1]University of Belgrade, Faculty of Organizational Sciences, makajic-nikolic.dragana@fon.bg.ac.rs
[2]University of Belgrade, Faculty of Organizational Sciences, marija.kuzmanovic@fon.bg.ac.rs
[3]University of Belgrade, Faculty of Organizational Sciences, panic.biljana@fon.bg.ac.rs

---

Abstract: *In the last two decades, one of the main risk at tourism destinations is related to the treats of terrorism attacks. Therefore, it is necessary for the tourism industry and everybody who invests in tourism to assess the vulnerability of particular tourism destination both to actual terrorism and perceived insecurity in general. This paper shows how Kinney method, originally developed for occupational health and safety analysis, can be used for the risk assessment of terrorism at particular destinations. Kinney method rank the risk of observed entities based on the ratings of probability, consequence and exposure to hazard. In the proposed approach, probability ratings can be obtained according to the number of previous terrorist attacks, the consequence rating can be obtained based on the number of victims and the exposure ratings can be assigned based on the type of the destination. The approach will be illustrated on the example of nine tourism destinations.*

Keywords: *Risk, Terrorism, Tourism, Kinney method.*

1. INTRODUCTION

In the last years, political instability is increased all over the world whether by some local riots, wars etc. or global attacks as consequences of the influence of powerful countries. One of the phenomena with the biggest global influence is the increasing of the terrorist attacks. Terrorist attacks often happen on the non-strategic destinations, and most of the victims are not directly connected with the cause of the attack, but they are civilians throughout terrorists want to send some message. However, one of the main messages received by people is that certain destination is not safe or that safety is significantly decreased. Information placed by media influence to people to precipitate certain destinations in one or another way. This can be very harmful to the tourism industry.

Terrorism is usually defined as "politically motivated violence perpetrated against civilians and unarmed military personnel by subnational groups" (U.S. Department of State definition, cited in Sönmez 1998). Schmid (1992) distinguished between the academic discourse, the governmental position, the public perception, and the terrorist's own view in attempt to escape what he noted was the defeatist position that "one man's terrorist is another man's freedom fighter." Other authors according to Coaffee (2009) used a different classifications including cultural, ideological, criminal, nuclear and chemical, computer, moral, political, religious and state-sponsored terrorism. According to Merari (1993) there are three cornerstones of terrorism: violence, political motivation and installation of fear into the target population. Political violence, especially terrorism can negatively affect tourism in some country, but if the violence stops and the country manages to reverse its negative image in the international media, then tourism can bounce back (Neumayer 2004).

Many authors researched consumer behaviour in tourism and risk perception (Roehl and Fesenmaier 1992, Fuchs and Reichel 2008, Yang and Nair 2014, Monahan 2015, Morakabati and Kapuściński 2016, Kapuściński and Richards 2016, Cui et al. 2016, Karl 2018) because risk perception has an impact on the motivation to travel and on the choice of the way and destination to travel. They proved that buyers of touristic services are aware of risks and that risk perception can influence on decision to travel to some area.

Many authors researched influence of terrorism on tourism industry (Richter and Waugh 1986, Sonmez 1998, Leslie 1999, Willis et al. 2003, Neumayer 2004, Morakabati and Kapuściński 2016, Kapuściński and Richards 2016) because tourists have been the targets of terrorist attacks or incidental victims. If countries are dependent on tourism, some groups that want to harm the government can find it useful to decrease some of the government's finance by reducing tourism. Therefore, the first step is to assess the risk of terrorism of a particular destination. Ezel et al. (2010) suggested multiple approaches as combination of different risk assessment techniques and methoda: event trees, decision trees, fault trees, influence diagrams, causal loop diagrams and systems dynamic models, Bayesian belief networks, game theory and agent-based models.

In this paper, we propose the use of Kinney method to assess the risk of terrorism. Risk assessment using Kinney method is based on three factors: the probability of an accident, its consequences and the time of exposure. The main idea of our approach is that, using historical data, the probability of the accident in a particular destinationcan be estimated based on the terrorist attacks that previously occurred, the consequence of the possible future attacks can be estimated based on the number of victims and wounded people, and the exposure can be determined by the type of the destination. The paper consists of five sections. In Section 2, the review of the literature related to tourism and terrorism risks is given. Section 3 is devoted to the Kinney method. Section 4 shows an empirical study that illustrates the proposed approach. The conclusions and directions of future research are given in Section 5.

2. LITERATURE REVIEW

Roehl and Fesenmaier (1992) have begun to examine the tourism risk. They asked people about types of risk of their vacation and they classified risks as equipment risk, financial, physical, psychological, social, satisfaction and time risks. Shaw (2010) identified risks in tourism and proposed classification of risks to the natural, crime, health and safety risk, political factors risk, socio-demographic, technological and economic risks. Arturas et al. (2015) ofered clasification of tourism risks of: physical risk (injury), financial (invested money will be lost), activity (the product won't meet the customer expectations), social (do not meet the needs of selected purchase groups), psychological (the product will not be compatible with the image of consumer), time (the product use will be too long, and customer has to choose other alternatives).

According to Cui et al. (2016) the concept of ''tourism risk perception'' can be divided into three views:

- Subjective feelings of the negative consequences or negative impact that may occur during travel;
- Objective evaluation of the negative consequences or negative impact that may occur during travel;
- Cognitive of exceeding the threshold portion of the negative consequences or negative impact that may occur during travel.

After that they summarized objective factors of tourism risk perception in five to seven dimensions (Cui et al., 2016):

- five-dimension risk: psychological, financial, performance, health and social risk;
- six-dimension risk: performance, physical, financial, psychological, social and time risk;
- seven-dimension risk: physical, economic, equipment, social, psychological, time risk and opportunity loss.

It is difficult to evaluate risk in tourism taking into account all these factors. Arturas et al. (2015) gave the literature analysis of tourism risk assessment models and explained: Valsamakis risk assessment model, Burke risk assessment model, Gray and Larson assessment model, Osborne risk assessment model and Passenheim risk assessment model. Valsamakis risk assessment model is more appropriate to manage the risk arising from a financial loss. Burke risk assessment model indicate whether the business is ready to accept the exposure of business risk. Gray and Larson assessment model is very similar to Burke model but in Gray and Larson model risk control is in the final step, and Burke starts his model with a risk control. Osborne risk assessment model is five-step risk management process: (1) identify risks, (2) quantify the risk, (3) defining the risk probability and impact, (4) countermeasures reducing, controlling or eliminating the risk, (5) monitor and evaluate efficiency. This model could be adapted to different sized businesses, including inbound tourism assessment. Passenheim risk assessment model considers risk assessment till now are focusing on preventing recurrence, which was done in the past. Passenheim suggest SWOT analysis, which can be used to identify potential risks.

Cui et al. (2016) mentioned risk evaluation model where risks are weighted and quantified to evaluate tourism risk perception. That model includes the subjective feeling of tourists, uncertainty and dangerousness of consequence and tourism risk perception can be calculated:

$$TRP = \sum_{i=1}^{n} \alpha_i \times PL_{ri} \times IL_{ri} \tag{1}$$

where - the weight of each dimension of the tourism risk perception;

ri - the risk of each dimension, $i = 1, 2, 3, \ldots, n$

PLri - the possibility of each dimension risk

ILri - the harm of each dimension risk.

According to Cui et al. (2016) the latter considers the subjective and objective factors of tourism risk perception, called multi-dimensional model. This model has questionnaire designed according to different tourism scenarios, then is used Likert scale to quantify the factors and finally, the results are processed by statistical method.

In the paper (Arturas et al. 2015), IT is suggested that the risks of tourism could be analyzed in three ways: by a tourist point of view, in terms of tourism destination and by the regulatory authorities approach. Depending on the selected ways, creation of a risk assessment methodology can be made from the perspective of users, tour operators or government authorities.

Willis et al. (2003) suggested calculation of the terrorism risk as the product of threat, vulnerability, and consequences:

$$\text{Risk} = \text{Threat} \times \text{Vulnerability} \times \text{Consequence} \quad (2)$$

Threat is the probability that a specific target is attacked in a specific way during a specified time period – *P* (attack occurs)

Vulnerability is the probability that damages occur, given a specific attack type, at a specific time, on a given target. Damages involve fatalities, injuries, property damage, etc. Target's vulnerability can be articulated as the probability that an attack of a given type will be successful if it happened and, measures vulnerability to specific types of damages only – *P* (attack results in damage | attack occurs)

Consequence is the expected magnitude of damage (deaths, injuries, property damages), a specific attack type at a specific time, that results in damage to a specific target – *E* (damage | attack occurs and results in damage)

Threat is most complicate to estimate. Quantifying *P* (attack occurs) "requires knowledge, data, or modeling about the motivations, intent, and capabilities of terrorists (largely the domain of the intelligence community), in addition to or instead of knowledge about historical attacks and their relevance to current risk." (Ezel et al. 2010) If objectives of terrorists are understandable and the actions, thus, can be anticipated.

Analysis by Waugh (1983) shows that terrorist objectives can be long range or ideological, mid range or strategic, and short range or tactical. Richter and Waugh (1986) explained that there are patterns in the kinds of objectives that terrorists may have and those patterns indicate that tourists are logical and appropriate targets of terroristic violence. Aven and Guikema (2015) analyzed existing models for assessment of terrorism risk. They suggested that terrorism risk should be calculated taking into account threats, attacks, their consequences, and the uncertainties, but also interactions between the adversaries.

Morakabati and Kapuściński (2016) researched whether willingness to travel alters after a terrorist attack. They found that terrorism significantly reduces the willingness to travel but risk perception can be mitigated by the strength of the benefits sought.

3. KINNY METHOD

Kinney method was founded in 1971th in the research by Fine & Kinney (Fine & Kinney, 1971). In this method risk is calculated considering the potential consequences of an accident (C), the exposure factor (E) and the probability factor (P). Each of these factors is assessed on the basis of scale, whether universal or scale defined for a particular problem.

CONSEQUENCES C: The most probable results of a potential accident, including injuries and property damageis considered a consequece. This is based upon an appraisal of the entire situation surrounding the hazard, and accident experience as it is shown in Table 1.

Table 1. Consequences ratings (Fine & Kinney, 1971)

Rating (C)	Description	Damage (financially expressed)
1	Minor cuts, bruises, bumps	minor damage
5	Disabling injuries	<$100
15	Extremely serious injury (amputation, permanent disability)	$1000 - $100,000
25	Fatality	$100,000 - $500.000
50	Multiple fatalities	$500,000 - $1,000,000
100	Catastrophe: Numerous fatalities;); major disruption of activitiesof national significance	>$1,000,000

EXPOSURE E: Frequency of occurrence of the hazard-event (the undesired event which could start the accident-sequence). In many research, exposure is observed as the time in which the person is exposed to the hazard (Moraru, 2012), (Netro et al. 2018). Selection is based on observation, experience and knowledge of the activity concerned as it is shown in Table 2.

Table 2. Exposure ratings (Fine & Kinney, 1971)

Rating (E)	Description
0,5	Very rarely (not known to have occurred,but considered remotely possible)
1	Rarely (it has been known to occur)
2	Unusually (from once per month to onceper year)
3	Occasionally (from once per week to onceper month)
6	Frequently (approximately once daily)
10	Continuously (or many times daily)

PROBABILITY P: This is the likelihood that, once the hazard-event occurs, the complete accident-sequence of events will follow with the necessary timing and coincidence to result in the accident and consequences. This is determined by careful consideration of each step in the accident sequence all the way to the consequences, and based upon experience and knowledge of the activity, plus personal observation. Classifications andratings are shown in Table 3.

Table 3. Probability ratings (Fine & Kinney, 1971)

Rating (P)	Description
0,1	Practically impossible sequence or coincidence; a "one in a million" possibility. (Has never happened in spite of exposure over many years.)
0,5	Extremely remote but conceivably possible. (Has never happened after many years of exposure.)
1	Would be a remotely possible coincidence. (It has happened here)
3	Would be an unusual sequence or coincidence
6	Is quite possible, would not be unusual, has an even 50/50 chance
10	Is the most likely and expected result if the hazard-event takes place

Quantitative calculation of the risk is given in the following relation (Fine & Kinney, 1971; Kinney& Wiruth; Marhavilas & Koulouriotis, 2008):

$$R = C * E * P \tag{3}$$

This relation provides a logical system for safety management to set priorities for attention to hazardous situations. Quantity R can be expressed in the scale of 1 to 1000. Authors Kinney & Wiruth (1976) associated the gradation of the risk value R with the urgency level of required actions as it is shown in Table 4.

Table 4. Risk ranking scale (Kinney & Wiruth, 1976)

Risk score	Risk situation
> 20	Risk; perhaps acceptable
21 – 70	Possible risk; attention indicated
71 – 200	Substantial risk; correction needed
201–400	High risk; immediate correction required
Over 400	Very high risk; consider discontinuing operation

As it is difficult to forecast if some terrorist attack will happen on some exact location chosen by the customer, because there is no precise information for forecasting, it is necessary to identify terrorist attack risk parameters by interpreting statistical data and to estimate risk. Parameters should be quantified, so final estimation of risk is numerical.

4. EMPIRICAL STUDY

Subject of this empirical study is risks of terrorism assessment using Kinney mehods for nine tourism destinations: Sousse (Tunisia), Koh Krabey (Cambodia), Tibet (China), Cote d'Azur (France), Addis Ababa (Ethiopia), Muscat (Oman), Bangkok (Thailand), Rome (Italy), and Istanbul (Turkey). These specific locations are chosen after research on the Internet, taking into account type of destination and history of terrorist attacks on since 9/11 attack.

For this research, historical data about terrorist attacks between 2000 and 2016 years on the defined destinations were collected. Collected data are used for rating of risk estimations factors in the Kinney method. Descriptions of each factor are adjusted for this problem. Risk estimation factors are:

- **Probability**of the terrorist attack occurrence. Probability rates are defined in relation to the number of terrorist attacks between 2000 and 2016 years (Table 5).The assumption is that the greater the number of attacks in the past, the greater the likelihood that it will happen in the future.In addition, higher scores are assignedto the attacks occurred recently.
- Intensity of **consequences**is based on average number of victims and wounded people(Table 6).This scale is also relative and it is formed based on the number of victims and wounded in the chosen destinations.
- Durationof **exposure**, i.e. average time(number of days) spent on the destination is defined according to the type of destination(Table 7).

Table 5. Probability ratings

Probability			
Number of attacks	**Time period**	**Description**	**Rating**
0	2000 - 2016	No terrorist attacks	0.1
1	2000 – 2016	One attack happened in that period	0.5
2 – 4	2000 – 2014	2 - 4 attacks happened, most of them in that period	1
2 – 4	2015 - 2016	2 - 4 attacks happened, most of them in that period	3
≥ 5	2000 – 2014	More than 5 attacks happened, most of them in that period	6
≥ 5	2015 - 2016	More than 5 attacks happened, most of them in that period	10

Table 6. Consequence ratings

Consequence			
Average number of victims (died)	**Average number of wounded**	**Description**	**Rating**
0	0	No victims nor wounded	1
0	≥ 1	During attack is wounded at least 1 person	5
1 – 5	≥ 0	Number of victims during the attack is 1 - 5	15
6 – 14	≥ 0	Number of victims during the attack is 6 - 14	25
≥ 15	< 100	More than 14 people were died and less than 100 people are wounded	50
≥ 15	≥ 100	More than 14 people were died and more than 99 people are wounded.	100

Table 7. Exposure ratings

Exposure		
Average number of days spent on the destination	**Description**	**Rating**
<1 day	Shorter than 24 hours – "just passing through" places	0.5
1-3 days	1 - 3 days - smaller cities	1
4 - 7 days	4 - 7 days – big cities	2
8-10 days	8 - 10 days – summer holiday resorts	3
11-14 days	11 - 14 days - summer holiday resorts	6
> 14 days	More than 14 days – exotic destinations	10

Historical data for number of attacks, time period they occur and intensity of terrorist attacks are found on the page (List of non-state terrorist incidents, 2016) while the duration is calculated as the average number of days offered by touristic organisations. These data for nine chosen destinations are shown in Table 8.

Table 8. Historical data of risk factors (List of non-state terrorist incidents, 2016)

Destination	Number of attacks	Time period	Intensity		Duration
			victims	wounded	
Sousse, Tunisia	2	2011-2014	39	39	8-10
KohKrabey, Cambodia	2	2000-2010	9	0	>14
Tibet, China	1	2000-2010	18	623	>14
Cote d'Azur, France	4	2015-2016	86	454	8-10
Addis Ababa, Ethiopia	2	2000-2010	12	24	11-14
Muscat, Oman	0	2000-2016	0	0	8-10
Bangkok, Thailand	8	2011-2014	25	256	4-7
Rome, Italy	2	2011-2014	0	5	1-3
Istanbul, Turkey	23	2015-2016	183	1314	1-3

Based on historical data and scales given in Tables 5-7, ratings for probability, consequence and exposure are assigned to nine chosen destinations. Risk factors ratings and risk ranks are shown in Table 9.

Table 9. Risk factors ratings, risk estimation and ranks of destinations

Destination	Ratings				
	Probability	**Consequence**	**Exposure**	**Risk**	**Rank**
Sousse, Tunisia	1	50	2	150	5
KohKrabey, Cambodia	1	15	10	150	5
Tibet, China	0.5	100	10	500	4
Cote d'Azur, France	3	100	3	900	3
Addis Ababa, Ethiopia	1	25	6	150	5
Muscat, Oman	0.1	1	3	0.3	9
Bangkok, Thailand	6	100	2	1200	1
Rome, Italy	1	5	1	5	8
Istanbul, Turkey	10	100	1	1000	2

According to obtained ranks, destination Bankok can be considered the most risky, followed by Istanbul, Cote d'Azur and Tibet. Moreover, based on Risk ranking scale shown in Table 4, these four destination have very high risk since their risk value is grether than 400. It should be noted that all these destinations have the highest ratings for the consequences of attacks coused by large number of victims and wounded. Sousse, Koh Krabey and Addis Ababa have substantial risk with the same risk value (150). Rome and Muscat are the destinations with acceptable risk.

In the research published in (Katić, Kuzmanović, Makajić-Nikolić, 2017), a survey was performed in which the respondents evaluated the same nine destination in relation to their risk of terrorism perception. The obtained rank of chosen destination was: Sousse, Istanbul, Cote d'Azur, Addis Ababa, Tibet, Muscat, Bangkok, KohKrabey, and Rome. The difference in ranking indicates the fact that preferences towards certain risk factors influence on perceiving the riskiness of a particular destination.

5. CONCLUSION

In this paper, we have shown how Kinney method, originally developed for occupational health and safety analysis, can be applied in the assessment of risks of terrorism in tourism destinations. According to the proposed approach, the probability ratings can be obtained based on the number of terrorist attacks in the past, the consequence rating can be obtained based on the number of victims and wounded people in those attacks and the exposure ratings can be assigned based on the type of the destination i.e. the number of days that tourists usually spend on observed destinations. The approach was illustrated on the example of nine chosen destinations based on historical data.

Comparison of destination ranking obtained by Kinney method with ranking assigned by respondents who directly evaluated the risk of the observed destinations, shows that Kinney method can be adopted by incorporating the preferences towards particular risk factors into analysis. In the further research, Conjoint analysis will be used in order to obtain preferences to probability, consequences and exposure to risk.

REFERENCES

Arturas, S., Jasinskas, E. &Svagzdiene, B. (2015). Risk assessment models in the tourism sector.*Amfiteatru Economic Journal*, 17(39), 836-850.

Aven, T. &Guikema, S. (2015). On the Concept and Definition of Terrorism Risk, *Risk Analysis, 35*(*12*). DOI: 10.1111/risa.12518.

Coaffee, J. (2009). *Terrorism, Risk and the Global city, Towards Urban Resilience,* Ashgate Publishing, Ltd.

Cui, F., Liu, Y. & Chang, Y. (2016). An overview of tourism risk perception, *Nat Hazards*, 82, 643–658. DOI 10.1007/s11069-016-2208-1.

Ezell, B.C., Bennett, S.P., Winterfeldt, D.Von, Sokolowski, J. &Collins, A.J. (2010). Probabilistic Risk Analysis and Terrorism Risk, *Risk Analysis*, 30(4), 575-89. doi:10.1111/j.1539-6924.2010.01401.x.

Fine, W.T. (1971). Mathematical evaluation for controlling hazards. *Journal od Safety Research*.

Fuchs, G. &Reichel, A. (2008). Tourist destination risk perception: the case of Israel. *Journal of Hospitality & Leisure Marketing*, 14(2), 83-108.

Kapuściński, G. & Richards, B. (2016). News framing effects on destination risk perception, *Tourism Management*,57, 234-244. https://doi.org/10.1016/j.tourman.2016.06.017.

Karl, M. (2018). Risk and Uncertainty in Travel Decision-Making: Tourist and Destination Perspective. *Journal of Travel Research,* 57(1), 129–146. doi: 10.1177/0047287516678337.

Katić, A, Kuzmanović, M, Makajić-Nikolić, D (2017) Analysis of terorism risk influence on a touristic destination selection. Chapter in International Thematic Monograph - Thematic Proceedings: Modern Management Tools and Economy of Tourism Sector in Present Era, 457-471 (in Serbian)

Kinney, G.F. &Wiruth, A.D. (1976).*Practical Risk Analysis for Safety Management*. Naval Weapons Center, China Lake, California 93555.

Leslie, D. (1999). Terrorism and tourism: the Northern Ireland situation - a look behind the veil of certainty. *Journal of Travel Research*, 38(1), 37-40.

Marhavilas, P.K.& Koulouriotis, D.E. (2008). A risk estimation methodological framework usig quantitative assessment techniques and real accidents data: Application in an aluminum extrusion industry. *Journal of Loss Prevention in the Process Industries*, 21(6), 596 - 603.https://doi.org/10.1016/j.jlp.2008.04.009.

Merari, A. (1993). Terrorism as a strategy of insurgency. *Terrorism and Political Violence*, 5(4), 213-51.

Monahan, J. (2015). The Individual Risk Assessment of Terrorism: Recent Developments. Virginia *Public Law and Legal Theory Research Paper*, 57, http://dx.doi.org/10.2139/ssrn.2665815.

Morakabati, Y.&Kapuściński, G. (2016).Personality, Risk Perception, Benefit Sought and Terrorism Effect. *Internationa Journal of Tourism Research*,18(5),506–514. doi: 10.1002/jtr.2068.

Moraru, R. I. (2012). Current trends and future developments in occupational health and safety risk management. In *Risk Management for the Future-Theory and Cases*. InTech.

Netro, Z. G. C., Romero, E. D. L. T., & Flores, J. L. M. (2018). Adaptation of the fine-kinney method in supply chain risk assessment. *WIT Transactions on The Built Environment*, *174*, 43-55.

Neumayer, E. (2004). The Impact of Political Violence on Tourism: Dynamic Cross-National Estimation, *The Journal of Conflict Resolution*,48(2), 259-281. doi: 10.1177/0022002703262358.

Richter, L. & Waugh, W. (1986). Terrorism and tourism as logical companions. *Tourism Management* 7,230-238.

Roehl, W.S. &Fesenmaier, D.R. (1992). Risk perception and pleasure travel: an exploratory analysis. *Journal of Travel Research*, 30,17-26.

Schmid, A.P. (1992). The response problem as a definitional problem, in Schmid, A.P. and Crelinsten, R.D. (eds), *Terrorism and Political Violence, SpecialEdition on Western Responses to Terrorism*(London: Frank Cass), 7-13.

Shaw, G.K. (2010). *A risk management model for the tourism industry in South Africa*. PhD. North-West University. Available at: http://lib.agu.edu.vn/images_old_lib/image_content_lib/tourimsindustry.pdf.

Sönmez, S.(1998). Tourism, terrorism and political instability. *Annals of Tourism Research*, 25(2),416-456.

Waugh, L.Jr. (1983). The values in violence: organizational and political objectives of terrorist groups.*Conflict Quarterly*, 1-19.

Williams, A.M. & Baláž, V. (2015). Tourism Risk and Uncertainty: Theoretical Reflections. *Journal of Travel Research*, 54(3). 271–287. doi: 10.1177/0047287514523334.

Willis, H., Morral, A., Kelly, T, &Medby, J. (2003).*Estimating terrorism risk*. RAND Center for Terrorism Risk Management Policy.

Yang, C.L. & Nair, V. (2014). Risk perception study in tourism: Are we really measuring perceived risk? *Procedia - Social and Behavioral Sciences*, 144, 322 – 327. https://doi.org/10.1016/j.sbspro.2014.07.302.

UNCERTAINTY IN DESIGN AND MANAGEMENT OF HUMANITARIAN LOGISTICS NETWORKS: A LITERATURE REVIEW

ŞEBNEM YILMAZ BALAMAN[1], MELIKE DIRICAN[1]

[1] Department of Industrial Engineering, Dokuz Eylül University, Tınaztepe Campus, 35160,İzmir,Turkey

Abstract: *In this study, a comprehensive literature survey is conducted compiling the modelling approaches and methods in the context of Operations Research developed and employed to capture uncertainties in the design and management of the Humanitarian Logistics (HL) Networks in Disaster Management Life Cycle (DMLC). The trends in this research field are explored and the research gaps that shed light on future researches are revealed. Based on the findings, some directions for future research are suggested for developing methods and models to represent and solve real world problems practically.*

Keywords: *Uncertainty management, disaster management life cycle, design and management of humanitarian logistics networks.*

1. INTRODUCTION

Uncertainty on the timing, location and magnitude of a natural disaster, as well as how it impacts the disaster area pose serious challenges for disaster preparedness and mitigation (Salman & Yücel, 2015). According to Das & Hanaoka, (2014), uncertainty arises from several sources in different stages of DMLC. Firstly, there is the uncertainty that stems from the lack of knowledge about the disaster which results in the inability to predict the outcome of possible actions taken during the disaster. The second source of uncertainty is the inability to predict the effect of disaster and particularly identifying the damaged transport link after a hazard. Thirdly, there is the uncertainty arising from inability to predict the number and situation of victims and the degree of relief urgency. According to the literature on DMLC management (i.e. Arnold et al., 2005; Tomasini & Van Wassenhove, 2004; Sahay & Gupta, 2016; Amiri et al., 2013; Salmeron & Apte, 2010; Rawls & Turnsquist, 2010; Chang et al.,2007), uncertainties for HL is classified into two broad groups: disaster uncertainty and parameter uncertainty.

Global warming and climate change that the World have been facing in recent decades, have a significant impact on the frequency and severity of natural hazards. Some hazards can be predicted and this includes avalanches, droughts, famines, hurricanes, and tornadoes, among others. In contrast, some disasters cannot be predicted and this includes earthquakes. World Bank identifies natural disaster hotspots, areas that have relatively high risks of losses from one or more natural hazards (Arnold et al., 2005) and assigns hotspot index for each zone. Arnold et al., (2005) identified that some places in the world are vulnerable to multiple disasters for example, India and New Zealand, are subject to both earthquakes and meteorological disasters. The hotspot index for each location changes with the inclusion of different disasters in the analytical model. One common assumption of disaster occurrence is that a disaster will strike only a single place (Balcik & Beamon, 2008; Mete & Zabinsky, 2010; Huang et al., 2010) and other places will remain unaffected. However, this assumption is not always true, disasters may strike different places simultaneously, or several hazards may successively affect the same place in a short amount of time. For instance, the relief requests of the Pakistan flood (2010) and the Haiti earthquake (2010) overlapped and the relief operation during the Pakistan therefore faced a significant amount of shortages.

Parameter uncertainty, which is further subdivided into six groups: demand uncertainty, supplier uncertainty, demand location, affected area, transportation network and task uncertainty. Parameter uncertainty refers to the lack of knowledge about the system conditions, parameters or the model considered. After a disaster, there may be uncertainties about; the number and situation of affected people, who will supply the demand, where will be the affected area, how we will transport the needs to the affected area and where will the needs be stored. Different parameters can be categorized as; (1)Demand uncertainties, (2)Supply uncertainties, (3)Uncertainties about demand location, (4)Uncertainties about affected area, (5)Uncertainties about transportation network, (6)Task uncertainties.

Most of the studies in DMLC have focused on demand uncertainty. The demand for required relief goods (such as first aid suppliers, response equipment, potable water, medicines, disease treatments and required relief workers for handling commodities) in humanitarian operations can be determined regarding the number of people in affected population. In real life, humanitarian operations, it is often seen that the uncertainty of demand is higher during the first stage of disaster response (Mert & Adivar, 2010). Demand estimation is a crucial task after of a disaster. The complexity in demand assessment arises from determining which relief goods are needed, how much of each one is needed and who needs them. Demand uncertainty prevails in both pre-disaster and post-disaster phases. It is always difficult to identify the location of victims and requirements. Jia et al., (2007) states that the efficient planning after a disaster may be restricted by the stochastic situations arising from demand uncertainty. Uncertainty in the cost of operations generally arises from the uncertainty associated with routes, suppliers, etc. According to Davis et al. (1993), demand uncertainty, is the most important uncertainty source and is caused by inaccurate assessments and volatile structure of demand. For example, in the Haiti earthquake, during the first 2–3 days estimates of the numbers of victims ranged from 30,000 to 100,000 (Martinez et al., 2010).

Supply uncertainty mainly stems from the decrease in the supply capability of suppliers after the disaster or vagueness in the total amount of donation. Partly or entirely loss of supplied goods at storage locations has also been considered as a source of uncertainty. In supply category, uncertainties regarding product quality and production capacity are also included. Another reason of the uncertainty in supply is the variability in the suppliers' deliveries arising from the faults or delays in the supply operations. It is often unknown which resources are available, and even the involvement and contribution of suppliers are unpredictable (Tomasini & Van Wassenhove, 2004).

Uncertainties about the affected areas and demand locations are similar to each other. They include the magnitude of damages to supply stores which is related to the severity of the disaster. For disasters, such as hurricanes, more information based on historical data and models can help to predict the path of potential affected areas after it starts, but even a specific storm can change paths. Affected areas might also be dynamic in the case of an epidemic, hence the planning activities should account for this. Location uncertainty imposes additional challenges on preparedness activities such as relief supply, equipment pre-positioning and infrastructure investment (Sahay & Gupta, 2016). Transport network uncertainties represent the most common issue in relief distribution and are crucial for humanitarian operations. Network information may not be readily available after a disaster and therefore it may take several days to obtain route-maps. Transportation network uncertainties mainly arise from the deficiency in the capacity, reliability and availability of links.

Task uncertainty is defined as "the difference between the amount of information required to perform the task and the amount of information already possessed" and "the absence of information" to perform a task and interpretation (Galbraith, 2003). Becerra et al. (2008), assess task uncertainty into five dimensions: novelty, analysability, amount of information, urgency, and impact. Among these, amount of task information is an important dimension of task uncertainty because the multiplicity of meaning conveyed by information lends itself to different and conflicting interpretations about the work context. Hence, the right amount of information is needed to successfully perform the task by considering the intrinsic equivocality of the task (Becerra et al., 2008). Task uncertainty has a direct relationship with information and knowledge availability. In parallel with the increase in the information, uncertainty decreases. Task novelty is often described as unexpected and novel events that occur while performing a task. Task analysability is the degree to which the task is structured workers can easily follow unambiguous processes to solve task-related problems if the analysability of the task is high. Task urgency and task impact refers to the degree to which the job has a substantial impact on the lives of other people, whether those people are in the affected area or in the world. While task urgency focuses on the immediate priority and timeframe that the task is needed to be done, task impact refers to the analysis and assessment of the extent of potential repercussions to prioritize the tasks (Galbraith, 2003).

2. METHODS OF HANDLING UNCERTAINTY IN DMLC

2.1. Risk Mapping and Probabilistic Measures Representing Risk

Risk maps originated from the models of disease transmission based on spatial and temporal data. These models incorporate, to varying degrees, epidemiological, entomological, climatic and environmental information (Liberatore et al., 2013). As emphasized in Huang et al., (2007), a risk map of natural disasters is "an atlas of a community or geographical zone that identifies the places and the buildings that might sustain heavy damages caused by natural disasters".

Risk maps are crucial for pre-event activities as mitigation and preparedness, being usually inputs for decision. The objective of risk mapping is to identify the probability of the occurrence of a specific hazard, as well as the variability at risk, intensity and areas of impact, within a period of time (Liberatore et al., 2013). Risk maps also describe in detail all the areas of humanitarian aids in which the risk of corruption arises. The risk map is parameterized with the localization on the map and can be connected with the historical data of the risk events. Different risk events can be related (e.g. earthquakes and landslides) and must treated in an aggregated way. Integration has been addressed also regarding resource allocation at strategic, tactical and operational ways. Relief facilities must be designed and relief resource must be allocated often for different risks (D'Uffizi et al., 2015). Risk mapping is used in order to calculate the quantity of population needs based on the size of the city and risk maps at response activity. Historical data and the risk mapping conducted by municipalities that comprehend the number of households in a vulnerable situation and, the other axis, according to the magnitude of a disaster multiplied by a weighting value (Balcik and Beamon, 2008). Many countries have national regulations forcing to develop risk maps, which are obtained by governmental units (national or local) on environmental or safety considerations. Risk maps are crucial for pre-event activities such as mitigation and preparedness, and are usually utilized as inputs for decision aid models. They are used over the network, so a map is included, but they are useful in a post-event phase more than for pre-event risk management (Liberatore et al., 2013). Some researchers have used probabilistic measures to represent and include risk in theri studies. Some of the reviewed studies are summarized In the Appendix.

2.2. Probability Distributions and Stochastic Programming

Stochastic Programming (SP) is an approach for modeling optimization problems when the parameters are uncertain, but assumed to lie in some given set of possible values following a probability distribution. In such a case, the decision maker seek a solution that is feasible for all possible parameter choices according to the considered probability distribution and optimizes the expected value of a given function of the decisions and the random variables. The goal of Stochastic programming is to find a solution that will perform well under any possible realization of the random parameters (Synder et al., 2007). The most widely applied and studied stochastic programming models are two-stage linear programs. Stating that two-stage SP allows for modeling uncertainties and time-dependent decisions effectively, Grass and Fischer, (2016) presented a literature survey on the modeling and solution approaches for two-stage stochastic programming. Two-stage possibilistic-stochastic models can be utilized to design the relief network by deciding on pre-positioning and distribution of emergency supplies while taking into account pre- and post-disaster events (Tofighi et al.,2016). Stochastic programming models usually are large dimension models, implying the use of advanced mathematical programming techniques. Perhaps this is one of the reasons, in addition to its good fit to the problem characteristics, to explain the large number of academic articles devoted to this methodology in the last years into the area of disaster management (Liberatore et al., 2013). Using SP is meaningful only when a certain action can be repeated several times. However, due to special characteristics of disasters, in most cases there is not enough historical/objective data to model uncertain parameters within each scenario as random data (Tofighi et al., 2016).

SP is an appropriate tool for planning in the preparedness phase due to its ability to handle uncertainty by probabilistic scenarios representing disasters and their outcomes . In disaster management literature, SP has been a successful way to handle uncertain parameters and conditions in many applications. Some of the reviewed studies are summarized In the Appendix. Some of the other studies that applied SP in disaster management are; Barbarosoglu & Arda, 2004; Beraldi et al., 2004; Chang et al., 2007; Cormican et al., 1998; Morton, 2007; Pan et al., 2003.

2.3. Robust Optimization

In uncertainty situations, parameters are uncertain, and furthermore, the information about probabilities may be unknown. In this case, problems under uncertainty are known as robust optimization problems and to solve these problems, researchers often attempt to optimize the worst-case performance of the system (Synder et al., 2007). Robust Optimization (RO) is one of the predominant approaches to solving linear optimization problems with uncertain data (Najafi et al., 2013). In contrast with stochastic optimization which models the uncertainty through a probabilistic description, robust optimization models the possible set of values, but nothing is said about their probabilities (Najafi et al., 2013). By robust optimizaiton, the decision-maker constructs a solution that is admissible in some sense through a set of scenarios, instead of seeking to immunize the solution in some probabilistic sense to stochastic uncertainty. Robust optimization can be especially suitable in absence of data, or when there is no interest to give more importance to some values of the parameter than to others (Liberatore et al., 2013).

Mulvey et al., (1995) defined two measures of robustness: a solution to an optimization model is defined as solution robust if it remains 'close' to optimal for all scenarios of the input data, and is model robust if it remains 'almost' feasible for all data scenarios. They also emphasized that, robust optimization explicitly incorporates the conflicting objectives of solution and model robustness by using a parameter reflecting the decision maker's preference and incorporates control and design variables. Control variables are subject to adjustment once a specific realization of the data is obtained, while design variables are determined before realization of the uncertain parameters and cannot be adjusted once random parameters are observed. According to Jabbarzadeh et al., (2014) the constraints are divided into two groups: structural and control constraints. Structural constraints are typical linear programming constraints which are free of uncertain parameters, while the coefficients of control constraints are subject to uncertainty. Some of the reviewed studies are summarized In the Appendix.

2.4. Simulation Models

Simulation modeling is the process of creating and analyzing a digital model of a physical model to predict its performance in real case, by replicating the real system's behaviour. The most extended use of simulation models is to represent dynamic systems, but especially when including uncertainty, static systems also are studied through simulation in order to obtain a simulated sample of the output. Techniques to obtain values of random variables to represent risky parameters are known as Monte Carlo simulation. In spite of having several shortcomings, simulation models are still widely used in modelling some humanitarian logistic problems under uncertainty, such as the egress problem (Liberatore et al., 2013). Simulation models have also been widely used to incorporate the uncertainties related to evacuees' behaviours on the evacuation process. Traffic simulation tools are used to observe the impacts of lane reversals, intersection delay management, and traffic direction reversals on traffic flows. Some models consider contraflows of rescuers (Özdamar et al., 2015). Some of the reviewed studies are summarized In the Appendix.

2.5. Fuzzy Sets

Fuzzy programming is another tool available for addressing optimization problems in the presence of uncertainty. Fuzzy logic is a form of mathematics that can be used to represent imprecision in the variables of a mathematical model. Zadeh, (1965) first introduced the use of fuzzy relationships (known as fuzzy sets) to represent vague and/or imprecise concepts. From that point onwards, fuzzy logic has developed into a very robust and widely applied approach for concept representation. Fuzzy programming, uncertainty is modeled using fuzzy numbers and fuzzy sets (such as the ones we just described) rather than discrete or continuous probability functions (Falasca, 2009).

Fuzzy logic is a form of many-valued logic in which the truth values of variables may be any real number between 0 and 1. Using fuzzy logic, it is possible to process objects that partially belong to more than one category. There is no need to generalize reality to fit it into classes; rather the degree of membership to the category is given. Note that, due to linguistic and numeric requirements, the fuzzy-modeling process has generally to deal with an important trade-off between the accuracy and the interpretability of the model. In other words, the model is expected to provide high numeric precision while incurring into a loss of linguistic descriptive power as little as possible (Liberatore et al., 2013). For years, fuzzy modeling has been one of the most relevant issues in qualitative analysis, given the linguistic capabilities of fuzzy logic, but now it is the base of many decision support models whose output is quantitative. Some of the reviewed studies are summarized In the Appendix.

3. DISCUSSION AND CONCLUSIONS

Researchers mostly concentrated on the uncertain system parameters rather than uncertainty in disaster itself while handling uncertainty in HL models, because parameter uncertainty is more quantifiable and easy to capture in the analytic studies than disaster uncertainty. Mostly uncertainty in the demand and supply is considered in modelling the problems in pre-disaster phase, while the transportation network, demand and supply uncertainties are the mostly considered ones for post-disaster problems. Very few researchers have considered the uncertainty related to the affected area in their studies. Although, it is concluded that task uncertainty is not captured sufficiently by the modelling approaches in the articles that are covered in the this literature study, some researchers emphasized that task uncertainty is a crucial uncertainty source that cause chaos during and just after the disaster because the stakeholders does not exactly know what to do during and after a disaster. A total of 134 studies that include modelling approaches and methods are examined and recognised that some researchers develop and use modelling approaches on a case study to capture the demand uncertainty, with an exact method, by considering some assumptions based on the demand data from the historical disaster cases. Some researchers did not consider some important details in the transportation route and network design; roads, transportation vehicles and networks may also be damaged, especially during a natural disaster. There are very few studies that take into account this case to model the disaster response or their preparedness strategy.

In total 134 modelling studies, the number of models without handling uncertainty is 59 and the number of models that capture uncertainty is 75. The uncertainty is mostly handled by using fuzzy numbers, probability values, scenario approaches, Monte Carlo Simulation and sensitivity analysis. It is also concluded that uncertainty is mostly handled by stochastic studies with the probabilistic values and scenario approaches in the literature. Scenario approach is applied in 29 HL model and probability values are considered in 30 HL models. Fuzzy set theory is also utilized in the literature in order to capture uncertainty, fuzzy modelling is used in 8 HL modelling studies. Sensitivity analysis is used in 6 HL modelling studies, in order to understand the sensitivity of the model performance according to the change in demand. It is concluded that Monte Carlo simulation was not as much prefered as the other approaches to capture uncertainty, which is used in 2 HL modelling studies. The distribution of the methods of handling uncertainty in the reviewed studies is depicted in Figure 1.

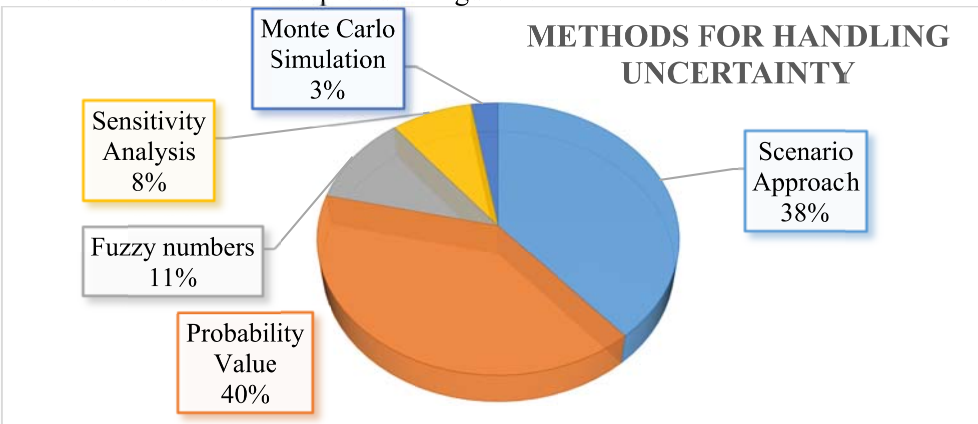


Figure 1. The distribution of the methods of handling uncertainty in the reviewed studies

REFERENCES

Adıvar, B.,& Mert, A.(2010). International disaster relief planning with fuzzy credibility. Fuzzy Optimization and Decision Making, 9(4), 413–433.

Arnold, M., Dilley, M., Deichmann, U., Chen, R.S., & Lern er-Lam, A.L. (2005) Natural disaster hotspots: A Global Risk Analysis. Washington: World Bank.

Balcik, B. & Beamon, B.M.(2008). Facility location in humanitarian relief. International Journal of Logistics Research and Applications,11(2), 101–121

Barbarosoglu, G. & Arda,Y.(2004) A two-stage stochastic programming framework for transportation planning in disaster response. Journal of the Operational Research Society, 55(1), 43–53.

Beraldi, P., Bruni, M. E., & Conforti, D. (2004). Designing robust emergency medical service via stochastic programming. European Journal of Operational Research, 158(1), 183-193.

Bozorgi-Amiri, A., Jabalameli, M. S., Alinaghian, M., & Heydari, M. (2012). A modified particle swarm optimization for disaster relief logistics under uncertain environment. The International Journal of Advanced Manufacturing Technology, 60(1-4), 357-371.

Bozorgi-Amiri, A., Jabalameli,M. S., & Mirzapour Ale-Hashem, S. M. (2013).A multi-objective robust stochastic programming model for disaster relief logistics under uncertainty. OR Spectrum, 35(4), 905–933.

Chang, M. S., Tseng, Y. L., & Chen, J. W. (2007). A scenario planning approach for the flood emergency logistics preparation problem under uncertainty. Transportation Research Part E: Logistics and Transportation Review, 43(6), 737-754.

Chen, X., & Zhan, F. B. (2003). Agent-based simulation of evacuation strategies under different road network structures. University Consortium of Geographic Information Science.

Chen, C. C., & Chou, C. S. (2009). Modeling and performance assessment of a transit-based evacuation plan within a contraflow simulation environment. Transportation Research Record: Journal of the Transportation Research Board, (2091), 40-50.

Church, R. L., & Sexton, R. M. (2002). Modeling small area evacuation: Can existing transportation infrastructure impede public safety?

Cormican, K. J., Morton, D. P., & Wood, R. K. (1998). Stochastic network interdiction. Operations Research, 46(2), 184-197.

Davis, E. Menegazzi, J. J., A., Sucov, A. N., & Paris, P. M. (1993). Reliability of the Glasgow Coma Scale when used by emergency physicians and paramedics. Journal of Trauma and Acute Care Surgery, 34(1), 46-48.

D'Uffizi, A., Simonetti, M., Stecca, G., & Confessore, G. (2015). A Simulation Study of Logistics for Disaster Relief Operations. Procedia CIRP, 33, 157-162.

Das, R., & Hanaoka, S. (2014). Relief inventory modelling with stochastic lead-time and demand. European Journal of Operational Research, 235(3), 616-623.

El-Anwar, O., El-Rayes, K., & Elnashai, A. (2009). An automated system for optimizing post-disaster temporary housing allocation. Automation in Construction, 18(7), 983-993.

Erick, M., Suppasri, A., Imamura, F., & Koshimura, S. (2012). Agent-based simulation of the 2011 great east japan earthquake/tsunami evacuation: An integrated model of tsunami inundation and evacuation. Journal of Natural Disaster Science, 34(1), 41-57.

Falasca, M. (2009). Quantitative decision models for humanitarian logistics (Doctoral dissertation, Virginia Tech).

Grass, E., & Fischer, K. (2016). Two-stage stochastic programming in disaster management: A literature survey. Surveys in Operations Research and Management Science.

Günneç, D., & Salman, F. S. (2011). Assessing the reliability and the expected performance of a network under disaster risk. OR spectrum, 33(3), 499-523.

Homma, S., Fujita, K., Ichimura, T., Hori, M., Citak, S., & Hori, T. (2014). A physics-based Monte Carlo earthquake disaster simulation accounting for uncertainty in building structure parameters. Procedia Computer Science, 29, 855-865.

Huang, T., Tzeng, G., &Cheng, H., (2007). Multi-objective optimal planning for designing relief delivery systems. Transportation Research Part E Logistics and Transportation Review, 43, 673–686.

Huang, R., Kim, S., & Menezes, M. B. (2010). Facility location for large-scale emergencies. Annals of Operations Research, 181(1), 271-286.

Jabbarzadeh, A., Fahimnia B., &Seuring, S. (2014). Dynamic supply chain network design for the supply of blood in disasters: A robust model with real world application, Transportation Research Part E 70, 225–244.

Jia, H., Ordóñez, F., & Dessouky, M. M. (2007). Solution approaches for facility location of medical supplies for large-scale emergencies. Computers & Industrial Engineering, 52(2), 257-276.

Jiang, W., Deng, L., Chen, L., Wu, J., & Li, J. (2009). Risk assessment and validation of flood disaster based on fuzzy mathematics. Progress in Natural Science, 19, 1419–1425.

Jotshi, A., Gong, Q., & Batta, R. (2009). Dispatching and routing of emergency vehicles in disaster mitigation using data fusion. Socio-Economic Planning Sciences, 43(1), 1–24.

Leskens, J. G., Brugnach, M., Hoekstra, A. Y., & Schuurmans, W. (2014). Why are decisions in flood disaster management so poorly supported by information from flood models?. Environmental Modelling & Software, 53, 53-61.

Liberatore, F., Pizarro, C., de Blas, C. S., Ortuño, M. T., & Vitoriano, B. (2013). Uncertainty in HLfor disaster management. A review. In Decision Aid Models for Disaster Management and Emergencies, 45-74.

Mert, A., & Adıvar, B. (2010). International disaster relief planning with fuzzy credibility. Fuzzy Optimization and Decision Making, 9(4), 413-433.

Mete, H. O., & Zabinsky, Z. B. (2010). Stochastic optimization of medical supply location and distribution in disaster management. International Journal of Production Economics, 126(1), 76-84.

Morton, J. F. (2007). The impact of climate change on smallholder and subsistence agriculture. Proceedings of The National Academy of Sciences, 104(50), 19680-19685.

Mulvey, J. M., Vanderbei, R. J., & Zenios, S. A. (1995). Robust optimization of large-scale systems. Operations Research, 43(2), 264-281.

Nagurney, A., & Nagurney, L. S. (2016). A mean-variance disaster relief supply chain network model for risk reduction with stochastic link costs, time targets, and demand uncertainty. In Dynamics of Disasters Key Concepts, Models, Algorithms, and Insights, 231-255. Springer International Publishing.

Najafi, M., Eshghi, K., & Dullaert, W. (2013). A multi-objective robust optimization model for logistics planning in the earthquake response phase. Transportation Research Part E: Logistics and Transportation Review, 49(1), 217-249.

Nolz, P. C., Semet, F., & Doerner, K. F. (2011). Risk approaches for delivering disaster relief supplies. OR spectrum, 33(3), 543-569.

Noreña, D., Akhavan-Tabatabaei, R., Yamín, L., & Ospina, W. (2011, December). Using discrete event simulation to evaluate the logistics of medical attention during the relief operations in an earthquake in Bogota. In Simulation Conference (WSC), Proceedings of the 2011 Winter, 2661-2673. IEEE.

Özdamar, L., & Ertem, M. A. (2015). Models, solutions and enabling technologies in humanitarian logistics. European Journal of Operational Research, 244(1), 55-65.

Pan, F., Charlton, W. S., & Morton, D. P. (2003). A stochastic program for interdicting smuggled nuclear material. In Network interdiction and stochastic integer programming, 1-19. Springer US.

Rath, S., Gendreau, M., & Gutjahr, W. J. (2015). Bi-objective stochastic programming models for determining depot locations in disaster relief operations. International Transactions in Operational Research.

Rawls, C., Turnquist, M. A. (2010). Pre-positioning of emergency supplies for disaster response. Transportation Research Part B, 44, 521-534.

Rennemo, S. J., Rø, K. F., Hvattum, L. M., & Tirado, G. (2014). A three-stage stochastic facility routing model for disaster response planning. Transportation research part E: logistics and transportation review, 62, 116-135.

Rezaei-Malek, M., & Tavakkoli-Moghaddam, R. (2014). Robust humanitarian relief logistics network planning. Uncertain Supply Chain Management, 2(2), 73-96.

Sahay, B. S., Gupta, S., & Menon, V. C. (2016). Managing Humanitarian Logistics. Springer India.

Salman, F. S., & Yücel, E. (2015). Emergency facility location under random network damage: Insights from the Istanbul case. Computers & Operations Research, 62, 266-281.

Salmero´n, J., Apte, A. (2010). Stochastic optimization for natural disaster asset prepositioning. Production and Operations Management, 19(5), 561-574.

Snyder, L. V., Daskin, M. S., & Teo, C. P. (2007). The stochastic location model with risk pooling. European Journal of Operational Research, 179(3), 1221-1238.

Tofighi,S., Torabia,S.A., & Mansourib, S.A. (2016). HLnetwork design under mixed uncertainty, European Journal of Operational Research, 250, 239–250.

Tomasini, R. M., & Van Wassenhove, L. N. (2004). Pan-American health organization's humanitarian supply management system: de-politicization of the humanitarian supply chain by creating accountability. Journal of public procurement, 4(3), 437.

Tzeng, G. H., Cheng, H. J., & Huang, T. D. (2007). Multi-objective optimal planning for designing relief delivery systems. Transportation Research Part E: Logistics and Transportation Review, 43(6), 673-686.

APPENDIX

Risk Mapping and Probabilistic Measures Representing Risk	
Günneç & Salman (2011)	Proposed eight probabilistic measures of connectivity and expected travel time/distance between critical elements of the network. The methodology is based on the generation of scenarios and a special algorithm is proposed.
D'Uffizi et al. (2015)	Worked on a national research project aimed at developing cloud-based systems and sensor networks for multi-risk management. The developed simulation system can provide crucial help to the rescuers in order to planning the best relief strategy for mitigate the effects of natural and man made disasters.
Jiang et al., (2009)	Focused on risk assessment and validation of flood disasters because of the difficult in the natural disaster field. Due to the lack of risk maps and division standards of flood disasters in Malaysia, they find an optimal assessment according to the proportion of validation data in the higher and highest risk zones.
Nolz et al., (2011)	Focused on the problem of designing a distribution system in a post-disaster context. The authors proposed a multi-

	objective model considering objective functions on risk, coverage, and total distribution time. In particular, risk measures represent the probability that a path might become.
Probability Distributions and Stochastic Programming	
Mete & Zabinsky, (2010)	Proposed a stochastic optimization approach for the storage and distribution problem of medical supplies to be used for disaster management under a wide variety of possible disaster types and magnitudes. They developed a stochastic programming model to select the storage locations of medical supplies.
Rath et al., (2016)	Proposed a two-stage bi-objective stochastic programming model for disaster relief operations considering a problem that occurs after a natural disaster. The transportation system for supplying disaster victims with relief goods is modeled by a discrete set of scenarios considering uncertainty in the acessibility of the road network.
Rawls & Turnquist, (2010)	Presented a two-stage stochastic programming model for determining pre-positioning storage locations and capacities, as well as allocating quantities of various emergency commodities to be stored at each location under demand uncertainty.
Nagurney et al., (2016)	Developed a mean-variance disaster relief supply chain network model with stochastic link costs and time targets for delivery of the relief supplies at the demand points, under demand uncertainty.
Rennemo et al., (2014)	Investigated a stochastic facility routing problem, and propose a three-stage mix integer stochastic programming model with a linear utility objective function to capture the fairness of emergency distribution.
Robust Optimization	
Jotshi et al., (2009)	Developed a robust optimization methodology for the dispatching and routing of emergency vehicles in a post-disaster environment.
Mas et al., (2010)	Presented a min–max robust multi-point, multi-vehicle transportation model to minimize the maximum rescue time for moving injured people. In the model, it is assumed that the distances between affected area and medical centers are uncertain.
Najafi et al., (2013)	Proposed a novel mathematical model for assisting disaster managers in scheduling the logistical activities for disaster relief materials and injured people according to demand and supply uncertainties, and solving it by means of a robust approach for stochastic models.
Bozorgi -Amiri et al., (2013)	Provided a multi-objective robust contingency planning model. They studied demand, supply and cost as non-deterministic parameters. The purpose of their model is to minimize the expected variance of relief chain costs and maximize coverage level of affected areas.
Rezaei-Malek & Tavakoli-Moghaddam, (2014)	Explained a bi-objective robust integer model for Humanitarian relief logistics and estimated optimal amount of relief commodities and optimal place of warehoused by taking the two variables of price and repose into account, simultaneously.
Simulation Models	
Chen & Chou, (2009)	Dealt with the assignment of buses to pickup points in a nonotice evacuation. Simulation with contraflows is used to solve the problem so that traffic speed, system travel time, and network clearance time are minimized.
Chen & Zhan, (2003)	Used an agent-based microsimulation model to estimate minimum evacuation clearance time and the number of evacuees who will need to be accommodated in case of the route disruption.
Church & Sexton, (2002)	They investigated the impact of different evacuation scenarios on evacuation time. They applied different traffic control plans considering evacuation scenarios including alternative exits, changing number of vehicles.
Leskens et al., (2014)	Analyzed the use of flood simulation models in flood disaster management, which takes place from about 1 to 5 days in advance of a potential flood. Specifically in this period, the potential consequences of a flood can be importantly reduced, for example, by reinforcements of dikes or evacuation of people.
Homma et al., (2014)	Described a Monte Carlo Simulation approach for physics-based earthquake disaster simulations, which accounts for uncertainty in the input parameters of the building structures. They created a stochastic structures model and examined the influence of the uncertainties in the parameters on the maximum story drift angle.
El-Anwar et al., (2009)	Developed an automated system to support decision-makers in optimizing postdisaster temporary housing arrangements. The system considers as objectives the minimization of negative socio-economic impacts, maximization of housing safety, minimization of negative environmental impacts and the minimization of public expenditures.
Norena, et al., (2011)	Developed a simulation model for the transfer of injured people to temporary and permanent hospitals after an earthquake scenario, considering triage classifications and hospital capacities.
Simpson & Hancock, (2009)	Applied simulation to the case of resource allocation in an emergency response system.
Erick et al., (2012)	Applied a newly developed evacuation model integrated with the numerical simulation of tsunami for casualty estimation. This tool is to support decisions in disaster management and disaster prevention education.
George, Paul, & Lin, (2010)	Proposed a simulation model to present the hospitals during an earthquake disaster scenario, and with the obtained results they fitted generalized regression equations in order to obtain steady-state hospital capacities during this event.
Fuzzy Sets	
Adivar & Mert, (2010)	Optimized a relief distribution system using fuzzy modelling in which relief items supplied from donor countries and collected at collection points and, then, shipped to points of delivery in disaster affected countries.
Tzeng et al., (2007)	Proposed a fuzzy multiple objective model that concentrates on the effectiveness and fairness of the overall distribution system, to avoid the oversight of critical but difficult-to-reach areas in the real world.
Xie & Hu, (2009)	Modeled an inventory routing problem in emergency logistics with fuzzy demands, which are converted to deterministic demands by Yager's fuzzy number ranking method and solved using a heuristic algorithm that coordinates inventory and route optimization for minimizing total cost.
Bozorgi-Amiri et al., (2012)	Investigated a relief chain design problem where the uncertainty in demands, supplies and the cost of procurement and transportation are considered. They formulated the problem as a multi-objective robust stochastic programming model, and developed a fuzzy PSO algorithm for solving the problem.

REVIEW OF FORECASTING AND P2P MODELS IN ELECTRICITY TRADING

MINJA MARINOVIĆ

University of Belgrade, Faculty of Organizational Sciences, marinovic.minja@fon.bg.ac.rs

Abstract: *The last three decades have been very progressive for electricity markets, since they have been transformed from being highly controlled to being deregulated. The main objective of the reforms during this period has been to introduce competition in the power industry and to give more options to the market participants. However, it also increases business risks. The literature on different aspects of this subject is extensive and gives a variety of different approaches and solutions that can be used in reducing business risks. The aim of this paper is to present some of the research aimed at forecasting the electricity price, as well as research into peer (P2P) electricity trading. The review in this paper covers the last five years of research.*

Keywords: *Electricity trade, Electricity price forecasting, P2P energy trading.*

1. INTRODUCTION

The traditional and integrated electric utility structure has been deregulated and replaced by a competitive market scheme in many countries around the world. With deregulation and increased competition, participants in the power market are facing new challenges every day. Today, electricity trading has indeed transformed from primarily being a technical business, to being one in which the product is treated in the same way as any other commodity. There are a variety of models, approaches and proposals for the electricity trade, but in this paper the focus is on electricity price forecasting and peer to peer electricity trading. This paper reviews the literature on several aspects and goals but with the same keyword, electricity trading.

The literature concerning issues related to energy trading and transmission is extensive. One of the first papers dealing with electricity trading and its economic prospects is by Dowlatabadi and Evans (1986). Some basic characteristics of electricity trading have been presented by Srinivasan (2013). Among the many approaches to and aspects of electricity trading, recent research shows that the current focus is based on: finding cleaner energy sources (Schusser and Jaraitė, 2018); finding a way to integrate variable renewable electricity into the electricity market (Hu et al., 2018) following the new regulations; and finding an optimal solution for maximizing profit by using a particular technique, for example game theory (Peng and Tao, 2018). Also, in a new era of blockchain technology, finding an application for this technology in energy trading has been inevitable (Mihaylov et al., 2014). The first application of blockchain technology in the field of electricity trading and the P2P market was presented by Sikorski et al. (2017).

The paper is organized as follows: after the introduction, Section 2 is devoted to forecasting electricity price models. Section 3 refers to peer to peer (P2P) approaches and models. The conclusions and directions for further research are given in Section 4.

2. FORECASTING THE ELECTRICITY PRICE

In the electricity market, the electricity price plays an inevitable role. In the context of the liberalized and deregulated electricity markets, price forecasting has become increasingly important for the plans and market strategies of energy companies.

While the energy demand is increasing in all regions of the world, fossil fuels are still the dominant option for world energy production (more than 80%). The rest is produced from renewable energy sources like small wind turbines and others. One of the challenges for optimizing electricity trading and price forecasting is to include the energy produced from Renewable Energy Sources (RESs) since this has become obligatory. This directly causes environmental risks, as well as degeneration and energy losses in power transmission due to the long physical distances between the generation and consumption sites (Jiang et al., 2016). The increasing integration of renewable energy sources (RES) into the energy system provides a solution to this environmental energy dilemma. Nevertheless, uncertainty and fluctuations in renewable generation of energy need to be taken into account.

Electricity price forecasting techniques and models can be divided into several categories (Girish, 2012):

- Long-term price forecasting - for investment profitability analysis and planning
- Medium-term forecasting - for balance sheet calculations, risk management and derivatives pricing
- Short-term price forecasting - for the participants of auction-type spot markets in which participants are requested to express their bids in terms of prices and quantities

Electricity price forecasting techniques in the literature can be broadly divided into (Girish, 2012):

- Production-cost (or cost-based) models
- Equilibrium (or game theoretic) approaches
- Fundamental (or structural) methods
- Quantitative (or stochastic, econometric, reduced-form) models
- Statistical (or technical analysis) approaches
- Artificial intelligence-based (or non-parametric) techniques

In the last few years many papers have considered the subject of electricity price forecasting. Two of the best review papers on this subject are by Weron (2014) and Murthy et al. (2014). This paper will only cover the papers on this subject from 2014 to 2018.

The price forecasting literature is typically divided into five areas: (1) multi-agent or game theory models simulating the operation of market agents;, (2) fundamental methods employing physical and economic factors; (3) reduced-form models using the statistical properties of the electricity trade for risk and derivatives evaluation; (4) statistical models comprising time series and econometric models; and (5) artificial intelligence methods (Weron, 2014). Table 1 presents the electricity price forecasting papers from 2014 to the present day.

The table 1 shows some of the latest research and models for electricity price forecasting. A one dimensional input featuring a feed-forward neural network model is presented in order to achieve an accurate prediction tool for forecasting the electricity price (Anbazhagan and Kumarappan, 2014). Itaba and Mori (2017) proposed a fuzzy-preconditioned ANN (Artificial Neural Network) model for electricity price forecasting. In contrast, Wang et al. (2014) created a hybrid forecasting model for the day-ahead electricity price. Neural networks are one of the most commonly used models in forecasting electricity prices. For example, Logoa et al. (2018) propose two different methods to incorporate market integration into electricity price forecasting and to improve the predictive performance. Abedinia et al., (2015) predict the future values of price data based on a Combinatorial Neural Network engine. Since there are many papers concerning day-ahead price forecasting via the application of artificial neural network based models, there is also a review paper on this subject (Panapakidis and Dagoumas, 2016). Athanasios et al. (2017) presented an integrated risk management model for electricity traders using ANN models and a clustering algorithm.

Quantitative models are also used for forecasting electricity prices. Ziel et al, (2015) created an econometric model for forecasting the price of electricity incorporating renewable energy. Based on the given model, the authors attempted to forecast the spot price of electricity on a given date. Spot price forecasting has also been considered by Girish et al. (2018) in order to forecast the electricity price for five regions in the Indian electricity market. Some authors (Heredia et al. 2018) used a stochastic programming model for determining a unit's optimal generation bid on the wholesale electricity market so that it maximizes the long-term profits of the utility within the given environmental restrictions. The prediction intervals for electricity demand and price, based on functional data and with use of the nonparametric autoregressive model and a partial linear semi-parametric model, are presented by Vilar et al. (2018). Radenkovic et al. (2018) proposed a business intelligence solution for providing a better flow of information for forecasting, data analysis and decision making for electricity market operators. Jiang et al. (2016) created a two stage model for forecasting electricity prices. The first stage is particle swarm optimization, namely core mapping with a self-organizing-map and fuzzy set, and the second stage is the selection rule.

In the framework of competitive electricity markets, power producers and consumers need accurate price forecasting tools. Price forecasts contain crucial information for producers and consumers when planning bidding strategies in order to maximize their benefits and utilities. For both spot markets and long-term contracts in an electricity market, price forecasts are necessary so that a firm/company can develop bidding strategies or negotiation skills in order to maximize its own profit. Competitors who can forecast prices correctly can also make an optimal plan for a production schedule accordingly and hence maximize their profit. Since the day-ahead spot market typically consists of auctions every 24 hours that take place simultaneously one day in advance, forecasting with a lead of anything from a few hours to a few days is of prime importance in day-to-day market operations (Girish, 2012).

Table 1. Latest Studies Related to Electricity Price Forecasting

Title	Authors	Year	Methodology
Day-ahead deregulated electricity market price forecasting using neural network input featured by DCT	Anbazhagan, S., & Kumarappan, N.	2014	Feed-forward neural network
A Hybrid Forecasting Model Based on Bivariate Division and a Backpropagation Artificial Neural Network Optimized by Chaos Particle Swarm Optimization for Day-Ahead Electricity Price	Wang, Z., Liu, F., Wu, J., & Wang, J.	2014	Bivariate division backpropagation artificial neural network method, chaos particle swarm optimization backpropagation artificial neural network method
Efficient modeling and forecasting of electricity spot prices	Ziel, F., Steinert, R., & Husmann, S.	2015	Econometric model
Electricity price forecast using Combinatorial Neural Network trained by a new stochastic search method	Abedinia, O., Amjady, N., Shafie-khah, M., & Catalao, J.P.S.	2015	Combinatorial neural network
Day-ahead electricity price forecasting via the application of artificial neural network based models	Panapakidis, I., P., & Dagoumas, A. S.	2016	Neural network
A Hybrid Multi-Step Model for Forecasting Day-Ahead Electricity Price Based on Optimization, Fuzzy Logic and Model Selection	Jiang, P., Liu, F., & Song, Y.	2016	Hybrid two-stage model
A Fuzzy-Preconditioned GRBFN Model for Electricity Price Forecasting	Itaba, S., & Mori, H.	2017	Artificial neural network, fuzzy logic
An integrated model for risk management in electricity trade	Athanasios, S. D., Koltsaklis, N. E., & Panapakidis, I. P.	2017	Clustering algorithm, neural network
Forecasting day-ahead electricity prices in Europe: The importance of considering market integration	Lago, J., Ridder, F., Vrancx, P., & Schutter, B.	2018	Neural network
Harnessing business intelligence in smart grids: A case of the electricity market	Radenković, M., Lukić, J., Despotović-Zrakić, M., Labus, A., & Bogdanović, Z.	2018	Business intelligence solution (Kimbal methodology)
Prediction intervals for electricity demand and price using functional data	Vilar, J., Aneiros, G., & Raña, P.	2018	Nonparametric autoregressive model, partial linear semi-parametric model
Spot electricity price discovery in Indian electricity market	Girish, G. P., Rath, B. N., & Akram, V.	2018	Unit root tests
Stochastic optimal generation bid to electricity markets with emissions risk constraints	Heredia, J. F., Cifuentes-Rubiano, J., & Corchero, C.	2018	Stochastic programming

3. PEER TO PEER ENERGY TRADING

Peer-to-peer energy trading is one of the most promising trends and innovations in this field. P2P is a network whose members (peers) share some of their own hardware resources and information in order to facilitate certain applications, like for instance file sharing or project collaboration (Schollmeier, 2001). Every participant (peer) is also a provider and receiver of resources and can directly communicate with the others participants without the mediation of an intermediary node, thus enabling the network to continue operations if one or more peers cease to function. When a producer has excess power, it can be stored using various methods (Vazquez, et al., 2010), but since this procedure is still quite expensive and the current technological level does not offer completely efficient storage, losses will occur. If a producer has an extra amount of energy, P2P systems and platforms are a great opportunity to sell the surplus. In order to be a part of a P2P energy trading system, peers must also produce energy. The amount of electricity generated from distributed renewable energy sources is constantly increasing. With the increasing installation of distributed generation on the demand side, more and more consumers are on both sides, since they can generate and consume energy. In order to start with electricity exchange, peers need to connect with each other through a microgrid or smart grid. A smart grid (SG) is an electricity network that can intelligently integrate the actions of all the users connected to it (generators, consumers and those that do both) in order to efficiently deliver sustainable, economic and secure electricity supplies. P2P energy trading cannot be implemented without SG technologies including Information and Communication Technologies (ICT), monitoring, and control functions. A microgrid is a network, in essence a smaller version of the smart grid previously described, of small-scale energy generation units (Markvart, 2006).

Energy sharing models, which specify how the participants exchange and trade energy with each other, are the core of the energy sharing projects. Many studies have been carried out in this field, which can be divided into three categories: 1) energy sharing conducted by one centralized authority; 2) energy sharing achieved by the interaction between an operator (price maker) and a group of prosumers (price takers); and 3) energy sharing achieved by the interaction of a group of prosumers, i.e. P2P energy sharing (Zhou et al, 2017). The following table presents papers relating to P2P electricity trading models and solutions.

Table 2. Latest Studies Related to P2P electricity trading

Title	Authors	Year	Methodology
Multiagent study of smart grid customers with neighborhood electricity trading	Kahrobaee, S., Rajabzadeh, R. A., Soh, L. L., & Asgarpoor, S.	2014	Agent-based model for electricity trading
Adaptive energy efficient scheduling in Peer-to-Peer desktop grids	Tchernykh, A., Pecero, J. E., Barrondo, A., & Schaeffer, E.	2014	Two-objective optimization analysis
A peer-to-peer approach to energy production	Giotitsas, C., Pazaitis, A., & Kostakis, V.	2015	Specific model for P2P energy grid
A rolling horizon optimization framework for the simultaneous energy supply and demand planning in microgrids	Silvente, J., Kopanos, G. M., Pistikopoulos, E. N., & Espuña, A.	2015	Mixed integer linear programming (MILP)
A Bidding System for Peer-to-Peer Energy Trading in a Grid-connected Microgrid	Zhang, C., Wu, J., Cheng, M., Zhou, Y., & Long, C.	2016	Four-layer architecture model
Design and evaluation of P2P overlays for energy negotiation in smart micro-grid	Amato, A., Di Martino, B., Scialdone, M., & Venticinque, S.,	2016	Service level agreement (SLA) based negotiation protocol
Comparative review and discussion on P2P electricity trading	Park, C., & Yong, T.	2017	Review of P2P electricity trading cases
Review of Existing Peer-to-Peer Energy Trading Projects	Zhang, C., Wu, J., Long, C., & Cheng, M.	2017	Review of P2P energy trading projects
Performance Evaluation of Peer-to-Peer Energy Sharing Models	Zhou, Y., Wu, J., Long, C., Cheng, M., & Zhang, C.	2017	Performance evaluation methodology
Feasibility of Peer-to-Peer Energy Trading in Low Voltage Electrical Distribution Networks	Long, C., Wu, J., Zhang, C., Cheng, M., Al-Wakeel, A.	2017	Linear programming
Culture, values, lifestyles, and power in energy futures: A critical peer-to-peer vision for renewable energy	Ruotsalainen, J., Karjalainen, J., Child, M., Heinonen, S.	2017	Energy transitions and social change relationship analysis
Designing microgrid energy markets: A case study: The Brooklyn Microgrid	Mengelkamp, E., Gärttner, J., Rock, K., Kessler, S., Orsini, L., Weinhardt, C.	2018	Blockchain technology
Peer-to-Peer energy trading in a Microgrid	Zhang, C., Wu, J., Zhou, Y., Cheng, M., Long, C.	2018	Hierarchical system architecture model with the use of game theory

Table 2 present a list of papers on the subject of the P2P electricity trade. Kahrobaee et al. (2014) present an agent based model for electricity trading in which the participants can decide whether to store the energy produced in their batteries, and they can manipulate their loads if necessary. It is assumed that this works in a P2P energy grid. Giotitsas, et al. (2015) proposed a specific model for efficient energy trading among those on a P2P energy grid. Tchernykh, et al. (2014) analyzed a variety of algorithms with different grids and workloads considering two objectives: the approximation factor and energy consumption. The main idea of their approach was to set replication thresholds, and dynamically adapt them to cope with different objective preferences, workloads and grid properties. Silvente et al. (2015) proposed a MILP model in order to make a reactive scheduling approach to deal with the presence of uncertainty associated with the production and consumption of energy in microgrids. A four-layer architecture model for standardizing the interactions among different technologies for P2P energy trading was produced by Zhang et al. (2016). The maximization of self-consumption within a neighborhood in a solar powered micro-grid is given in a solution presented by Amato et al. (2015). When it comes to current P2P energy trading projects, the papers presented by Zhang et al. and Park & Yong (2017) introduce all available P2P energy trading projects such as Piclo, Vandebron, PeerEnergyCloud, Smart Watts, Electron and others. The performance of three existing P2P energy sharing models was reviewed by Zhou et al. (2017), with the aim of identifying the potential value, estimating the energy bill and finally giving the performance index value of the P2P energy sharing models. One of the approaches in which the authors try to find an optimal balance between demand and supply with the introduction of a P2P index is presented by Long et al. (2017). Since we aware that for every change there must be a social effect on people, Ruotsalainen et al. (2017) tried to describe the relationship between energy transitions and social change. The authors offered a plausible socio-cultural vision of an era of renewable energy. When it comes to modern (future) suggestions for electricity trading, the blockchain-based microgrid energy market presented by Mengelkamp et al. (2017) is definitely worth reading. Zhang et al. (2018) presented an architecture model in order to identify and categorize the key elements and technologies involved in P2P energy trading. They also suggested a trading platform that was simulated using game theory.

Considering the small number of researchers that are investigating and exploring this field, the literature is quite limited. On the other hand, as the P2P energy market grows over time, these numbers will increase. Microgrids and Supergrids are the future in energy production and consumption.

4. CONCLUSION

In this paper, a variety of research papers considering the subject of forecasting electricity prices, as well as papers considering P2P energy trading have been analyzed. The analysis covers the period from 2014 to 2018. A variety of different approaches and techniques can be noticed that optimize profit and use on the electricity market. The goal of the majority of this research is to reduce risks related to electricity trading. A review of the literature shows around 26 publications in refereed journals and conference prooceedings just in the last five years, with seven of these being published in the last year. However, while numerous studies have been documented, some of them are just descriptive and classificatory. The future of electricity trading is unpredictable, but one of the directions will surely be the implementation and application of blockchain technology in the field of electricity trading and on the P2P market.

REFERENCES

Schollmeier, R. (2001). A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications. In: Proceedings of the First International Conference on peer-to- peer computing (P2P2001). IEEE computer Society; 2001. p.101e2

Schusser, S., & Jaraitė, J. (2018). Explaining the interplay of three markets: Green certificates, carbon emissions and electricity, Energy Economics, Volume 71, Pages 1-13, ISSN 0140-9883, https://doi.org/10.1016/j.eneco.2018.01.012.

Anbazhagan,S., & Kumarappan, N. (2014). Day-ahead deregulated electricity market price forecasting using neural network input featured by DCT, Energy Conversion and Management, Volume 78, Pages 711-719, ISSN 0196-8904, https://doi.org/10.1016/j.enconman.2013.11.031.

Itaba, S., & Mori, H. (2017). A Fuzzy-Preconditioned GRBFN Model for Electricity Price Forecasting, Procedia Computer Science, Volume 114, Pages 441-448, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2017.09.010.

Peng, X., & Tao, X. (2018). Cooperative game of electricity retailers in China's spot electricity market, Energy, Volume 145, Pages 152-170, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2017.12.122.

Park, C., & Yong, T. (2017). Comparative review and discussion on P2P electricity trading, Energy Procedia, Volume 128, Pages 3-9, ISSN 1876-6102, https://doi.org/10.1016/j.egypro.2017.09.003.

Markvart, T. (2006). Microgrids: power systems for the 21st century? Refocus, 7(4):44e8.

Abedinia, O., Amjady, N., Shafie-khah, M., & Catalão, J. P. S. (2015). Electricity price forecast using Combinatorial Neural Network trained by a new stochastic search method, Energy Conversion and Management, Volume 105, Pages 642-654, ISSN 0196-8904, https://doi.org/10.1016/j.enconman.2015.08.025.

Vilar, J., Aneiros, G., & Raña, P. (2018). Prediction intervals for electricity demand and price using functional data, International Journal of Electrical Power & Energy Systems, Volume 96, Pages 457-472, ISSN 0142-0615, https://doi.org/10.1016/j.ijepes.2017.10.010.

Heredia, J. F., Cifuentes-Rubiano J., & Corchero, C. (2018). Stochastic optimal generation bid to electricity markets with emissions risk constraints, Journal of Environmental Management, Volume 207, Pages 432-443, ISSN 0301-4797, https://doi.org/10.1016/j.jenvman.2017.11.010.

Lago, J., De Ridder, F., Vrancx, P., & De Schutter, B. (2018). Forecasting day-ahead electricity prices in Europe: The importance of considering market integration, Applied Energy, Volume 211, Pages 890-903, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2017.11.098.

Amato, A., Di Martino, B., Scialdone, M., & Venticinque, S. (2016). Design and evaluation of P2P overlays for energy negotiation in smart micro-grid, Computer Standards & Interfaces, Volume 44, Pages 159-168, ISSN 0920-5489, https://doi.org/10.1016/j.csi.2015.04.004.

Srinivasan, S. (2013). Electricity as a traded good, Energy Policy, Volume 62, Pages 1048-1052, ISSN 0301-4215, https://doi.org/10.1016/j.enpol.2013.08.004.

Mengelkamp, E., Gärttner, J., Rock, K., Kessler, S., Orsini, L., & Weinhardt, C. (2018). Designing microgrid energy markets: A case study: The Brooklyn Microgrid, Applied Energy, Volume 210, Pages 870-880, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2017.06.054.

Hu, J., Harmsen, R., Crijns-Graus, W., Worrell, E., & Van Den Broek, M. (2018). Identifying barriers to large-scale integration of variable renewable electricity into the electricity market: A literature review of market design, Renewable and Sustainable Energy Reviews, Volume 81, Part 2, Pages 2181-2195, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2017.06.028.

Sikorski, J. J., Haughton, J., & Kraft, M. (2017). Blockchain technology in the chemical industry: Machine-to-machine electricity market, Applied Energy, Volume 195, Pages 234-246, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2017.03.039.

Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future, International Journal of Forecasting, Volume 30, Issue 4, Pages 1030-1081, ISSN 0169-2070, https://doi.org/10.1016/j.ijforecast.2014.08.008.

Mihaylov, M., Jurado, S., Avellana, N., Van Moffaert, K., de Abril, I. M., Nowe, A. (2014). Nrgcoin: virtual currency for trading of renewable energy in smart grids. In:2014 11th International conference on the European energy market (EEM). IEEE, p. 1–6

Ruotsalainen, J., Karjalainen, J., Child, M., & Heinonen, S. (2017). Culture, values, lifestyles, and power in energy futures: A critical peer-to-peer vision for renewable energy, Energy Research & Social Science, Volume 34, Pages 231-239, ISSN 2214-6296, https://doi.org/10.1016/j.erss.2017.08.001.

Dagoumas, A. S., Koltsaklis, N. E., & Panapakidis, I. P. (2017). An integrated model for risk management in electricity trade, Energy, Volume 124, Pages 350-363, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2017.02.064.

Silvente, J., Kopanos, G. M., Pistikopoulos, E. N., & Espuña, A. (2015). A rolling horizon optimization framework for the simultaneous energy supply and demand planning in microgrids, Applied Energy, Volume 155, Pages 485-501, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2015.05.090.

Jiang, P., Liu, F., & Song, Y. (2016). A Hybrid Multi-Step Model for Forecasting Day-Ahead Electricity Price Based on Optimization, Fuzzy Logic and Model Selection. Energies, 9(8), 618. doi:10.3390/en9080618

Wang, Z., Liu, F., Wu, J., & Wang, J. (2014). A Hybrid Forecasting Model Based on Bivariate Division and a Backpropagation Artificial Neural Network Optimized by Chaos Particle Swarm Optimization for Day-Ahead Electricity Price, vol. 2014, Article ID 249208, 31 pages, doi:10.1155/2014/249208

Vazquez, S., Lukic, S. M., Galvan, E., LG, Carrasco JM. Energy storage systems for transport and grid applications. IEEE Trans Ind Electron 2010;57(12):3881e95

Radenković, M., Lukić, J., Despotović-Zrakić, M., Labus, A., & Bogdanović, Z. (2018). Harnessing business intelligence in smart grids: A case of the electricity market, Computers in Industry, Volume 96, Pages 40-53, ISSN 0166-3615, https://doi.org/10.1016/j.compind.2018.01.006.

Giotitsas, C., Pazaitis, A., & Kostakis, V. (2015). A peer-to-peer approach to energy production, Technology in Society, Volume 42, Pages 28-38, ISSN 0160-791X, https://doi.org/10.1016/j.techsoc.2015.02.002.

Tchernykh, A., Pecero, J. E., Barrondo, A., & Schaeffer, E. (2014). Adaptive energy efficient scheduling in Peer-to-Peer desktop grids, Future Generation Computer Systems, Volume 36, Pages 209-220, ISSN 0167-739X, https://doi.org/10.1016/j.future.2013.07.011.

Girish, G. P. (2012). Modeling and Forecasting Day-Ahead Hourly Electricity Prices: A Review, Proceedings of the International Conference on Business Management & Information Systems

Kahrobaee, S., Rajabzadeh, R. A., Soh, L. K., & Asgarpoor, S. (2014). Multiagent study of smart grid customers with neighborhood electricity trading, Electric Power Systems Research, Volume 111, Pages 123-132, ISSN 0378-7796, https://doi.org/10.1016/j.epsr.2014.02.013.

Girish, G. P., Rath, B. N., & Akram, V. (2018). Spot electricity price discovery in Indian electricity market, Renewable and Sustainable Energy Reviews, Volume 82, Part 1, Pages 73-79, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2017.09.009.

Murthy, G. G. P., Sedidi, V., Panda, A. K., & Rath. B. N., (2014). Forecasting Electricity Prices in Deregulated Wholesale Spot Electricity Market - A Review. 2014, International Journal of Energy Economics and Policy, Vol. 4, No. 1, pp.32-42, ISSN: 2146-4553

Ziel, F., Steinert, R., & Husmann, S. (2015). Efficient modeling and forecasting of electricity spot prices, Energy Economics, Volume 47, Pages 98-111, ISSN 0140-9883, https://doi.org/10.1016/j.eneco.2014.10.012.

Long, C., Wu, J., Zhang, C., Cheng, M., & Al-Wakeel, A. (2017). Feasibility of Peer-to-Peer Energy Trading in Low Voltage Electrical Distribution Networks, Energy Procedia, Volume 105, Pages 2227-2232, ISSN 1876-6102, https://doi.org/10.1016/j.egypro.2017.03.632.

Zhang, C., Wu, J., Cheng, M., Zhou, Y., & Long, C. (2016). A Bidding System for Peer-to-Peer Energy Trading in a Grid-connected Microgrid, Energy Procedia, Volume 103, Pages 147-152, ISSN 1876-6102, https://doi.org/10.1016/j.egypro.2016.11.264.

Zhang, C., Wu, J., Long, C., & Cheng, M. (2017). Review of Existing Peer-to-Peer Energy Trading Projects, Energy Procedia, Volume 105, Pages 2563-2568, ISSN 1876-6102, https://doi.org/10.1016/j.egypro.2017.03.737.

Zhang, C., Wu, J., Zhou, Y., Cheng, M., & Long, C. (2018). Peer-to-Peer energy trading in a Microgrid, Applied Energy, Volume 220, Pages 1-12, ISSN 0306-2619, https://doi.org/10.1016/j.apenergy.2018.03.010.

Zhou, Y., Wu, J., Long, C., Cheng, M., & Zhang, C. (2017). Performance Evaluation of Peer-to-Peer Energy Sharing Models, Energy Procedia, Volume 143, Pages 817-822, ISSN 1876-6102, https://doi.org/10.1016/j.egypro.2017.12.768.

B20
Statistics, Stochastics & Simulation

BALCOR 2018
XIII Balkan Conference on Operational Research

SIMULATION MODEL FOR DETERMINING THE LOCATION OF RUNWAY EXIT

STEFAN ŠLJUKIĆ[1], MARKO ĐOGATOVIĆ[2], BOJANA MIRKOVIĆ[3]

[1] University of Belgrade, Faculty of Transport and Traffic Engineering, stefans-bg@hotmail.com
[2] University of Belgrade, Faculty of Transport and Traffic Engineering, m.djogatovic@sf.bg.ac.rs
[3] University of Belgrade, Faculty of Transport and Traffic Engineering, b.mirkovic@sf.bg.ac.rs

Abstract: *A simulation model to determine a runway exit location is proposed in the paper. The location of the runway exit is found with the objective to minimize the cost of taxiing to the terminal complex after landing, in the case when aircraft pass by the terminal during landing. This problem is relevant for airports with terminal complex located close to the landing threshold, especially when runway saturation exists, but also under lower runway loads. Simulation analysis is performed for the case of Airport "Nikola Tesla", in Belgrade. Current case (one runway exist) is evaluated and location of additional runway exit is determined for the given fleet mix.*

Keywords: *Runway exit location, airport, simulation, modelling.*

1. INTRODUCTION

Runway throughput is influenced by wake-vortex separations, on one hand, and on the other, by taxiway system configuration which dictates runway occupancy times (ROT). The runway occupancy time of an arriving aircraft is defined as the time between moment when the aircraft touches down on the runway and the moment when vacate the runway (De Neufville and Odoni 2013). At airports with mature taxiway system (taxiway parallel to runway), ROT can be additionally decreased with adequate planning of runway exits – their type, number and location. Apart from being important for gaining additional capacity large congested airports, issue of runway exit location can also be important for small and medium airports for different reason. This paper addresses the issue of airports with terminal complex located near the threshold of dominant runway direction in use. The problem is modeled on example of Airport "Nikola Tesla" (ANT), Belgrade, Serbia.

At ANT approximately 70 percent of total movements are performed through runway 12 (RWY12). RWY12 has two exits - one is located approximately two-thirds of runway length (D) and another at the runway end (E), see Figure 1. Terminal complex is located near landing threshold (Figure 1). It means that, after landing on RWY12, aircraft decelerate along the runway, and vacate it using one of available exits (D or E). Then, aircraft need to taxi back to the terminal complex using parallel taxiway. Main problem caused by inadequate location of runway exit (apart from increased ROT), with such airport layout, is unnecessary taxiing distance which results in increased fuel consumption and increased block time per flight (time from engine started, till time when engine is shutdown). The problem addressed in this paper - minimization of unnecessary taxiing, can bring fuel savings and decrease block time both having an impact on airline costs.

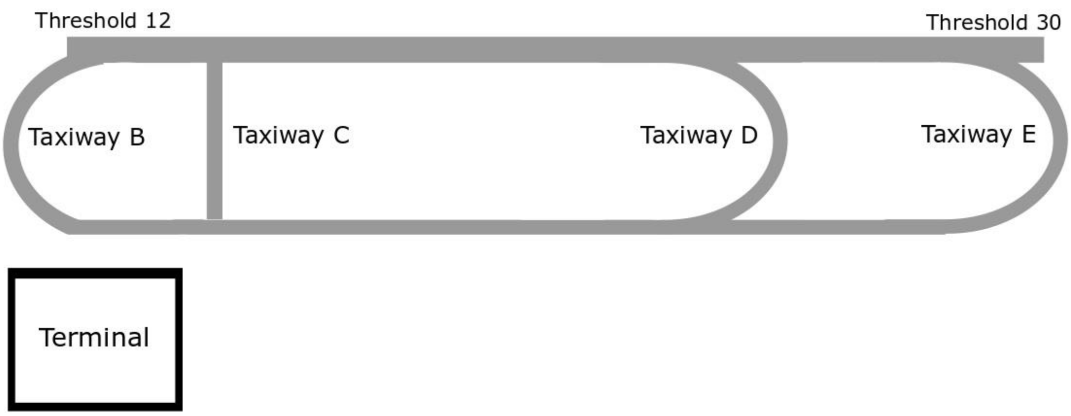


Figure 1: Graphical representation of Belgrade Airport

In order to analyze effects of different runway exit locations and types on runway occupancy time and costs, simulation model is proposed. Simulation is a set of techniques, methods and tools for developing a simulation model of the real system and using the model to describe behavior of the system and conduct experiments for the purpose of simulation analysis of the particular problem (Radenković et al. 1999).

Runway operations description is given in Section 2. In Section 3 the simulation model is presented, together with required input data and followed by the results obtained. Section 4 summarizes main findings of the simulation analysis.

2. RUNWAY OPERATIONS

Runway system is illustrated in Figure 2. To assure safe runway operations there are two main rules. Firstly, aircraft need to be safely separated in the air (between FIX and threshold), and secondly - only one aircraft can occupy the runway at one time. FIX present point typically 5 to 8 NM from the runway threshold, where aircraft start descending along final approach path until touchdown on the runway (De Neufville and Odoni 2013).

Distance based separation in approach is commonly used at airports. Minimum separation (recommended by International Civil Aviation Administration – ICAO) is determined for the pair of aircraft and depends on the possible impact of the leading aircraft wake-vortex to trailing aircraft, although air traffic control can apply larger than minimum separation due to local conditions. Nowadays, some congested airports also apply time-based separation concept, used primarily as a mean to increase runway throughput in conditions of strong head winds.

As the two landing aircraft should not be at the same time on the runway, model prevents trailing aircraft to touchdown before the leading aircraft vacates the runway (De Neufville and Odoni 2013).

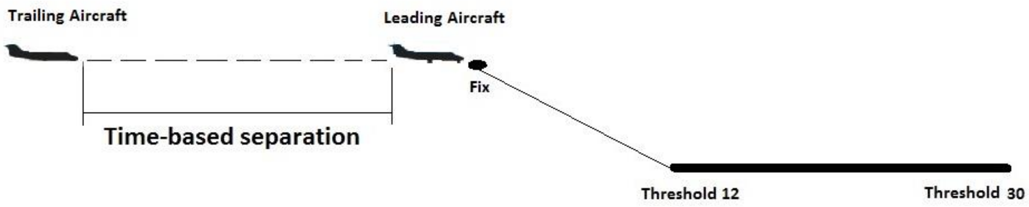


Figure 2: Runway system

Runway operations are highly dependent on traffic characteristics - fleet mix (share of different aircraft types), separation minimal and landing performance for each aircraft type.

General rule is that an aircraft can vacate the runway when it reaches velocity equal or lower than maximum acceptable velocity of the runway exit. Maximum acceptable velocity of the runway exit depends on the runway exit type, i.e. runway exit design. Aircraft velocity at a certain point on runway depends on the touchdown position and velocity above the threshold, as well as deceleration rate along the runway. Velocity above threshold depends on aircraft type and wind.

Figure 3 shows mean velocity measured along runway on example of medium aircraft type DC-9, based on empirical data collected at ANT (Tošić and Živković 1984).

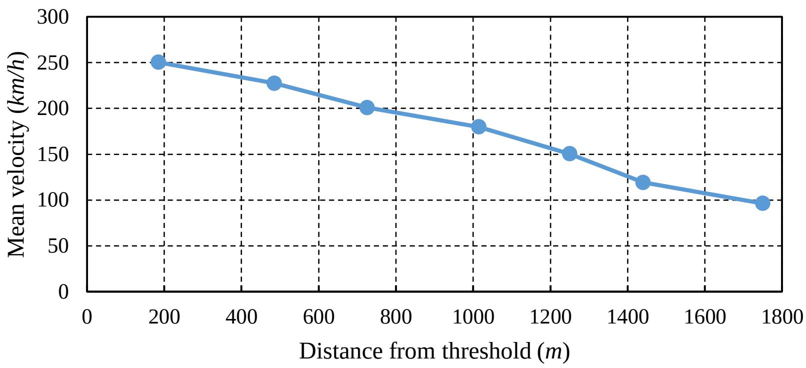


Figure 3: Mean velocity along runway

The location of touchdown is also important for the problem od runway exit location. Touchdown is correlated with the velocity. Knowing these variables, the location of runway exit to minimize unnecessary taxiing for the given fleet mix can be determined.

According to collected empirical data, mean deceleration measured on runway has bimodal distribution with two peaks, as presented on Figure 4. The first peak is result of the main landing gear contact with the runway surface while the second is from using reverse thrust (Tošić and Živković 1984).

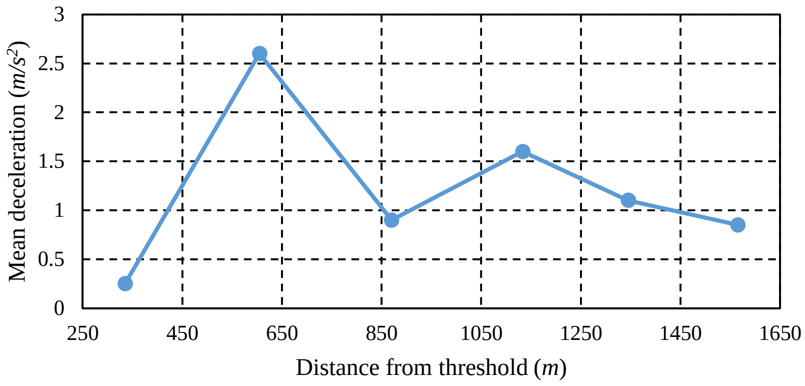


Figure 4: Mean deceleration along runway

Aircraft decelerates along the runway till it reaches a sufficiently low velocity to vacate the runway. However, runway can be vacated only at the location of runway exits. Aircraft can reach that velocity much earlier than the runway exit location. Keeping that in mind, unnecessary taxiing is defined as the difference between the taxiing distance/time for a given exit configuration and taxiing distance/time for an ideally located exit (Tošić and Živković 1984). Ideally located exit would be at the point where aircraft is reaching the velocity of runway exit acceptance. Due to position of the terminal complex, unnecessary taxiing is doubled because in addition to unnecessary distance taxied along the runway the same distance has to be crossed along a parallel taxiway. The objective of this study is to determine the location for runway exit and additional rapid exit taxiway at ANT for given input data.

Runway exit at ANT is located at 1800 m from the RWY12 threshold and can be accepted at the velocity of about 60 km/h (see Figure 1). It is angled runway exit not usable for landings on RWY30 (like it is the case with 90 degree exit C), adapted to allow runway vacation under higher velocity for more frequently used RWY12. Exit at runway end is located about 2800m from threshold 12.

Advanced runway exit type is so called high-speed exit or rapid exit taxiway, is designed for velocity limit of about 90 km/h, according to ICAO recommendations. Intersection angle between runway and rapid exit taxiway should be 30 degrees. It is important that rapid exit taxiway includes a straight distance after the turn-off curve sufficient for an exiting aircraft to come to a full stop, clear of any intersecting taxiway, see Figure 5 (ICAO 2016).

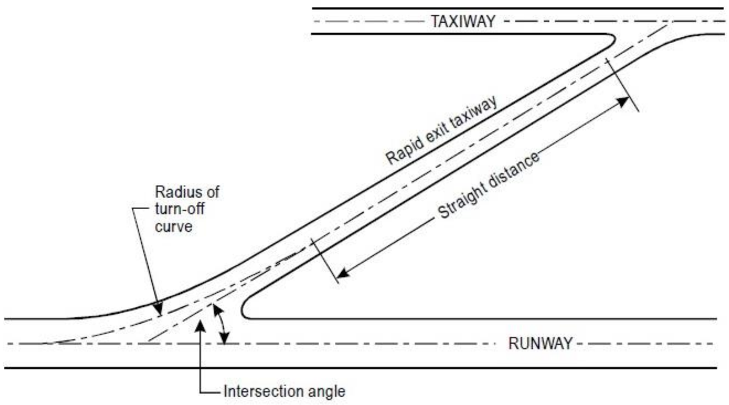


Figure 5: High-speed exit or rapid exit taxiway (ICAO, 2016)

3. SIMULATION MODEL AND RESULTS

The aim of simulation analysis is twofold - to evaluate location of the current runway exit (taxiway D) at ANT and to determine location of the additional runway exit. For additional taxiway two cases are compared – conventional and rapid-exit taxiway. Analysis is performed using current layout of ANT, empirical data for landing performance for two different aircraft types and assumed fleet mix.

Discrete event simulation (DES) is a specific simulation methodology used for modeling over time of a system that can be described with a set of events. Event is a discrete change of state for the system and events occur at discrete points in time. There is no change of state between two consecutive events. In DES time is changing from one instantaneous event to the next event (Radenković et al. 1999). In the paper we are using SimEvents for modeling purposes. SimEvents is a discrete-event simulation engine and Simulink component library for analyzing event-driven system models and optimizing performance characteristics such as latency, throughput, and packet loss. With SimEvents, it is possible to study the effects of task timing and resource usage on the performance of distributed control systems, software and hardware architectures, and communication networks ("SimEvents," 2018).

In the simulation model two aircraft types in the fleet are assumed – medium and heavy with share of 60% and 40%, respectively. Aircraft arrive at the airport between 80 sec and 160 sec apart, uniformly distributed. For departure, aircraft come to the runway between 110 sec and 170 sec apart, uniformly distributed. When runway is occupied, aircraft for departure will wait for permission to enter the runway. In this paper runway operations model uses separation based on time calculated from distance separations and constant aircraft velocity with zero head wind in approach.

Table 1 shows cumulative distribution of touchdown and mean initial deacceleration along runway. This table is based on empirical data collected at ANT (Tošić and Živković 1984).

Table 1: Cumulative distribution of touchdown and mean initial deceleration along runway

Distance from threshold (m)	Cumulative probability	Deceleration (m/s^2)
0 – 250	0	2.1000
250 - 350	0.04	2.1875
350 - 450	0.21	2.2750
450 - 550	0.51	2.3625
550 - 650	0.72	2.4500
650 - 750	0.89	2.5375
750 - 850	0.95	2.6250
850 - 950	0.97	2.7125
950 - 1050	1.00	2.8000

Table 2 shows range of touchdown velocities on four runway segments for medium aircraft. When medium aircraft touches down the runway, depending on the runway segment where touchdown occurred, touchdown velocity is calculated using uniform distribution for the appropriate range. In the model, velocity for heavy aircraft is calculated as value of velocity for medium aircraft increased by 25 km/h.

Table 2: Range of touchdown velocities on different runway segments for medium aircraft type

Segment	Distance from threshold (m)	Range of touchdown velocity (km/h)
I	350	233 – 265
II	600	210 – 243
III	870	180 – 225
IV	1100	158 – 205

So, aircraft have different initial deceleration value along runway, correlated with the location of touchdown. In the paper mean initial deceleration value is growing linearly depending on touchdown (see Table 1).

Value of the initial deceleration, d_0, is obtained using uniform distribution with the mean initial deceleration value and spread 0.2 sec. Deceleration is changing along the runway depending on the distance from threshold, l, according to following equation:

$$d(l) = k(l - l_0) + d_0, \quad (1)$$

where l_0 is touchdown and k is slope coefficient of deceleration. In the model the slope coefficient is constant with a value of -0.0018 s^{-2}. Value of the velocity depending on the distance from the threshold is calculated using the following equation:

$$v(l) = \sqrt{v_0^2 - 2\int_{l_0}^{l} d(\lambda)d\lambda}, \quad (2)$$

where v_0 is velocity on touchdown. With use of Equation (2) it is possible to calculate the runway occupancy time according to the expression:

$$t = \int_{l_0}^{l} \frac{d\lambda}{v(\lambda)}. \tag{3}$$

Values of deceleration, velocity and ROT time are numerically calculated. The trapezoidal rule is used for integration. Figure 6 shows comparison of deceleration used in simulation model, and deceleration collected from empirical data (see Figure 4).

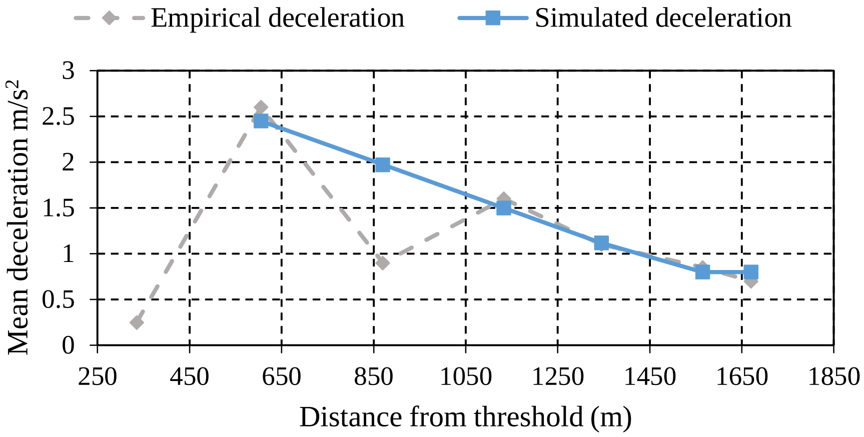


Figure 6: Comparison of deceleration used in simulation model and the mean deceleration from empirical data

In the model the minimal adopted value of deceleration is 0.8 m/s^2, while the minimal adopted value of velocity is 30 km/s, for taxiing along runway. Minimal average time for aircraft to safely vacate the runway is 10 sec.

Simulation experiment is performed 50 times. For every simulation run different seeds are passed to Uniform Random Number blocks in the model (see Figure 7). Simulink Functions are blocks preconfigured as starting points for graphical defining of functions with Simulink blocks. These fuctions are used to generate touchdown, velocity and deceleration and to calculate ROT time. The simulation model is created with SimEvents blocks for entity (aircraft) generation, queueing, serving and termination.

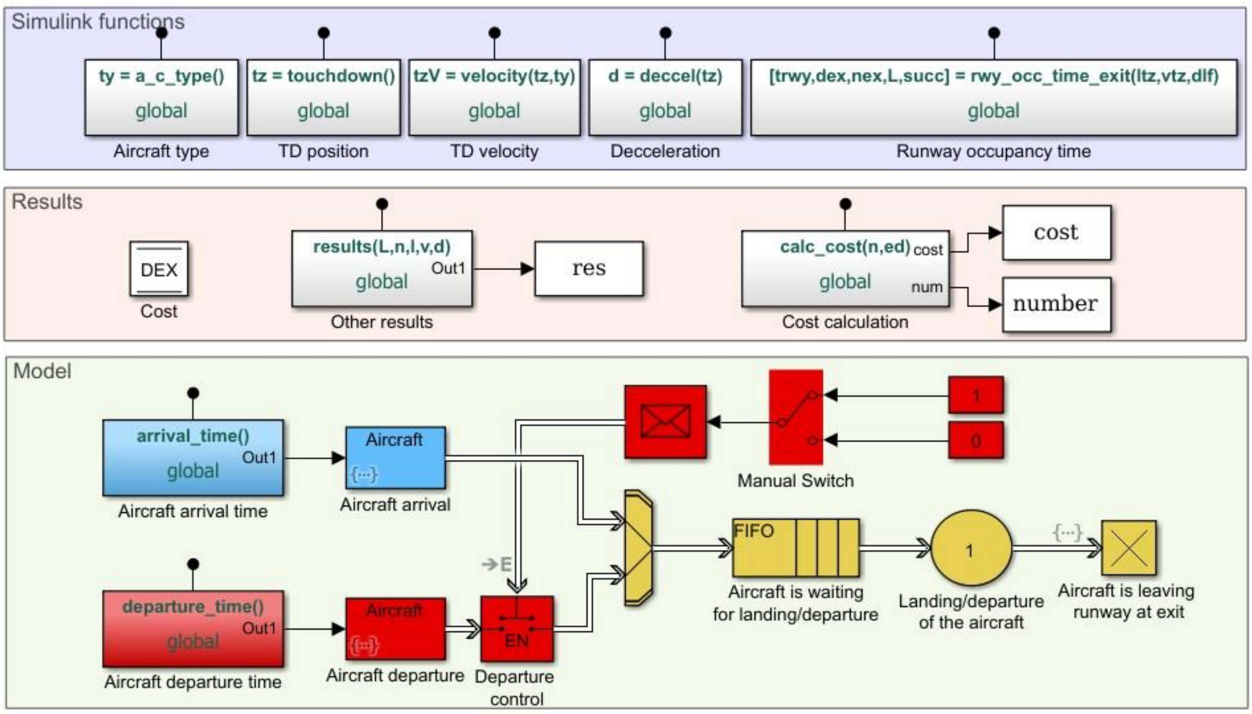


Figure 7: DES model realized in Simulink

According to following equation (Tošić et al. 1985), unnecessary taxiing distance for one aircraft type $n = 1$, and one exit ($m = 1$) located at l_1 from the landing threshold, can be expressed as:

$$C(L_1) = C(l_1) = \begin{cases} \sum_{l=a_1}^{b_1} 2 \cdot (l_r - l) \cdot p_1(l) \cdot c_1 = \text{const.} \quad \text{for } l_1 < a_1 \\ \sum_{l=a_1}^{l_1} 2 \cdot (l_1 - l) \cdot p_1(l) \cdot c_1 + \sum_{l=l_1+1}^{l_r} 2 \cdot (l_r - l) \cdot p_1(l) \cdot c_1 \quad \text{for } l_1 > a_1 \end{cases} \tag{4}$$

where:

l_r – runway length i.e. location of the runway exit at the end of the runway,
L_1 – set of one exit,
c_1 – the cost of taxiing for aircraft type 1, per unit length,
a_1 – lower bound of probability distribution of exit acceptance for aircraft type 1,
b_1 – upper bound of probability distribution of exit acceptance for aircraft type 1,
$p_1(l)$ – probability of aircraft exit on segment at distance l from the landing threshold, given that it did not exit earlier.

Unnecessary taxiing distance for the general case when n aircraft types and m exits located at $l_1, l_2, \ldots, l_i, \ldots, l_m$ where $l_i > l_{i-1}$, i.e., where exits are ordered starting from the landing threshold towards the runway end, can be expressed as (Tošić et al. 1985):

$$C\,(L_m) = \sum_{l=a_{min}-m+1}^{l_1} 2(l_1 - l) \sum_{j=1}^{n} p_j(l) p_j c_j + \sum_{l=l_1+1}^{l_2} 2(l_2 - l) \sum_{j=1}^{n} p_j(l) p_j c_j + \cdots + \sum_{l=l_{m-1}+1}^{l_m} 2(l_m - l) \sum_{j=1}^{n} p_j\,(l) p_j c_j + \sum_{l=l_m+1}^{l_r} 2(l_r - l) \sum_{j=1}^{n} p_j(l) p_j\, c_j, \tag{5}$$

where:

l_i – location of exit i,
c_j – the cost of taxiing of aircraft type j, per unit length,
a_{min} – the lowest value from the lower bounds of probabilities of exit acceptance for all aircraft j,
$a_{min} = \min_j \{a_j\}$,
p_j – the proportion of aircraft type j in the fleet at a given airport
L_m – set of m exits
$p_j(l)$ – the probability of aircraft type j exiting on the segment at distance l from the landing threshold, given that it did not exit earlier.

The expression for $C\,(L_m)$ is valid for the runway range in which at least some of the aircraft can use at least one of the considered exits

$$l_m > a_{min},$$

where:

$$l_m = \max_i (l_i).$$

Unnecessary cost of taxiing is presented here as unnecessary taxiing distance (in meters) i.e. cost of taxiing is assumed to be equal 1 in the formulas above.

Savings from the introduction of the additional k runway exit is calculated as the difference between unnecessary cost of taxiing for initial layout with m runway exits and unnecessary cost of taxiing for the layout with m+k runway exits. In this paper, current case with one runway exit (m = 1) is observed, and upgraded layout with two runway exits (k = 1).

Figure 8 shows probability distribution of exit acceptance at a given distance from threshold using simulation experiment, discreetly divided into segments of 200 m. As we can see, probability distribution for only medium aircraft type in the fleet has different allocation along runway, than probability distribution for the assumed mix of 60% medium and 40% heavy aircraft.

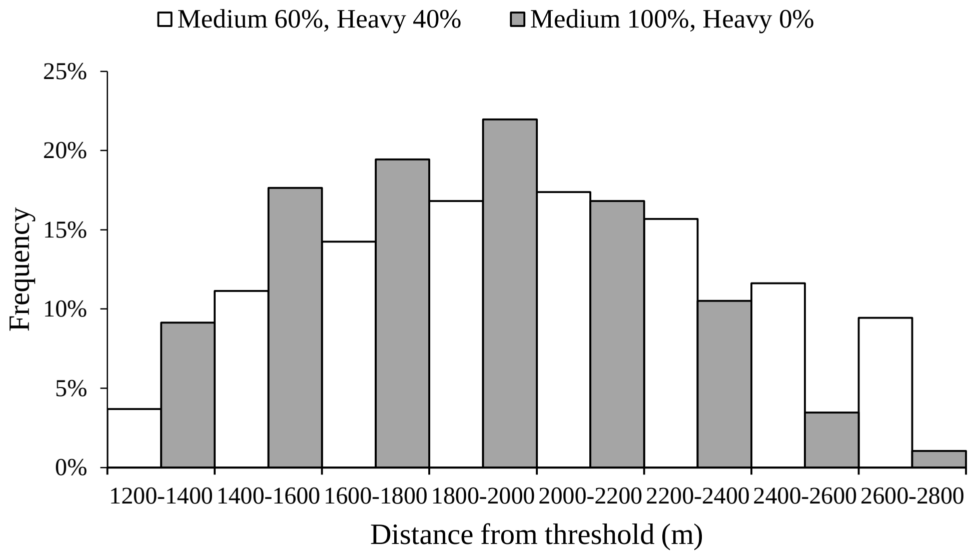


Figure 8: Probability distribution of exit acceptance

Results from the model are used to discuss current layout and additional runway exit location. Current ANT layout - one runway exit located approximately two-thirds of runway length (D) and another at the runway end (E), is discussed using one aircraft type (medium aircraft). Additional runway exit location is analyzed for the case with two aircraft types (60% medium and 40% heavy aircraft), using segments of 200 m from 1600 m to 2400 m.

Table 3 presents average (of 50 runs) unnecessary taxiing distance for the case of medium aircraft. I is unnecessary taxiing of all aircraft that can accept exit at the given distance (upper part of formula 4, $l_1 < a_1$). II refers to aircraft that miss runway exit and vacate the runway at its end (lower part of the formula 4, $l_1 > a_1$). It can be seen that the best location of the runway exit if all medium aircraft would be using it is at 2000 m with unnecessary taxing of 23515,7 m.

Table 3: Average unnecessary taxiing distance for determining only one runway exit, for medium aircrafts

Exits at	I	II	Total unnecessary taxiing distance (m)
1600	1511.16	28785.92	30297.1
1800	5444.16	20520.52	25964.7
2000	12250.24	11265.48	23515.7
2200	20338.28	4021.96	24360.2
2400	26723.52	728.32	27451.8

For the fleet mix of 60% medium and 40%, heavy aircraft with, additional runway exit should be at the location with the greatest savings in unnecessary taxiing distance in comparison to current layout. Tables 4 and 5 show the results obtained - unnecessary taxiing distance and savings. For the case of rapid exit taxiway, Table 4, velocity limit of 90 km/h is assumed (ICAO 2016). Table 5 show the savings achieved if the second runway exit would be of the same type as the current exit, i.e. velocity limit of 60 km/h.

For the rapid exit taxiway, the worst case is exit at 1600 m, savings that we can get is very small with average of only 923,5 m. Rapid exit taxiway should be located at the distance of 2000 m from the RWY12 threshold, because of significant savings of 9999,2 m (see Table 4).

Table 4: Average unnecessary taxiing distance for current runway exit location (bold) and potential new locations of rapid exit taxiway

Exits at	Unnecessary taxiing distance (m)	Savings from added new exit (m)
*1600,**1800***	23805.28	923.5
***1800**,2000*	14729.6	9999.2
***1800**,2200*	16112.2	8616.6
***1800**,2400*	20590.12	4138.7

It can be seen that the worst case is exit at 1600 m, savings that we can get is very small with average of only 1779,5 m. On the other side, added exit at 2200 m has significant savings of 9125,8 m, and most aircraft will vacate runway on that exit (see Table 5).

Table 5: Average unnecessary taxiing distance for current runway exit location (bold) and potential new locations of runway exit

Exits at	Unnecessary taxiing distance (m)	Savings from added new exit (m)
*1600,**1800***	22628.16	1779.5
***1800**,2000*	18492.64	5914.6
***1800**,2200*	15281.36	9125.8
***1800**,2400*	15520.8	8886.4

4. CONCLUSION

Simulation model described in the paper is designed to analyze unnecessary taxiing distance by changing the location and number of exits along runway, likewise to see where could be additional rapid exit taxiway at ANT for given input data.

Simulation analysis is performed on example of ANT using available empirical data. Obtained results show preference location of the runway exit(s) for the assumed landing traffic characteristics at RWY 12, (proportion of aircraft type in the fleet, distribution of touchdown along runway and deceleration techniques).

Probability distribution of runway exit acceptance at a given distance from threshold (discreetly divided into segments of 200 m) is obtained using simulation experiment, for two cases - only medium aircraft and mix of two aircraft types (60% medium and 40% heavy aircraft).

Location of the runway exit is evaluated based on the unnecessary taxing distance that is calculated as the the difference between ideal location at which an aircrfat could have vacated the runway and given location of the runway exit.

Simulation results show that the best location of one runway exit, if all medium aircraft would be using it, is at 2000m. For the case with two aircraft types (60% medium and 40% heavy aircraft) new (conventional) runway exit at 2200 m will enable significant savings since the most aircraft would vacate the runway via that exit. In the case of rapid exit taxiway, with velocity limit of 90 km/h, the best location would be at the distance of 2000 m from the RWY12 threshold.

One of the problems for model can be input data, due to the fact that part of the sample used here was very small, and it is with low measurement precision.

Available data were used for the purpose of the simulation model testing. Further research include collection of the current traffic characteristics at ANT and perfoming the same analysis with updated inputs to get more reliable results.

REFERENCES

De Neufville, R. & Odoni, A. (2013). *Airport Systems - Planning, Design and Management*, New York, NY: McGraw-Hill.

ICAO (2016). *Aerodrome Design Manual (Doc 9157), Part 2, Taxiways* (7th ed), Montreal, Quebec: International Civil Aviation Organization.

Radenković, B., Stanojević, M. & Marković, A. (1999). *Computer Simulation* (in Serbian). Belgrade, Serbia: University of Belgrade – Faculty of Organizational Sciences and Faculty of Transport and Traffic Engineering.

SimEvents documentation. (2018, April 7). Retrieved from: https://www.mathworks.com/products/simevents.html.

Tošić, V., Teodorović, D. & Babić, O. (1985). Optimum runway exit location. *Transportation Planning and Technology*, 10, 135-145.

Tošić, V., Živković, Z. (1984). *Pilot study on landing traffic characteristics on runway 12 at the Belgrade Airport*. Unpublished paper.

BALCOR 2018
XIII Balkan Conference on Operational Research

SIMULATION ANALYSIS OF QUALITY OF BUSINESS IN IP NETWORKS

VESNA RADONJIĆ ĐOGATOVIĆ[1], MARKO ĐOGATOVIĆ[2], MILORAD STANOJEVIĆ[3]

[1]University of Belgrade, Faculty of Transport and Traffic Engineering, v.radonjic@sf.bg.ac.rs
[2]University of Belgrade, Faculty of Transport and Traffic Engineering, m.djogatovic@sf.bg.ac.rs
[3]University of Belgrade, Faculty of Transport and Traffic Engineering, milorad@sf.bg.ac.rs

---

Abstract: *In the paper, we propose a simulation model for mapping Quality of Service (QoS) parameters to Quality of Business (QoBiz) in IP networks. We assume Internet Service Provider (ISP) offers tariff packages based on the proposed QoS to QoBiz mapping and users' requirements. Available bit rate and security are chosen as key QoS parameters and price is selected as main QoBiz parameter from users' perspective, while revenue singles out as main QoBiz parameter from ISP's perspective. We also assume ISP applies hybrid pricing in a manner that price reduction is performed depending on QoS vioaltion, which is defined through network load. The simulation model is conducted using agent-based simulation methodology. Agents can be seen as autonomous units that mutually interact in the environment. In this research, users and ISP are observed as simulation agents and IP market is seen as an environment. Output parameters in the simulation analysis are ISP's revenue and service price.*

Keywords: *Quality of Service, Quality of Business, pricing, agent-based simulation, modelling.*

1. INTRODUCTION

Considering highly diverse traffic in Internet Protocol (IP) networks, Quality of Service (QoS) differentiation has become crucial in ensuring proper support for different QoS requirements. QoS describes technical ability of network to provide a service with an assured service level (ITU-T E.800 2008). In order to provide a suitable economic framework for trade-off between quality offered and delivered by Internet Service Provider (ISP) and its financial benefit from service provisioning, Quality of Business (QoBiz) is being increasingly used. It deals with financial aspects of service provisioning that are important from both service providers' and users' perspective. Generally, QoBiz refers to all those parameters that are expressed in monetary units, such as cost, service provider's revenue, profit, service price, etc. (Wolter and Moorsel 2001).

One of the main factors affecting business operation of an ISP is pricing. Therefore, selection of an appropriate pricing scheme is very important for an ISP. A pricing scheme should fulfil compromise between providing satisfied users' and ISP's revenue goals. Considering that users' satisfaction mainly depends on changes in QoS and price, an efficient pricing scheme should include both aspects. A wide range of different pricing schemes have been applied in IP networks (Radonjić Đogatović and Kostić-Ljubisavljević 2015). Applying static pricing scheme is the simplest but it means that users are charged independently of the resource consumption and QoS delivered. Unlike static pricing, dynamic pricing is a process of allocating the price as a cost per unit of resource consumption and according to level of QoS guarantees provided for the particular service class. In this research we use combination of static and dynamic pricing which is known as hybrid pricing. It is defined as a process of applying the static price in regular network operation mode while during congestion the dynamic pricing is enforced allowing deviations from contracted tariff (Radonjić Đogatović 2018).

In order to analyze implications on QoBiz parameters when hybrid pricing is applied, we propose QoBiz model for which we have conducted simulation analysis. Computer simulation is a set of techniques, methods and tools used for modelling of a real system in the form of the computer program. Developed program is the simulation model that imitates behavior of the real system (Radenković et al. 1999). Agent-based simulation (ABS) is a form of computer simulation whereby a phenomenon is simulated and modeled in terms of agents and their interactions. An agent is an autonomous computational individual or object with particular properties and actions (Wilensky and Rand 2015). In the paper simulation analysis of the proposed QoBiz model is performed using ABS techniques, focusing on the interaction between participant (users and ISP) in an IP network. For that purpose a computer program is developed with a display of animation in order to understand behavior and interactions between components of the simulation model.

The rest of the paper is organized in the following way. In Section 2, spiral solution for mapping QoS to QoBiz is proposed as well as tariff packages (TPs) based on QoS parameters combinations and users' requirements. Simulation model is proposed and explained in detail in Section 3. Concluding remarks are given in Section 4.

2. MODELLING QUALITY OF BUSINESS IN IP NETWORKS

Quality of IP services is characterized by combined aspects of service support performance, service operability performance, service security performance and other factors specific to each service. Therefore it can be estimated based on different aspects which describe and distinguish service provisioning (Stankiewicz et al. 2011). Considering that ISP strives to maximize its revenue while providing users the required QoS at the acceptable price, in this research we focus on mapping QoS to QoBiz parameters.

In the proposed model we take into account only requirements regarding QoS that are significant and transparent enough to users and thereby relevant for ISP's QoBiz. Steps preceding development of the model include: selection of QoS parameters that significantly affect QoBiz requirements, identification of key QoBiz parameters and finding the most appropriate solution for mapping QoS to QoBiz (Radonjić Đogatović et al. 2013).

First, we choose available bit rate and security as key QoS parameters for mapping to QoBiz. Regarding available bit rate, we consider download and upload speed as separated QoS parameters. Security comprises data confidentiality, data integrity and availability of the system and its information (ITU-T X.1031 2008).

Considering price is one of the most significant parameters for users while the main goal for ISP is to maximize its revenue, we focus on a service price and ISP's revenue, as key QoBiz parameters.

For the purpose of this research, we propose the spiral solution for mapping QoS to QoBiz with different settings for each QoS parameter (Figure 1):

- download speed (s_{d1}, s_{d2}, s_{d3}, s_{d4}, s_{d5}), where $s_{d1} > s_{d2} > s_{d3} > s_{d4} > s_{d5}$
- upload speed (s_{u1}, s_{u2}, s_{u3}, s_{u4}, s_{u5}), where $s_{u1} > s_{u2} > s_{u3} > s_{u4} > s_{u5}$ and
- security (s_1, s_2, s_3, s_4, s_5), where s_1 indicates the highest level of security while each subsequent level indicates a bit lower security, so s_5 point out to the lowest level of security.

Each of these values corresponds to a specific price, as it is illustrated in Figure 1. Service with the best performances (s_{d1}, s_{u1}, s_1) is charged with maximum price, denoted as p_0. It is assumed that a price linearly and equally depends on each QoS parameter, i.e. it decreases for 5% with any QoS degradation, which implies that a service with the worst performances by all QoS parameters (s_{d5}, s_{u5}, s_5) is charged with 60% lower price compared to a service with the best performances.

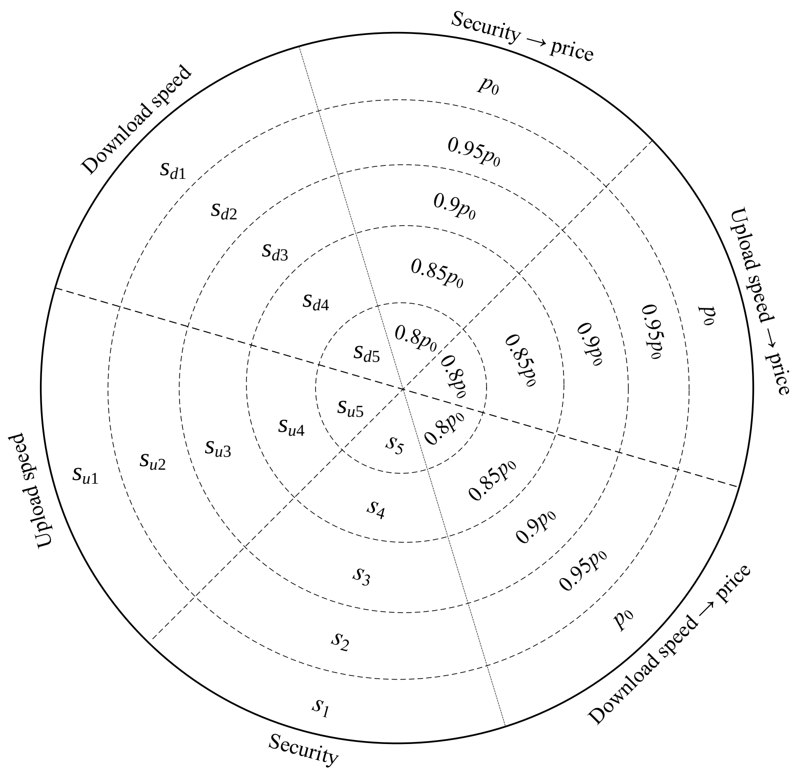


Figure 1: Spiral QoS to QoBiz mapping

Further, users are classified according to their preferences regarding to QoS and price. We assume that there are eight different types of users:

1. Users who prefer high speed Internet access with maximum download speed,
2. Users who prefer high speed Internet access with maximum upload speed,
3. Users who prefer maximum security protection,
4. Users who can tolerate slightly lower download speed
5. Users who can tolerate slightly lower upload speed
6. Users who can tolerate slightly lower level of security
7. Users who are willing to pay no more than p_1 ($p_1 > 0.4\ p_0$) for monthly Internet usage and
8. Users who are willing to pay no more than p_2 ($p_2 \geq 0.4\ p_0$) for monthly Internet usage (the assumption is that $p_2 < p_1$).

Users can belong to more than one type but some types are mutually exclusive. For example, type 1 users can be at the same time type 2 and/or type 3 users but this excludes the possibility of belonging to types 4, 7 and 8. We assume type 7 and 8 don't have any QoS preferences specified for previous types, so they will be assigned lower QoS values than previous types and consequently they will be charged with lower prices.

Based on QoS parameters combinations and users' requirements, ISP defines TPs, which is shown in Table 1.

Table 1: Tariff packages

QoS/ QoBiz	*Download speed*	*Upload speed*	*Security*	*Price*
TP1	s_{d1}	s_{u1}	s_1	p_{TP_1}
TP2	s_{d1}	s_{u1}	s_2	p_{TP_2}
TP3	s_{d1}	s_{u2}	s_1	p_{TP_3}
TP4	s_{d2}	s_{u1}	s_1	p_{TP_4}
TP5	s_{d1}	s_{u1}	s_3	p_{TP_5}
TP6	s_{d1}	s_{u3}	s_1	p_{TP_6}
TP7	s_{d3}	s_{u1}	s_1	p_{TP_7}
TP8	s_{d1}	s_{u3}	s_3	p_{TP_8}
TP9	s_{d3}	s_{u1}	s_3	p_{TP_9}
TP10	s_{d3}	s_{u3}	s_1	$p_{TP_{10}}$
TP11	s_{d4}	s_{u4}	s_4	$p_{TP_{11}}$
TP12	s_{d5}	s_{u5}	s_5	$p_{TP_{12}}$

Acordingly, ISP's revenue can be determined as:

$$\sum_{i=1}^{k} p_{TP_i} N_{TP_i} \quad (1)$$

where p_{TP_i} is monthly price per tariff package *TPi*, N_{TP_i} is number of users per tariff package *TPi* and *k* is the number of TPs (in this case *k*=12).

Monthly prices are determined according to spiral QoS to QoBiz mapping. Thus, $p_{TP_1} = p_0$, $p_{TP_2} = p_{TP_3} = p_{TP_4} = 0.95p_0$, $p_{TP_5} = p_{TP_6} = p_{TP_7} = 0.9p_0$, $p_{TP_8} = p_{TP_9} = p_{TP_{10}} = 0.8p_0$, $p_{TP_{11}} = 0.55p_0$, $p_{TP_{12}} = 0.4p_0$.

Finally, we assume ISP applies hybrid pricing scheme, which is implemented in a fashion that the monthly price is reduced by a certain percentage if users, due to excessive network load (NL), experience significantly lower speed than declared during period longer than defined time interval (typically several minutes).

3. SIMULATION MODEL AND RESULTS

In the paper the ABS methodology is used for development of the simulation model. The main components of any ABS are: *agents*, *environment* and *interactions*. Agents are the basic units of the simulation model, while the environment is the surrounding world in which the agent exists. Interactions are mutual actions that can occur between agents or between agents and the environment (Wilensky and Rand 2015).

The simulation model is coded in C++ programming language with intensive use of Standard Template Library (STL) (Prata 2012). Since it is necessary to simulate and animate the model at the same time a *fixed increment time advance simulation mechanism* is used. Graphical user interface of the simulator is developed using Qt application framework. Qt consists of several libraries and applications and is the compelling framework for cross-platform software development (Lazar and Birminghan 2016). All input parameters for the simulation are located in a relational database. A database file located on a local computer is accessed through the SQLite embedded relational database management system.

In the simulation model following input parameters are chosen:

- total number of users,
- QoS parameters, i.e. download speed, upload speed and security,
- TPs based on QoS parameters combinations and users' requirements,
- monthly price for each TP,
- percentage of users per each TP,
- NL and NL probability,
- variable speed depending on NL,
- time intervals with speed less than declared,
- price reduction depending on number of time intervals with reduced speed and
- simulation period.

In this simulation model, we assume the following QoS setting: download speed (200 Mbit/s, 150 Mbit/s, 100 Mbit/s, 60 Mbit/s, 40 Mbit/s), upload speed (10 Mbit/s, 8 Mbit/s, 6 Mbit/s, 4 Mbit/s, 2 Mbit/s) and security (very high, high, medium, low, very low). Declared bit rates are maximal values for download and upload speed and it is assumed that up to 10% lower bit rates do not mean QoS violation. Maximum price is set to 40 EUR. Thus, for each defined TP in Table 1, concrete values for QoS parameters and prices are specified in Table 2.

Table 2: Tariff packages with input QoS parameters and prices

QoS/ QoBiz	*Download speed* (Mbit/s)	*Upload speed* (Mbit/s)	*Security*	*Price* (EUR)
TP1	200	10	very high	40
TP2	200	10	high	38
TP3	200	8	very high	38
TP4	150	10	very high	38
TP5	200	10	medium	36
TP6	200	6	very high	36
TP7	100	10	very high	36
TP8	200	6	medium	32
TP9	100	10	medium	32
TP10	100	6	very high	32
TP11	60	4	low	22
TP12	40	2	very low	16

Percentage of users per each TP is shown in Figure 2.

We define NL as a percentage of total number of users using Internet at the same time and propose the following NL setting observed over a period longer than 10 minutes:

- *NL1* - If NL < 60%
- *NL2* - If 60% ≤ NL < 80%
- *NL3* - If 80% ≤ NL < 90%
- *NL4* - If NL ≥ 90%

In case of *NL1* network is functioning normally, meaning all QoS parameters are within defined level for each TP and there is no reduction in price. In case of *NL2*, *NL3* and *NL4*, QoS violation occurs, i.e. bit rates are lower than declared for more than 10% during period longer than 10 minutes for:

- *TP1-TP9* and consequently monthly price is reduced for 1% for all users belonging to *TP1-TP9* when *NL2* occurs,
- *TP1-TP10* and consequently monthly price is reduced for 1% for all users belonging to *TP1-TP10* when *NL3* occurs and

- *TP1-TP11* and consequently monthly price is reduced for 1% for all users belonging to *TP1-TP11* when *NL4* occurs.

Price reductions are performed each time *NL2*, *NL3* or *NL4* lasts more than 10 minutes. Only for *TP12* there is no reduction on monthly price regardless of network load.

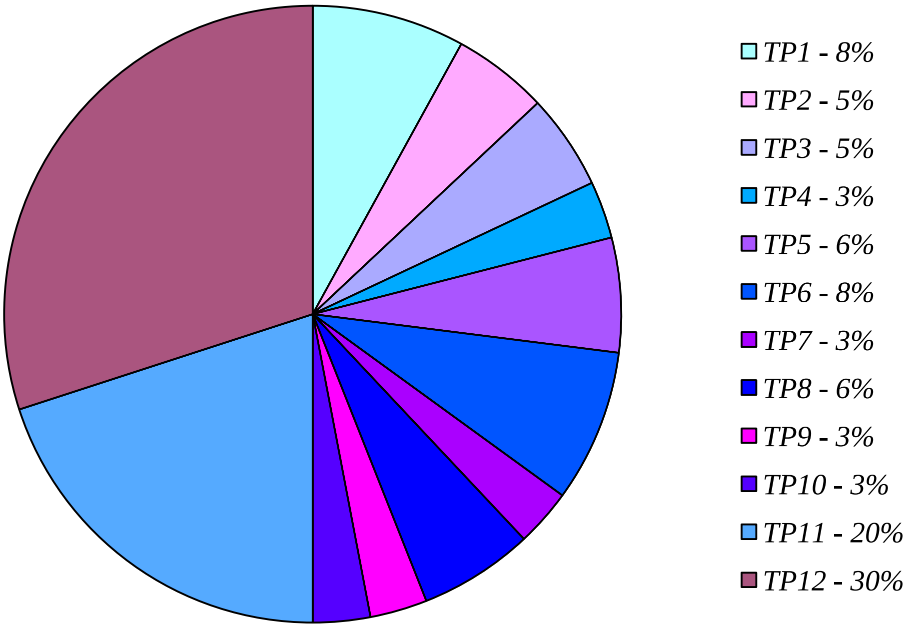


Figure 2: Tariff packages distribution

We assume that daily NLs are different for working days and weekends. Daily NL probabilities for working days and weekends are given in Table 3 and Table 5, respectively. Duration of daily NL for working days and for weekends is exponentially distributed with mean times presented in Table 4 and Table 6, respectively.

Table 3: Daily network load probabilities for working days

Network Load	*00-06*	*06-08*	*08-16*	*16-18*	*18-22*	*22-00*
NL1	0.99	0.95	0.9	0.85	0.75	0.8
NL2	0.01	0.05	0.1	0.1	0.1	0.1
NL3	-	-	-	0.05	0.1	0.05
NL4	-	-	-	-	0.05	0.05

Table 4: Mean time of daily network load for working days

NL / Average duration (min)	*00-06*	*06-08*	*08-16*	*16-18*	*18-22*	*22-00*
NL1	180	55	235	50	100	50
NL2	4	5	6	7	8	8
NL3	-	-	-	5	7	4
NL4	-	-	-	-	5	3

Table 5: Daily network load probabilities for weekends

Network Load	*00-06*	*06-08*	*08-16*	*16-18*	*18-22*	*22-00*
NL1	0.98	0.99	0.85	0.85	0.7	0.75
NL2	0.02	0.01	0.15	0.1	0.15	0.1
NL3	-	-	-	0.05	0.1	0.1
NL4	-	-	-	-	0.05	0.05

Table 6: Mean time of daily network load for weekends

NL / Average duration (min)	*00-06*	*06-08*	*08-16*	*16-18*	*18-22*	*22-00*
NL1	180	60	230	50	95	45
NL2	5	3	5	6	9	9
NL3	-	-	-	5	6	6
NL4	-	-	-	-	5	5

In this simulation we have two types of agents: user and ISP. Two main aspects that define agents are their properties and the methods that they can execute. Properties of user agents are: user's TP, TP price and TP price after reduction. User agent has one method: *reduce TP price*. This method is used for interaction with the ISP which calls this method each time *NL2*, *NL3* or *NL4* lasts more than 10 minutes. The ISP agent has two properties: *maximal revenue* and *revenue with hybrid pricing*. Methods of the ISP agent are: *examine network load* and *initiate reduction of users' prices* (prices per TP). The environment is an IP network where different network loads can happen depending on day and time during a day.

Screenshot of simulator's main window with running animation is shown in Figure 3. Users are presented as circles with color that depends on TP (colors of TPs are the same as in Figure 2). The background color defines network load (*NL1* - green, *NL2* - yellow, *NL3* - orange and *NL4* - red). When *NL2*, *NL3* or *NL4* duration is larger than 10 minutes, ISP initiates price reduction for users in particular TPs. These users are presented as circles with black outline.

Figure 3: Screenshot of simulator software with running animation

Simulation experiment is performed 50 times with 1000 users. For every simulation run different seeds are used for random numbers. Selected simulation period is one year, starting from January 1st to December 31st 2019. Selected time increment is 30 seconds. Simulation results for one run are shown in Figure 4. ISP's revenue is presented annually (Figure 4a) while service prices are shown monthly for different TPs (Figure 4b). We present prices for users belonging to *TP1*, *TP2*, *TP5*, *TP8*, *TP10*, *TP11* and *TP12*. *TP3* and *TP4* are not observed as the price of these tariff packages is the same as *TP2* and they experience equivalent price reduction based of network load. For the same reason we don't present results for *TP6* and *TP7*, which have the same price and price reduction as *TP5*. The same applies for *TP9* which has the same price and price reduction as *TP8*. Although *TP10* has the same price as *TP8*, users belonging to *TP10* experience price reduction if *NL3* or *NL4* occurs, but not in case of *NL2*, which distinguishes them from *TP8* and *TP9*. Service prices are calculated for all users during the period of 600 months. On average *TP1-TP9* users achieve 11.5% price reduction on monthly price, *TP10* users achieve 4.2% monthly price reduction, while *TP11* users achieve only 0.7% monthly price reduction on average.

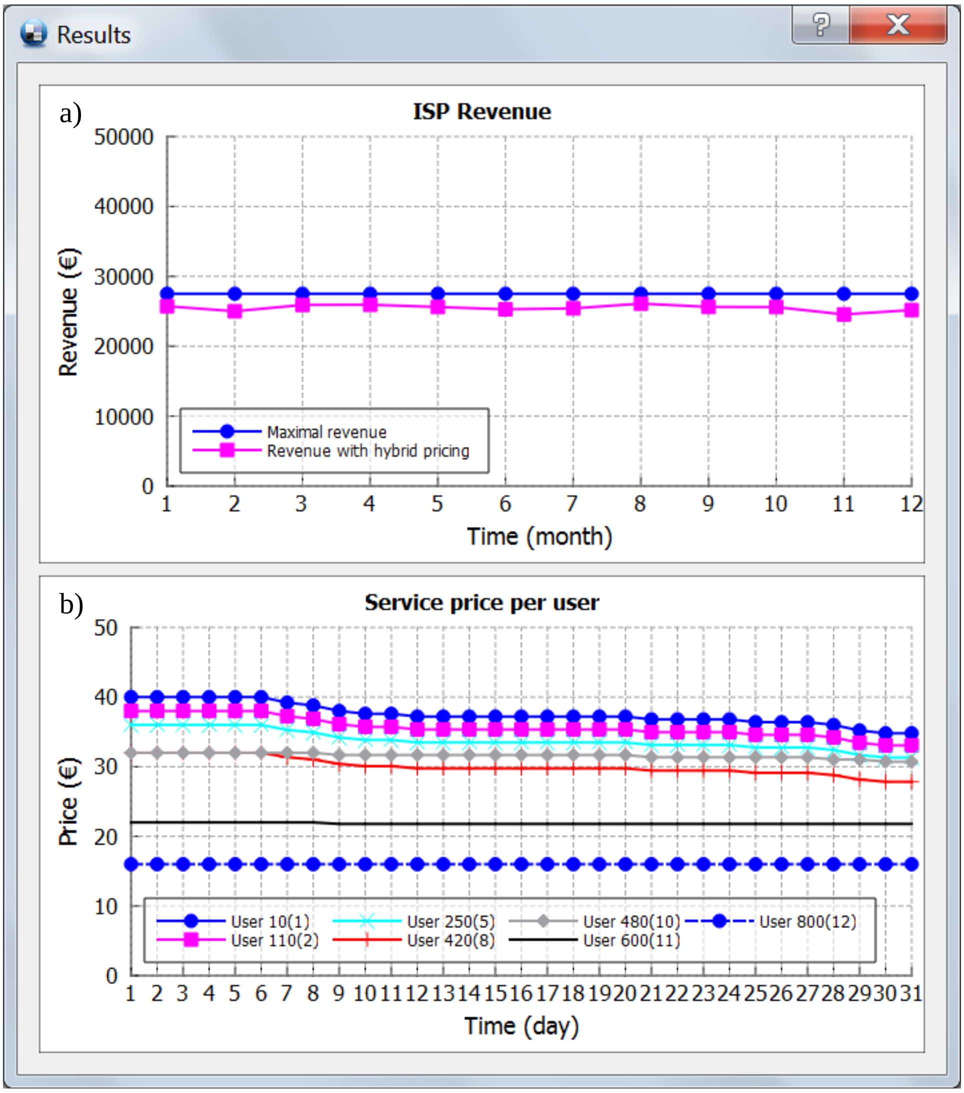


Figure 4: Simulation results: a) Maximal ISP revenue and ISP revenue with hybrid pricing, b) Reduction of service price for selected users

Applying hybrid pricing, ISP expects to gain more users and to achieve maximal revenue at the same time. With this aim, based on simulation results, it is necessary for ISP to improve QoS parameters, which can be performed by expanding the network. Further cost optimization can be conducted in order to reduce costs and to maximize its revenue.

4. CONCLUSION

In the paper we focus on ISP's revenue and service prices, which are observed as main QoBiz parameters from ISP's and users' perspective, respectively. We propose QoBiz model which takes into account QoS parameters and assume ISP applies hybrid pricing scheme. Simulation analysis of the proposed model is conducted using ABS techniques with focus on the interaction between users and ISP in IP network.

One of the advantages of the proposed model is transparent mapping QoS parameters to QoBiz including direct reflection of QoS violation on service prices and consequently on ISP's revenue. Users are charged according not only to declared QoS but also with regards to achieved QoS. Simulation results show prices decrease for most tariff packages. Applying hybrid pricing scheme, ISP is likely to attract new users, but in order to achieve maximal revenue, ISP should tend to improve QoS parameters, which can be performed by additional investments in the network.

In future research, cost optimization for ISP can be performed in order to optimize trade-off between cost investments and expected revenue from new users. Further, the observed ISP's revenues can be simulated in cases of different tariff packages offers. The proposed model can be expanded to include the impact of service prices variations offered by other ISPs in the same market.

Acknowledgement

This work is partially supported by Ministry of Education, Science and Technological Development of the Republic of Serbia under No. 32025 and No. 44004.

REFERENCES

ITU-T Recommendation E.800 (2008). Definitions of terms related to quality of service.

ITU-T Recommendation X.1031 (2008). Roles of end users and telecommunications networks within security architecture.

Lazar, G. & BirminghanR. P. (2016). Mastering Qt 5.Packt Publishing

Radenković, B., Stanojević, M. & Marković, A. (1999). *Computer Simulation* (in Serbian). Belgrade, Serbia: University of Belgrade – Faculty of Organizational Sciences and Faculty of Transport and Traffic Engineering

Prata, S. (2012). C++ Primer Plus (6thed.). New York: Addison-Wesley.

Radonjić Đogatović, V., Kostić-Ljubisavljević, A. & Stojanović, M. (2013). Quality of Business Considerations in Telecommunication Networks, In *Proceedings of the 7th International Quality Conference*, Kragujevac, Serbia, 201-205.

Radonjić Đogatović, V. & Kostić-Ljubisavljević, A. (2015). *Telecommunications Pricing Fundamentals*. Belgrade, Serbia: University of Belgrade – Faculty of Transport and Traffic Engineering.

Radonjić Đogatović, V. (2018). Improving Quality of Business in Next Generation Telecom Networks. In Mehdi Khosrow-Pour (Eds.), *Encyclopedia of Information Science and Technology, Fourth Edition*, IGI Global, 6546-6555.

Stankiewicz, R., Cholda, P. & Jajszczyk, A. (2011). QoX: What is It Really? *IEEE Communications Magazine*, 49(3), 148-158.

Wilensky, U. & Rand, W. (2015). An Introduction to Agent-Based Modeling. London, England: Massachusetts Institute of Technology.

Wolter, K. & Moorsel, A. (2001). The Relationship between Quality of Service and Business Metrics: Monitoring, Notification and Optimization. HP Labs Technical Report, HPL-2001-96.

BALCOR 2018
XIII Balkan Conference on Operational Research

NUMERICAL PERFORMANCE OF TWO MULTILEVEL AMERICAN MONTE CARLO METHODS

HITOSHI INUI[1], KATSUNORI ANO[2]

[1] Waseda University, Institute for Business and Finance, h-inui@y.aoni.waseda.jp, h.inui.08@gmail.com

[2] Shibaura Institute of Technology, kano2@mac.com

Abstract: *A number of American options pricing methods via Monte Carlo simulation have been developed. These methods can be constructed under the multilevel Monte Carlo framework proposed in Giles(2008). In this paper, we construct multilevel pricing algorithms of both Least Squares Monte Carlo and Grant et al.(1996) methods. We test the two algorithms for an American put option pricing. Numerical results show that the former algorithm is efficient and the latter is not efficient in terms of variance reduction.*

Keywords: *Monte Carlo, multilevel Monte Carlo, variance reduction, American option*

1. INTRODUCTION

The Monte Carlo method is already a useful computational tool in finance; however, its computational complexity becomes too large for achieving the required accuracy. Giles (2008) proposed a multilevel Monte Carlo (MLMC) method to reduce computational complexity. The complexity reduction is achieved by the variance reducing effect of the MLMC method. Giles (2008) study tests the MLMC method for pricing European style options.

There are two main styles of options: European style options and American style options. Europeann style options are options that can only be exercised at their maturity date, whereas American style options can be exercised anytime until their maturity. American options pricing is more complicated and more challenging than European options, since at each time step one has to calculate options value and has to see whether options should be exercised or not. Therfore, many simulation-based methods for pricing American style options, such as Grant-Vora-Weeks method (GVW, Grant et al (1996)), Least Squares Monte Carlo method (LSM, (Longstaff and Schwartz(2001)), Stochastic Tree method (ST, Broadie and Glasserman (1997)), Stochastic mesh method (SM, Broadie and Glasserman (2004)), have been developed after Tilley (1993), which is the first research of American options pricing via Monte Carlo simulation, proposed a bundling algorithm. We note that these methods were originally proposed under a standard Monte Carlo (SMC) framework. Under the SMC framework, sample paths generated by only one given time step are used for options pricing. The details are as follows. We consider a stochastic differential equation (SDE) of the form

$$dS_t = \mu(S_t,t)\,dt + \sigma(S_t,t)\,dW_t, \qquad 0 \le t < T, \tag{1}$$

where $S_t \in \mathbb{R}^m$, $S_0 = s$ is given, $T < \infty$, $W_t \in \mathbb{R}^d$ is a standard Brownian motion, and $\mu : \mathbb{R}^m \to \mathbb{R}^m$ and $\sigma : \mathbb{R}^m \to \mathbb{R}^{m\times d}$ are drift and volatility coefficients, respectively. We denote $t_0 = 0$ and $t_D = T$ and divide the interval $[0,T]$ in D subintervals of equal lengths, that is, $[t_0,t_1],[t_1,t_2],\ldots,[t_{D-1},t_D]$; $\Delta t \equiv t_n - t_{n-1} = T/D$ for any $1 \le n \le D$. $t_0,t_1,\ldots,t_D$ are discrete time grids. Then the discretization of $\{S_t\}_t$ using an Euler scheme with Δt is given by

$$\widehat{S}_{t_{n+1}} - \widehat{S}_{t_n} = \mu\left(\widehat{S}_{t_n},t_n\right)\Delta t + \sigma\left(\widehat{S}_{t_n},t_n\right)\Delta W_{t_n}, \quad n = 0,1,\ldots,D-1. \tag{2}$$

Where $\Delta W_{t_n} = W_{t_n} - W_{t_{n-1}}$. Under the SMC framework, we generate sample paths using (2).

On the contrary, the MLMC framework needs multiple time steps

$$h_\ell = T/M^\ell, \ell = L_0, L_0+1,\ldots,L. \tag{3}$$

The MLMC framework needs sample paths generated by a discretization of (1) with each time step h_ℓ are used for computing the MLMC estimate of option price.

The MLMC method has been studied in recent years (e.g., Antithetic variates method: see Giles and Szpruch (2013a) and Giles and Szpruch (2014), Control variates method: Nobile and Tesei (2015), Importance sampling method: Kebaier and Lelong (2017), Quasi Monte Carlo method: Dick, Kuo, Gla, and Scwab (2016), American options: Belomestny et al (2013), introduction: Higham (2015), survey: Giles (2015) and Giles and Szpruch

(2013b)). Belomestny et al (2013) developed a multilevel version of a dual simulation algorithm proposed in Andersen and Broadie (2004). A key feature of the dual algorithm at each time grid is to estimate a conditional expectation, which is a martingale, by using a number of sub-sample paths. We note that Belomestny et al (2013) applies the multilevel treatment to a sequence of martingales rather than time grids; Belomestny et al (2013) uses same time step on all levels of generating sample paths. Our research motive was to investigate whether applying the multilevel idea itself in Giles (2008) to American options pricing methods based on Monte Carlo simulation is effective or not.

We organize this paper as follows. In Section 2, we briefly introduce SMC and MLMC estimators and discuss the optimal number of sample paths. In Section 3, we construct a multilevel version of GVW and LSM algorithms. In Section 4, we illustrate the numerical performance of the constructed multilevel algorithms.

2. MLMC AND SMC METHODS

2.1. SMC and MLMC estimators

We introduce SMC and MLMC estimators. Let $P(\cdot)$ be a payoff function. Under the SMC framework, we can compute the standard estimator Y, that is, the approximation of $E[P(S_t^*)]$

$$Y = N^{-1} \sum_{i=1}^{N} \widehat{P}\left(\widehat{S}_{t^*}^{(i)}\right), \tag{4}$$

where N is the number of simulation paths and $\widehat{P}(\widehat{S}_{t^*}^{(i)})$ is the approximation of $P(S_{t^*})$ with a time step Δt. If we price European style options, t^* is T. If we treat American options pricing, t^* is τ (exercise time). Set $D = M^L$. Under the MLMC method, the option price is constructed by

$$\mathrm{E}[\widehat{P}_L] = \mathrm{E}[\widehat{P}_{L_0}] + \sum_{\ell=L_0}^{L} \mathrm{E}[\widehat{P}_\ell - \widehat{P}_{\ell-1}], \tag{5}$$

where each $\widehat{P}_\ell$ is the approximation of $P(S_{t^*})$ on level ℓ; $\widehat{P}_\ell$ means the discretization of $P(S_{t^*})$ with a time step $h_\ell = T/M^\ell$. The coarsest level and the finest level are zero and L, respectively. The MLMC method uses all levels from zero to L. The MLMC estimator is uniquely constructed by

$$\widehat{Y} = \sum_{\ell=L_0}^{L} \widehat{Y}_\ell,$$

where

$$\widehat{Y}_\ell = \begin{cases} N_{L_0}^{-1} \sum_{i=1}^{N_{L_0}} \widehat{P}_{L_0}^i, & (\ell = L_0), \\ N_\ell^{-1} \sum_{i=1}^{N_\ell} \left(\widehat{P}_\ell^i - \widehat{P}_{\ell-1}^i\right), & (L_0 < \ell \le L). \end{cases}$$

Note that $\widehat{Y}_{L_0}$ is the estimator of $E[\widehat{P}_{L_0}]$ using N_{L_0} simulation paths and that each $\widehat{Y}_\ell$ is the estimator of $E[\widehat{P}_\ell - \widehat{P}_{\ell-1}]$ using N_ℓ paths for $\ell = L_0+1, L_0+2, \ldots, L$.

2.2. Discussion of the optimal number of sample paths

The mean squared error (MSE) of the MLMC estimator is

$$\mathrm{MSE} = \mathrm{E}[(\widehat{Y} - \mathrm{E}[P])^2] = \mathrm{V}[\widehat{Y}] + (\mathrm{E}[\widehat{Y}] - \mathrm{E}[P])^2,$$

where the first term on the right-hand side is the variance of the estimator and the second term is the square of its bias due to discretization. Giles (2008) proves the MLMC complexity theorem. It claims that the computational cost to attain MSE $< \varepsilon^2$ is reduced from $O(\varepsilon^{-3})$ to $O(\varepsilon^{-2}(\ln \varepsilon)^2)$ for a simple case. If the theorem holds, both the variance and the square of bias error have the same upper bound, $\varepsilon^2/2$ (See the proof in Giles (2008)).

The MLMC method needs the optimal number of sample paths for all levels. We discuss the optimal number for a fixed computational complexity

$$C = \sum_{j=L_0}^{L} N_\ell M^\ell = \sum_{j=L_0}^{L} N_\ell T/h_\ell. \tag{6}$$

Note that two algorithms, which are constructed in Section 3 for an American put option pricing, always generate all sample paths from initial date to option's maturity date. We regard $V[\widehat{Y}] = \sum_{\ell=L_0}^{L} N_\ell^{-1} V_\ell$ as a partially differentiable function of $N_\ell, \ell = L_0, L_0+1, \ldots, L$. We want to require the optimal number of sample paths, minimizing $V[\widehat{Y}]$.

To minimize $V[\widehat{Y}]$, we apply the Lagrange Multipliers method. We create the Lagrange equation as follows:

$$\mathcal{L} := \mathcal{L}(N_{L_0}, N_{L_0+1}, \ldots, N_L) = \sum_{\ell=L_0}^{L} N_\ell^{-1} V_\ell - \lambda \left(C - \sum_{\ell=L_0}^{L} N_\ell T / h_\ell \right).$$

Set the partial derivative $\mathcal{L}_{N_{L_0}}, \mathcal{L}_{N_{L_0+1}}, \ldots, \mathcal{L}_L$ equal to zero,

$$\mathcal{L}_{N_\ell} = -N_\ell^{-2} V_\ell + \lambda T / h_\ell = 0, \quad \ell = L_0, L_0+1, \ldots, L.$$

Thereby,

$$N_\ell = \sqrt{\frac{V_\ell h_\ell}{\lambda T}}, \quad \ell = L_0, L_0+1, \ldots, L. \tag{7}$$

If $V[\widehat{Y}] < \varepsilon^2/2$, it holds that

$$V[\widehat{Y}] = \sum_{\ell=L_0}^{L} N_\ell^{-1} V_\ell = \sum_{\ell=L_0}^{L} \sqrt{\frac{\lambda T}{V_\ell h_\ell}} V_\ell < \varepsilon^2/2. \tag{8}$$

Due to (7) and (8), we set

$$N_\ell = 2\varepsilon^{-2} \sqrt{V_\ell h_\ell} \left(\sum_{k=L_0}^{L} \sqrt{V_k / h_k} \right), \quad \ell = L_0, L_0+1, \ldots, L. \tag{9}$$

Therefore, we get

$$\dot{N}_\ell = [N_\ell], \quad \ell = L_0, L_0+1, \ldots, L.$$

where $[n]$ is the least integer greater than or equal to n. If $L_0 = 0$, we note that each $\dot{N}_\ell$ is the optimal simulation times used in Section 5 of Giles (2008). Due to (6) and (9)

$$2\varepsilon^{-2} \sum_{k=L_0}^{L} \sqrt{V_k / h_k} = \frac{C}{\sum_{j=L_0}^{L} T \sqrt{V_j / h_j}}.$$

Using (9) and the above equality, for a fixed C, we have

$$N_\ell = \frac{C \sqrt{V_\ell h_\ell}}{\sum_{j=L_0}^{L} T \sqrt{V_j / h_j}}, \quad \ell = L_0, L_0+1, \ldots, L, \tag{10}$$

Therefore, we have the optimal number of paths for each level, minimizing the variance of the MLMC estimator

$$N_\ell^* = [N_\ell], \quad \ell = L_0, L_0+1, \ldots, L. \tag{11}$$

3. MULTILEVEL ALGORITHMS

We construct a multilevel version of both GVW and LSM pricing algorithms.

For simplicity, we consider an American put option pricing and set multiple time steps $h_\ell = T/2^\ell, \ell = L_0, L_{0+1}, \ldots, L$. The discounted exercise value at time t is

$$f(S_t) = \exp(-rt) \max(K - S_t, 0).$$

3.1. Multilevel version of GVW pricing algorithm

We construct a multilevel version of GVW pricing algorithm (ML-GVW pricing algorithm). First, we execute the following A.1 and B.1.

A.1 Estimate optimal exercise boundaries, $S^*_{t_1}, S^*_{t_2}, \ldots, S^*_{t_D}$, (see Grant et al (1996)).
B.1 Estimate the optimal number of sample paths, $N^*_{L_0}, N^*_{L_0+1}, \ldots, N^*_L$, in (11).

Second, we lord the exercise boundaries and the optimal number of sample paths. We set $N_\ell = N^*_\ell, \ell = L_0, L_0+1, \ldots, L$, and execute the following:

- Estimation of $\widehat{Y}_\ell, \ell = L_0+1, L_0+2, \ldots, L$
  If level $\ell = L$, we note that finer level and coarser level sample paths are $\{\widehat{S}^{f,n}_{t_{D/2^L \cdot 1}}, \widehat{S}^{f,n}_{t_{D/2^L \cdot 2}}, \ldots, \widehat{S}^{f,n}_{t_D}\} = \{\widehat{S}^{f,n}_{t_1}, \widehat{S}^{f,n}_{t_2}, \ldots, \widehat{S}^{f,n}_{t_D}\}, n = 1, 2, \ldots, N_L$ and $\{\widehat{S}^{c,n}_{t_{D/2^{L-1} \cdot 1}}, \widehat{S}^{c,n}_{t_{D/2^{L-1} \cdot 2}}, \ldots, \widehat{S}^{c,n}_{t_D}\} = \{\widehat{S}^{c,n}_{t_2}, \widehat{S}^{c,n}_{t_4}, \ldots, \widehat{S}^{c,n}_{t_D}\}, n = 1, 2, \ldots, N_L$, respectively.

C.1 Generate N_ℓ finer level sample paths $\{\widehat{S}^{f,n}_{t_{D/2^\ell \cdot 1}}, \widehat{S}^{f,n}_{t_{D/2^\ell \cdot 2}}, \ldots, \widehat{S}^{f,n}_{t_D}\}\}, n = 1, 2, \ldots, N_\ell$ and coarser level sample paths $\{\widehat{S}^{c,n}_{t_{D/2^{\ell-1} \cdot 1}}, \widehat{S}^{c,n}_{t_{D/2^{\ell-1} \cdot 2}}, \ldots, \widehat{S}^{c,n}_{t_D}\}, n = 1, 2, \ldots, N_\ell$.
C.2 Set finer level exercise times, $\tau^f_n = \min\{t_j \in \{t_{D/2^\ell \cdot 1}, t_{D/2^\ell \cdot 2}, \ldots, t_D\} | \widehat{S}^{f,n}_{t_j} \leq S^*_{t_j}\}, n = 1, 2, \ldots, N_\ell$, and coarser level exercise times, $\tau^c_n = \min\{t_j \in \{t_{D/2^{\ell-1} \cdot 1}, t_{D/2^{\ell-1} \cdot 2}, \ldots, t_D\} | \widehat{S}^{c,n}_{t_j} \leq S^*_{t_j}\}, n = 1, 2, \ldots, N_\ell$.
C.3 Compute $\widehat{Y}_\ell = N_\ell^{-1} \sum_{n=1}^{N_\ell} \left(\widehat{P}^n_\ell - \widehat{P}^n_{\ell-1}\right) = N_\ell^{-1} \sum_{n=1}^{N_\ell} \left(f(\widehat{S}^{f,n}_{\tau^f_n}) - f(\widehat{S}^{c,n}_{\tau^c_n})\right)$.

- Estimation of Y_{L_0}

D.1 Generate N_{L_0} sample paths $\{\widehat{S}^{f,n}_{t_{D/2^{L_0} \cdot 1}}, \widehat{S}^{f,n}_{t_{D/2^{L_0} \cdot 2}}, \ldots, \widehat{S}^{f,n}_{t_D}\}\}, n = 1, 2, \ldots, N_{L_0}$.
D.2 Set exercise times, $\tau_n = \min\{t_j \in \{t_{D/2^{L_0} \cdot 1}, t_{D/2^{L_0} \cdot 2}, \ldots, t_D\} | \widehat{S}^n_{t_j} \leq S^*_{t_j}\}, n = 1, 2, \ldots, N_{L_0}$.
D.3 Compute $\widehat{Y}_{L_0} = N_{L_0}^{-1} \sum_{n=1}^{N_{L_0}} \widehat{P}^n_{L_0} = N_{L_0}^{-1} \sum_{n=1}^{N_{L_0}} f(\widehat{S}^n_{\tau_n})$.

- Calculation of American option price

E.1 Compute $\widehat{Y} = \sum_{\ell=L_0}^{L} Y_\ell$.

3.2. Multilevel version of LSM pricing algorithm

We construct a multilevel version of LSM pricing algorithm (ML-LSM pricing algorithm). First, we execute the following A.1 and B.1.

A.1 Estimate Longstaff-Schwartz regression coefficients, $a_i, i = 1, 2, \ldots, D$ (see (Longstaff and Schwartz(2001)).
B.1 Estimate the optimal number of sample paths, $N^*_{L_0}, N^*_{L_0+1}, \ldots, N^*_L$, in (11).

Second, we lord the Longstaff-Schwartz regression coefficients and the number of sample paths. We set $N_\ell = N^*_\ell, \ell = L_0, L_0+1, \ldots, L$, and execute the following:

- Estimation of $\widehat{Y}_\ell, \ell = L_0+1, L_0+2, \ldots, L$
  If level $\ell = L$, we note that finer level and coarser level sample paths are $\{\widehat{S}^{f,n}_{t_{D/2^L \cdot 1}}, \widehat{S}^{f,n}_{t_{D/2^L \cdot 2}}, \ldots, \widehat{S}^{f,n}_{t_D}\} = \{\widehat{S}^{f,n}_{t_1}, \widehat{S}^{f,n}_{t_2}, \ldots, \widehat{S}^{f,n}_{t_D}\}, n = 1, 2, \ldots, N_L$ and $\{\widehat{S}^{c,n}_{t_{D/2^{L-1} \cdot 1}}, \widehat{S}^{c,n}_{t_{D/2^{L-1} \cdot 2}}, \ldots, \widehat{S}^{c,n}_{t_D}\} = \{\widehat{S}^{c,n}_{t_2}, \widehat{S}^{c,n}_{t_4}, \ldots, \widehat{S}^{c,n}_{t_D}\}, n = 1, 2, \ldots, N_L$, respectively.

C.1 Generate N_ℓ finer level sample paths $\{\widehat{S}^{f,n}_{t_{D/2^\ell \cdot 1}}, \widehat{S}^{f,n}_{t_{D/2^\ell \cdot 2}}, \ldots, \widehat{S}^{f,n}_{t_D}\}\}, n = 1, 2, \ldots, N_\ell$ and coarser level sample paths $\{\widehat{S}^{c,n}_{t_{D/2^{\ell-1} \cdot 1}}, \widehat{S}^{c,n}_{t_{D/2^{\ell-1} \cdot 2}}, \ldots, \widehat{S}^{c,n}_{t_D}\}, n = 1, 2, \ldots, N_\ell$.
C.2 Lord finer level regression coefficients, $a_{D/2^\ell \cdot 1}, a_{D/2^\ell \cdot 2}, \ldots, a_D$, and coarser level regression coefficients, $a_{D/2^{\ell-1} \cdot 1}, a_{D/2^{\ell-1} \cdot 2}, \ldots, a_D$.
C.3 Compute finer level continuation values, $\{C^f_{t_j} \sum_{i=1}^{B} a_j b_i(\widehat{S}^{f,n}_{t_j}), j = D/2^\ell \cdot 1, D/2^\ell \cdot 2, \ldots, D\}, n = 1, 2, \ldots, N_\ell$, and coarser level continuation values, $\{C^c_{t_j} \sum_{i=1}^{B} a_j b_i(\widehat{S}^{c,n}_{t_j}), j = D/2^{\ell-1} \cdot 1, D/2^{\ell-1} \cdot 2, \ldots, D\}, n = 1, 2, \ldots, N_\ell$, where $b_i, i = 1, \ldots, B$ are Laguerre polynomials.
C.4 Compute finer level discounted exercise values (discounted payoff function values), $\{f(\widehat{S}^{f,n}_{t_j}), j = D/2^\ell \cdot 1, D/2^\ell \cdot 2, \ldots, D\}, n = 1, 2, \ldots, N_\ell$, and compute coarser level discounted exercise values, $\{f(\widehat{S}^{c,n}_{t_j}), j = D/2^{\ell-1} \cdot 1, D/2^{\ell-1} \cdot 2, \ldots, D\}, n = 1, 2, \ldots, N_\ell$.
C.5 Set finer level exercise times, $\tau^f_n = \min\{t_j \in \{t_{D/2^\ell \cdot 1}, t_{D/2^\ell \cdot 2}, \ldots, t_D\} | f(\widehat{S}^{f,n}_{t_j}) \geq C^f_{t_j}\}, n = 1, 2, \ldots, N_\ell$, and coarser level exercise times, $\tau^c_n = \min\{t_j \in \{t_{D/2^{\ell-1} \cdot 1}, t_{D/2^{\ell-1} \cdot 2}, \ldots, t_D\} | f(\widehat{S}^{c,n}_{t_j}) \geq C^c_{t_j}\}, n = 1, 2, \ldots, N_\ell$.
C.6 Compute $\widehat{Y}_\ell = N_\ell^{-1} \sum_{n=1}^{N_\ell} \left(\widehat{P}^n_\ell - \widehat{P}^n_{\ell-1}\right) = N_\ell^{-1} \sum_{n=1}^{N_\ell} \left(f(\widehat{S}^{f,n}_{\tau^f_n}) - f(\widehat{S}^{c,n}_{\tau^c_n})\right)$

- Estimation of Y_{L_0}

D.1 Generate N_{L_0} sample paths $\{\widehat{S}^{f,n}_{t_{D/2^{L_0}\cdot 1}}, \widehat{S}^{f,n}_{t_{D/2^{L_0}\cdot 2}}, \ldots, \widehat{S}^{f,n}_{t_D}\}\}, n = 1,2,\ldots,N_{L_0}$.

D.2 Lord regression coefficients, $a_{D/2^{L_0}\cdot 1}, a_{D/2^{L_0}\cdot 2}, \ldots, a_D$.

D.3 Compute finer level continuation values, $\{C_{t_j} \sum_{i=1}^{B} a_j b_i(\widehat{S}^n_{t_j}), j = D/2^{L_0}\cdot 1, D/2^{L_0}\cdot 2, \ldots, D\}, n = 1,2,\ldots,N_{L_0}$.

D.4 Compute level discounted exercise values (discounted payoff function values), $\{f(\widehat{S}^n_{t_j}), j = D/2^{L_0}\cdot 1, D/2^{L_0}\cdot 2, \ldots, D\}, n = 1,2,\ldots,N_{L_0}$.

D.5 Set exercise times, $\tau_n = \min\{t_j \in \{t_{D/2^{L_0}\cdot 1}, t_{D/2^{L_0}\cdot 2}, \ldots, t_D\} | f(\widehat{S}^n_{t_j}) \geq C_{t_j}\}, n = 1,2,\ldots,N_{L_0}$,

D.6 Compute $\widehat{Y}_{L_0} = N_{L_0}^{-1} \sum_{n=1}^{N_{L_0}} \widehat{P}^n_{L_0} = N_{L_0}^{-1} \sum_{n=1}^{N_{L_0}} f(\widehat{S}^n_{\tau_n})$.

- Calculation of American option price

E.1 Compute $\widehat{Y} = \sum_{\ell=L_0}^{L} Y_\ell$.

4. NUMERICAL EXPERIMENTS

We test the ML-GVW and ML-LSM algorithms for an American put option pricing. We compare the ML-GVW algorithm with the GVW algorithm and compare the ML-LSM algorithms with the LSM algorithm in terms of variance reduction.

We set $M = 2$ (see (3)), $K = 1$, $L_0 = 1$, $L = 6$, $C = 640,000$ (ML-LSM) and $M = 2$ (see (3)), $K = 1$, $L_0 = 1$, $L = 4$, $C = 160,000$ (ML-GVW and ML-GVW with a Brownian bridge interpolation). We present each result for

$$dS_t = 0.06 S_t dt + 0.4 S_t dW_t, \quad 0 \leq t < T,$$

where $S_0 = 1$ and $T = 1$. We note that the results of LSM are based on the parameter sets (N in (4) : 10,000 paths, D in (2) : 2^6 time-steps) and the results of GVW are based on the parameter sets (N in (4) : 10,000 paths, D in (2) : 2^4 time-steps).

In comparison to LSM, Table 1 shows that ML-LSM achieves variance reduction of about 38%. By contrast, Table 2 shows that ML-GVW has larger variance and does not achieve variance reduction. Table 3 shows that the use of ML-LSM get high correlation coefficients between $\widehat{P}^n_\ell$ and $\widehat{P}^n_{\ell-1}$. On the other hand, Table 4 shows that the use of ML-GVW get low correlation coefficients. Tables 4 and 5 show that coarser level interpolate-paths using a Brownian bridge method achieve high correlation coefficients between $\widehat{P}^n_\ell$ and $\widehat{P}^n_{\ell-1}$. Table 2 also shows that the interpolate-paths using a Brownian bridge method achieve smaller variance; in contrast to usual ML-GVW, ML-GVW with a Brownian bridge method achieves variance reduction of about 57 %.

Table 1: comparison of variances

	ML-LSM	LSM
option price	0.1302	0.1295
variance	1.5337E-06	2.4750E-06

Table 2: comparison of variances

	ML-GVW	ML-GVW with Brownian bridge	GVW
option price	0.131	0.134	0.133
variance	4.163E-06	1.780E-06	1.851E-06

Table 3: comparison of correlations (ML-LSM)

$\widehat{P}_\ell$	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}_5$	$\widehat{P}_6$
$\widehat{P}_{\ell-1}$	-	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$	$\widehat{P}_5$
correlation	-	0.9681	0.9717	0.9764	0.9807	0.9871

Table 4: comparison of correlations (ML-GVW)

$\widehat{P}_\ell$	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$
$\widehat{P}_{\ell-1}$	-	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$
correlation	-	0.8904	0.9138	0.9370

Table 5: comparison of correlations (ML-GVW with Brownian bridge)

$\widehat{P}_\ell$	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$	$\widehat{P}_4$
$\widehat{P}_{\ell-1}$	-	$\widehat{P}_1$	$\widehat{P}_2$	$\widehat{P}_3$
correlation	-	0.9814	0.9746	0.9784

5. CONCLUSION

The numerical results show that the ML-LSM algorithm performs better than a standard LSM and show that the ML-GVW algorithm is not efficient in terms of variance reduction. However, the use of the Brownian bridge interpolation method achieves smaller variance. Future research can have one direction. It can apply MLMC method to other well-known American options pricing methods via Monte Carlo simulation, such as the SM method, the ST method and Tilley's bundling algorithm.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 17K18182.

REFERENCES

Andersen, L. and M. Broadie. (2004). "A primal-dual simulation algorithm for pricing multidimensional American options." *Management Sciences*, **50** (9), pp. 1222-1234.

Belomestny, D., J. Schoenmakers, and F. Dickmann. (2013). "Multilevel dual approach for pricing American type derivatives." *Finance and Stochastics*, **17**(4), pp. 717-742

Broadie, M. and P. Glasserman. (1997). "Pricing American-Style Securities Using Simulation." *Journal of Economic Dynamics and Control*, **21** (8)-(9), pp. 1323-1352.

Broadie, M. and P. Glasserman. (2004). "A Stochastic Mesh Method for Pricing High-Dimensional American Options." *The Journal of Computational Finance*, **7** (4), pp. 35-72.

Dick, J., F. Y. Kuo, T. Le. Gla, and C. Schwab. (2016). "Multilevel Higher Order QMC Petrov–Galerkin Discretization for Affine Parametric Operator Equations." *SIAM Journal on Numerical Analysis*, **54** (5), pp. 2541-2568.

Giles, M. B. (2008). "Multilevel Monte Carlo path simulation." *Operations Research*, **56** (3), pp. 607-617.

Giles, M. B. (2015). "Multilevel Monte Carlo methods." *Acta Numerica*, **24**, pp. 259-328.

Giles, M. B. and L. Szpruch. (2013a). "Antithetic multilevel Monte Carlo estimation for multidimensional SDEs." In Monte Carlo and Quasi-Monte Carlo Methods 2012, (Edited by Dick, J., F. Y. Kuo, G. W. Peters, and I. H. Sloan), pp. 367-384, Springer.

Giles, M. B. and L. Szpruch. (2013b). "Multilevel Monte Carlo methods for applications in finance." In Recent Developments in Computational Finance, (Edited by Gerstner, T. and P. E. Kloeden), World Scientific.

Giles, M. B. and L. Szpruch. (2014). "Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation." *Finance Stoch*, **24** (4), pp. 403-413.

Grant, D., G. Vora, and D. Weeks. (1996). "Simulation and the Early-Exercise Option Problem." *Journal of Financial Engineering*, **5** (3), pp. 211-227.

Higham, D. J. (2015). "An introduction to multilevel Monte Carlo for option valuation." *International Journal of Computer Mathematics*, **92** (12), pp. 2347-2360.

Kebaier, A. and J. Lelong. (2017). "Coupling Importance Sampling and Multilevel Monte Carlo using Sample Average Approximation." *Methodology and Computing in Applied Probability*, https://doi.org/10.1007/s11009-017-9579-y.

Longstaff, F. A. and E. S. Schwartz. (2001). "Valuing American options by simulation: A simple least squares approach." *Review of Financial Studies*, **14**, pp. 113-147.

Nobile, F. and F. Tesei. (2015). “A Multi Level Monte Carlo method with control variate for elliptic PDEs with log-normal coefficients.” *Stochastic Partial Differential Equations: Analysis and Computations*, **3** (3), pp. 398-444.

Tilley, J. A. (1993). “Valuing American Options in a Path Simulation Model.” Transactions of the Society of Actuaries, **45**, pp. 83-104.

B21
Structural Optimization

XIII Balkan Conference on Operational Research

ONE APPROACH FOR THE GROUP SYNTHESIS OF RECOGNITION AND CLASSIFICATION TASKS

KALIMOLDAYEV MAKSAT[1], AMIRGALIYEV YEDILKHAN[1,5], BERIKOV VLADIMIR[2,3] , CHERICBAYEVA LYAILA[4], LATUTA KONSTANTIN[5], KALYBEK UULU BEKTURGAN[6]

[1]Institute of Information and Computational Technologies, SC MES RK, Almaty, mnk@ipic.kz, amir_ed@mail.ru
[2] Sobolev Institute of Mathematics, SB RAS, Novosibirsk, berikov@math.nsc.ru
[3]Novosibirsk State University, berikov@math.nsc.ru
[4]Alfarabi Kazakh National University, Almaty, lyailya_sh@mail.ru
[5]Suleyman Demirel University, Almaty, konstantin.latuta@sdu.edu.kz
[6]Institute of Automation and Information Technology of the NAN KR, Kyrgyzstan, ukuevb@gmail.com

Abstract: *In this work semi-supervised learning was considered. To solve the problem of semi-supervised learning CASVM and CANN algorithms were developed. The algorithms are based on combination of collective cluster analysis and kernel methods. Probabilistic model of classification with use of cluster ensemble was proposed. Within the model, error probability of CANN was studied. Assumptions that make probability of error converge to zero were formulated. The proposed algorithms were experimentally tested on a hyper spectral image. It was shown that CASVM is more noise resistant than standard SVM.*

Keywords: *Recognition, Classification, Hyper spectral image, Semi-supervised learning.*

1. INTRODUCTION

At present, a sufficiently large number of algorithms for cluster analysis have been developed (Berikov 2013, Amirgaliev & Mukhamedgaliev 1985, Aidarkhanov et al. 2001) . The problem of cluster analysis can be formulated as follows. There are many objects described by a set of some variables (or a distance matrix). These objects are to be broken down into a relatively small number of clusters (groups, classes) so that the grouping criterion would take its "best" value. The number of clusters can be either selected in advance or not specified at all (in the latter case, the optimal number of clusters must be determined automatically). A quality criterion usually means a certain function, depending on the scatter of objects within the group and the distances between groups (Berikov 2013, Amirgaliev & Mukhamedgaliev 1985) .

Recently cluster analysis has been actively developing an approach based on collective decision-making. It is known that algorithms of cluster analysis are not universal: each algorithm has its own specific area of application: for example, some algorithms can better cope with problems in which objects of each cluster are described by "spherical" regions of multidimensional space; other algorithms are designed to search for "tape" clusters, etc. In the case when the data are of a heterogeneous nature, it is advisable to use not one algorithm but a set of different algorithms to allocate clusters. The collective (ensemble) approach also makes it possible to reduce the dependence of grouping results on the choice of parameters of the algorithm, to obtain more stable solutions in the conditions of "noisy" data, if there are "omissions" in them (Berikov, 2013, Amirgaliev & Mukhamedgaliev 1985, Aidarkhanov et al. 2001) .

In classification problems, group methods are widely used. They consist in the synthesis of results obtained by applying different algorithms to a given source information, or in selection of optimal, in some sense, algorithms from a given set. There are various ways of defining group classifications. The formation of recognition as an independent scientific theory is characterized by the following stages:

- The appearance of a large number of various incorrect (heuristic) methods and algorithms to solve practical problems, oftentimes applied without any serious justification.
- The construction and research of collective (group) methods, providing a solution to the problem of recognition based on the results.
- processing of initial information by separate algorithms (Joydeep & Ayan, 2011) .

Ensemble approach allows improving the quality of clustering. There are several main directions in the methods of constructing ensemble solutions of cluster analysis: based on the consensus distribution, on the co-associative matrices, on the models of the mixture of distributions, graph methods, and so on. as well as the main methods for obtaining collective cluster solutions: the use of a pairwise similarity/difference matrix;

maximization of the degree of consistency of decisions (normalized mutual information, corrected Rand index, etc.) Each cluster analysis algorithm has some input parameters, for example, the number of clusters, the boundary distance, etc. In some cases, it is not known what parameters of the algorithm work best. It is advisable to apply the algorithm with several different parameters rather than one specific parameter.

1.1. Ensemble cluster analysis

The group ensemble consists of different partitions. Such partitions can be obtained from several applications of any one algorithm with different parameters, or from applying different algorithms to one data set. The proposals of cluster ensembles solve the problems inherent in clustering: they can provide more reliable and stable solutions using consensus on several clustering results. The orthogonal problem associated with clustering is a large dimension. Large data is a complex task for the clustering process. Different algorithms of clustering can process data of low dimensionality, but as data dimension increases, these algorithms tend to collapse. In large-size spaces, it is very likely that for any given pair of points within one cluster, there are at least several dimensions at which the points are far apart. As a consequence, distance functions that are using equally all input functions can be inefficient. The technique of a group ensemble is characterized by two components: a mechanism for creating a variety of sections, as well as a consensus function for combining the input of partitions into finite clustering. A variety of partitions are usually generated using different clustering algorithms, or by applying a single algorithm with different parameters, possibly in combination with a selection of data or functions. One popular methodology for constructing a consensus function uses a co-associated matrix. Such a matrix can be considered as a similarity matrix, so it can be used with any clustering algorithm that works directly on similarities.

Let set S = {x_1, x_2, ..., x_n} has n points. An ensemble is a collection of m clustering solutions: G = {G_1, G_2, ..., G_m}. Each solution G_L for L = 1, ..., m, is a partitioning of set S, that is, G_L = {G_L^1, G_L^2 , …, $G_L^{K_l}$}, where $\cup_K\ G_L^K$=S. Given the set of cluster solutions C and the desired number of clusters k, the goal is to combine different solution clusters and calculate a new partition of S into k disjoint clusters. The task of cluster ensembles is to develop an appropriate consensus function that integrates cluster component solutions into "improved" final clustering.

In this work semi-supervised learning is considered. In semi-supervised learning the classes are known only for a subset of objects in the sample. The problem of semi-supervised learning is important for the following reasons:

- unlabeled data is cheap
- labeled data may be difficult to obtain
- using unlabeled data along with some labeled data may increase the quality of learning

There are many algorithms and approaches to solve the problem of semi-supervised learning (Joydeep & Ayan, 2011) . The goal of the work is to devise and test a novel approach to semi-supervised learning. The novelty lies in the combination of algorithms of collective cluster analysis (Domeniconi & Al-Razgan 2009, Berikov 2014) and kernel methods (support vector machines *SVM* (Berikov & Pestunov, 2017) and nearest neighbor *NN*), as well as in theoretical analysis of the error of the proposed method. In the coming sections a more formal problem statement will be given, some cluster analysis and kernel methods will be reviewed, the proposed methods will be described and its theoretical and experimental ground will be provided.

Cluster ensembles combine multiple clusters of a set of objects into one consolidated clustering, often called a consensus solution.

2. FORMAL PROBLEM STATEMENT OF SEMI-SUPERVISED LEARNING

Suppose we have a set of objects X to classify and finite set of class labels Y . Features describe all the objects. By a feature of an object we mean the following mapping $f: X \to D_f$, where D_f — set of values of a feature.

Depending on D_f features can be of the following types: Binary features: D_f ={0,1}; Numerical features: $D_f = R$; Nominal features: D_f — finite set; Ordered features: D_f — finite ordered set.

For a given feature vector $f_1,.., f_m$, vector $x = (f_1(\alpha),.., f_m(\alpha))$ is called feature descriptor of object $\alpha \in X$. Further, in the text we do not distinguish between an object and its feature descriptor. In the problem of semi-supervised learning at the input we have a sample $X_N = \{ x_1,.., x_N \}$ of objects from X .

There are two types of objects in the sample:

- $X_c = \{x_1,..,x_k\}$ - labeled objects with the classes they belong to: $Y_c = \{y_1,..,y_k\}$
- $X_u = \{x_{k+1},..,x_N\}$ - unlabeled objects

Conduct so-called inductive learning — build a classification algorithm $a : X \to Y$, which minimizes probability of error and match objects to their X_u, and new objects to X_{test}, which were unavailable at the time of building of the algorithm.

The second is so-called transductive learning. Here we get labels only for objects from X_u with minimal error. In this work, we consider the second variant of problem statement.

The following example shows how semi-supervised learning differs from a supervised learning.

Example: Label objects are given at the input $X_c = \{x_1,..,x_k\}$ with their respective classes $Y_c = \{y_1,..,y_k\}$, where $y_i \in \{0,1\}, i = 1,..,k$. The objects have two features and their distribution is shown in Figure 1 a). Unlabeled data is also given $X_u = \{x_{k+1},..,x_N\}$ as shown in Figure 1b).

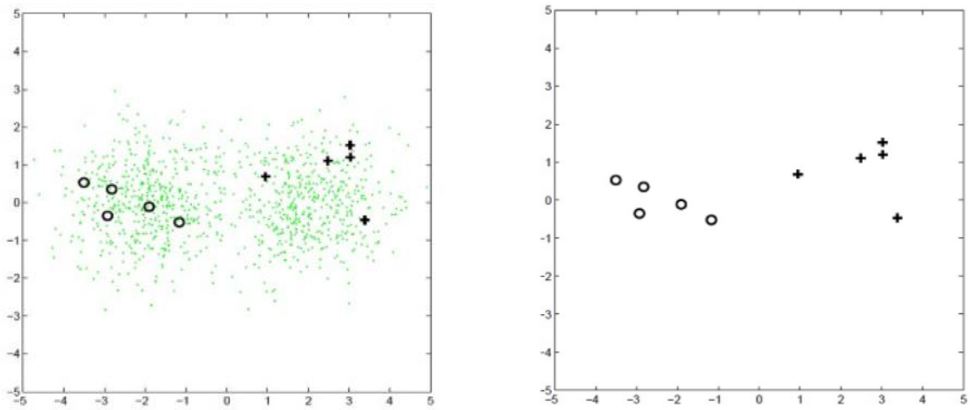


Figure 1: a) Features of objects; b) Labeled objects X_c with unlabeled objects X_u

Suppose that a sample from a mixture of normal distributions is given. Let's estimate the density of the classes throughout the data set at only on the labeled data, after which we construct the separating curves. Then, from Figure 2 it can be seen that the quality of the classification using the full set of data is higher.

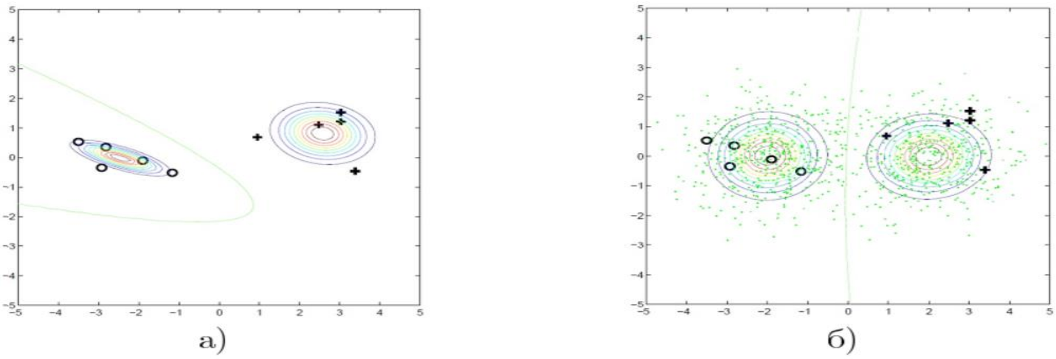


Figure 2: Obtained class densities: a) by labeled data; b) by unlabeled data

3. COLLECTIVE SOLUTIONS IN CLUSTER ANALYSIS

3.1. On the reasons for the development of the collective approach

The task of cluster analysis is to split the sample into disjoint subsets, called clusters, so that each cluster represents a group of similar objects, and objects in different clusters differ significantly. The solution of the clustering problem may be ambiguous for several reasons:

- There is no best criterion for the quality of clustering. A large number of reasonable heuristic criteria and algorithms are known that do not have an explicitly defined criterion, but quite a decent clustering;
- The number of clusters is very often unknown in advance and is set either manually or during the operation of the algorithm;
- The results of clustering depend very much on the metric, which is chosen by the expert and the specifics of the application domain.

3.2. The matrix of average differences

To construct a matrix of average differences, clustering of all available objects $X = \{x_1,..,x_N\}$ is done by an ensemble of several different algorithms $\mu_1,..,\mu_M$. Each algorithm gives L_m variants of partition, $m = 1,..,M$. Based on the results of the algorithms, a matrix H of average differences is built for objects of X. The matrix elements are equal to:

$$h(i,j) = \sum_{m=1}^{M} \alpha_m \frac{1}{L_m} \sum_{l=1}^{L_m} h_{lm}(i,j), \qquad (1)$$

where $i, j \in \{1,..,N\}$ - objects' numbers $(i \neq j), \alpha_m \geq 0$ - initial weights so that $\sum_{m=1}^{M} \alpha_m = 1; h_{lm}(i,j) = 0$, if pair (i,j) belong to different clusters in l-th variant of partition, given by algorithms μ_m and 1, if it belongs to the same cluster.

Weights α_m may be same or, for example, may be set with respect to quality of each clustering algorithm. The selection of optimal weights is researched in Domeniconi & Al-Razgan (2009) .

4. KERNEL METHODS OF CLASSIFICATION

To solve the classification problem, kernel methods are widely used, based on the so-called "kernel trick". To demonstrate the essence of this "trick", consider the support vector machine method (SVM) - the most popular kernel method of classification. SVM is a binary classifier, although there are ways to refine it for multiclassification.

4.1. Binary classification with SVM

In the problem of dividing into two classes (the problem of binary classification), a training sample of objects $X = \{x_1,..,x_n\}$ is at the input with classes $Y = \{y_1,..,y_n\}$, $y_i \in \{+1,-1\}$, for $i = 1,..,n$, where object are points in m-dimensional space of feature descriptors. We are to divide the points by hyperplane of dimension ($m-1$). In the case of linear class separability, there exist an infinite number of separating hyperplanes. It is reasonable to choose a hyperplane, the distance from which to both classes is maximized. An optimal separating hyperplane is a hyperplane that maximizes the width of the dividing strip between classes. The problem of the support vector machine method consists in constructing an optimal separating hyperplane. The points lying on the edge of the dividing strip are called support vectors.

A hyperplane can be represented as $\langle w, x \rangle + b = 0$, where $\langle \, , \rangle$ — scalar product, w — vector perpendicular to separating hyperplane, and b — an auxiliary parameter. Support vector method builds decision function in in the form of

$$F(x) = sign(\sum_{i=1}^{n} \lambda_i c_i \langle x_i, x \rangle + b)$$

It is important to note that the summation goes only along support vectors for which $\lambda_i \neq 0$. Objects $x \in X$ with $F(x) = 1$ will be assigned one class, and objects with $F(x) = 0$ another.

With linear inseparability of classes, one can perform a transformation $\varphi : X \to G$ of object space X to a new space G of a higher dimension. The new space is called is called "rectifying", because the objects in the space can already be linearly separable.

Decision function $F(x)$ depends on scalar products of objects, rather that the objects themselves. That is why scalar products $\langle x, x' \rangle$ can be substituted by products of $\langle \varphi(x), \varphi(x') \rangle$ kind in the space G. In this case the decision function $F(x)$ will look like this:

$$F(x) = sign(\sum_{i=1}^{n} \lambda_i c_i \langle \varphi(x_i), \varphi(x) \rangle + b)$$

Function $K(x, x') = \langle \varphi(x), \varphi(x') \rangle$ is called kernel. The transition from scalar products to arbitrary kernels the "kernel trick". Selection of the kernel determines the rectifying space and allows to use linear algorithms (like SVM) to linearly non-separable data.

5. PROPOSED METHOD

The idea of the method is to construct a similarity matrix (1) of all objects from the input sample X. The matrix will be compiled by applying different clustering algorithms to X. The more a pair of objects are classified as belonging to one class the more similar they will be. Two possible variants of prediction for unlabeled classes X_u will be proposed using similarity matrix. Further the idea of the algorithms will be described in detail. The following theorem holds:

Theorem 1. Let $\mu_1, .., \mu_M$ — be algorithms of clustering analysis , each algorithm gives L_m variants of partition, $m = 1, .., M$, $h_{lm}(x, x') = 0$, if a pair of objects (x, x') belongs to different clusters in l-th variant of partition, given by algorithm μ_m and 1, if it belongs to the same cluster. $\alpha_m \geq 0$ - initial weights such that $\sum_{m=1}^{M} \alpha_m = 1$. Then function $H(x, x') = \sum_{m=1}^{M} \alpha_m \frac{1}{L_m} \sum_{l=1}^{L_m} h_{lm}(x, x')$ satisfies the condition of Mercer theorem.

Proof. It is obvious that function $H(x, x')$ symmetric. Let C_r^{lm} - be the set of indices of objects that belong to r-th cluster, given by m-th algorithmin l-th variant of partition. Let's show that $H(x, x')$ nonnegatively defined.

Let take arbitrary $z \in R^p$ and show that $z^T Hz \geq 0$

$$z^T Hz = \sum_{i,j=1}^{p} \sum_{m=1}^{M} \alpha_m \frac{1}{L_m} \sum_{l=1}^{L_m} h_{lm}(i, j) z_i z_j = \sum_{m=1}^{M} \alpha_m \frac{1}{L_m} \sum_{l=1}^{L_m} \sum_{i,j=1}^{p} h_{lm}(i, j) z_i z_j =$$

$$\sum s_{m=1}^{M} \alpha_m \frac{1}{L_m} \sum_{l=1}^{L_m} (\sum_{i,j \in C_1^{lm}} z_i z_j + ... + \sum_{i,j \in C_{K_{lm}}^{lm}} z_i z_j) = \sum_{m=1}^{M} \alpha_m \frac{1}{L_m} \sum_{l=1}^{L_m} ((\sum_{i inC_1^{lm}} z_i)^2 + ... + (\sum_{i \in C_{K_{lm}}^{lm}} z_i)^2) \geq 0.$$

Thus, function $H(x, x')$ can be used as a kernel in kernel methods of classification. For instance, in support vector machines (SVM) and in nearest neighbor method (NN). Further, the two variants of the algorithm that implement the proposed method are described:

Algorithm CASVM

Input: objects X_c with their classes Y_c and objects X_u, number of clustering algorithms M, number of clustering L_m by each algorithm $\mu_m, m = 1, ..., M$.

Output: classes of objects X_u.

1. Cluster objects $X_c \cup X_u$ by algorithms $\mu_1, .., \mu_M$, and get L_m variants of partitions from each algorithm $\mu_m, m = 1, ..., M$.

2. Computer matrix H for $X_c \cup X_u$ by formula (1).

3. Train *SVM* with labeled data X_c, using matrix H as kernel.

4. By means of *SVM* predict classes of unlabeled data X_u.

End of algorithm

Algorithm CANN

Input: objects X_c with given classes Y_c and objects X_u, number of clustering algorithms M, number for clusters L_m by each algorithm $\mu_m, m = 1,...,M$.

Output: classes of objects X_u.

1. Cluster objects $X_c \cup X_u$ by algorithms $\mu_1,..,\mu_M$, get L_m variants of partitions from each algorithm $\mu_m, m = 1,...,M$.

2. Compute H for $X_c \cup X_u$ by formula (1).

3. Use *NN*: for each unlabeled object $x \in X_u = \{x_{k+1},..,x_N\}$ assign the most similar class in sense $H(x, x')$ of labeled object $x' \in X_c = \{x_1,..,x_k\}$.

Formally written: $x_i = \arg\max_{j=1,..,k} H(x_i, x_j),\ i = k+1,..,N$.

End of algorithm

Note that in the proposed algorithms there is no need to store matrix H in memory $N \times N$ entirely: it is enough to store the clustering matrix of size $N \times L$, where $L = \sum_{l=1}^{M} L_m$, in this case H can be computed dynamically. In practice, $L << N$, for example, when working with image pixels.

6. THEOETICAL ANALYSIS OF CANN

Let's recall the problem statement: At the input we have sample of objects $X_N = \{x_1,..,x_N\}$. There are two types of objects in the sample:

- $X_c = \{x_1,..,x_k\}$ - labeled objects with classes $Y_c = \{y_1,..,y_k\}$, $I_c = \{1,...,k\}$ - object indices
- $X_u = \{x_{k+1},..,x_N\}$ - unlabeled objects, $I_u = \{k+1,...,N\}$ - indices of the objects

For simplicity assume that he classes do not intersect, run L clustering by one algorithm μ with random parameters $\Omega_1,...,\Omega_L$.

Let's introduce the following notations for $i, j \in I_u$:

$$h_l(x_i, x_j) = \begin{cases} 1, algorithm\ \mu in\ variant\ l\ combined\ pair\ of\ x_i, \quad x_j\ into\ one\ cluster \\ 0, \quad otherwise \end{cases}$$

and $L_1(i,j) = \sum_{l=1}^{L} h_l(x_i, x_j),\quad L_0(i,j) = L - L_1(i,j)$, which represent the number of variants of clustering, in which algorithms has voted to combine x_i, x_j, or against, respectively .

Let $Y(x)$ - hidden true labels of the classes of unlabeled objects $x \in X_u$. Let's introduce a quantity for $i, j \in I_u$:

$$z(x_i, x_j) = \begin{cases} 1, if\ Y(x_i) = Y(x_j) \\ 0, if\ Y(x_i) \neq Y(x_j). \end{cases}$$

In the nearest neighbor methods (NN) for all $i \in I_u$ we assign label y_i a values y', where y' - class of object, $x' = \underset{x_j \in X_c}{arg\ max}\ H(x_i, x_j)$

The following theorem holds

Theorem 1. Let $\forall l \in \{1,...,L\}$, $P[h_l(x_i,x_j)=1 \mid z(x_i,x_j)=1] > \frac{1}{2}$, 1and $L_0(i,j)=const \quad \forall i,j \in I_u$. Then in the algorithm CANN for object $x_i \in X_u$ the probability of incorrect classification $P_{er}(x_i) = P[Y(x_i) \neq y'] \to 0$ for $L \to \infty$.

The theorem shows that the probability of a classification error by the CANN algorithm tends to zero under the assumptions that the classes of objects do not intersect and that the algorithms of cluster analysis that are used correctly classify pairs of objects to one or different clusters with probability more $1/2$, that is, they do not act at random.

7. EXPERIMENTAL ANALYSIS

A typical RGB image contains three channels: the intensity values for each of the three colors. In some cases, this is not enough to get complete information about the characteristics of the object being shot. To obtain data on the properties of objects that are indistinguishable by the human eye, hyper spectral images are used.

For an experimental analysis of the developed algorithm, we used a picture of Pavia University scene with size of 610 x 340 pixels, which contains 103 spectral channels. The spatial resolution of the image is 1.3 m. Figure 5a shows the RGB composite images (channels 40, 50 and 70), and in Figure 5b) the standard image partition into thematic classes is given.

а) *б*)

Рис. 4. Гиперспектральное изображение Pavia University scene (RGB композит) (а) и размеченные данные (б).

Figure 3: Hyper spectral image of Pavia University scene a) RGB composite images b) Marked data

Note that the image has unmapped pixels that are not assigned to any of the nine classes. These pixels were excluded from consideration in the analysis.

In an experimental analysis of the algorithm, 1% of the pixels selected at random for each class made up the labeled sample; the remaining ones were included in the unlabeled set.

To study the effect of noise on the quality of the algorithm, randomly selected r% of the spectral brightness values of the pixels in different channels were subjected to a distorting effect: the corresponding value was replaced by a random variable from interval

To study the effect of noise on the quality of the algorithm, randomly selected r % of the spectral brightness values of the pixels in different channels were subjected to a distorting effect: the corresponding value x was replaced by a random variable from interval $[x(1-p),\ x(1+p)]$, where r, p - initial parameters. The noisy data table containing the spectral brightness values of the pixels across all channels was fed to the input of the CASVM algorithm, and the K-means algorithm was chosen as the basic algorithm for constructing the cluster ensemble. Different variants of partitioning were obtained by varying the number of clusters in the interval $[30, 30+L]$, where L was equal to 120.

In addition, for the construction of each variant of the solution, channels were randomly chosen, the number of which was set to 2. To speed up the operation of the K-means algorithm and to obtain more diverse groupings, the number of iterations was limited to 1.

Since the proposed algorithm implements the idea of distance metric learning, it would be natural to compare it with a similar algorithm (SVM method), which uses the standard Euclidean metric, under similar conditions (the algorithm parameters recommended by default in Matlab environment).

Table 1 shows the accuracy values of the classification of the unlabeled pixels of the Pavia University scene for some values of the noise parameters. The running time of the algorithm was about 2 minutes on a dual-core Intel Core i5 processor with a clock speed of 2.8 GHz and 4 GB of RAM. As it is shown in the table, CASVM algorithm has better noise resistance than SVM algorithm.

Table 1: Accuracy of CASVM and SVM under various noise values

Noise parameters r, p	0%, 0	10%, 0.1	20%, 0.2	30%, 0.3
CASVM	0.82	0.80	0.78	0.77
SVM	0.83	0.75	0.66	0.64

8. CONCLUSION

The paper considers one of the variants of the problem of pattern recognition - the task of semi-supervised learning. The algorithms CASVM and CANN were developed to solve this problem. They are based on a combination of methods of collective cluster analysis and kernel classification methods.

A probabilistic classification model using a cluster ensemble was proposed. Within the model, the behavior of error probability of the CANN algorithm was analyzed. The assumptions are formulated, in which the error probability tends to zero.

An experimental study of the proposed algorithm on a hyperspectral image was performed. It was shown that the CASVM algorithm is more noise-resistant than the standard method of SVM.

REFERENCES

Aidarkhanov, M.B., Amirgaliev, E.N., La, L.L. (2001). Correctness of algebraic extensions of models of classification algorithms. In *Cybernetics and Systems Analysis*, 37(5), 777-781.

Amirgaliev, E.N., Mukhamedgaliev, A.F. (1985). On optimization model of classification algorithms. In *USSR Computational Mathematics and Mathematical Physics*, 25(6), 95-98.

Berikov, V.B., Lbov, G.S. Current trends in cluster analysis. http://biocomparison.ucoz.ru/_ld/0/49_berikov_lbov.pdf.

Berikov, V.B. (2013). A team of algorithms with weights in a cluster analysis of heterogeneous data. *Bulletin of Tomsk State University*, 2 (23), 22-31.

Berikov, V.B. (2014) Weighted ensemble of algorithms for complex data clustering.*Pattern Recognition Letters*, 38, 99-106.

Berikov, V., Pestunov, I. (2017). Ensemble clustering based on weighted co-association matrices: Error bound and convergence properties. *Pattern Recognition*, 63, 427-436.

Domeniconi, C., Al-Razgan, M. (2009). Weighted cluster ensembles: Methods and analysis. *ACM Transactions on Knowledge Discovery from Data*, 2(4):17.

Joydeep, G., Ayan, A. (2011). Cluster ensembles. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1(4), 305–315.

B22
Traffic & Transportation

XIII Balkan Conference on Operational Research
BALCOR 2018

AIRLINE PROFITABILITY CYCLES: AN UNDAMPED SYSTEM MODEL APPROACH

NEMANJA MIJOVIĆ[1], MILICA KALIĆ[1], JOVANA KULJANIN[1], MANUEL RENOLD[2]

[1] University of Belgrade, Faculty of Transport and Traffic Engineering, nemanja.mijovic91@gmail.com, m.kalic@sf.bg.ac.rs, j.kuljanin@sf.bg.ac.rs
[2] Zurich University of Applied Sciences, reno@zhaw.ch

Abstract: *The airline industry exhibits highly cyclical behavior in terms of profitability. The fluctuations of profits are mainly market driven by various numbers of endogenous and exogenous factors that tend to exert the strong competitive pressure on airlines. The airline is keen to know the period of anticipated downturns in order to tailor the appropriate strategy. The paper employs the undamped system to model the world airline profit cycles for the period between 1980 and 2016. In addition to the fundamental cycle period T obtained by the model, the paper also proposes the modification of general model to include the paradigm change occurred in profitability cycles as the effect of 2008 world financial crisis.*

Keywords*: Profitability Cycles, Undamped system, Discrete Fourier transform, Exogenous shock.*

1. INTRODUCTION

The Airline Deregulation Act stipulated in United States in 1978 is deemed as an outstanding event that dramatically changed the outlook of airline industry in the years to come. The airline industry that had previously been fully controlled by the state operating in regulated market evolved into market-based sector driven by the principle of liberalized economy. In other words, the balance between supply and demand becomes an essential force that influences the profitability of airline industry. In such new circumstances, the airline industry had boosted through rapid growth in capacity as well as in traffic. For example, as stated in Jiang and Hansman (2006) the annual growth of domestic scheduled traffic of the U.S. airlines between 1978 and 2002 averaged 11.7 million Revenue Passengers Miles (RPMs) per year, more than doubled the average growth between 1954 and 1978 that was 5.8 million RPMs per year, and the operating capacity of the industry grew on average 4% per year between 1980 and 2000. In line with these institutional changes, between 1978 and 1985 the U.S. airline industry has experienced the fundamental reorganization of network structure when “trunkline” carriers endeavoured to configure their point-to-point system into hub-and-spoke system (Viscusi et al., 1998). About ten years later after Deregulation in U.S., the wave of profound changes have spread in European Union (EU) as one of the largest market in the world. Although the liberalization of aviation market in EU have similar effects as those in U.S, their scope and size were not so radical since most of the EU carriers had previously adopted HS structure in order to efficiently manage their intercontinental flights.

Despite such a tremendous growth in traffic generating the substantial growth in operating revenue, in certain periods the airline industry has seriously suffered from substantial financial losses on global level. According to Franke (2007), the airlines tend to make up for all losses of past downturns in good time and lose more money than they have ever gained in each subsequent crisis. As previously discussed in relevant literature, the airlines’ profitability cycle length mainly stems from endogenous causes such as inability to adjust their capacity to market conditions, slump in demand, inadequate pricing strategy and excessive costs mainly driven by increasing price of jet oil and skilled labour. According to Jiang and Hansman (2006), theses causes directly affect the fundamental cycle periods of profitability, while exogenous factors such deregulation and liberalization as well as terrorist attack of September 11 2001 could be responsible for the oscillation amplitude.

Although previous studies showed that the cycle length is endogenously driven, it is evident that effect of the post 2008 global recession caused the paradigm change in profitability cycles. As stated in Franke and John (2011), the crisis came more quickly and struck more deeply than most experts had anticipated. The paper applies the methodology based on signal detection approach to identify the fundamental cycle periods in the period between 1980 and 2016. The results reveal that the cycle has been unexpectedly distorted by exogenous shock of severe economic crisis that went along with dropping demand and skyrocketing fuel prices. In order to overcome the difficulties in cyclic behaviour of airlines’ profit occurred as a result of 2008 economic crisis, the paper investigates the introduction of opposite signals to efficiently manage the disturbances in the model and increase their replicative efficiency. In other words, applying the opposite signal that corresponds to exogenous shocks will mitigate the differences between actual and estimated values.

The paper is organized as follows: after the brief Introduction, Section 2 provides an overview of the anatomy of crisis that occured in last several decades (i.e. from 1980 until 2016). The methodology part is described in Section 3, followed by the results in Section 4. Finally, the Section 5 concludes the results.

2. THE FACTORS INFLUENCING AIRLINES' PROFITABILITY

The airline industry has seen positive trends in both global consolidated traffic (Revenue Passenger Kilometers - RPK) and capacity (Available Seat Kilometres-ASK) since deregulation in U.S.. In the first two decades after this event, the annual growth rate of traffic was approximately 4.6% (1980 to 1999), while this value has increased to 4.8%, mainly as a consequence of liberalized European market (2000-2016). The regulatory changes in Europe have become an indispensable condition for the introduction of low-cost carriers that gradually increased their market share in total movements in Europe from less than 1% in 1995 to almost 20% in 2008 (Cento, 2009). It is worth mentioning that the "low-cost" business model had already been well established in U.S. with Southwest being a pioneering carrier that dramatically changed the "rules of the game" and became a serious threat to full-service carriers. As it can be seen from Figure 1, the airline industry had negative growth (in terms of both passengers and capacity) three times in last three decades: in 1991 linked to the First Gulf War, 2001/2003 associated with 9/11 terrorist attack and 2003 Iraq war together with SARS epidemic, and 2008 as a consequences of global financial crisis.

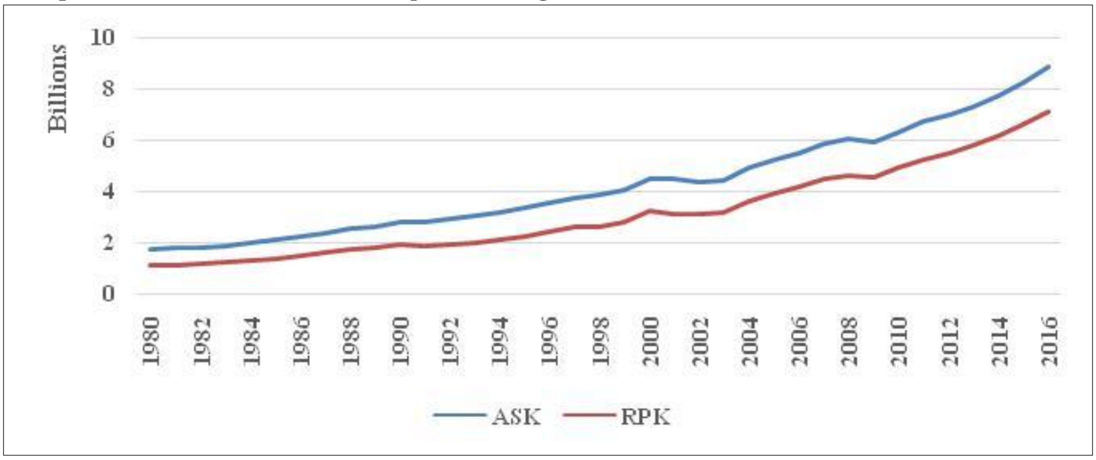


Figure 1: Annual traffic and capacity of scheduled service of world airlines

However, the airlines' financial balance seems to be a significantly worse compared to traffic records in the observed period. As seen from Fig. 2, total operating revenue and costs have grown in line with traffic, whereas total profits have oscillated around zero with amplitude that tends to increase over time.

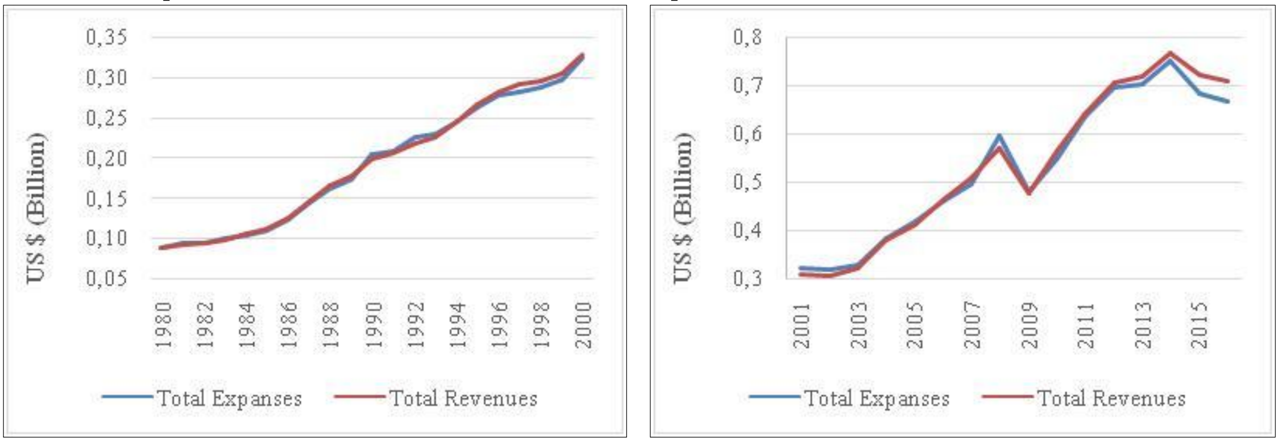


Figure 2: Annual Operating Revenue and Costs (in 2010 US $) for global airline industry (1980-2016)

The airline industry tends to exert the cyclical behaviour as many other industries such as paper industry, real estate markets, shipbuilding industry etc. The periods of high profits alternate the periods of financial losses, the process that occurs in seemingly periodically order. To understand the cyclic behaviour of airlines' profits, one has to take a look at a broader context in which the airline industry persists. In other words, it is of vital importance to explore underlining factors that affect the success of airline in both short-term and long-term periods. Although the various level and factors are obviously interlinked in a complex and multi-faced ways (Franke and John, 2011), some of the factors seem to be crucial to airline's cost structure and cyclical behaviour of airline profits. They will be roughly divided into two sets. The first set contains those factors that can be fully or partially controlled by the airlines (endogenous), whereas the second set consists of those called "disruptive events" or "external shocks" (exogenous).

2.1. Endogenous factors

1. Demand

The air travel demand remains the focus in airline strategy since it is the main factor that accelerates the development of the airline industry, but also the entire economy. According to the definition, the demand for air travel in certain markets accounts all potential passengers that are willing to travel upon defined fares. Historically, two types of demand growth can be distinguished: underlying growth that occurs naturally over time and is driven by factors external to the industry (such as Gross Domestic Product - GDP, Foreign Direct Investments-FDI), and induced growth which occurs as a response to actions that airline have taken over time (such as creating brand image, product quality, customer service etc.). Bearing in mind that air travel demand is characterised by high fluctuations, consumer heterogeneity and uncertainty about journey (Cento, 2009), the airline management have to carefully consider passengers' requirements.

2. Yield Management

It is very well known that airline product is non-storable perishable goods and the remaining empty seats cannot be sold once the flight departs. Thus, yield management stands out in the focus of airline strategies as an efficient tool to minimize the costs by offering adequate service to different customer segments by properly designed tariffs. Despite the adopted airline business model, both full-service and low-cost carriers struggle to increase the seat load factor on their flights, as this constitutes the basic mechanism of generating the revenue.

3. Capacity and Fleet Management

Capacity management is perceived as one of the most important leverage in cycle management (Liehr et al., 2001). The overcapacity can lead to underutilization of a fleet that further implies lower seat load factors and increase in total airline costs. On the other hand, the shortage in capacities disables the airline to meet current demand while missing the opportunity to gain profit. Acquiring new types of aircraft through either purchasing or leasing can have large impact on airline costs since new generation of aircraft are more fuel efficient and generally provide more comfort to passengers. During the process of fleet modernization, the airline has to take into consideration the lag between aircraft orders and deliveries that often accounts between 18 and 24 months. Thus, the careful and proper planning on the number and types of airplane in the fleet become a key precondition for airline success.

4. Costs

In order to sustain profitability on the market, the airlines endeavour to keep their costs as low as possible. This challenging task is often very difficult to manage since some of the costs (such as fuel – Figure 3) are exogenously driven and airlines have very little influencing power.

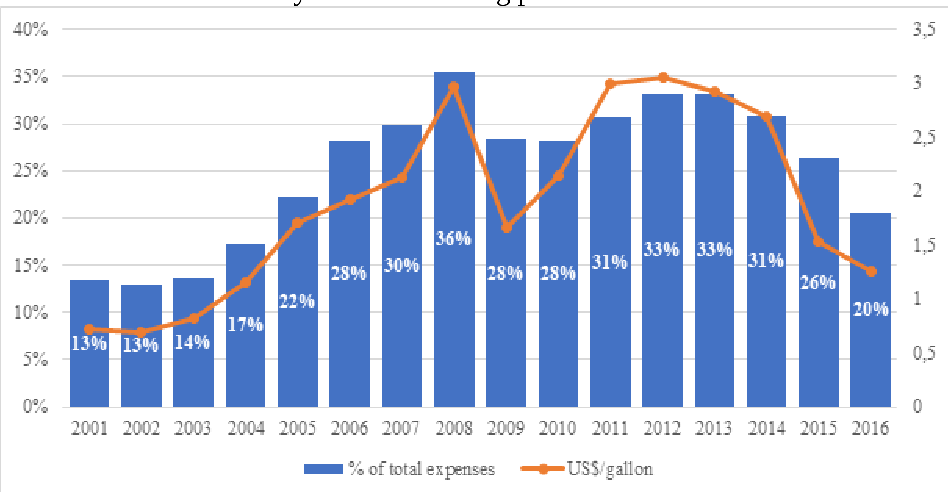


Figure 3: Annual Fuel Costs and its share in airline total expenses

As mentioned, the airline industry is highly susceptible to jet fuel price fluctuation as its cost encompasses substantial portion of airline's total operating costs.

For example, the fuel costs constitute 36% of the total operating cost in 2008 overlapping with severe economic crisis that put a burden on the airline industry. Despite its unpredictable behaviour, many airlines nowadays have adopted so called "fuel hedging" programme to mitigate their exposure to future fuel prices that may be higher than current.

Additionally, the labour seems to be the key driving force for any company, but their costs tend to vary over the years. According to IATA (2010) the share of labour costs in total operating costs for major airlines in the world has fallen from 28.3% in 2001to 20.1% in 2008. In the periods of downturns, the airlines used a variety of mechanism to reduce the labour costs such as shorter hours of working, reducing staff through divestment or outsourcing, voluntary redundancy or simply cutting staff.

2.2. Exogenous factors

In addition to these endogenous factors that can be handled to a certain extent, the unexpected events (i.e. economic recessions, wars, epidemics etc.) remain out of airline control and can severely jeopardize the airlines profits.

1. Economic growth

It is very well known that the airline industry is highly linked to economic environment and very sensitive to its fluctuations. Personal income, the rate of unemployment and foreign direct investments are some among many economic factors that may have impact on air travel demand. The economic growth (expressed through GDP growth) has traditionally been one of the major catalysts to historic growth rates in aviation. The demand for air travel tends to grow at a multiple of GDP growth (as observed in Fig. 4), although this elasticity could vary across different economy and different level of market maturity. For example, the airline industry has experienced the biggest boom in its history between mid-1990s to the beginning of new millennium, the period that coincides with worldwide increase in GDP combined with a greater demand for travel as a result of globalization. This certainly boosts the airline's profits across different continents that reassert the importance of economic activity and its benefits on airline industry growth.

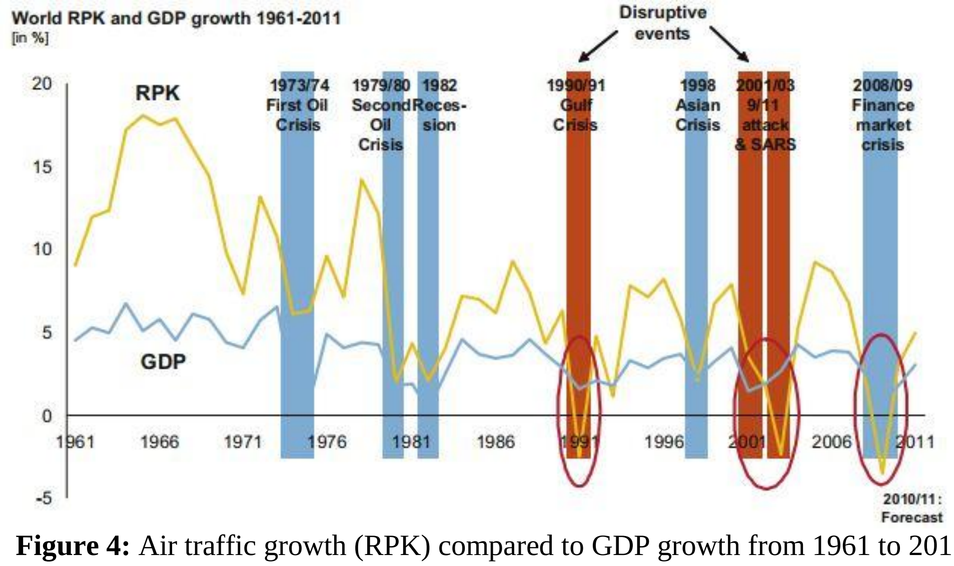


Figure 4: Air traffic growth (RPK) compared to GDP growth from 1961 to 2011

2. Disruptive events

As depicted in Fig. 1, the airline industry has been hit by several crises in last four decades, the events that have had tremendous impact on airline profitability. The crisis occurred in 1973, 1979 and 1981 are mainly driven by the unrest in the Middle East and subsequent oil crises which had dramatically raised the prices of crude oil.

The onset of new millennium has started with the worst airline crisis ever seen. The crisis initially started as demand slowdown followed by the cost impact of overcapacity from the supply side (Cento, 2009). The bursting of the "dot.com" bubble has additionally exacerbated the initial drop of global air travel demand, since a vast number of Internet-based companies reported huge losses while others went bankrupt. This had particularly strong impact on the lower demand in premium class products as well as competitive pricing pressure among airlines in the market.

However, the terrorist attack on the Twin Towers in New York and on the Pentagon in Washington D.C. on September 11 2001 was one the most shocking events that immediately devastated the air traffic demand globally. As an exogenous demand shock, this event generated fear of air travel and caused the diminishing of RPK between Europe and North America by 33% (November, 2001) compared to the period before the attack, when this market had zero annual growth. Shortly after, the world has seen the serious threat from SARS epidemic in East Asia which began in February 2003. This double dip scenario has slowed down the recovery of airlines' profitability which reached the former level of 2004.

In the year following 9/11, globally airline profitability fell by \$22.5 billion, with operating margins down by just over 7% points (Morrell, 2011). As it can be seen from Fig. 5, 2006 was the first year in profitability cycles when profitability went upwards.

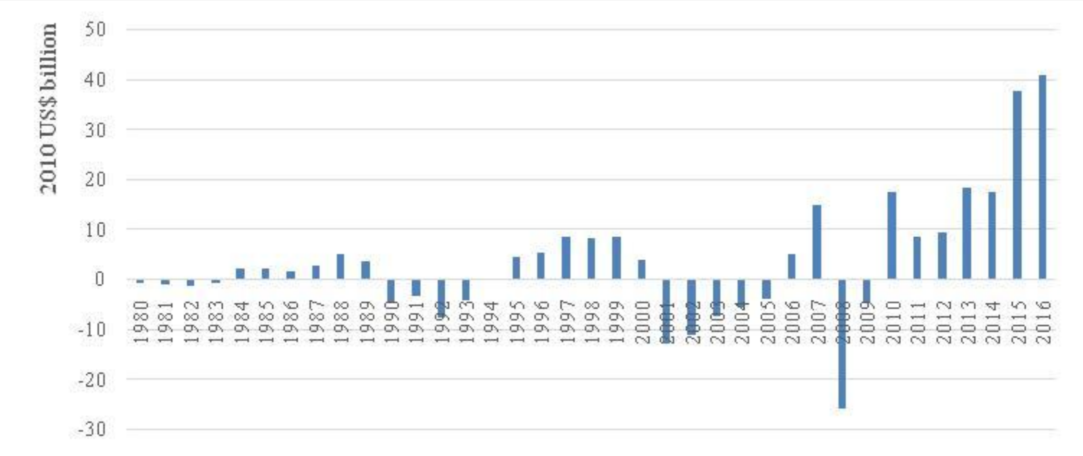


Figure 5: Annual Net Profit of the World Airlines

After two years of positive records, the rising profitability was severely hit by global economic crisis occurred in 2008. The crisis is perceived as the worst financial crisis since the Great Depression of the 1930s. The post 2008 downturn had the most extreme effect on airlines' profitability compared to effects of any previous crisis. In addition to decreasing demand that was part of a cyclical downturn, the airline industry had faced skyrocketing fuel prices in the same year that significantly reduce the profits as a direct outcome of the cost increase. Although, fuel prices has dropped by 18% in 2009 (Morrell, 2011) compared to 2008, the airline industry could not get back to the profitable track until 2010 mainly due to the impact of the financial sector that adversely affected the air travel demand.

In order to understand the cyclical behaviour of airlines' profitability, it is crucial to investigate the profitability cycles. The next section employs the signal detections methods to identify the fundamental cycle period T.

3. METHODOLOGY

3.1. Mathematical formulation

As previously mentioned, the paper uses the methodology proposed by Jiang and Hansman (2004) who modelled the profit cycles as a second-order system. The system equation is mathematically formulated as:

$$\ddot{x} + 2\xi\omega_n\dot{x} + \omega_n^2 x = 0 \quad (1)$$

where:

$x(t)$ - industry profit/loss,

ω_n - natural frequency of the system and

ξ - damping ratio

The parameter ξ can take different values that can cause the various behaviour of the system. The aviation market after the Deregulation can be described as an undamped system with damping ration between -1 and 0. System with these characteristics oscillates at the damped frequency $\omega_d = \omega_n\sqrt{1-\xi^2}$.

The analytical solution of Eq. 1 for $\xi \in (-1,1)$ has the general form:

$$x(t) = A_0 e^{-\xi\omega_n t}\sin(\omega_d\, t + \phi) \quad (2)$$

A_0, φ are determined by initial conditions. If we define $\omega_d = \frac{2\pi}{T}$ and $\tau = -\frac{1}{\xi\omega_n}$ the Eq. 2 can be transformed into:

$$x(t) = Ae^{\frac{(t-t_0)}{\tau}} \sin(\frac{2\pi(t-t_0)}{T}) \tag{3}$$

where:

τ - e-folding time indicating how fast the amplitude grows,

T - the fundamental cycle period of the system,

t - the chronicle year,

t_0 - the time instant the system crosses zero

The general form of airline profitability cycle model is specified by Eq. 3 and as such will be used in subsequent analysis.

3.2. Identifying fundamental cycle period

There are four parameters A_0, T, τ, t_0 that have to be estimated in Eq. 3. Among them four, the period T is of upmost importance since it characterizes the system behaviour. In order to determine the fundamental frequency of the system, a Discrete Fourier transform (DFT) was applied on the data sample of annual profits for the period between 1980 and 2016.

The N-point DFT was specified as follows (Eq. 4):

$$X[k] = \sum_{n=0}^{N-1} x(n) e^{-\frac{2\pi jkn}{N}}, \qquad 0 \leq k \leq N-1 \tag{4}$$

where:

$x(n)$ - the data sample of annual profits,

N - total number of observations (years)

The relative magnitude of $X[k]$ expressed in decibel is specified as follows (Eq. 5):

$$|X[k]|_{dB} = 20 \log_{10} \frac{|X[k]|}{\max(|X[k]|)} \tag{5}$$

The frequency bin with respect to $X[k]$ is formulated as follows:

$$f[k] = \frac{k}{N} F_s, \qquad 0 \leq k \leq N-1 \tag{6}$$

where:

F_s - the sampling frequency of input data, defined as $F_s = 1/year$, because the data sample is given on annual basis.

According to DFT theory, the largest magnitude corresponds to the most significant fundamental frequency in the system. In order to improve the resolution of DFT, the frequency bin was narrowed through the procedure *"zero padding"*, the concept that refers to adding zeros to end of data series to increase its length.

3.3. Model estimation

The nonlinear least square regression is applied on the model specified in Eq. 3 to estimate the parameters. The objective function is given:

$$Minimize \sum_{i=1}^{N} (x_i - \hat{x}_i)^2 \tag{7}$$

where:

x_i - the actual value of $x(t)$, i.e. airline profit in year t

N - total number of observations (years),

$\hat{x}_i$ - the fitted value of $x(t)$, i.e. airline profit in year t

The annual profit data for world airline industry are given in constant 2000 U.S. dollars and were derived from the Internet site (http://airlines.org/). The fundamental cycle periods obtained as described in the previous subsection was used as an initial value of T to initiate the iteration and assure the convergence of the solution. The statistical software *R* was used to estimate the parameters of regression model.

4. RESULTS

This Section provides the results obtained by the methodology described above. The world airline profit is analysed under three scenarios: from 1980 to 2007 and from 1980 to 2016 with the model specified in Eq. 3 and from 1980 to 2016 with opposite signal added to Eq. 3 to capture the effect of world financial crisis since that advent permanently disturbed the cyclical trend observed so far.

4.1. Model estimation for the period from 1980 and 2007

The results of the model specified in Eq. 3 are provided in Table 1. The correlation coefficient is 0.858. The estimates of all parameters from Eq. 3 are significant at 5% of significance.

Table 1: Regression results of world airline net profit between 1980 and 2007 (parameters from Eq. 3)

Variable	Estimate	Standard Error	*t* statistics
A	1.536	0.630	2.439
T	10.727	0.266	40.405
t_0	1973.063	0.703	2804.949
τ	14.481	3.020	4.796

As it can be seen from Fig. 6, the model fits the actual values of world airline profit in accurate manner, while deviations from the actual values are evident only in the case of 1999 and 2001. As it was already discussed above, this period coincides with the onset of new millennium which intensified the globalization through Internet business expansion and subsequently the shock occurred as a result of 2001 terrorist attack.

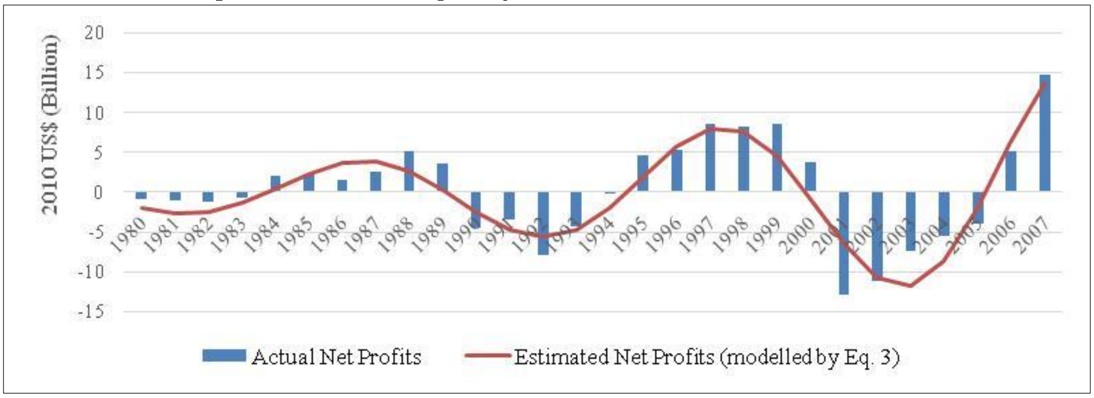


Figure 6: Net Profit Analysis Results of World Airline Industry in the period from 1980 to 2007

4.2. Model estimation for the period from 1980 and 2016

This subsection investigates the introduction of opposite signal in Eq. 3 to include the impact of 2008 world financial crisis. As seen from Fig. 7, the model initially specified in Eq. 3 leads to very poor results if the data set is extended to 2016. Namely, the recovery of 2008 post economic crisis effects came more quickly than anyone could expect representing the paradigm change in airline industry.

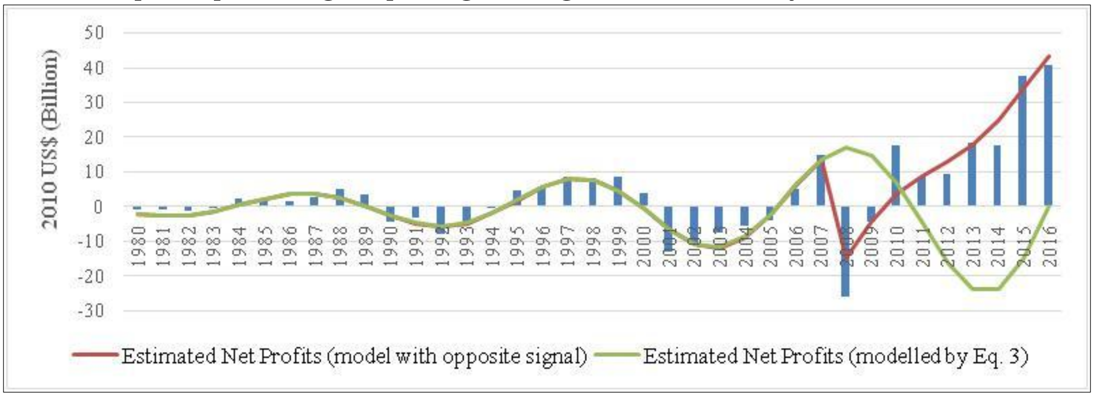


Figure 7: Net Profit Analysis Results of World Airline Industry in the period from 1980 to 2016

Such rapid recovery stems from the knowledge and experience that airlines had already gained through the 2001/2003 "double dip" economic crises. As the post-9/11 crisis gave an impetus to the reinforcing the low-cost model to the fore on intercontinental routes in Europe and U.S. (Franke and John, 2011), the 2008 crises brought the number of innovations in airlines' strategy which allows them to efficiently cope the effects of crisis. The cost-cutting strategy and adjustment of their capacity was seen as the most common strategy that many airlines adopted. Finally, all these circumstances accelerate the emergence of new business models such as the hybrid business model encompassing the characteristics of both full-service and low-cost business model.

In order to overcome the poor replicative characteristics of the model previously defined by Eq. 3, the model is extended by cosines signal that describes the 2008 shock and specified as follows (Eq. 8):

$$x(t) = \sum_{1980}^{n} A_1 e^{\frac{(t-t_{01})}{\tau_1}} \sin(\frac{2\pi(t-t_{01})}{T_1}) + \sum_{2008}^{n} A_2 e^{\frac{(t-t_{02})}{\tau_2}} \cos(\frac{2\pi(t-t_{02})}{T_2}) \quad (8)$$

Table 2: Regression results of world airline net profit between 1980 and 2016 (model with opposite signal)

Variable	Estimate	Standard Error	*t* statistics
A	-42.128	32.857	-1.282
T	16.942	7.721	2.194
t_0	2005.955	2.183	918.989
τ	50	249.596	0.200

Table 2 gives the summarized regression results of net profit modelled by Eq. 8. The ultimate equation presents the collusion of two opposite signals, the procedure that can facilitate the researcher to model the unexpected external shocks that influence the airline profitability. The correlation coefficient is 0.896.

5. CONCLUSION

The paper extends the work of Jiang and Hansman (2004) who efficiently applied undamped system to model the profitability cycles. This basic model has been extended in this paper by including the time period from 2005 to 2014. The model estimated on data between 1980 and 2007 revealed that fundamental cycle period for world airline industry was approximately 11 years (i.e. 10.727), the results identically obtained in Jiang and Hansman (2004) for U.S. airlines. As significant changes have occurred in airline industry in the post 2008 global financial crisis, the undamped model successfully applied to net profits could no longer achieve the satisfactory results in terms of its replicative feature.

Thus, our model has been enhanced by including opposite signal (modelled by cosine) as a way to describe the exogenous shock of 2008 economic crisis after which the airline industry has rapidly recovered. The results of the modified model are significantly improved compared to the replicative characteristics of the basic model and can help airlines to eventually anticipate the effects of future similar crisis. Still, it certainly cannot explain the driving force which underlies the profitability cycles. Thus, the further researches should pay subtle effort to reveal the factors that influence the behaviour of overall system.

Acknowledgement

This research has been supported by the Ministry of Education, Science and Technological Development, Republic of Serbia, as a part of the project TR36033 (2011–2018).

REFERENCES

Airlines for America (2018). Internet site: http://airlines.org/ (accessed on February, 2018)

Cento, A. (2009). *The Airline Industry: Challenges in the 21st Century*. Heidelberg: Physica-Verlag.

Franke, M., & John, F. (2007). Innovation: The winning formula to regain profitability in aviation?. *Journal of Air Transport Management*, 13, 23-30.

Franke, M. (2011). What comes next after recession? – Airline industry scenarios and potential end games. *Journal of Air Transport Management*, 17, 19-26.

IATA ECONOMIC BRIEFING (2010). Airline Fuel and Labour Cost Share. Internet site: https://www.iata.org/whatwedo/Documents/economics/Airline_Labour_Cost_Share_Feb2010.pdf (accessed on February, 2018).

Liehr, M., Größler, A., Klein, M., & Milling, M.P. (2001). Cycles in the sky: understanding and managing business cycles in the airline market. *System Dynamics Review*, 17, 311-332.

Morrell, P. (2011). Current challenges in a 'Distressed' industry. *Journal of Air Transport Management*, 17, 14-18.

Viscusi, W.K., Vernon, J.M., Harrington, J.E., (1998). *Economics of Regulation and Antithrust. (4th ed.)*. MIT Press, Cambridge.

XIII Balkan Conference on Operational Research
BALCOR 2018

INVESTIGATION OF THE FACTORS AFFECTING THE SEVERITY OF TRAMWAY ACCIDENTS IN ESKISEHIR

KADIR BERKHAN AKALIN[1], HUSEYIN ONUR TEZCAN[2], SAFAK BILGIC[3]

[1] Eskisehir Osmangazi University, Department of Civil Engineering, kbakalin@ogu.edu.tr
[2] Istanbul Technical University, Department of Civil Engineering, tezcanhu@itu.edu.tr
[3] Eskisehir Osmangazi University, Department of Civil Engineering, safakb@ogu.edu.tr

---

Abstract: *Urban rail systems are considered as safe, fast and sustainable systems because traffic and pedestrian intersection areas are less than those of rubber-tired public transportation systems. Although metro and light rail systems with high capacity and completely independent lines in the city are the safest and reliable systems because they do not interact with any external factors, their investment and operation costs are considerably higher than other systems. On the other hand, trams with their adaptability to environment and their ability to turn along small radii curves are seen as very suitable systems for downtowns. Along with all its advantages, the accident risk of trams is higher compared with other rail systems. In this study, 264 tramway accidents in Eskişehir between 2005 and 2013 were examined and the factors affecting tram accidents were investigated. According to the obtained data; weather conditions, daytime conditions, number of vehicles and persons involved in the accidents, lighting conditions in the night, accident type (hitting vehicle or pedestrian), road geometry and the condition of the road surface were considered as factors affecting the tram accidents. The severity of the accidents have been examined by separating into three classes: material damage, injury, and fatal, as commonly used in the literature. Each of the factors affecting the severity of the accidents was investigated in terms of the number of accidents, the elements determined to have more effects on the accident were identified, and their solution proposals were given.*

Keywords: *Tramway accidents, Urban rail systems, Severity of accidents, Factors affecting accidents, Traffic accidents.*

1. INTRODUCTION

The use of public transport systems is recommended as a solution to the increasing congestion and the increasing pollution caused by fossil fuels in developed and developing countries. Urban rail systems are constantly increasing their importance in public transport, due to their comfort and high capacity as well as the use of more environmentally friendly energy sources compared to rubber-tired public transport systems. In addition, rail systems are considered as safer systems because traffic and pedestrian intersection areas are less than other public transportation systems. Metro and light rail systems are the safest and most reliable systems due to their fully protected lines, but the initial investment and operating costs are quite high. Therefore, for cities with smaller city centers and less travel demand, the use of trams would be most appropriate when the compatibility with the environment and the ability to turn along small radii curves are considered.

Trams were first used in 1776 for coal transport in the UK, but for the first time in 1832, they began to be used in New York City for passenger transport. The first urban rail system in Turkey, Istanbul Dersaadet Tramway Company and Tunnel Facility was founded in 1869. Until 1960, the construction and use of new railway system lines continued to increase. However, at the beginning of the 1960s, tram services were stopped and even the rail lines were dismantled. After this period, importance has been given to the use of highways and rubber-tired vehicles, but in the 1980s, construction and use of rail systems due to increased congestion, pollution and travel demand started again.

Today, it is seen that, in response to the constantly increasing travel demand, public transportation gain more attention. From the point of view of capacity, it is possible to carry 1200 passengers by heritage trams, 3000 passengers by streetcars (trams) and 10000 passengers by fast trams in one hour per one-way. In tram systems, headway can be reduced to 3 minutes. However, it is usually operated with a longer than 5 minutes headway, because traffic can be adversely affected by the system with shorter headway. Comparisons of the public transportation systems concerning capacity and headway are shown in Table 1.

Table 1: Comparisons of the public transportation systems concerning capacity and headway

Category	System Type	Passenger Capacity of Vehicle	Headway (minute)	Passenger Capacity per Hour per Oneway
Rubber-tired Vehicles	Dolmush	4 – 10	≥10	24 – 60
	Minibus	10 – 18	≥10	60 – 108
	Bus	50 – 200	≥10	300 – 1 200
	Bus Line	50 – 200	≥3	2 000 – 4 000
	Bus Rapid Transit	150 – 400	≥1	9 000 – 24 000
Rail Systems	Heritage Tram	50 – 100	≥5	600 – 1 200
	Streetcar (Tram)	150 – 250	≥5	1 800 – 3 000
	Fast Tram	300 – 500	≥5	3 600 – 10 000
	Light Rail System	333 – 667	≥2	10 000 – 20 000
	Light Metro	667 – 1 000	≥2	20 000 – 30 000
	Metro	≥1 000	≥2	≥30 000

In the last 80 years, the question of how and why traffic accidents have taken place has been the subject of numerous investigations. As a result, there are a large number of types of accident models and theories in the literature (Katsakiori et al. 2009). Accident cause-and-effect models are being developed to investigate and prevent accidents. In the literature, accident causality surveys and analyzes, which have been performed numerous times for road vehicles, are seldom found for urban rail systems.

When the literature is reviewed, it is seen that the factors causing the accidents are examined at different viewpoints in different countries. Kim and Yoon (2013) who investigated the 73 rail systems' incident reports in the UK, found that human errors, technical failures, and external influences resulted 56 accidents while 17 near miss events played nearly equal roles.

When the results of accident causality models are examined, almost none of the accidents originate from a single cause (Sklet 2004). Accidents are caused by multiple factors that are related or unrelated to each other. For this reason, this complex situation must be reflected in the investigation of the causes of accidents and supported by analytical techniques to support researchers.

Factors affecting the severity of accidents were investigated in an accident analysis conducted by Shankar (1996) in a rural area of the United States of America in Washington State. Results of the analysis carried out are as follows: The absence of night lighting increases the likelihood of accident resulting in material damage. The presence of wet and icy road surfaces increases the likelihood of accident resulting in material damage and accident involving personal injury. Driving at a speed faster than the speed limit increases the likelihood of accident involving death and personal injury at a very high rate.

As a result of investigations made for the traffic accident literature and various cities of the world, it has been seen that the severity of accidents is generally evaluated in three classes. According to the severity, this classification which is made as fatal, injury and material damage has been adopted within the scope of this study.

In the scope of the study, the data about the traffic accidents which the streetcar system in Eskişehir were involved examined and the factors affecting the tram accidents were investigated by using the data. Weather conditions, daytime lighting conditions, number of vehicles and persons involved in the accident, lighting conditions at night, accident type (hitting car or pedestrian), road geometry, and road surface condition are considered as factors affecting the severity of the accidents according to the information obtained from the data. Each of these factors, which are thought to have an effect on the severity of the accident, were examined in terms of the number of accidents and the ones that were found to have more effects were identified and solution proposals are presented in the results section.

2. INTRODUCTION OF RESEARCH FIELD

In this study where the factors affecting the severity of the tram accident were investigated, the data belonging to the accidents that the ESTRAM streetcar operated in Eskişehir was considered. In Turkey, streetcar systems are also operated in Antalya, Bursa, Gaziantep, Istanbul, Kayseri, Konya and Samsun provinces, but accident reports for these systems were not shared to analyze.

As a result of the decisions taken in 1999, Eskisehir Metropolitan Municipality initiated work on the pedestrianization of the district at the shopping and entertainment center of the city and the only way to arrive at this area would be by tram. As a result of these studies, the streetcar system called ESTRAM, whose route includes the main campuses of the two universities, hospitals, some schools, public institutions, intercity bus terminal and major shopping centers, was opened on December 24, 2004. Until August 2014, this tram system, which has a total length of 16 km, has started to provide services to seven lines with additional lengths of 37 km.

The Bombardier Flexity Outlook model tramcar, which has a maximum speed of 70 km/h is operated at an average speed of 18 km/h in the ESTRAM lines. ESTRAM tramcar fleet has 33 units of these vehicles with lengths of 29.50 m and widths of 2.30 m. The capacity determined by the manufacturer is 101 passengers standing (4 passenger/m^2) and 58 passengers seated, but capacity by ESTRAM was arranged as 151 passengers standing (6 passenger/m^2) and 58 passengers seated. According to the data from 2014, on average 824 journeys were made and 37 294 241 passengers were carried daily. By the year 2015, on average 1100 journeys were made and 41 642 437 passengers were carried by the ESTRAM system daily.

3. INTERPRETING DATA

The density map of all tram accidents between 2005 and 2013 of Eskişehir province is shown in Figure 1 (Akalın, 2016).

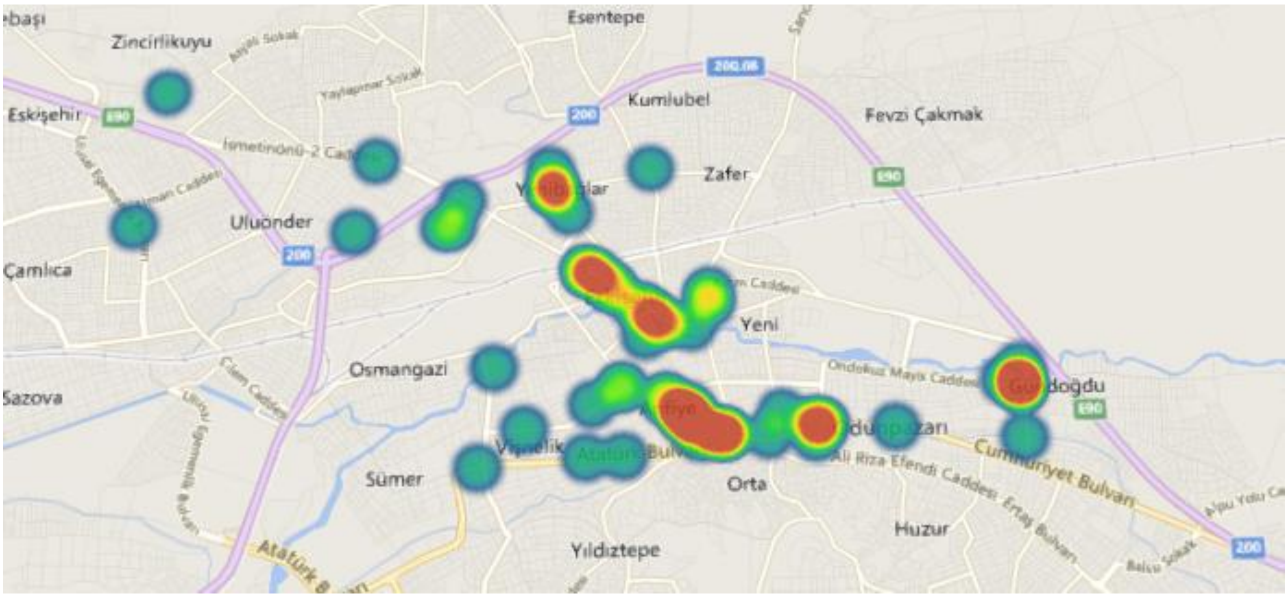


Figure 1: Density map of tram accidents in Eskişehir between 2005 and 2013

When the tram accident density map is examined, it is seen that the accidents spread across different spots along the line. However, it can be said that the number of accidents in the central region where pedestrian and vehicle traffic are intense are considerably high.

Table 2 shows number of accidents by accident cause factors such as the weather conditions, daytime lighting conditions, number of vehicles and persons involved in the accident, lighting conditions at night, accident type (hitting car or pedestrian), road geometry, and road surface condition and accident severity obtained from the reports of 264 tram accidents.

Table 2: Number of accidents by factors causing accidents and accident severity

Factors		Number of Accidents Involving Death	Number of Accidents Involving Injury	Number of Accidents Resulting in Material Damage	Total Number of Accidents
Weather	**Fair**	2	95	145	**242**
	Rainy or Snowy	0	5	17	**22**
Time of day	**Daytime**	1	62	105	**168**
	Night	1	38	57	**96**
Number of persons involved in the accident (Excluding tram driver)	**1**	2	86	156	**244**
	2	0	10	6	**16**
	3+	0	4	0	**4**
number of vehicles involved in the accident (Including tramcar)	**1**	2	47	1	**50**
	2	0	53	156	**209**
	3+	0	0	5	**5**
Lighting conditions at night	**Operating**	0	27	46	**73**
	Not available	1	11	11	**23**
Accident Type	**Hitting pedestrian**	2	82	1	**85**
	Hitting motorized vehicle	0	12	154	**166**
	Hitting non-motorized vehicle	0	6	3	**9**
	Inside tramcar	0	0	4	**4**
Road Geometry	**Divided road**	2	90	32	**124**
	Undivided road	0	8	34	**42**
	Roundabout	0	2	14	**16**
Road Surface Condition	**Dry**	2	88	137	**227**
	Wet	0	12	25	**37**

When the number of accidents was examined according to the time of day, it was observed that 63.6% of the accidents occurred during daytime. The reason for the low number of nightly accidents is because fewer tram services occur at night.

When the number of accidents is examined according to the weather conditions, it is seen that 91.7% of the accidents occur in the fair weather. This is due to the fact that drivers are more careful and there are fewer pedestrians in rainy weather.

According to the literature on accidents, it is seen that the severity of accidents is usually evaluated in three classes. According to the severity, this classification, which is made as fatal, injury and material damage, has been adopted within the scope of this study. However, since only two of the 264 accident data were reported as fatal, the accidents involving death and personal injury were collected and interpreted in the same group.

The number of accidents according to the number of vehicles including tramcar involved in the accident and the severity of the accidents are given in Table 3. The data are tabulated according to day and night lighting conditions. When the table is examined, there are no accidents involving three or more vehicles including tramcar in the severe (fatal and injury) accidents category. Both two cars and tramcar are involved in only 3.1% of the accidents resulting in material damage. 61.8% of the accidents involving death and personal injury and 64.8% of the accidents resulting in material damage occurred in daytime. In the case when lightings are operating at night, the proportion of the accidents involving death and personal injury is 26.4%, while in the absence of lightings this rate is 11.7%. It is seen that 28.4% of the accidents resulting in material damage occurred in case of lightings are operating, 6.8% of this kind of accidents occurred while in the absence of lightings at night.

Table 3: The number of accidents, including the number of vehicles including tramcar involved in the accident and the severity of the accident

Lighting condition	**Daytime**			**Night**					
				Lightings are operating			**Lightings are not available**		
Number of vehicles involved in the accident (including tramcar)	**1**	**2**	**3+**	**1**	**2**	**3+**	**1**	**2**	**3+**
The accidents involving death and personal injury	33 (%32,4)	30 (%29,4)	0 (%0,0)	13 (%12,7)	14 (%13,7)	0 (%0,0)	3 (%2,9)	9 (%8,8)	0 (%0,0)
The accidents resulting in material damage	1 (%0,6)	103 (%63,6)	1 (%0,6)	0 (%0,0)	43 (%26,5)	3 (%1,9)	0 (%0,0)	10 (%6,2)	1 (%0,6)

The number of accidents according to the number of persons except tram driver involved in the accident and the severity of the accidents are given in Table 4. Similar to Table 3, the data are tabulated according to day and night lighting conditions. When the data were examined, it was observed that 53.9% of the accidents involving death and personal injury occurred in the daytime and one person was involved. When the accidents resulting in material damage were examined, it was observed that 96.3% of the accidents occurred with a person except for the tram driver. It can be said that there is usually one person in the car in an accident involving collision with the vehicle.

Table 4: The number of accidents according to the number of persons except tram driver involved in the accident and the severity of the accidents

Lighting condition	**Daytime**			**Night**					
				Lightings are operating			**Lightings are not available**		
Number of persons involved in the accident (excluding tram driver)	**1**	**2**	**3+**	**1**	**2**	**3+**	**1**	**2**	**3+**
The accidents involving death and personal injury	55 (%53,9)	6 (%5,9)	2 (%2,0)	22 (%21,6)	3 (%2,9)	2 (%2,0)	11 (%10,8)	1 (%1,0)	0 (%0,0)
The accidents resulting in material damage	103 (%63,6)	2 (%1,2)	0 (%0,0)	43 (%26,5)	3 (%1,9)	0 (%0,0)	10 (%6,2)	1 (%0,6)	0 (%0,0)

The number of accidents by type of accident and severity are given in Table 5. Since the numbers of collision accidents with non-motorized vehicles are quite small, collisions with motorized and non-motorized vehicles were combined as a collision with the vehicle while the table was being created. When the table was examined, it was observed that 82.3% of the accidents involving death and personal injury occurred in the form of hitting pedestrian, no death or injury occurred inside tramcar accidents but a few material damage records. It has been seen that in an accident that occurred during daytime, there is a pedestrian hit but there is no injury or death reported, but it was reported as material damage. It is believed that this is the result of injury or death, but it is thought that this accident was reported as a near miss but tramcar got damaged because of something that pedestrian was carrying.

Table 5: The number of accidents by type of accident and severity

Lighting condition	**Daytime**			**Night**					
				Lightings are operating			**Lightings are not available**		
Accident Type	**Hitting pedestrian**	**Hitting vehicle**	**Inside tramcar**	**Hitting pedestrian**	**Hitting vehicle**	**Inside tramcar**	**Hitting pedestrian**	**Hitting vehicle**	**Inside tramcar**
The accidents involving death and personal injury	54 (%52,9)	9 (%8,8)	0 (%0,0)	19 (%18,6)	8 (%7,8)	0 (%0,0)	11 (%10,8)	1 (%1,0)	0 (%0,0)
The accidents resulting in material damage	1 (%0,6)	102 (%63,0)	2 (%1,2)	0 (%0,0)	46 (%28,4)	0 (%0,0)	0 (%0,0)	9 (%5,6)	2 (%1,2)

The number of accidents according to road geometry and accident severity are given in Table 6. In 82 of the 264 accident reports, no information was given about road geometry. When the table is examined, it is seen that 90.2% of the accidents involving death and personal injury occur on the divided roads. When the accidents resulting in material damage are examined, it is seen that 40.0% is on divided roads, 42.5% is on the undivided roads, and 17.5% is on the roundabout.

Table 6: The number of accidents according to road geometry and accident severity

Lighting condition	**Daytime**			**Night**					
				Lightings are operating			**Lightings are not available**		
Road Geometry	**Divided**	**Undivided**	**Roundabout**	**Divided**	**Undivided**	**Roundabout**	**Divided**	**Undivided**	**Roundabout**
The accidents involving death and personal injury	58 (%56,9)	4 (%3,9)	1 (%1,0)	24 (%23,5)	2 (%2,0)	1 (%1,0)	10 (%9,8)	2 (%2,0)	0 (%0,0)
The accidents resulting in material damage	17 (%21,3)	22 (%27,5)	13 (%16,3)	11 (%13,8)	9 (%11,3)	1 (%1,3)	4 (%5,0)	3 (%3,8)	0 (%0,0)

The number of accidents by surface condition and severity of accidents are given in Table 7. Here, the road surface condition was examined in two groups, dry and wet. The wet data includes all wet surface conditions such as by rain, snow, ice, mud or any other reason. Looking at the table, 88.3% of the accidents involving death and personal injury and 85.0% of the accidents resulting in material damage occurred on dry road surface. Consequently, this situation indicates that people drive more careful and slowly when they see wetness on the road.

Table 7: The number of accidents by surface condition and severity of accidents

Lighting condition	Daytime		Night			
			Lightings are operating		Lightings are not available	
Surface Condition	Dry	Wet	Dry	Wet	Dry	Wet
The accidents involving death and personal injury	58 (%56,9)	5 (%4,9)	22 (%21,6)	5 (%4,9)	10 (%9,8)	2 (%2,0)
The accidents resulting in material damage	85 (%52,9)	20 (%12,3)	41 (%25,3)	5 (%3,1)	11 (%6,8)	0 (%0,0)

According to the number of accidents by type of accident and severity data, risk matrix in risk assessment of different types of accidents is applied as shown in Table 8.

Table 8: Risk matrix in risk assessment of different types of accidents.

Likelihood / Severity	Minor	Moderate	Severe
Likely		**Hitting vehicle**	
Possible	**Inside tramcar accident**		**Hitting pedestrian**
Unlikely			

4. CONCLUSION

In this study, factors affecting the severity of accidents involving the streetcar (tram) system, which has high compatibility with the environment and the ability to turn along small radii curves are investigated. In the literature, accident causality analyzes which have been performed numerous times for road vehicles, are rarely found for urban rail systems. Within the scope of this study, the relationship between the factors that could cause the accident and the severity of the accident which is classified into three classes are examined by number of accidents.

When the number of accidents are examined according to the road geometry, it is observed that 40.0% of the accidents resulting in material damage occurred on the divided roads, 42.5% of them occurred on the undivided roads, and 17.5% of them occurred on the roundabout. In this case, it is seen that most of the accidents happen on the roads where roadways are not physically separated from tramways.

When the accidents involving death and personal injury were investigated, it was observed that 90.2% of these accidents occurred on the divided road. When these ratios are examined, it is less likely that the accidents resulting in material damage would occur on divided roads, when the roadways are separated from the tramways by plastic delineators or kerbs which can prevent vehicles to enter the tramway. However, it is understood that the accidents involving death and personal injury cannot be reduced by separating the roadway from tramway with the existing methods, because people can easily stride over the borders or pass through plastic delineators. In order to reduce such accidents, it may be more appropriate to separate the roadways from tramways with continuous metal fences instead of kerbs or delineators. Pedestrian violations are frequently encountered in areas where tramways are separated by yellow lines which are often applied in pedestrianized areas and are only visual separators. This situation creates an environment for the accidents involving death and personal injury. In order to prevent this, it is possible to consider the application of metal fences along the tramway line with crossing areas by certain intervals.

REFERENCES

Adams, E. (1976). Accident causation and the management system. *Professional Safety*, 21(10), 26-29.

Akalın, K.B. (2016). Investigation of Factors Affecting Tram Accident Severity with Mutinomial Logit Model. *Master of Science Thesis, Eskişehir Osmangazi University, Graduate School of Natural and Applied Science, Department of Rail Systems,* Eskişehir.

Bird, F. E. (1974). Management Guide to Loss Control, *17 Institute Press*. Atlanta, Georgia.

Estram (2017), http://www2.estram.com.tr/Anasayfa, *last accessed: 13.02.2017.*

Heinrich, H. W. (1931). Industrial Accident Prevention. A Scientific Approach.

Katsakiori, P., Sakellaropoulos, G., & Manatakis, E. (2009). Towards an evaluation of accident investigation methods in terms of their alignment with accident causation models. *Safety Science,* 47(7), 1007-1015.

Kim, D.S., & Yoon, W. C. (2013). An accident causation model for the railway industry: Application of the model to 80 rail accident investigation reports from the UK. *Safety science*, 60, 57-68.

Kim, J. K., Ulfarsson, G. F., Kim, S., & Shankar, V. N. (2013). Driver-injury severity in single-vehicle crashes in California: a mixed logit analysis of heterogeneity due to age and gender. *Accident Analysis & Prevention,* 50, 1073-1081.

Reason, J. (1990). Human error. *Cambridge University Press*, New York.

Shankar, V., Mannering, F., & Barfield, W. (1996). Statistical analysis of accident severity on rural freeways. *Accident Analysis & Prevention*, 28(3), 391-401.

Sklet, S. (2004). Comparison of some selected methods for accident investigation. *Journal of hazardous materials*, 111(1-3), 29-37.

Weaver, D. A. (1971). Symptoms of Operational Error. *Professional safety*, 104(2), 39-42.

BALCOR 2018
XIII Balkan Conference on Operational Research

EVALUATION OF TRAFFIC PERCEPTION WITH ROBUST REGRESSION ANALYSIS

KADIR BERKHAN AKALIN[1], ARZU ALTIN YAVUZ[2], MURAT KARACASU[3], BARIS ERGUL[4]

[1] Eskisehir Osmangazi University, Department of Civil Engineering, kbakalin@ogu.edu.tr
[2] Eskisehir Osmangazi University, Department of Statistics, aaltin@ogu.edu.tr
[3] Eskisehir Osmangazi University, Department of Civil Engineering, muratk@ogu.edu.tr
[4] Eskisehir Osmangazi University, Department of Statistics, bergul@ogu.edu.tr

---

Abstract: *The ownership and use of motor vehicles is becoming increasingly widespread due to the constant increase in the world population. However, these developments bring some problems. Especially in Turkey where serious road transport accident occurs, it is an important issue to conduct researches which are necessary for detection of the problems. Firstly, it is significant to determine the traffic perception of the individual in terms of determining the deficiencies in implementation. This research was planned and carried out to measure the traffic perceptions of the individuals living in Eskisehir. For this purpose, a questionnaire was prepared to determine the demographic and socio-economic characteristics of the research subjects and their attitudes towards traffic perception. The obtained data were analyzed by Robust Regression Analysis techniques. The traffic perceptions of the individuals in Eskişehir province were tried to be modeled by Robust Regression Analysis by creating the traffic perception index with the help of dimensions determined from analysis.*

Keywords: *Robust Regression, factor analysis, traffic perception, traffic awareness.*

1. INTRODUCTION

Traffic safety imply the methods and measures used to prevent users of roads from being killed or injured and loss of property. According to the World Health Organization 2018 fact sheet on road traffic injuries, More than 1.25 million people die each year becasue of road traffic accidents and between 20 and 50 million more people suffer non-fatal injuries, with many incurring a disability (WHO 2018).

There are such studies about traffic perception in the literature: In the study by Dressler et al. (2017), the development of items to evaluate a broad of traffic perception and hazard avoidance skills for use identification of skill level in driver training and testing. Pre-school children traffic perceptions in Turkey, with the help of a questionnaire conducted to 804 children aged between 3 and 6 in the study by Hatipoglu (2011) were tried to be determined.

2. MATERIAL AND METHOD

This study is part of a survey conducted to measure traffic perceptions of individuals living in Eskisehir. A survey form was used to determine the demographic and socio-economic characteristics of the individuals in the Eskişehir Province and their attitudes towards traffic perception. Questionnaires are prepared as simple as possible so that individuals can easily answer. All questions are close-ended multiple-choice. In the questionnaire form, 41 questions have been prepared to determine the attitudes of individuals towards traffic perception. The first 16 of the 41 questions were prepared to investigate the demographic characteristics of the individuals and the remaining 25 items to investigate their attitudes towards traffic perception. Within the scope of the research, the expressions of the individuals were rated according to the 5-point Likert scale type and they were asked to evaluate with one of the following: "I absolutely agree (5) I agree (4) I do not understand (3) I do not agree (2) I absolutely do not agree".

The aim of this study is to reveal the factors that constitute the traffic perception in Eskişehir with the help of dimensions and to model mathematically the relation between the related dimensions and traffic perception index. The data related to 5100 parents of students were reached among the students living and studying at elementary schools in Eskişehir were selected using stratified sampling technique. The attitudes and behaviors of these individuals towards traffic perceptions were analyzed by Factor Analysis and these attitudes and behaviors were analyzed by using Principal Axis Factoring Technique. In the next step, the traffic perception index was established and related attitudes and behaviors were modeled using Robust Regression Analysis.

Factor analysis is defined as the process of obtaining the functional definitions of concepts using the factor loadings of the relevant variables (questionnaire items). Thus, it is possible to reduce many variables to several dimensions. Each of these dimensions is called factor. The purpose of factor analysis is to determine the pattern of variables that are not directly observable, based on directly observed variables. Another aim is to maximize the variance of variables and to derive a set of variables called dimension (Khalaf 2007, Rummel 1970).

There are two important factor analysis techniques, Explanatory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). CFA tries to verify the theoretical model and identifies factors using path analysis. EFA identifies the factors and the structure of unobserved variables based on the data set and tests the expectations (Child 2006).

The main purpose of regression analysis is to determine the functional structure of an observed event. When regression analysis is performed, it is necessary to express the event with a mathematical function using observation values. This model is called regression model (David 1981). The multiple linear regression model for k independent variables is given in Equation 1.

There are some basic assumptions for the regression coefficients to be used for estimation purposes. These assumptions generally relate to error terms. Error terms are independent and identical, mean zero, variance normal distribution.

$$Y = X_i b_i + \varepsilon \quad (1)$$

Y; the dependent variable vector,

X_i; independent variables (or X data matrix),

b_i; regression coefficients,

ε; the random error vector.

b_i regression coefficients are estimated using the Equation 2 with the help of the least squares method (LSM).

$$\hat{b} = (X'X)^{-1}X'Y \quad (2)$$

The Least Squares (LSM) method is the most widely used technique for parameter estimation in regression analysis. The aim in this technique is to minimize the squares of error terms (Equation 3). The $\hat{b}$ estimates that minimize the sum of error squares are the LSM estimators of the b coefficients in the regression model.

$$min \sum_{i=1}^{n} \left(\varepsilon_i^2\right) \quad (3)$$

ε; error terms

There are some basic assumptions for the regression coefficients to be used for estimation purposes. These assumptions are generally related to error terms. Error terms are independent and identical, mean zero, variance normal distribution. The error terms are independent and identically distributed with a mean value of zero and a constant variance. If these assumptions are distorted, inconsistent and biased predictions can be made (Mosteller and Tukey 1977). However, if the distribution of error terms does not have normal distribution, the techniques to be applied are Robust Regression techniques are methods which are not affected by values that are called outliers and which show considerable differences from the rest of the data. Approaches that are not dependent on assumptions, especially insensitive to the assumption of normality, are called robust (Ergul 2006, Ozturk 2003).

In this study, M-Regression technique, which is one of Robust Regression techniques in the literature, will be introduced. It is not correct to apply the LSM and use the obtained model for prediction when the error terms normality assumption for the main group is invalid. M-regression which is based on this assumption was developed as an alternative to LSM. In the M-regression technique, a function of error terms (Huber, Andrews, Tukey) is minimized as shown in Equation 4 (Birkes and Dodge 2011, Ergul 2006). For the parameter estimation, the weighted LSM is used.

$$min \sum_{i=1}^{n} \rho(\varepsilon_i) \quad (4)$$

ρ; objective function

3. RESULTS AND DISCUSSIONS

Frequency tables of the demographic structure of the individuals living in Eskisehir City and participating in the survey are as shown in Table 1-3. 76% of the 5100 individuals who participated in the survey were male and 24% were female. 59% of the respondents are in the age range of 36-50 and 27% are in the age range of 31-35. 39% of the respondents were high school graduates and 37% were college graduates.

Table 1: Frequency distribution of gender

	Frequency	Cumulative
Men	3882 (76.1%)	76.1%
Women	1218 (23.9%)	100.0%
TOTAL	5100 (100.0%)	

Table 2: Frequency distribution of age

	Frequency	Cumulative
18-20	45 (0.9%)	0.9%
21-23	82 (1.6%)	2.5%
24-26	93 (1.8%)	4.3%
27-30	260 (5.1%)	9.4%
31-35	1399 (27.4%)	36.8%
36-50	3019 (59.2%)	96.0%
51-60	172 (3.4%)	99.4%
60+	30 (0.6%)	100.0%
TOTAL	5100 (100.0%)	

Table 3: Frequency distribution of education

	Frequency	Cumulative
Primary	967 (19.0%)	19.0%
High school	2002 (39.3%)	58.2%
University	1892 (37.1%)	95.3%
Post graduate	239 (4.7%)	100.0%
TOTAL	5100 (100.0%)	

As a result of the Reliability Analysis of the questionnaire items consisting of 25 questions and mentioned in Table 4, the Cronbach alpha value was found to be 0.682. Four questions were removed from the questionnaire because they disqualified the reliability of the questionnaire. After then the Reliability Analysis was applied again and the Cronbach alpha value was found to be 0.711. Since the Cronbach Alpha value is in the range of 0.7-1.0, the reliability of the questionnaire is high.

Factor analysis was applied to the questionnaire items by using the Principal Axis technique and the dimensions of the questionnaire items were tried to be determined. The results of the Kaiser-Meyer-Olkin (KMO) and Bartlett test for the Principal Axis technique are shown in Table 5. The KMO test provides information on whether the sample will be sufficient for factor analysis. The Bartlett test gives information on whether the variables are suitable for factor analysis. As can be seen from Table 5, the value of the KMO test was found to be 0.830 which is an indication that the sample is sufficient for factor analysis. Likewise, the significance value for the Bartlett test value was found to be 0.000 which means that the hypothesis H_0 will not be rejected because the variables are not suitable for factor analysis.

The percentages of explained variance of the Principal Axis technique are as given in Table 6. Accordingly, there are seven factors with an eigenvalue greater than 1. The first factor explains 23.301% of the total variance. The second factor explains 11.493% of the total variance. The third factor explains 11.135% of the total variance. The fourth factor explains 7.361% of the total variance. The fifth factor explains 6.149% of the total variance. The sixth factor explains 5.090% of the total variance and the seventh factor explains 4.387% of the total variance. Seven factors together explain 68.915% of the variance.

Table 4: Traffic perception items (TPI)

	Traffic Perception Items
1	The permitted alcohol level for driving is too much.
2	Tougher penalties for drivers who break the rule should be implemented.
3	Drivers who doesn't maintain a safe following distance are not aware of the danger.
4	Distance between police checks on motorways does not attract the attention of the driver is away.
5	Those who are stopped by the police due to improper passing are unlucky.
6	I make inspection and maintenance of the vehicle myself.
7	Vehicle inspection reduces traffic accidents caused by vehicle defects.
8	Vehicle inspection increases traffic safety.
9	Advanced and safe driving training must abe mandatory.
10	Getting advanced and safe driving training does not make the drivers experienced.
11	Advanced and safe driving training is a waste of time.
12	Advanced and safe driving training is the personal choice.
13	If all drivers are trained in advanced and safe driving, the roads would be safer.
14	Advanced and safe driving training to reduce traffic accidents.
15	It would be nice if drivers get a discount on automobile insurance after advanced and safe driving training.
16	I do not drive at night.
17	I do not drive in bad weather conditions.
18	I am afraid of being involved in the accident as a driver.
19	I am afraid of being wounded in traffic accidents as a driver.
20	I am afraid to die in a traffic accident.
21	I understand the mistakes of other drivers.
22	I do not go into situations that will create risk in traffic.
23	Vehicle insurance fees are prohibitive.
24	A car is a necessity.
25	A person's car gives information about the person.

Table 5: KMO and Bartlett's test results

Test	**Value**
KMO Test	0.830
Bartlett’s Test	
Chi-Square	37451.724
d.f.	210
Sig.	0.000

Factor 1-7 are determined as shown in Table 7. Table 8 shows the dimensions expressed by the variables. As a result of the Principal Axis technique, it has been found that 4% of residual values have a value greater than 0.05. This indicates that the model and data fit is very high.

In the next stage, the Traffic Perception Index (TPI) was constructed and used as a dependent variable in this study. Seven factor scores were taken as independent variables and LSM was applied. The results for the LSM are as shown in Table 9. After then, it was suspected that there were outliers and it is seen that the error terms are not normally distributed (Table 10). The graph of the outliers is shown in Figure 1. Multicollinearity and autocorrelation are not detected in the data set. Since the distribution of error terms are not normally distributed, M-Regression which is the one the technique of Robust Regression was applied and the results are shown in Table 11. The Akaike Information Criterion (AIC) used for determining to choose better technique.

Table 6: Total variance explained

Factor	Initial Eigenvalue Total	% of Total	Cumulative %	Rotation Loadings Total	% of Total	Cumulative %
1	4.893	23.301	23.301	3.602	17.152	17.152
2	2.414	11.493	34.794	2.251	10.720	27.872
3	2.338	11.135	45.929	2.147	10.224	38.096
4	1.546	7.361	53.290	1.996	9.502	47.599
5	1.291	6.149	59.439	1.533	7.301	54.900
6	1.069	5.090	64.529	1.487	7.081	61.981
7	0.921	4.387	68.915	1.456	6.934	68.915
8	0.727	3.463	72.378			
9	0.679	3.233	75.610			
10	0.617	2.940	78.550			
11	0.556	2.650	81.200			
12	0.519	2.471	83.671			
13	0.516	2.458	86.129			
14	0.489	2.327	88.456			
15	0.469	2.235	90.691			
16	0.402	1.916	92.607			
17	0.354	1.683	94.290			
18	0.333	1.585	95.875			
19	0.320	1.524	97.399			
20	0.277	1.321	98.721			
21	0.269	1.279	100.000			

Table 7: Determined factors

Factor	Explanation
1	Traffic rules perception questions
2	Fear of traffic accident questions
3	Advanced and safe driving training questions
4	Accident prevention questions after advanced and safe driving training
5	Vehicle maintenance questions
6	Accident risk questions
7	Vehicle ownership perception questions

Table 8: Rotated factor matrix by PA

TPI	Factor						
	1	2	3	4	5	6	7
TPI4	0.848						
TPI2	0.822						
TPI3	0.806						
TPI5	0.773						
TPI1	0.771						
TPI19		0.901					
TPI18		0.853					
TPI20		0.800					
TPI10			0.796				
TPI11			0.772				
TPI9			0.743				
TPI15			0.490				
TPI13				0.754			
TPI12				0.735			
TPI14				0.704			
TPI7					0.824		
TPI6					0.777		
TPI22						0.876	
TPI21						0.801	
TPI24							0.853
TPI23							0.828

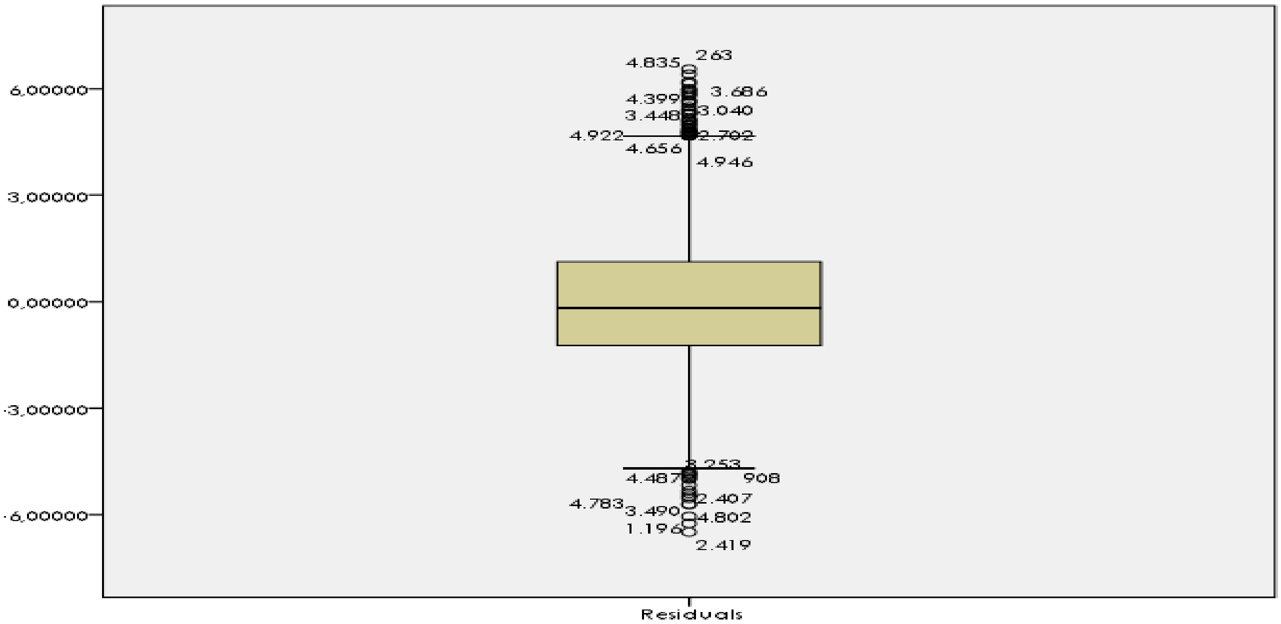


Figure 1: Graph of residuals

Table 9: Result of regression analysis by OLS

Parameter	b	S.E.	t	Sig.	VIF
Traffic rules perception questions (x_1)	5.736	0.026	223.791	0.000	1.000
Fear of traffic accident questions (x_2)	3.659	0.026	142.769	0.000	1.000
Advanced and safe driving training questions (x_3)	3.442	0.026	134.313	0.000	1.000
Accident prevention questions after advanced and safe driving training (x_4)	1.392	0.026	54.303	0.000	1.000
Vehicle maintenance questions (x_5)	2.360	0.026	92.094	0.000	1.000
Accident risk questions (x_6)	1.672	0.026	65.253	0.000	1.000
Vehicle ownership perception questions (x_7)	1.489	0.026	58.096	0.000	1.000
Constant	70.200	0.026	2739.311	0.000	

Table 10: Test of normality

Test Statistics	Statistic	Kolmogorov-Smirnov d.f.	Sig.
Residuals	0.048	5100	0.000*

** Not normal*

Table 11: Result of Robust Regression

Technique	Equation	AIC
OLS	$70.200+5.736x_1+3.659x_2+3.442x_3+1.392x_4+2.360x_5+1.672x_6+1.489x_7$	6172.743
Huber	$70.140+5.729x_1+3.644x_2+3.439x_3+1.407x_4+2.338x_5+1.674x_6+1.499x_7$	**6169.728**
Andrews	$70.138+5.728x_1+3.645x_2+3.442x_3+1.410x_4+2.340x_5+1.672x_6+1.499x_7$	6170.319
Tukey	$70.138+5.728x_1+3.645x_2+3.442x_3+1.410x_4+2.340x_5+1.672x_6+1.499x_7$	6170.319
Cauchy	$70.138+5.727x_1+3.645x_2+3.437x_3+1.410x_4+2.341x_5+1.672x_6+1.502x_7$	6170.433
Logistic	$70.131+5.727x_1+3.646x_2+3.435x_3+1.410x_4+2.342x_5+1.672x_6+1.502x_7$	6170.433

4. CONCLUSION

It is aimed to determine the traffic perceptions of the individuals living in Eskisehir province and their behavior in traffic within the scope of this traffic perception study. Some results that the institutions and organizations seeking for solutions to the traffic problem in Eskisehir province can benefit about the attitudes and behaviors of people about traffic are as follows:

- Police checkpoints on the highway do not attract drivers' attentions.
- More strict penalties must be given to drivers who violate traffic rules.
- The drivers are afraid of being wounded or dying in a traffic accident.
- Even after the advanced and safe driving training is given, there is the perception that this training can not provide experience to drivers in traffic.
- Advanced and safe driving training and regular vehicle meintanence provide decrease in number of traffic accidents.
- The drivers understand the mistakes of other drivers.
- It has been concluded that vehicle insurance fees are hindering the individuals from driving

In the study, the model among seven factors that can affect the traffic perception index was established with the help of regression analysis. However, the biggest problem while working with real data set in regression analysis is that the error terms are not normally distributed. One of the most important assumptions that enables the use of the LSM technique for statistical inference purposes is that the distribution of error terms is normal distribution.

When using the LSM, it is determined that one unit increase in the traffic rules perception value cause the average increase in the traffic perception index to be 5.736 units. Because the error terms are not normally distributed, the parameter values, confidence intervals and hypothesis tests related to the LSM will not reflect the truth. Similarly, parameter values, confidence intervals, and hypothesis for all factors will be biased. However, when the Robust Regression technique was applied, Huber M-Regression technique had the smallest AIC value. According to the Huber M-Regression technique, it can be said that one unit increase in the traffic rules perception value cause the average increase in the traffic perception index to be 5.729 units.

When the error terms are not normally distributed, it is expected that the AIC will be reduced by applying Robust Regression techniques. In addition, this study introduces alternative regression techniques to the regression analysis used in many researches in the field of traffic and provides a basis for new studies to be performed.

REFERENCES

Birkes, D., & Dodge, Y. (2011). *Alternative methods of regression* (Vol. 190). John Wiley & Sons.

Child, D. (2006). *The essentials of factor analysis*. A&C Black.

David, A. (1981). Introduction to statistics. *St. Paul: West Publishing Company*.

Dressler, A., Bredow, B., Rossger, L., Rudel, M., & Sturzbecher, D. (2017). Development and Trial of an Item Set for Testing Traffic Perception and Hazard Avoidance BT - Advances in Human Aspects of Transportation. In N. A. Stanton, S. Landry, G. Di Bucchianico, & A. Vallicelli (Eds.) (pp. 1003–1013). Cham: Springer International Publishing.

Ergul, B. (2006). *Robust regression and applications*. Eskisehir Osmangazi University.

Hatipoglu, S. (2011). Okul öncesi çağı çocuklarının trafik bilgi ve algıları. *Pegem Eğitim ve Öğretim Dergisi*, *1*(3), 23–28.

Khalaf, K. (2007). Factor Analysis and an Application. *Gazi University Institute of Science, Ankara*.

Mosteller, F., & Tukey, J. W. (1977). Data analysis and regression: a second course in statistics. *Addison-Wesley Series in Behavioral Science: Quantitative Methods*.

Ozturk, L. (2003). *Robust estimation techniques in linear regression and their comparisons*. Mimar Sinan University.

Rummel, J. (1970). Applied Factor Analysis (Evanston, Illinois; Northwestern Universtiy Press).

WHO. (2018). WHO Road Traffic Injuries Fact Sheet. Retrieved February 24, 2018, from http://www.who.int/mediacentre/factsheets/fs358/en/

BALCOR 2018
XIII Balkan Conference on Operational Research

A BI-LEVEL AIRPORT CHOICE MODEL (BACM) IN A MULTI-AIRPORT CONTEXT. THE CASE OF ROME.

STEFANIA SINESI[1], MARIA GIOVANNA ALTIERI[2], MAURO DELL'ORCO[3]

[1] Technical University of Bari, stefania.sinesi@poliba.it
[2] Technical University of Bari, mariagiovanna.altieri@poliba.it
[3] Technical University of Bari, mauro.dellorco@poliba.it

Abstract: *In the last decade, the researchers' interest in airport ground planning grew with the growth in the number of the air passengers. The aim of this paper is to set up a model of the airports' competitiveness, considering the connections between airports and territory and the passengers' perception of airports' services. We proposed a Bi-level Airport Choice Model (BACM) in a multi-airport context, able to find the best combination airport/transport mode from the users' point of view, considering also the airport ground accessibility. We also drew a graphical user interface (GUI), where the users can insert the inputs relevant to their needs and visualize the result suggested by BACM. Finally, we have applied BACM to the case of Rome; the results underline the wide necessity to increase the connections to and from airports and reflect the current competitiveness between the considered airports.*

Keywords: *Airport Ground Accessibility, Airports' competitiveness, Utility, Users' choice model.*

1. INTRODUCTION

In the last decade the number of people traveling by planes, for small or long distances, grew up. The constant growth of air transport importance in the world is the reason for the interest of researchers on this topic; in particular, on issues related to airport ground access planning.

The connection between airport and territory is a part of a wider network, including not only air connections operated by air carriers but also urban transfers to and from the airport itself. Therefore, an easy and quick access for users coming from the surrounding urban areas is a crucial factor for the airports' competitiveness.

Until a few decades ago, the user's choice of the airport of departure was often forced, due to the lack of alternatives. Instead, nowadays travelers often have a range of available alternatives, since various airports could be at their disposal. In this case, the choice of the departure airport is a result of a set of choices regarding, for example, departure flight and time, destination airport, arrival time, airline company, type of air carrier (whether charter or scheduled), ticket price, transport mode to reach the airport, etc. In some cases, this process is simplified in relation to the individual traveler's needs and preferences; some users prefer to choose the departure/arrival time windows, while others would prefer the cheapest alternative.

To calculate the attractive power of an airport in a multi-airport context, in this paper we studied the major factors influencing the users' decision-making process and set up a Bi-level Airport Choice Model (BACM) considering the passengers' perception of the airport's ground access, and how this perception influences their successive connected choices.

2. LITERATURE REVIEW

In the study of transport systems, the Random Utility Models are the most used for estimating the transport demand (Cascetta and Papola, 2001). The concept of "Utility" is a theoretical artifice that allows associating a numerical value at the user's satisfaction level of a given attribute, to quantify it (Lancaster, 1966).

Usually, a user chooses in a rational way, associating a measure of perceived utility to every alternative and choosing the one maximizing his/her utility, excluding all the others (Domenich and McFadden, 1975); the choice of a specific alternative is made on the basis of the users' tastes, needs, and constraints, but this choice can vary over the time and can be different even among users who have the same socio-economic characteristics (McFadden, 1981).

The first studies on the choice models of the departure airport, based on utility measures, date back to the 1970s; the most sought-after approach associates the airport infrastructure with attributes representing the service level offered, and takes into account airline ticket prices, airlines type, flight frequency and accessibility from the users' point of view.

In the specific literature, the above models are classified into:

- • Models simulating only the airport choice;
- • Models simulating a combination of two or three dimensions of choice (for example, airport and mode of access, or type of carrier and destination).

Among the studies on this topic, we recall the one by Harvey (1987), carried out in San Francisco area, which states that access time and frequency of flights are the determining factors in the user selection process.

In a recent study, De Luca et al. (2012) have deepened a choice model for three airports in Southern Italy: Naples Capodichino, Rome Ciampino, and Rome Fiumicino, having different accessibility and service levels.

In this study, they found that the airport access time is the only attribute that can influence significantly the travelers' decision-making process.

To summarize, we can conclude that the airport choice depends not only on-air traffic management services, on the number of reachable destinations, on ticket prices or on the frequency of flights, but also on connections of the air station with the surrounding territory and on interactions of the airport with the metropolitan areas through the transport network. In fact, according to De Luca et al., these are the factors that mostly influence the users' perception of utility associated with an airport.

The concept of accessibility has long been discussed in the transportation planning literature. From a conceptual point of view, accessibility reflects either the ease of a traveler to reach a place –"active accessibility" – or the ease an activity to be reached by potential users –"passive accessibility".

Depending on the considered variables, in the literature the different accessibility measures are divided into three categories:

1. Measures considering only travel time;
2. Measures considering both time and travel costs, parking costs, travel time reliability, level of service, etc.;
3. Measures considering different transport modes to complete the journey.

During the years, a lot of authors have discussed the concept of accessibility and involved models based on the three measures mentioned above; see, for example, Ben-Akiva and Lerman (1979), Pirie (1979), Miller (1999) and Cascetta (2009). The travelers' decision for or against an airport depends, to a certain extent, on its accessibility. On its turn, the accessibility of an airport is related both to the users' perception of the airport level of service, and to the users' choice.

In literature, there are several papers studying the variables affecting the airport ground accessibility, proposing models, and finding factors influencing travelers' choices, such as airport of origin or destination for their trips, type of travel, whether traveling by plane or not, etc.

One of the first research on the ground airport accessibility was by Ellis et al. (1974), underlining how the travel cost and time are the most sensitive factors in passengers' airport ground access mode choice in Baltimore-Washington airport.

In 2008, Gupta et al. developed a joint airport and ground access mode choice model for the New York City metropolitan region. They formulated a bi-level logit model with airport choice at the upper level, and ground access mode choice at the lower level. Results showed that air passengers' travel behavior is significantly different for business and non-business travelers. Moreover, willingness to pay for trips to and from the airport is much higher than for regular intracity trips. In addition, they found that access time, access cost, and individuals' socioeconomic characteristics are significant factors affecting the resulting mode choice.

Keumi and Murakami (2012) have considered waiting time, delay cost and service frequency. Roh (2013) studied the effects of the gasoline cost and the trip origin.

Hesse et al. (2013) affirmed that the mode choice for access to airports also depends on the structure of airport itself and on the number of airlines operating; the destination airport, the airline, the aircraft type, the on-time performances, and the attributes of the service may have a relevant effect on airport ground access. According to the authors, if all other variables are constant between two airports, accessibility is the only determinant in selecting airport by the users; consequently, in the same time, the choice of the departure airport may influence the passengers' choice of modes to reach the airport.

Bao D. et al. (2016) studied the airport accessibility for nine large airports in Jiangsu province, China. They proposed a cost-based airport accessibility model to calculate airport's level of accessibility by generating a graph showing the relationship between airport accessibility and related services. The results showed that the airport accessibility is significantly affected by passengers' traffic and airlines frequency; the higher airport accessibility is perceived by users, the higher is the number of passengers that choose that airport. Thus, the measure of ground airport accessibility is an important factor for competition between airports, regarding passengers' choices and, consequently, airlines preferences.

In conclusion, the accessibility as perceived by users affects their choices of the departure/arrival airport, and the mode to reach it. At the same time, the travelers' airport choice is influenced by many other variables, such as time and monetary cost to reach the airport, socio-economic conditions of travelers, number of airlines operating at the airport, frequency of flights, number of destinations available from the selected airport of origin. These last variables also influence the airlines-perceived accessibility.

3. THE PROPOSED MODEL

In the Bi-level Airport Choice Model (BACM) the I level finds, through a Utility measure, the airport having the greatest utility, as perceived by users, among those belonging to a certain region. The II level finds, through an Accessibility measure, the transport mode (car, bus, train, or a combination of these modes) having the highest accessibility to the airport found in the I level. Furthermore, the overlap of the two levels provides a single result finding the airport/transport mode combination with a greater overall utility perceived by the user, also considering the airport ground accessibility.

Finally, we set up a graphical user interface (GUI), in which the users can insert the inputs relevant to their needs and immediately visualize the result of the BACM. This GUI is a useful support for user's decisions and allows changing the inputs, to know the different alternatives and to evaluate the choice according to the individual needs.

In the I level, we have considered the following dimensions of choice:

- the city origin of trips to reach the airport;
- the destination city of the flight;
- the preferred departure day of week;
- the preferred departure time window.

In a multi-airport context, I level returns the name of the airport *j* having the highest utility perceived by the users; the user, by entering the inputs, chooses to reach the destination city *d* by air, starting from the city *i*, in the weekly day *g* and in the time window *f*.

The mathematical model of the I level is:

$$I = max\left\{I_{ijdgf}\right\} = max\left\{\gamma_{jgf} \bullet \left[\beta D_{ij} + \left(\frac{\alpha_{jdgf} \bullet \beta C_{jdgf}}{AF_{jgf}}\right)\right] \bullet n_{jdgf}\right\} \quad (1)$$

where:

i = i-th city from which the user leaves to reach the airport;
j = j-th departure airport;
d = d-th destination city;
g = g-th weekly day;
f = f-th time window of the takeoff. In the proposed model we have considered four time windows: 0.00 – 6.00; 6.00 – 12.00; 12.00 – 18.00; 18.00 – 24.00;
I_{ijdgf} = Utility measure perceived by the user who leaves from the city i, goes to the airport j to reach the destination city of the flight d, in the weekly day g and departing in the preferred time slot f;
γ_{jgf} = Dummy variable showing if the airport *j* is working within the weekly day g and in time window *f*, selected by the user as input;
βD_{ij} = Distance from the city *i* to the airport *j*;
α_{jdgf} = Dummy variable showing the possibility to reach the desired destination d with a direct flight from *j*, within the weekly day *g* and the preferred time window *f*;
βC_{jdgf} = Minimum flight ticket price that the user has to pay to reach *d*, leaving from *j*, during the day *g*, in the time window *f*. In the proposed model we have assumed that the user buys the ticket one month before the departure;
AF_{jgf} = Crowding index of the airport *j*, in the weekly day *g* and time window *f*;
n_{jdgf} = Number of flights taking off from the airport *j* and directed to *d*, during the day *g* and time window *f* selected by the user.

After finding, in the I level, the best airport j to leave from, perceiving the maximum utility, the II level finds out the transport mode *m* to reach the airport *j*, maximizing also the perceived ground accessibility.
The model used in the II level is the following:

$$A = max \{A_{ijmdgf}\} = max \left\{ \alpha_{jdgf} \bullet AF_{jgf} \bullet \gamma_{jgf} \bullet \frac{\text{ч}_{ijm} \bullet S_{ijm}}{C_{ijm}^2} \right\} \quad (2)$$

where:

m = m-th transport mode;

A_{ijmdgf} = Airport ground accessibility measure perceived by the user traveling from the city *i* to the airport *j* (determined in the I level) using the transport mode *m*, during the day *g* and in the time window *f*;

ч_{ijm} = Dummy variable showing the possibility to travel from the city *i* to the airport *j* using the transport mode *m*;

S_{ijm} = Service index associated to m used to reach the airport. It is a function of the individual characteristics of the transport mode *m*, such as, for cars: number of parking spaces for cars near the airport, hourly parking fees; for bus or train: distance from the bus or train stop to the j's departures terminal, number of means of transport needed to reach *j*, frequency of those means, etc.;

C_{ijm} = Generalized Cost of the trip from *i* to *j*. It is a function of travel time, additional waiting times, tickets cost, gasoline cost, possible tolls, etc.;

The BACM proposed in this paper consists of the overlap of the two levels exposed above.
The final formulation of the proposed model is the following:

$$V = max \{V_{ijmdgf}\} =$$

$$= \max \left\{ \left\{ \gamma_{jgf} \bullet \left[\beta D_{ij} + \left(\frac{\alpha_{jdgf} \bullet \beta C_{jdgf}}{AF_{jgf}} \right) \right] \bullet n_{jdgf} \right\} + \alpha_{jdgf} \bullet AF_{jgf} \bullet \gamma_{jgf} \bullet \frac{\text{ч}_{ijm} \bullet S_{ijm}}{C_{ijm}^2} \right\} \quad (3)$$

Considering the model inputs selected by the user, V is a Utility Indicator that maximized simultaneously both the usefulness perceived in reaching one of the airports present in a multi-airport system and the airport ground accessibility resulting from the use of the transport mode *m*.

The BACM does not always return the same result as the two levels applied one at a time; this is because sometimes an airport presents a low value of I (Eq.1) due, for example, to a high ticket price, or a short distance between the city *i* and the airport *j*; on the other hand, the same airport could have a high value of A (Eq. 2) for a particular mean of transport, so that the result of the BACM is different from the two single levels' results.

4. THE CASE STUDY

In this research we applied the proposed BACM to the multi-airport context of Rome, in Italy, consisting of two airports: Fiumicino and Ciampino. The first one operates more than 200 destinations all over the world, while the second around 60.

The Fiumicino airport is approximately 32 km south-west far from Rome; in 2016, with almost 42 million of passengers, it has been the main Italian airport; instead, the Ciampino airport, at the south-east of Rome, initially operated as a military airport or for charter flights. After the low-cost companies' birth, has become one of the most important Italian airports and today it counts more than 5 million of passengers per year. Together, these two airports receive the largest share of traffic in Central Italy, which represents 30% of Italian international traffic and 30% of low-cost traffic. As defined in the National Plan of Airports, in 2012, it is possible to consider that both airports have the same influence area, in which the users can choose one airport or the other, according to their needs.

A large number of models defining the air transport demand based on the socio-economic users' characteristics is present in literature. In our study, to quantify the demand depending on the socio-economic conditions of the potential users, we calculated an Indicator of the economic situation for each city, expressed as a function of both the average income and the number of inhabitants.

At first, we considered the five regions surrounding the Lazio region in the survey area to elaborate a more detailed analysis but, after some accurate evaluations, we decided to consider only the 392 Municipalities belonging to the Lazio region, as well as the 15 Rome's districts.

Through a cumulative function of an *Economic situation indicator***,** we found that 90% of the air transport demand is generated from only 55 Lazio's municipalities, including the 15 districts of Rome. Moreover, we found that only 17 destinations are reachable by direct flights from both airports.

To evaluate the outputs of the proposed BACM model, we first applied separately the two levels to the case study and then the BACM. Considering the 55 municipalities previously found, the 17 destinations reachable by direct flights from both airports, seven weekly days, and four time windows, we obtained 26,180 different cases per week, representing the possible user's inputs.

The following figure shows the GUI created for this study, in which the results of the BACM are reported; along the light blue row, the inputs set by the user are listed.

Complete the blue fields by choosing the items from the drop-down menu.

City i	Destination d	Weekly day g	Time window f	BACM
Latina	BERLINO	MONDAY	6.00 - 12.00	FIUMICINO USING CAR

Figure1: Example of a GUI.

Instead, for the same case reported in Fig. 1, in Fig. 2 are shown the numerical results given by the application of I and II levels, separately, and application of the BACM.

j= F (Fiumicino)

i	d	g	f	βD_{ij}	α_{djgf}	βC_{djg}	n_{dgjf}	AF_{jgf}	γ_{jgf}	Iijdgf (I Level)	Aijmdf (II Level)	Vijmdgf (BACM)
Latina	BERLINO	MONDAY	6.00 - 12.00	0,579418	1	0	2	0,809322	1	1,431861	82,35928	83,791141

m	Oijm	Iijdgf	Aijmdf	Vijmdgf
CAR	101,7633	1,431861	82,35928	83,791141
BUS	0	1,431861	0	1,431861
TRAIN	0	1,431861	0	1,431861
MULTIMODAL	2,787499	1,431861	2,255984	3,6878453

j= C (Ciampino)

i	d	g	f	βD_{ij}	α_{djgf}	βC_{djg}	n_{dgjf}	AF_{jgf}	γ_{jgf}	Iijdgf (I Level)	Aijmdf (II Level)	Vijmdgf (BACM)
Latina	BERLINO	MONDAY	6.00 - 12.00	0,665129	1	1	1	0,799435	1	2,082882	15,69131	17,774196

m	Oijm	Iijdgf	Aijmdf	Vijmdgf
CAR	6,278927	2,082882	5,019594	7,102476
BUS	19,628	2,082882	15,69131	17,774196
TRAIN	0	2,082882	0	2,0828816
MULTIMODAL	1,882405	2,082882	1,50486	3,587742

Figure 2: Numerical results given by the application of I and II levels, separately, and application of the BACM, for the same case of the Fig.1.

As can be observed in the Fig.1, the BACM does not provide the same output as the two levels applied separately; in fact, the I level indicates Ciampino as the airport whose utility perceived by the user is maximum, and the II level suggests the bus as the mode providing the highest accessibility. Instead, the BACM indicates as the best alternative traveling to Fiumicino by car. In this condition, according to the proposed model, both utility and accessibility measures are maximized simultaneously.

5. RESULTS

After applying the I level to 26,180 cases, we have found that for 26.88% of cases users perceive the maximum utility choosing Fiumicino as the departure airport, while for 25.40% Ciampino is favored; in 47.72% of the cases, the users perceive the same utility choosing one of the two airports.

However, applying the II level we got a confirmation of the general idea that an infrastructure is more accessible if reached by car, due to the lack of public transport offer; in fact, in 45% of cases, both airports are perceived more accessible by car, while in the remaining 55% of cases the perception of airports accessibility is distributed among the others considered transport modes.

Furthermore, we have verified that in 92.20% of cases, the results obtained by the BACM and the application of the two levels separately are coincident. Applying the BACM and considering both the users' needs and the airport ground accessibility, the results showed that overall utility is almost the same: 49.87%

Fiumicino and 50.13% Ciampino. Therefore, it is possible to understand how in this multi-airport context, the results provided by the proposed model found a clear evidence of the real competitiveness between the two Rome airports.

6. CONCLUSIONS

Starting from the study of the utility and accessibility measures proposed by many authors over the years, in this paper we proposed a Bi-Level airport choice model based on the perceived utility on the first level, and ground accessibility on the second level. Then, we applied the proposed model to the case of Rome airports: Ciampino and Fiumicino. The main objective of the study was to understand if the competitiveness between airports could be connected to the users' perception, regarding the services offered by the two airports and the way they are included into the transport network.

Moreover, the application of the BAMC to the case of Rome has shown that the overall utility is perceived, for the two airports, in the same measure by users.

The results obtained from the proposed research find real support in the future scenarios presented in the "MasterPlan for 2044" document, whose objective is to tackle the development of air traffic, increasing the level of ground accessibility of the Rome's airports.

In fact, in the "MasterPlan for 2044", both the planning of new road infrastructures connecting the airports to the neighboring territory and the concept of ground accessibility with a view to sustainable mobility assume particular importance. Therefore, to reduce the road congestion, the MasterPlan proposes the increasing of the rail supply and encourages people to use public transport systems.

REFERENCES

Bao D., Hua S., Gu J. (2016). Relevance of airport accessibility and airport competition. *Journal of Air Transport Management*, 55, 52-60.

Ben-Akiva, M. E., Lerman, S. R. (1979). Disaggregate travel and mobility-choice models and measures of accessibility. *Behavioural Travel Modelling, edited by D. A. Hensher and P. R. Storper*, 654–79, (London).

Cascetta E., Pappola A. (2001). Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research*, 9C, 249-263.

Cascetta, E. (2009). *Transportation Systems Analysis: Models and Applications*. 2nd edition. Springer.

Domenich T., McFadden D. (1975). *Urban Travel Demand – A Behavioural Analysis, North Holland, Amsterdam*.

De Luca S. (2012). Modelling airport choice behaviour for direct flights, connecting flights and different travel plans. *Journal of Transport Geography* 22, 148-163.

Ellis, R. H., Bennett, J. C. & Rassam, P. R. (1974). Approaches for improving airport access. *Journal of Transportation Engineering*, 100.

Gupta, S., Vovsha, P. & Donnelly, R. (2008). Air Passenger Preferences for Choice of Airport and Ground Access Mode in the New York City Metropolitan Region. *Transportation Research Record: Journal of the Transportation Research Board*, 2042, 3-11.

Harvey G. (1987). Airport Choice in a Multiple Airport Region. *Trasportation Research A*, 21 (6), 439-449.

Hesse C., et al. (2013). Accessibility Measures and Flight Schedules: An Application to the European Air Transport. *European Transport Issue*, 55, Paper n° 6, ISSN 1825-3997.

Keumi, C. & Murakami, H. (2012). The role of schedule delays on passengers' choice of access modes: A case study of Japan's international hub airports. *Transportation Research Part E: Logistics and Transportation Review*, 48, 1023-1031.

Lancaster K.J. (1966). A New Approach to Consumer Theory. *Journal of Political Economy*, 74, 132-157.

McFadden D. (1981). Econometrics Models of Probabilistic Choice, Structural Analysis of Discrete Data. *MIT Press, Cambridge, Mas.*

Miller, J.H. (1999). Measuring space–time accessibility benefits within transportation networks: basic theory and computational procedures. *Geogr. Anal.* 31 (2), 187–212.

Pirie, G.H. (1979). Measuring accessibility: a review and proposal. *Environ. Plan. A* 11, 299–312.

Roh, H.-J. (2013). Mode Choice Behavior of Various Airport User Groups for Ground Airport Access. *The Open Transportation Journal*, 7.

YUJOR - Yugoslav Journal of Operations Research is an international journal dealing with all aspects of operations research, systems science, and management science. **YUJOR** publishes refereed papers describing significant results in the above areas whether theoretical or empirical, mathematical or descriptive. The international composition of **YUJOR**'s Editorial Board particularly welcomes manuscripts devoted to applications involving novel ideas and relations with future information technologies.

YU ISSN 0354-0243

YUJOR publishes contributed papers, survey papers, tutorial expositions, and technical notes. The Journal also has pages devoted to book reviews, calendars of the forthcoming events, and other short messages about the related activities in the professional fields.

It is hoped that **YUJOR** will appeal to interdisciplinary oriented scientists and practitioners including but not restricted to OR specialists, applied mathematicians and engineers, economists and business administrators, military specialists and entrepreneurs.

YUJOR is included in IFOR's list of primary journals and regularly reviewed in:

- International Abstracts in Operations Research,
- Mathematical Reviews,
- Zentralblatt für Mathematik and
- Referativnyi Zhurnal
- EBSCO
- Scopus.

YUJOR Editorial Office:

University of Belgrade, Faculty of Organizational Sciences
Laboratory for Operations Research "Jovan Petrić", Jove Ilića 154
11000 Belgrade, Serbia
Phone: (381 11) 3950 835,
Fax: (381 11) 3950 868
e-mail: yujor@fon.bg.ac.rs
http://yujor.fon.bg.ac.rs