The Application of Gravity Model in the Investigation of Spatial Structure

Áron Kincses, Géza Tóth

Hungarian Central Statistical Office
Keleti K. u. 5-7, 1024 Budapest, Hungary
E-mail: aron.kincses@ksh.hu; geza.toth@ksh.hu

Abstract: In this paper the authors wish to introduce an application of the gravity model through a concrete example. In their investigation the gravity model was transformed to analyse the impact of accessibility in a way, that not only the size of gravitational forces but their direction can also be measured. Displacements were illustrated by a bi-dimensional regression. The aim of this paper to give a new perspective to the investigation of spatial structure through a Hungarian example. This makes easier the transport planning and land development modeling. In our work accessibility analysis, gravity modeling, bi-dimensional regression calculations and GIS visualization were performed.

Keywords: gravity model; bi-dimensional regression; accessibility; spatial structure; Hungary

1 Introduction

The overall goal of modelling is to simplify reality, actual processes and interactions and on the basis of the obtained data to draw conclusions and make forecasts. Models based on gravitational analogy are the tools of spatial interactions of classical regional analyses. They were first applied in the 19th Century [2, 3, 4, 11, 19, 21, 31].

The application of the geographical gravity is confirmed by the theory of experience according to which (just as in time) the things that are closer to each other in space are more related than distant things. This is called the "first law of geography” [25].

There are two basic areas of the application of gravity models based on physical analogy: the spatial flow analysis [7, 16], and the demarcation of catchment areas [15, 17]. The potential models based on gravitational analogy are the most important groups of accessible models. In general, it can be stated that they are accessible approaches according to which models show potential benefits of the region compared to other regions where the benefits are quantified [20].

The use of accessible models in transport-geographical studies is very common. However, when models are used, it is not entirely clear what is actually modelled; because of their complexity their interpretations may be difficult [13]. It should be stressed that accessibility has no universally accepted definition; in empirical studies different methodological background indicators are used (see [9, 10, 29, 30].The gravitational theory is a theory of contact, which examines the territorial interaction between two or more points in a similar way as correlations are analysed in the law of gravitation in physics. According to Dusek [5], despite the analogy, there are significant differences between gravity models used by social sciences and the law of gravitation used in physics. It is worth bearing in mind that "the gravity model is not based on the gravitational law”. It is a fundamental statement based on the experience of undeniable statistical character that takes into consideration spatial phenomena. According to this statement, phenomena interact with each other. The phenomena, which are closer to each other in space, are more related than distant phenomena [5 p. 45].

There are a number of differences between the law and the model. In this study, we wish to highlight a new point of view. As a consequence of the spatial interaction, classical gravitational potential models show the magnitude of potential at spatial points. Regarding the law of gravity in physics, the direction of forces cannot be evaded. In our approach each unit area is assigned an attraction direction. That is, in the case of the gravity model (although such spaces are free of vortex) the space is characterized using vectors.

2 Method

The universal gravitational law, Newton's gravitational law, states that any two point-like bodies mutually attract each other by a force, whose magnitude is directly proportional to the product of the bodies’ weight and is inversely proportional to the square of the distance between them [1] (1st formula):

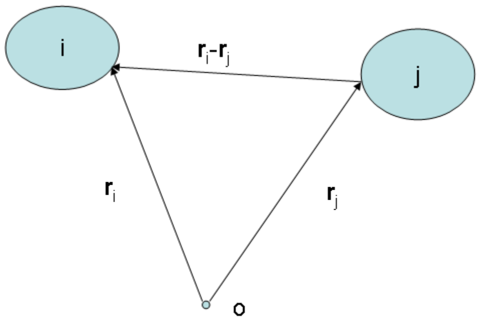
$$F = \gamma \cdot \frac{m_1 \cdot m_2}{r^2} \tag{1}$$

Where γ is the proportional factor, the gravitational constant (independent of time and place).

If r indicates the radius vector drawn from the mass point number 2 to a mass point number 1, then r/r is the unit vector drawn from 1 towards 2, thus, the impact of gravitational force on mass point 1 from mass point 2 is in (Equation 2) (Figure 1), which reads

$$\vec{F}_{1,2} = -\gamma \cdot \frac{m_1 \cdot m_2}{r^2} \cdot \frac{\vec{r}}{r} \tag{2}$$

A gravitational field is set if the gradient (K) can be specified by a direction and magnitude in each point of the range. Since K is a vector quantity, three numbers (two in plane) are required to be known at each point, for example the right angle components of gradient K_x, K_y, K_z, which are functions of the site. However, many fields, including the gravitational field, can be characterized in a more simple way. They can be expressed by a single scalar, the so-called potential function, instead of three values.



The impact of force on i by j : $\vec{\mathbf{F}}_{\mathbf{ij}} = -\boldsymbol{\gamma} \cdot \frac{\mathbf{m_i} \cdot \mathbf{m_j}}{\mathbf{r^2}} \cdot \frac{\vec{\mathbf{r}}_{\mathbf{i}} - \vec{\mathbf{r}}_{\mathbf{j}}}{\mathbf{r}}$

Figure 1

Gravitational power

The potential is in a similar relation with gradient as work and potential energy with force.

Taking advantage of this, gravity models are also applied in most social sciences where space is usually described by a single scalar function, [13], whereas in gravitational law vectors characterizing the space are of great importance. The primary reason for this is that arithmetic operations calculated with numbers are easier to handle than with vectors. Perhaps, we could say that by working with potentials we can avoid calculation problems in problem solving. The potential completely characterizes the whirl-free gravity gravitational field, because there is a definite relationship between the field strength and the potential:

$$\vec{K} = -gradU \Rightarrow \qquad K_x = -\frac{\partial U}{\partial x}; \qquad K_y = -\frac{\partial U}{\partial y}. \tag{3}$$

In other words, the potential (as mathematical functional) is the negative gradient of field strength. Various types of potentials and models, which are different from the ones directly based by the gravitational analogy, but in this case, force effects among space power sources are quite different. In fact, these models differ from each other since the attractive forces remain above a predetermined limit value and within set distances.

The force in a general form is:

$$\left|\vec{F}\right| = C\frac{m_1^{\alpha} m_2^{\beta}}{r^{\gamma}} \tag{4}$$

where C, α, β, γ are constants. [14].

However, how they describe actual power relations between social masses is another question.

Although potential models often characterize concentration on focal points of areas and spatial structures, they fail to provide information in which direction and with how much force the social attributes of other areas attract each delimited area unit. Thus, we attempt to use vectors in order to show the direction where the Hungarian micro regions tend to attract micro regions (LAU1) in the economical space compared to their actual geographical position. This analysis can demonstrate the most important centres of attraction, or discrepancies, and the differences between the gravitational orientations of micro-regions can be displayed on a map after the evaluation of data from 2000, 2005 and 2010 has been performed. In the study the geometric centres of specific micro-regions were the co-ordinates of Hungarian micro-regions, which were determined in the EOV co-ordinate system by (Geographical Information System) GIS software.

Our goal can be reached by using Equation (3) to potentials or directly with the help of forces. We chose the latter one.

In the conventional gravity model [21] D_{ij} is the "demographic force" between i and j where W_i and W_j are the population size of the settlements (regions), d_{ij} is the distance between i and j, and finally, g is the empirical constant (Equation 5).

$$D_{ij} = g \cdot \left(\frac{W_i \cdot W_j}{d_{ij}^2}\right) \tag{5}$$

In this study, the W_i and W_j weight factors represent personal income, which is the base of the income tax in small communities, dij is the actual distance between i and j regional centres measured on road by a minute (regardless of the traffic conditions and only the maximum speed depending on the road type is taken into consideration).

By generalizing the aforementioned formula, we write the following equation (Equation 6 and 7):

$$D_{ij} = \left|\vec{D}_{ij}\right| = \frac{W_i \cdot W_j}{d_{ij}^c} \tag{6}$$

$$\vec{D}_{ij} = -\frac{W_i \cdot W_j}{d_{ij}^{C+1}} \cdot \vec{d}_{ij} \tag{7}$$

Where W_i and W_j are the masses, d_{ij} is the distance between them, c is a constant, which is the change of the intensity of inter-regional relations as a function of distance. As the exponent, c, increases and the intensity of inter-regional relations becomes more sensitive to distance, this significance of masses gradually decreases [5]. The minus sign express mathematically, that the masses attract each other (see Figure 1).

With the extension of the above Equation we cannot only measure the strength of the force between the two regions, but its direction as well.

While performing calculations, it is worth dividing the vectors into x and y components and summarize them separately. To calculate the magnitude of this effect (vertical and horizontal forces of components) the following Equations are required (Equation 8 and 9), which follow from (6):

$$D_{ij}^X = -\frac{W_i \cdot W_j}{d_{ij}^{c+1}} \cdot (x_i - x_j) \tag{8}$$

$$D_{ij}^Y = -\frac{W_i \cdot W_j}{d_{ij}^{c+1}} \cdot (y_i - y_j) \tag{9}$$

Where x_i, x_j, y_i, y_j are the coordinates of the i and j regions.

However, if we perform the calculation on all unit areas involved, we will know in which direction their forces exactly act and how strongly they affect the given unit area. (Equation 10)

$$D_i^X = -\sum_{j=1}^{n} \frac{W_i \cdot W_j}{d_{ij}^{c+1}} \cdot (x_i - x_j)$$

$$D_i^Y = -\sum_{j=1}^{n} \frac{W_i \cdot W_j}{d_{ij}^{c+1}} \cdot (y_i - y_j) \tag{10}$$

It should be noted that while in potential models, the results are modified by the introduction of "self-potential", in the examination of forces we disregard the introduction of "self- forces".

Thus, it is possible to determine the magnitude and direction of force in which other areas affect each territorial unity. The direction of the vector, which is assigned to the region, determines the attraction direction of other unit areas, while the length of the vector is proportional to magnitude of force. For the sake of mapping and illustration, we transformed the received forces into shifts proportional to them in the following way (Equation 11 and 12):

$$x_i^{mod} = x_i + \left(D_i^X \frac{x^{max}}{x^{min}} k \frac{1}{\frac{D^{X\,max}}{D^{X\,min}}} \right) \quad (11)$$

$$y_i^{mod} = y_i + \left(D_i^Y \frac{y^{max}}{y^{min}} k \frac{1}{\frac{D^{Y\,max}}{D^{Y\,min}}} \right) \quad (12)$$

Where X_i mod and Y_i mod are the coordinates of new points modified by the gravitational force, x and y are the coordinates of the original point set, the extreme values of which are x_{max}, y_{max}, a x_{min}, y_{min}, D_i is the force along the x and y axes, k is a constant and in this case it is 0.5. This has the effect of normalizing the data magnitudes.

We assume that in our model the amount of interactions between the „masses" is the same as in Equation 7, and based on the superposition principle, it can be calculated for a given region by Equation 10. The new model cannot directly be compared with transport-geographical data, but the results compared with traffic data in potential models to verify our model [13].

Our model is a kind of complement to the potential models that ensures a deeper insight into them. In the following sections of this study we intend to communicate some significant results of this model.

3 Application of Bi-Dimensional Regression

The point set obtained by the gravitational calculation (W_i is the population size of the i. micro-region in Hungary, d_{ij} is the distance between i and j micro-regions), is worth comparing with the baseline point set, that is, with the actual real-world geographic coordinates and examining how the space is changed and distorted by the field of force. The comparison, of course, can be done by a simple cartographic representation, but with such a large number of points, it is not really promising good results. It is much better to use a bi-dimensional regression.

The bi-dimensional regression is one the methods of comparing partial shapes. The comparison is possible only if one of the point coordinates in the coordinate systems differing from each other is transformed to another coordinate system by an appropriate rate of displacement, rotation and scaling. Thus, it is possible to determine the degree of local and global similarities of shapes as well as their differences that are based on the unique and aggregated differences between the points of the shapes transformed into a common coordinate system. The method was developed by Tobler, who published a study describing this procedure in 1994 after the precedents of the 60s and 70s [23, 24, 26, 27, 28]. There are many examples using this procedure, which are not necessarily motivated by geographic issues [12, 22, 18].

As for the equation relating to the calculation of the Euclidean version, see [27, 8, 6].

Table 1
The equation of the two dimensional regression of Euclidean

1. Equation of the regression	$\begin{pmatrix} A' \\ B' \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \beta_1 & -\beta_2 \\ \beta_2 & \beta_1 \end{pmatrix} * \begin{pmatrix} X \\ Y \end{pmatrix}$
2. Scale difference	$\Phi = \sqrt{\beta_1^2 + \beta_2^2}$
3. Rotation	$\Theta = \tan^{-1}\left(\frac{\beta_2}{\beta_1}\right)$
4.Calculation of β_1	$\beta_1 = \frac{\sum(a_i - \bar{a}) * (x_i - \bar{x}) + \sum(b_i - \bar{b}) * (y_i - \bar{y})}{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2}$
5.Calculation of β_2	$\beta_2 = \frac{\sum(b_i - \bar{b}) * (x_i - \bar{x}) - \sum(a_i - \bar{a}) * (y_i - \bar{y})}{\sum(x_i - \bar{x})^2 + \sum(y - \bar{y})^2}$
6. Horizontal shift	$\alpha_1 = \bar{a} - \beta_1 * \bar{x} + \beta_2 * \bar{y}$
7. Vertical Shift	$\alpha_2 = \bar{b} - \beta_2 * \bar{x} - \beta_1 * \bar{y}$
8. Correlation based on error terms	$r = \sqrt{1 - \frac{\sum\left[(a_i - a_i')^2 + (b_i - b_i')^2\right]}{\sum\left[(a_i - \bar{a})^2 + (b_i - \bar{b})^2\right]}}$
9. Breakdown of the square sum of the difference	$\sum\left[(a_i - \bar{a})^2 + (b_i - \bar{b})^2\right] =$ $\sum\left[(a_i' - \bar{a})^2 + (b_i' - \bar{b})^2\right] + \sum\left[(a_i - a_i')^2 + (b_i - b_i')^2\right]$ SST=SSR+SSE
10. Calculation of A’	$A' = \alpha_1 + \beta_1(X) - \beta_2(Y)$
11. CALCULATION of B’	$B' = \alpha_2 + \beta_2(X) + \beta_1(Y)$

Source: [27] and [8] based on [6 p. 14]

Where x and y are the coordinates of independent shapes, a and b are the coordinates of dependent shapes, and represent the coordinates of dependent shapes in the system of independent shapes. α_1 determines the measure of horizontal shift, while α_2 determines the measure of vertical shift. β_1 and β_2 are the scalar difference and (Φ) and (Θ) determine the angle of shifting.

SST is the total square sum of difference. SSR is the square sum of difference explained by regression. SSE is the square sum of difference not explained by the regression (residual). Further details about the background of the two-dimensional regression can be seen in [6 pp. 14-15].

Table 2

Two-dimensional regression between the gravitational and the geographical space

Year	r	α_1	α_2	β_1	β_2	Φ	Θ
2000	0,942	6304,48	2017,44	0,99	0,00	0,99	0,00
2005	0,942	6030,56	2012,23	0,99	0,00	0,99	0,00
2010	0,941	8026,79	2632,29	0,99	0,00	0,99	0,00

Year	SST (%)	SSR (%)	SSE (%)
2000	100,00	98,73	1,27
2005	100,00	98,74	1,26
2010	100,00	98,69	1,31

Source: own calculation

Our results show that there is a strong relation between the two point systems; the transformed version from the original point set can be obtained without using rotation (Θ = 0). Essential ratio difference between the two shapes was not observed. Comparing the obtained results, it is obvious that the set of points behaves like a single-centre mid-point similarity, when it is diminished. This means that only the attractive force of Budapest can be determined at a national level.

4 Map Display and Direction Analyses

The aforementioned statement can be illustrated by a map display of a two-dimensional regression. The Darcy program can be used in the application (http://www.spatial-modelling.info/Darcy-2-module-decomparaison).

The square grid attached onto the shape-dependent coordinate system and its interpolated modified position further generalizes the information received from the participating points.

The arrows in Figures 2 and 3 show the directions of the shifts, while the darker shades illustrates the type of distortion. The darker zones express the divergent forces of the area, which are considered to be the most important gravitational displacements.

The data in Table 2 shows that the space shaped by the gravity model causes only a slight distortion compared to the geographic space. The magnitude of vertical and horizontal displacements increased slightly in 2010.

Practically, the maps produced by Darcy software verify this (Figure 2 and Figure 3). It can be seen that the capital of Hungary is Hungary's main centre of gravity, the centre towards which the largest power is attracted. The regional centres like Győr, Pécs, Szeged, Debrecen are also gravity nodes. The national role of regional centres is weak. In the area of Budapest a gravity fault line emerges.

The reason for this phenomenon is that the Hungarian capital attracts all the micro-regions, while very weak forces are applied to Budapest compared to its mass.

The map also illustrates that the regular force fields are the major transport corridors, namely they are slightly distorted due to highways. Between 2000 and 2010, the significance of green-marked gravitational nodes increased. The comparison of the two maps clearly shows the intensification of regional differences.

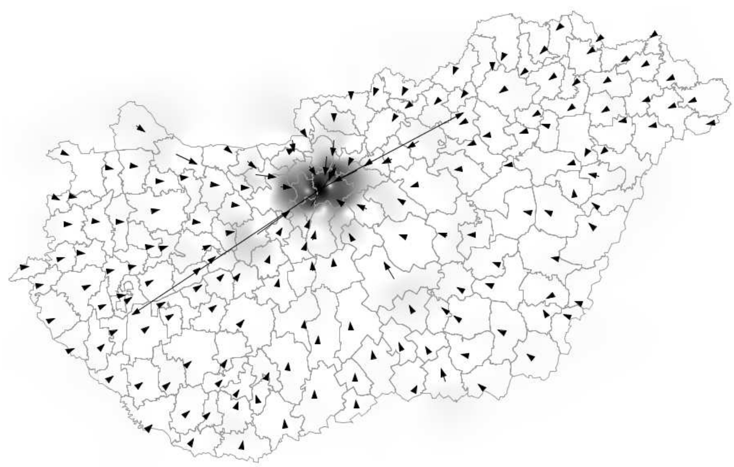


Figure 2

Results in 2000

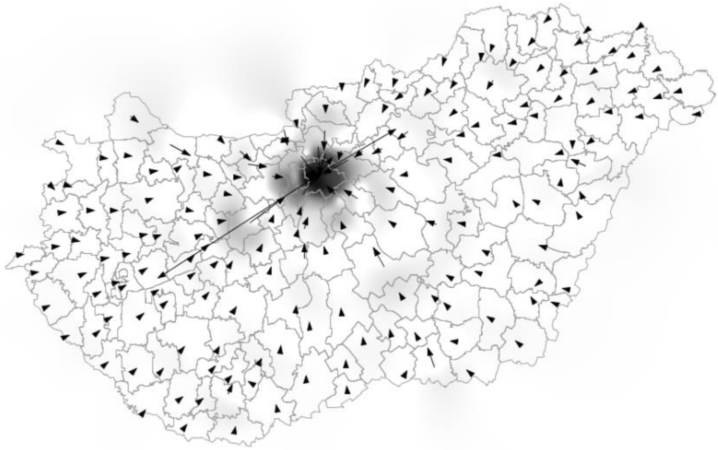


Figure 3
Results in 2010

Based on the obtained results, micro-regions can be grouped by the direction of force applied to them by other micro-regions. Four groups can be formed. They are as follows:

South and West,

North and West,

North and East,

South and East.

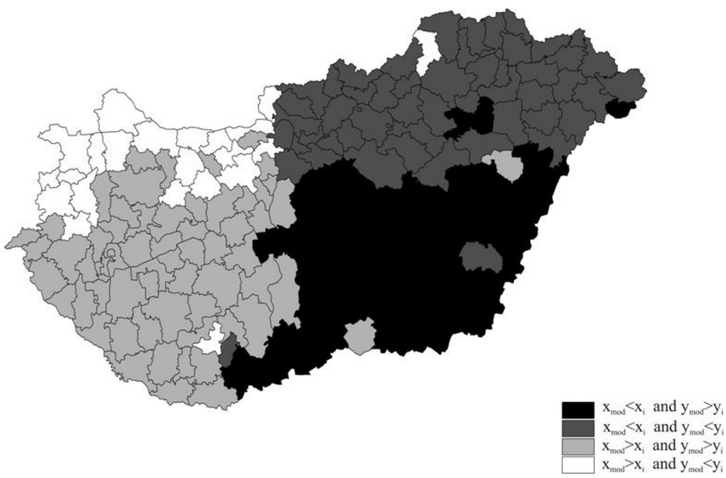


Figure 4
Results in 2000

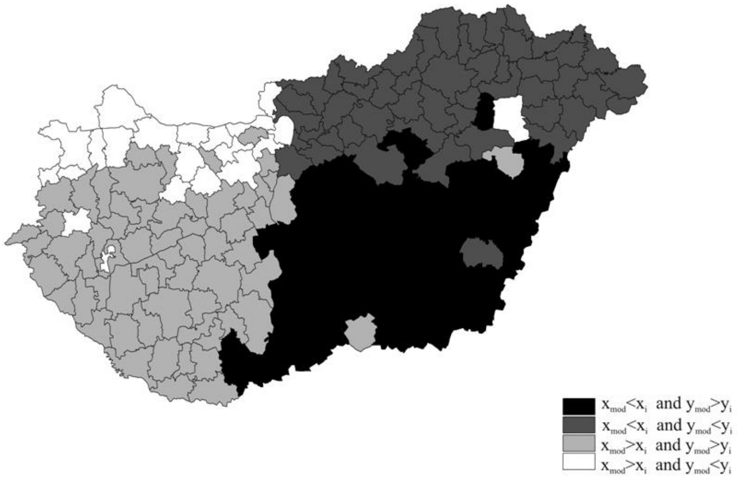


Figure 5
Results in 2010

All micro-regions can be placed in one of the aforementioned groups. The results are shown in Figure 4 and Figure 5. On the maps the North-South segmentation appears more significantly than the East-West one. This statement is especially true in Western Hungary where the developed micro-regions of Gyor-Moson-Sopron and Komárom-Esztergom Counties attract other micro- regions of Transdanubia. The effect of the east-west gradient can hardly be demonstrated because of the central location of Budapest and its strong impact on the whole country.

The impact of regional centres is clearly seen in the areas where the direction of forces differs from its environment, i.e., in the neighbouring areas where colours are different from the ones in their environment. According to the results of 2000 and 2010 years, stable local centres can be seen in the micro-regions of Debrecen, Miskolc, Nyíregyháza, Szeged and Pécs.

In the advanced territories of northwestern Hungary a similar phenomenon, though less distinctive, is seen when micro-regions differentiate from their environment, since several micro-regional groups tend to be similar in character in these areas.

5 Magnitude of Force Per Unit Mass

In a micro-region other regions apply different forces. However, the same force strength fails to result the same impact due to the difference of masses. It is possible to calculate the forces acting on a unit of internal mass by the Equation 13:

$$F = \frac{\sqrt{D_i X^2 + D_i Y^2}}{w_i} \qquad (13)$$

Where F is the force per unit mass, D_i is the vertical and horizontal forces and w_i is the own weight of point i.

As it is shown in Figure 6, the most significant forces compared to their own mass are the micro-regions located in Budapest agglomeration, especially the Budaörs micro-region. Outside the agglomeration, Tata micro-region shows an outstanding magnitude. The micro-regions that represent higher than the average value are mainly located further from Budapest, generally along motorways. The long arrows around Budapest were formed due to the impact of the motorways.

Figure 6 shows the changes between 2000 and 2010. The most significant changes are the outcomes of the construction of motorways. A perfect example for this is Tiszavasvári and Komló micro-regions, which are located near M3 and M6 motorways. Somewhat different examples are the famous Sarkadi and Sárvár micro-regions. In their cases, the decrease of their own weight caused a specific power growth. There is a micro-region, which emerged primarily due to the population growth in its environment and to its gravitational force growth. It is Pilisvörösvár micro-region.

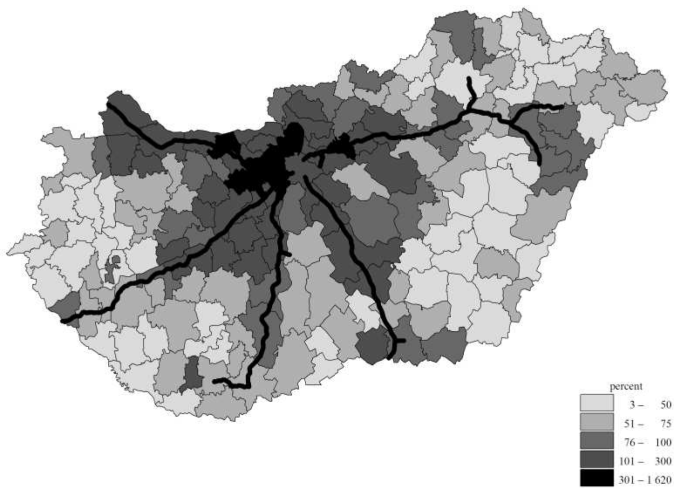


Figure 6

Forces per unit mass (population) as a percentage of the national average in 2010

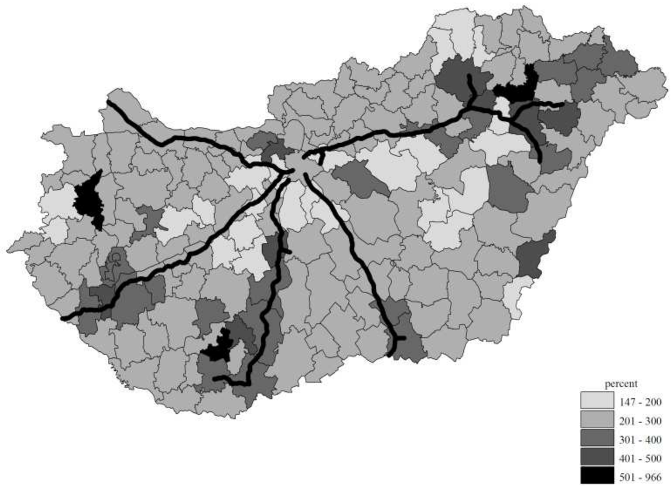


Figure 7
Changes in forces per unit mass (population) in 2000 and in 2010

Conclusions

In our study we made an attempt to introduce the potential and unexplored areas of gravity models and problems of their interpretation by expanding and extending the methodology. The forces applied were illustrated by using the income tax base as weights in micro-regions of Hungary. On the basis of the model, the result in line with the experience illustrated that Budapest has no counterweight in Hungary and the local central areas are weak. However, the presence of stable local centres is detectable in Debrecen, Miskolc, Nyíregyháza, Szeged and Pécs micro-regions.

It is important to note that compared to their weight, micro-regions affected by the most significant forces are located in the broader vicinity of Budapest, mainly along highways. Most significant changes compared to unit mass are caused by highways, but in many cases the population decline in micro-regions is also a determining factor.

Acknowledgement

This work was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

References

[1] Budó, Á.: Experimental Physics I. National Textbook Press, Budapest, 2004

[2] Bhattacherjee, A.: Social Science Research: Principles, Methods, and Practices. USF Tampa Bay Open Access Textbooks Collection. Book 3, 2012

[3] Converse P. D. New Laws of Retail Gravitation. Journal of Marketing, Vol. 14, Issue 3, 1949, pp. 379-384

[4] Dodd S. G.: The Interactance Hypothesis. A Gravity Model Fitting Physical Masses and Human Groups. American Sociological Review, Vol. 15, Issue 2, 1950, pp. 245-256

[5] Dusek, T. :A gravitációs modell és a gravitációs törvény összehasonlítása. Tér és Társadalom, Vol. 17, Issue 1, 2003, pp. 41-57

[6] Dusek. T.: Kétdimenziós regresszió a területi kutatásokban. Területi Statisztika, Vol. 51, Issue 1, 2011, pp. 11-22

[7] Filippo, S.,Maritan, A., Néda, Z: Continuum Approach for a Class of Mobility Models, [physics.soc-ph] 19 Jun 2012, http://arxiv.org/pdf/1206.4359.pdf

[8] Friedman, A., Kohler, B.: Bidimensional Regression: Assessing the Configural Similarity and Accuracy of Cognitive Maps and Other Two-Dimensional Data Sets. Psychological Methods, Vol. 8, Issue 4, 2003, pp. 468-491

[9] Geurs, K. T, van Wee, B.: Accessibility Evaluation of Land-Use and Transport Strategies: Review and Research Directions. Journal of Transport Geography, Vol. 12, Issue 2, 2004, pp. 127-140

[10] Gutierrez, J.: Location, Economic Potential and Daily Accessibility: an Analysis of the Accessibility Impact of the High-Speed Line Madrid–Barcelona–French border. Journal of Transport Geography, Vol. 9, Issue 4, 2001, pp. 229-242

[11] Rodrigue, J. P.: The Geography of Transport Systems. Routledge, New York, 2013

[12] Kare, S., Samal, A., Marx, D.: Using Bidimensional Regression to Assess Face Similarity. Machine Vision and Applications Vol. 21, Issue 3, 2010, pp. 261-274

[13] Kincses, Á., Tóth, G.: Geometry of Potential Modells. Regional Statistics Vol. 52, Special Issue, 2012, pp. 74-89

[14] M. Barthélemy: Spatial Network. Physics Reports, Vol. 499, 2011, pp. 1-101, http://arxiv.org/abs/1010.0302

[15] Vieira, M. S., Vasconcelos, G., Nagel, S.: Dynamics of Spring-Block Models: Tuning to Criticality Phys. Rev. E., Vol. 47, Issue 4, 1993, R2221-R2224

[16] Martinez-Zarzoso, I., Nowak-Lehmann, F.: Augmented Gravity Model. An Empirical Application to Mercosur-European Union trade flows. Journal of Applied Economics, Vol. 6, Issue 2, 2003, pp. 291-316

[17] Mate, G., Neda, Z., Benedek, J.: Spring-Block Model Reveals Region-like Structures. PLoS ONE, Vol. 6, Issue 2, 2011, e16518, doi:10.1371/journal.pone.0016518

[18] Nakaya, T.: Statistical Inferences in Bidimensional Regression Models. Geographical Analysis, Vol. 29, Issue 2, 2010, pp. 169-186

[19] Reilly, W. J.: Methods for the Study of Retail Relationships. University of Texas Bulletin, No. 2944, 1929

[20] Schürmann, C.–Spiekermann, K.–Wegener, M.: Accessibility Indicators. Berichte aus dem Institüt für Raumplanung 39, Dortmund, IRPUD, 1997

[21] Stewart J. Q.: Demographic Gravitation: Evidence and Applications. Sociometry, Vol. 11, Issues 1-2, 1948, pp. 31-58

[22] Symington, A., Charlton, M. E., Brunsdon, C. F.: Using Bidimensional Regression to Explore Map Lineage Computers, Environment and Urban Systems, Vol. 26, Issue 2, 2002, pp. 201-218

[23] Tobler, W. R.: Map Transformations of Geographic Space. PhD dissertation, Seattle, University of Washington, 1961

[24] Tobler, W. R. Computation of the Correspondence of Geographical Patterns. Papers of the Regional Science Association, Vol. 15, Issue 1, 1965, pp. 131-139

[25] Tobler, W. R.: A Computer Model Simulating Urban Growth in the Detroit Region. Economic Geography, Vol. 46, Issue 2, 1970, pp. 234-240

[26] Tobler, W. R.: Comparisons of Plane Forms. Geographical Analysis, Vol. 10, Issue 2, 1978, pp. 154-162

[27] Tobler, W. R.: Bidimensional Regression. Geographical Analysis, Vol. 26, Issue 3, 1994, pp. 187-212

[28] Tobler, W. R.: On the First Law of Geography: A Reply Annals of the Association of American Geographers, Vol. 94, Issue 2, 2004, pp. 304-310

[29] Tschopp, M., Fröhlich, P., Keller, P., Axhausen, K. W.: Accessibility, Spatial Organisation and Demography in Switzerland through 1920 to 2000: First Results. 3rd Swiss Transport Research Conference Monte Verità / Ascona, March 19-21. 2003

[30] Willigers, J., Floor, H., van Wee, B.: Accessibility Indicators for Locational Choice of Offices: An Application to the Intraregional Distributive Effects of High-Speed Rail in the Netherlands. Environment Planning. A Vol. 39, Issue 9, 2007, pp. 2086-2098

[31] Krings, G., Calabrese, F., Ratti, C., Blondel, V. D.: Urban Gravity: a Model for Inter-City Telecommunication Flows. Journal of Statistical Mechanics: Theory and Experiment, Vol. 9, Issue 7, 2009, pp. 1-8

Applications of Signatures to Expert Systems Modelling

Claudiu Pozna[1,2], Radu-Emil Precup[3]

[1]Department of Informatics, Széchenyi István University, Egyetem tér 1, H-9026 Győr, Hungary

[2]Department of Automation and Information Technology, Transilvania University of Brasov, Str. Mihai Viteazu 5, RO-500174 Brasov, Romania
E-mail: cp@unitbv.ro

[3]Department of Automation and Applied Informatics, Politehnica University of Timisoara, Bd. V. Parvan 2, RO-300223 Timisoara, Romania
E-mail: radu.precup@aut.upt.ro

Abstract: This paper offers a new approach that applies the signatures to expert systems modelling. Signatures and their operators, viewed as a generalization of fuzzy signatures, represent a convenient framework for the symbolic representation of data. The models are derived by a three-step algorithm that maps the signatures onto expert systems. An expert systems modelling algorithm is given. Our algorithm has two inputs, the knowledge base, i.e., the rules, and the data base, i.e., the facts, and it constructs the signatures which represent models of expert systems. The algorithm is advantageous because of its systematic and general formulation allowing for the modelling of uncertain expert systems. The theoretical results are exemplified by a case study which produces models of a Bayesian expert system with mechatronics applications.

Keywords: expert systems; mechatronics; modelling; operators; rules; signatures

1 Introduction

Expert systems are products of artificial intelligence that aim the goal of problem solving. The problems belong to various domains, and expert level knowledge is used with this regard. Successful implementations of expert systems include evolving classification and fuzzy systems [1, 2], multi-agent systems [3, 4], data mining [5, 6], and meta-heuristic engines [7-9]. As mentioned in [10, 11], the better understanding of how natural cognitive processes can effectively co-evolve with processes in artificially cognitive systems in the framework of Cognitive Infocommunications (CogInfoCom) [12-14]. Rule-based expert systems will be addressed in this paper as a step in the context of cognitive processes modelling.

Several approaches to expert systems modelling are currently reported in the literature. Multi-expert models are discussed in [15]. The combination of fuzzy logic, neural networks, genetic algorithms and statistic analysis is analysed in [16]. Reliable rule-based systems with uncertainty are obtained in [17] on the basis of semantic data integration. Model-driven engineering techniques for the development of multi-agent systems are proposed in [18]. A comparison of structure learning approaches to structure learning for belief rule base expert system is conducted in [19].

The new idea of this paper in the context of the above analysed literature is a new approach to expert systems modelling based on signatures. Signatures and their operators are defined in [20] as a convenient framework for the symbolic representation of data as a generalization of fuzzy signatures [21, 22]. Applications of signatures are given in [10, 23, 24].

The new modelling approach is formulated as an original three-step algorithm that maps the signatures onto expert systems. The algorithm has two inputs represented by the knowledge base which are the rules, and by the data base which represents the facts. The algorithm constructs the signatures which stand for a class of expert systems models.

Our expert systems modelling approach is important and advantageous with respect to the state-of-the-art because of the following reasons:

- The systematic formulation in terms of an algorithm offers transparency and relatively simple modelling.
- The formulation of the algorithm is general and applicable to both certain and uncertain expert systems.

This paper is organized as follows: a short overview on signatures and on their operators is presented in the next section. Section 3 is dedicated to the new modelling approach. An illustrative example is included and the modelling algorithm is given. Section 4 validates the theoretical approach by a case study focused on the construction of signature-based models of a Bayesian expert system defined in [25]. The conclusions are finally outlined.

2 Overview on Signatures and on Their Operators

Signatures and operators on signatures are defined and analysed in [20]. A part of the definitions which enable the modelling of expert systems is presented in this section, and the reader is invited to address [20] for examples to understand the concept of signatures. Let $S^{(n)}$ be a set defined recursively as

$$S^{(n)} = \prod_{i=1}^{n} S_i,\ S_i = \mathbf{R} \text{ or } S_i = S^{(m)},\ m \geq 1,\ i = 1...n, \tag{1}$$

where $\mathbf{R}$ is the set of real numbers, and $\prod$ is the Cartesian product.

Definition 1. Let X be a nonempty set. The *collection of signatures* is defined as the function $A: X \to S^{(n)}$, the *signature* of the element $x \in X$ is $A(x) \in S^{(n)}$ and the *transposition* of the signature $A(x)$ is represented by $A^T(x)$ given as

$$A(x) = \begin{bmatrix} \cdots \\ a_i \\ \begin{bmatrix} a_{i+1,1} \\ a_{i+1,2} \end{bmatrix} \\ \begin{bmatrix} a_{i+2,1} \\ \begin{bmatrix} a_{i+2,2,1} \\ a_{i+2,2,2} \end{bmatrix} \end{bmatrix} \\ \cdots \end{bmatrix}, \tag{2}$$

$$A^T(x) = [\ldots \quad a_i \quad [a_{i+1,1} \quad a_{i+1,2}] \quad [a_{i+2,1} \quad [a_{i+2,2,1} \quad a_{i+2,2,2}]] \quad \ldots].$$

The following notations are introduced in [20] to simplify the characterization of signatures:

- A signature $A(x)$ with values $a_1, a_2, ..., a_n, a_{i,1}, a_{i,2}, ..., a_{i,m}, ...a_{j,k,l}, ...$, is indicated by $a^{\cdots}$.
- $A(x) = a^{1,\ldots,n}$ is used if $\exists x \in X$ and $A^T(x) = [a_1 \quad \ldots \quad a_n]$.
- If $\exists y \in Y$ and $A^T(y) = [a_1 \quad \ldots \quad a_{i-1} \quad [a_{i,1} \quad \ldots \quad a_{i,m}] \quad a_{i+1} \quad \ldots \quad a_n]$, then we will use the notation $A(y) = a^{1,\ldots,[1,\ldots,m]_i,\ldots,n}$. The sets are defined here as $S_1 = S_2 = \ldots = S_{i-1} = S_{i+1} = \ldots = S_n = \mathbf{R}$, and their Cartesian product is expressed as $S_i = \prod_{i=1}^{m} \mathbf{R} = \mathbf{R}^m$.
- A signature of type $[\ldots \quad [[a_1]] \quad \ldots]$ is equivalent to the signature $[a_1]$, where $a_1 \in \mathbf{R}$.

As shown in [20], signatures can be used in complex data representation. Some definitions of operators on signatures will be exemplified in the sequel.

Definition 2. The *contraction* of a signature is defined as one of the three functions

$$
\begin{aligned}
&{}^{f}@ : \mathrm{S} \to \mathrm{S},\ {}^{f}@(a^{1,2,\dots,n}) = a^{1} = [a],\\
&{}^{f}@_{i} : \mathrm{S} \to \mathrm{S},\\
&\begin{cases} {}^{f}@_{i}(a^{1,\dots[1,\dots,m]_i,\dots,n}) = a^{1,\dots,i,\dots,n}, & \text{if } i \le n,\\ {}^{f}@_{i}(a^{\cdots}) = a^{\cdots}, & \text{otherwise,} \end{cases}\\
&{}^{f}@_{\mathrm{i,j,k}} : \mathrm{S} \to \mathrm{S},\\
&\begin{cases} {}^{f}@_{\mathrm{i,j,k}}(a^{1,\dots[1,\dots[1,\dots[1,\dots q]_k,\dots,r]_j,\dots,m]_i,\dots,n}) = a^{1,\dots[1,\dots[1,\dots,k,\dots,r]_j,\dots,m]_i,\dots,n}, & \text{if } i \le n, j \le m, k \le r,\\ {}^{f}@_{i,j,k}(a^{\cdots}) = a^{\cdots}, & \text{otherwise,} \end{cases}
\end{aligned}
\tag{3}
$$

where $a = f(a_1, a_2, \dots, a_n)$, $f : \mathbf{R}^n \to \mathbf{R}$, $a_i = f(a_{i1}, \dots, a_{im})$, $f : \mathbf{R}^m \to \mathbf{R}$, and $a_{i,j,k} = f(a_{i,j,k1}, \dots, a_{i,j,kq})$, $f : \mathbf{R}^q \to \mathbf{R}$. Extra indices can be inserted after *i*, *j* and *k*, to generalize this definition. We use the following notation for the absolute value of a contraction if a has the first form in (3):

$$
|{}^{f}@(a^{1,2,\dots,n})| = a. \tag{4}
$$

Definition 3. The *extension* of a signature is defined as one of the functions

$$
\begin{aligned}
&{}^{g}\overline{@}_{\mathrm{i}}^{\mathrm{p}} : \mathrm{S} \to \mathrm{S},\\
&\begin{cases} {}^{g}\overline{@}_{i}^{p}(a^{1,2,\dots,i,\dots,n}) = a^{1,2,\dots[1,\dots,p]_i,\dots,n}, & \text{if } i \le n,\\ {}^{g}\overline{@}_{i}^{p}(a^{\cdots}) = a^{\cdots}, & \text{otherwise,} \end{cases}\\
&{}^{g}\overline{@}_{\mathrm{i,j,k}}^{\mathrm{q}} : \mathrm{S} \to \mathrm{S},\\
&\begin{cases} {}^{g}\overline{@}_{i,j,k}^{q}(a^{1,\dots[1,\dots[1,\dots,k,\dots,r]_j,\dots,m]_i,\dots,n}) = a^{1,\dots[1,\dots[1,\dots[1,\dots q]_k,\dots,r]_j,\dots,m]_i,\dots,n}, & \text{if } i \le n, j \le m, k \le r,\\ {}^{g}\overline{@}_{i,j,k}^{q}(a^{\cdots}) = a^{\cdots}, & \text{otherwise,} \end{cases}
\end{aligned}
\tag{5}
$$

where two forms of the function *g* are used, $g : \mathbf{R} \to \mathbf{R}^p$, $g(a_i) = [a_{i1}, \dots, a_{ip}]$, $g : \mathbf{R} \to \mathbf{R}^q$, $g(a_{i,j,k}) = [a_{i,j,k1}, \dots, a_{i,j,kq}]$, and we can continue to generalize it by adding extra indices after *i*, *j* and *k*. The *zero-step extension* of a signature is defined as the function

$$
\begin{aligned}
&{}^{g}\overline{@}^{\mathrm{p}} : \mathrm{S} \to \mathrm{S},\\
&{}^{g}\overline{@}^{p}(a^{1,2,\dots,n}) = {}^{g}\overline{@}_{1}^{p}({}^{g}\overline{@}_{2}^{p}(\dots({}^{g}\overline{@}_{n}^{p}(a^{1,2,\dots,n})))) = a^{[1,\dots,p],[1,\dots,p],\dots[1,\dots,p]},
\end{aligned}
\tag{6}
$$

where $g : \mathbf{R} \to \mathbf{R}^p$, $g(a_j) = [a_{j1}, \dots, a_{jp}]$, $j = 1 \dots n$.

Definition 4. The *pruning* of a signature is defined as one of the three functions

$$
\begin{aligned}
&\varnothing_i : \mathrm{S} \to \mathrm{S},\\
&\begin{cases} \varnothing_i(a^{1,2,\ldots i,\ldots n}) = a^{1,2,\ldots i-1,i+1,\ldots n}, & \text{if } i \le n,\\ \qquad \varnothing_i(a^{\cdots}) = a^{\cdots}, & \text{otherwise,} \end{cases}\\
&\varnothing_i : \mathrm{S} \to \mathrm{S},\\
&\begin{cases} \varnothing_i(a^{1,\ldots[1,\ldots m]_i,\ldots n}) = a^{1,\ldots i-1,i+1,\ldots n}, & \text{if } i \le n,\\ \qquad \varnothing_i(a^{\cdots}) = a^{\cdots}, & \text{otherwise,} \end{cases}\\
&\varnothing_{i,j,k} : \mathrm{S} \to \mathrm{S},\\
&\begin{cases} \varnothing_{i,j,k}(a^{1,\ldots[1,\ldots[1,\ldots[1,\ldots q]_k,\ldots,\ldots r]_j,\ldots m]_i,\ldots n}) & \\ = a^{1,\ldots[1,\ldots[1,\ldots k-1,k+1,\ldots r]_j,\ldots m]_i,\ldots n}, & \text{if } i \le n,\ j \le m, k \le r,\\ \qquad \varnothing_{i,j,k}(a^{\cdots}) = a^{\cdots}, & \text{otherwise.} \end{cases}
\end{aligned}
\tag{7}
$$

This definition can be generalized as well by adding extra indices after *i*, *j* and *k*, where $a^{\cdots}\varnothing_i = b^{\cdots}$ is the notation for the pruning of a signature.

Definition 5. The *addition* of two signatures is defined as the function

$$
\begin{aligned}
&{}^f\oplus_i : \mathrm{S}\times\mathrm{S} \to \mathrm{S},\\
&{}^f\oplus_i(a^{1,2,\ldots[1,\ldots m]_i,\ldots n}, b^{1,2,\ldots m}) = c^{1,2,\ldots[1,\ldots m]_i,\ldots n},
\end{aligned}
\tag{8}
$$

where $c_k = a_k \ \forall\, k \ne i,\ c_{ij} = f(a_{ij}, b_j)\ \forall\, j = 1 \ldots m,\ f : \mathbf{R}^2 \to \mathbf{R}$. If $a^{1,2,\ldots n},\ b^{1,2,\ldots m} \in \mathrm{S}$, then

$$
\begin{cases} a^{1,2,\ldots n}\, {}^f\oplus_i b^{1,2,\ldots m} = {}^g\overline{@}_i^m (a^{1,2,\ldots n})\, {}^f\oplus_i b^{1,2,\ldots m} & \\ = c^{1,2,\ldots[1,2,\ldots m]_i,\ldots n}, & \text{if } i \le n,\\ \qquad a^{\cdots}\, {}^f\oplus_i b^{\cdots} = a^{\cdots}, & \text{otherwise,} \end{cases}
\tag{9}
$$

where $c_{ij} = f(a_{ij}, b_j),\ [a_{i1}, \ldots, a_{im}] = g(a_i),\ j = 1 \ldots m,\ c_k = a_k\ \forall\, k \ne i$. We can generalize this definition by adding extra indices after *i*, where $a^{\cdots}\, {}^f\oplus_i b^{\cdots} = c^{\cdots}$ is the notation for the addition of two signatures.

All these operators presented above refer to structural transformations and data transformations. Several examples concerning the application of these operators and of other operators (e.g., grafting and multiplication) are presented in [20]. These operators can be conveniently implemented as software objects. The extension to inequalities is also of interest [26-28] with focus on mechatronics applications [29-36].

3 Modelling Approach

The structure of a rule-based expert system is presented in Figure 1 (a) which points out the following subsystems: the knowledge base which contains the "If...Then..." rules, the database which contains the facts, the inference engine where the goal of the expert system is computed, and the user interface where the user interacts with the expert system. Several internal elements can be added to this structure; they include explanation facilities where the results are explained systematically, and the developer interface where the expert system interacts with the developer. External elements can be included as well like external databases or programs which support the inference engine.

The core of the expert system is the inference engine, where the rules are fired using the known facts. After firing a rule a new fact is inferred; this can fire in turn a new rule. This process is cyclic, and it can be represented by the schema illustrated in Figure 1 (b). The end of the cycle is obtained when no more rules can be fired and the knowledge on the goal is obtained.

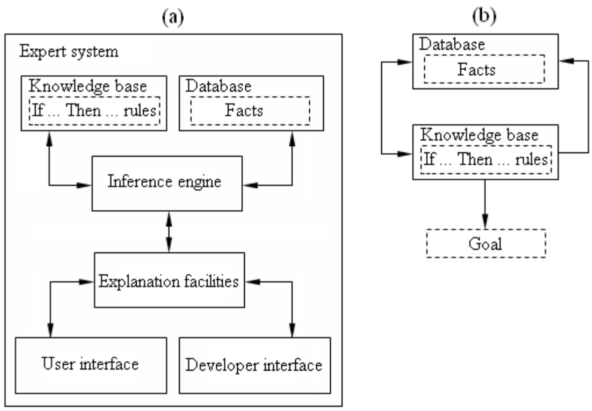


Figure 1
Structure of rule-based expert system (a), and cycles of inference engine (b)

The cycle presented in Figure 1 (b) suggests that the rules are interconnected in an inference chain which starts with facts and ends with the problem goal. Two ways are rules to execute the rules, forward chaining and backward chaining. The forward chaining is data-driven reasoning expressed as: the reasoning starts with the known data, each time only the topmost rule is executed, when fired the rule adds a new fact in the database, and the cycle stops when no further rules can be fired. Contrarily, the backward chaining is goal-driven reasoning, where: the knowledge base is first searched to find the rules that might have the desired solution, and if this rule contains facts which are not observed then these facts are next replaced with related rules until all rules from the decision chain contain observable facts.

Signatures will be used as follows as data structures to model the processes specific to expert systems. Our idea is to transform an inference chain into a signature and thus to map signatures onto expert systems. The inference chains can be different because of the observed facts. This means that we can divide the operating processes of expert systems in two steps:

- I. First, carry out the backward construction of the signature starting with the goal and replacing the unobserved facts with rules until all rules contains observed facts.
- II. Second, apply certain operators to the already constructed signature and compute the goal of the expert system.

Each "*If...Then...*" rule contains a relation between two types of facts. The first part of rule, "*If...*", concerns the antecedent facts, and the second part of the rule, "*Then...*", is related to the consequent facts. The terms "observed / unobserved facts" mentioned several times in the previous section refer to the antecedent facts. If such a fact is not observed the rule cannot be fired. In this case the only solution is to find another rule (from the knowledge base) which contains the unobserved fact like a consequent fact.

The construction of the inference chain is exemplified by means of the following example. If the goal involves *Z* and the knowledge base contain the rule

$$\begin{aligned} &\text{Rule } 1: \mathit{If}\ Y\ \mathit{Then}\ Z, \\ &\text{Rule } 2: \mathit{If}\ X\ \mathit{Then}\ Y, \end{aligned} \tag{10}$$

we will start (backward) with the rule 1. If *Y* is observed, then rule 1 will be fired and the knowledge on *Z* will be obtained. If not, *Y* will be replaced by rule 2:

$$\mathit{If}\ (\mathit{If}\ X\ \mathit{Then}\ Y)\ \mathit{Then}\ Z, \tag{11}$$

and we will ask about the observability of the fact *X*.

We propose the following typology of facts: observable facts: the facts which can be inputted to the expert system, unobservable facts: facts which are not observable, inferable facts: facts which are established by rules, and uninferable facts: facts which are not inferable. An uninferable fact must be observable and that the inferable facts can be observable or unobservable. Observable facts are not observed in certain situations. An observed or inferred fact can be modelled with the value 1; contrarily, if the fact is not observed or inferred, it can be modelled with the value 0.

Since a rule can de modelled by a signature, the construction of the inference chains can modelled by operators on signatures as well. In this regard, two definitions are suggested as follows. The following definition concerns a rule which is a dependency between two types of facts, viz. the antecedents and the consequences.

Definition 8. A rule is modelled by signatures in one of the three forms

$$\begin{aligned}
&\text{Rule } r_1 : \textit{If } A \textit{ Then } Z \text{ is equivalent to } r_1 = [A]^T \text{ and } @(r_1) = [Z]^T, \\
&\text{Rule } r_1 : \textit{If } A \wedge B \wedge C \textit{ Then } Z \text{ is equivalent to} \\
&\qquad r_1 = [A, B, C]^T \text{ and } {}^f@(r_1) = [Z]^T, \\
&\text{Rule } r_1 : \textit{If } A \textit{ Then } Z \vee \text{Rule } r_2 : \textit{If } B \textit{ Then } Z \text{ is equivalent to} \\
&\qquad r = [A, B]^T \text{ and } {}^g@(r) = [Z]^T,
\end{aligned} \tag{12}$$

where $\wedge$ stands for the conjunction, $\vee$ stands for the disjunction, and f and g are functions related to the conjunction and to the disjunction, respectively:

$$\begin{aligned}
&f : \{0,1\}^n \to \{0,1\},\ f(x_1, ..., x_n) = \min_{i=1...n} x_i, \\
&g : \{0,1\}^n \to \{0,1\},\ g(x_1, ..., x_n) = \max_{i=1...n} x_i.
\end{aligned} \tag{13}$$

The addition of signatures replaces an antecedent fact with a rule in terms of the following definition.

Definition 7. Let the two rules be

$$\begin{aligned}
&\text{Rule } r_1 : \textit{If } X \wedge Y \textit{ Then } Z \text{ is equivalent to} \\
&\qquad r_1 = [X, Y]^T \text{ and } {}^f@(r_1) = [Z]^T, \\
&\text{Rule } r_2 : \textit{If } V \wedge W \textit{ Then } X \text{ is equivalent to} \\
&\qquad r_2 = [V, W]^T \text{ and } {}^f@(r_2) = [X]^T.
\end{aligned} \tag{14}$$

The replacement of an antecedent fact with these two rules is modelled as

$$r = r_1 \oplus_1 r_2 = [[V, W], Y]^T, \tag{15}$$

where r is obtained as an inference of r_1 and r_2.

The application of Definitions 6 and 7 to expert systems modelling is exemplified as follows by a short example. Let us suppose that five rules are defined as

$$\begin{aligned}
&\text{Rule } r_1 : \textit{If } F_1 \vee F_2 \textit{ Then } Z, \\
&\text{Rule } r_2 : \textit{If } C \wedge D \wedge E \textit{ Then } F_1, \\
&\text{Rule } r_3 : \textit{If } A \wedge B \textit{ Then } \text{C}, \\
&\text{Rule } r_4 : \textit{If } H \wedge I \textit{ Then } F_2, \\
&\text{Rule } r_5 : \textit{If } J \wedge K \wedge L \textit{ Then } I.
\end{aligned} \tag{16}$$

These rules can be represented as the inference chain presented in Figure 2 (a).

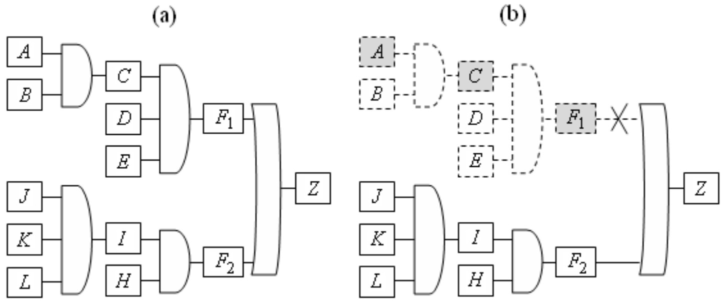


Figure 2
Inference diagram (a), which illustrates the inference chain corresponding to (30), and pruning the diagram (b)

Using the previous proposed typology it can be concluded that *A*, *B*, *J*, *K*, *L*, *D*, *E* and *H* are not inferable facts so they are observable, and that *C*, *I*, *H*, F_1 and F_2 are inferable facts. This aspect does not indicate that *C*, *I*, *H*, F_1 or F_2 are not observable. Definitions 6 and 7 lead to the following step-by-step construction of a signature:

$$\begin{aligned} &r_1=[F_1,F_2]^T,\ {}^g@(r_1)=[Z]^T,\ r_2=[C,D,E]^T,\ {}^f@(r_2)=[F_1]^T,\\ &r_3=[A,B]^T,\ {}^f@(r_3)=[C]^T,\ r_4=[I,H]^T,\ {}^f@(r_4)=[F_2]^T,\\ &r_5=[J,K,L]^T,\ {}^f@(r_5)=[I]^T,\ r=r_1\oplus_1 r_2\oplus_{1,1} r_3\oplus_2 r_4\oplus_{2,1} r_5\\ &=[[[A,B],D,E],[[J,K,L],H]]]^T. \end{aligned} \tag{17}$$

Inferring the expert system output means

$$[Z]^T = @(r) = {}^g@({}^f@_1({}^f@_{1,1}({}^f@_2({}^f@_{2,1}(r))))), \tag{18}$$

where *f* and *g* indicate the conjunction and the disjunction, respectively.

Since in this case the function *g* must output not only the degree of truth but also the rule to be applied (F_1, F_2 or Φ), the following function is proposed:

$$g:\{0,1\}^2\to\{F_1,F_2,\Phi\},$$
$$g(x_1,x_2)=\begin{cases}F_i, & \text{if} \quad x_i=1,\ x_j=0,\\ \Phi, & \text{if} \quad x_i=0,\ x_j=0,\\ F_i, & \text{if} \quad x_i=1,\ x_j=1,\ p(i)>p(j),\end{cases} \tag{19}$$

where $p(i)$ is the priority of rule r_i, and Φ indicates no rules to apply.

The signature presented in (17) is constructed accepting the assumption that *A*, *B*, *J*, *K*, *L*, *D*, *E* and *H* are observed. If this assumption is not accepted, i.e., at least one of the observable facts is not observed, the signature is simplified. The pruning operation can be used in order to simplify the signatures. In such situations since the conjunction operator $\wedge$ links all facts, we can erase all branches until we reach a disjunction $\vee$. Figure 2 (b) points out that A, C and F_1 are not observed, and this situation is modelled by the pruning operation:

$$r = \varnothing_1(r) = [[J, K, L], H]]^T . \tag{20}$$

If C is observable and it is observed the rule r_3 is no more needed, the signature is transformed into

$$r = r_1 \oplus_1 r_2 \oplus_2 r_4 \oplus_{2,1} r_5 = [[C, D, E], [[J, K, L], H]]]^T . \tag{21}$$

A conflict management strategy should be considered if F_1, F_2, and F_3 are inferred (observed) and each one is a possible output of the expert system: The conflict management strategy should be focused on establishing a priority order. Our approach to expert systems modelling based on signatures can manage easily this situation by including the priority order in the contraction law g, where g is defined in (19).

Our modelling approach is supported by a three-step algorithm. The first step concerns the construction of the signature of the inference engine, the second step computes the inference result represented by the expert system output, and the third step ensures the iteration. As suggested in Figure 1 (a), the algorithm uses two inputs, the knowledge base (the rules) and the database (the facts). The expert systems modelling algorithm consists of the following steps:

Step 1. Select from the knowledge base those rules which are related to the expert system goal, use equation (12) to construct the signature, and memorize the contraction law of the signature.

Step 2. Develop the signature by the one-by-one investigation of the facts contained in the signature:

- If the fact is unobservable, select all rules from the knowledge base which refer this fact as a consequence, replace them using equation (15), and memorize the contraction law,
- If the fact is observable, search the database to find out if the fact has actually been observed:
    - If yes, replace it with the observed value,
    - If not, search the database to find the rules which refer this fact as a consequence:
        - If a rule is found, replace the fact with the rule using equation (15), and memorize the contraction law,
        - If a rule is not found, prune the signature from this fact (leaf) to the first branch which supposes a disjunctive contraction.

Step 3. Continue with step 2 until all facts of the signature are replaced with data.

The flowchart of the algorithm is presented in Figure 3. This algorithm can be simplified if the rules which contain unobservable facts are identified. The idea is to compute a priori composed rules. Equation (14) is employed in such cases in order to generate a signature which can be used directly at step 2. The application of our algorithm is exemplified in the next section.

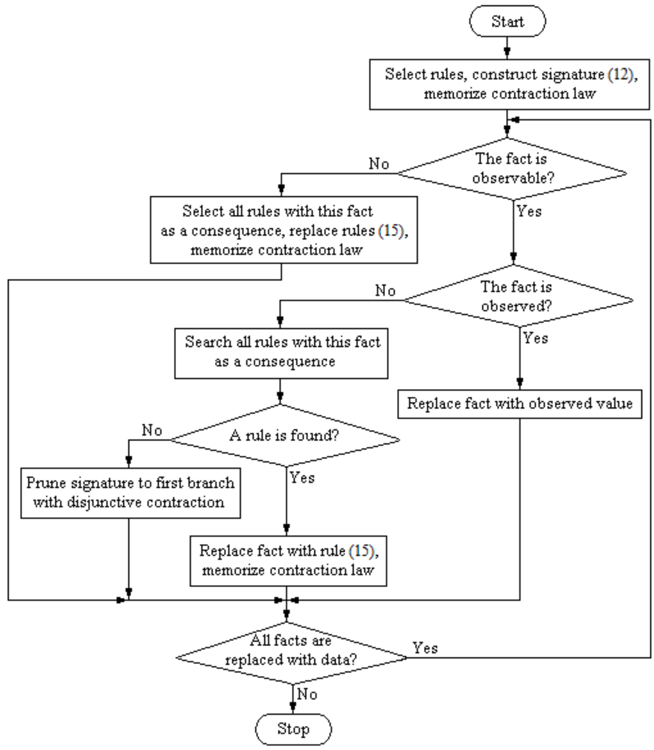


Figure 3
Flowchart of expert systems modelling algorithm

4 Case Study

This case study will apply the results presented in the previous section to derive models of a Bayesian rule-based expert system. The new models are expressed as signatures defined in Section 2.

The case study considers the signature-based modelling of an uncertain rule-based expert system represented by a Bayesian expert system. This rule-based expert system is also taken from [25], and it predicts the weather for the next day, viz. the "tomorrow" weather (tomorrow will rain or not). Two signatures will be constructed as follows using the algorithm defined in Section 3; the first one will ignore the accumulation of evidence and the second one will consider that.

Prior to the presentation of the application of our expert systems modelling algorithm, some details on the firing of Bayesian rules are given as follows. Let us consider the rule

$$\text{Rule 1}: \textit{If}\ A\{\text{LS} \equiv x, \text{LN} \equiv y\}\ \textit{Then}\ B\{\text{prior}\, u\}, \tag{22}$$

where $\text{LS} \equiv x$ is the likelihood of sufficiency of fact *A*, $\text{LN} \equiv y$ is the likelihood of necessity of fact *A*, and *u* is the prior probability of fact *B*. The application of Definition 6 leads to

$$r_1 = [A]^T,\ {}^{h}@(r_1) = [B]^T, \tag{23}$$

where the definition of the function *h* is

$$h(A) = \begin{cases} \dfrac{O(B\,|\,A)}{1+O(B\,|\,A)} & \textit{if} \quad A, \\ \dfrac{O(B\,|\,\neg A)}{1+O(B\,|\,\neg A)} & \textit{if} \quad \neg A, \end{cases} \tag{24}$$

$$\begin{cases} O(B\,|\,A) = x\,O(B), \\ O(B\,|\,\neg A) = y\,O(B), \end{cases} \quad O(B) = \frac{u}{1-u},$$

$O(B)$ is the prior evidence of fact *B*, $O(B\,|\,A)$ is the posterior evidence of fact *B* given the fact *A* (true); *and*; $h(A)$ is the posterior probability of fact *B* given the fact *A*.

The rules and signatures of this expert system are synthesized in Table 1, *f* is defined in (12), and the general notation $h \circ f = h(f)$ is used.

If the accumulation of evidence is ignored the application of the three steps of our expert systems modelling algorithm are first presented as follows.

Step 1. The signature is

$$r = [TwD, TwR]^T, \tag{25}$$

and the algorithm memorizes

$$\begin{aligned} &{}^{g}@(r) = [Tw]^T, \\ &Tw = g(TwD, TwR) = \begin{cases} TwD, & \text{if} \quad TwD > TwR, \\ TwR, & \text{if} \quad TwR > TwD. \end{cases} \end{aligned} \tag{26}$$

Table 1
Rules, symbols and signatures of the Bayesian rule-based expert system

Nr.	Rule	Symbol	Signature
1	If [Today is rain] {LS=2.5; LN=0.6} Then [Tomorrow is rain] {prior=0.5}	[Today is rain] = *TyR* [Tomorrow is rain]=*TwR*	$r_1 = [TyR]^T$, $^{h_1 \circ f} @(r_1) = [TwR]^T$
2	If [Today is dray] {LS=1.6; LN=0.4} Then [Tomorrow is dray] {prior=0.5}	[Today is dray] = *TyD* [Tomorrow is dray] = *TwD*	$r_2 = [TyD]^T$, $^{h_2 \circ f} @(r_2) = [TwD]^T$
3	If [Today is rain] and [Rainfall is low]{LS=10; LN=1} Then [Tomorrow is dray] {prior=0.5}	[Today is rain] = *TyR* [Rainfall is low] = *RaL* [Tomorrow is dray] = *TwD*	$r_3 = [TyR, RaL]^T$, $^{h_3 \circ f} @(r_3) = [TwD]^T$
4	If [Today is rain] and [Rainfall is low] and [Temperature is cold] {LS=1.5; LN=1} Then [Tomorrow is dray] {prior=0.5}	[Today is rain] = *TyR* [Rainfall is low] = *RaL* [Temperature is cold] = *TeC* [Tomorrow is dray] = *TwD*	$r_4 = [TyR, RaL, TeC]^T$, $^{h_3 \circ f} @(r_3) = [TwD]^T$
5	If [Today is dray] and [Temperature is warm]{LS=2; LN=0.9} Then [Tomorrow is rain] {prior=0.5}	[Today is dray] = *TyD* [Temperature is warm] = *TeW* [Tomorrow is rain]= *TwR*	$r_5 = [TyD, TeW]^T$, $^{h_5 \circ f} @(r_5) = [TwR]^T$
6	If [Today is dray] and [Temperature is warm] and [sky is overcast]{LS=5; LN=1} Then [Tomorrow is rain] {prior=0.5}	[Today is dray] = *TyD* [Temperature is warm] = *TeW* [Sky is overcast] = *SyO* [Tomorrow is rain] = *TwR*	$r_6 = [TyD, TeW, SyO]^T$, $^{h_6 \circ f} @(r_6) = [TwR]^T$

Step 2. The signature is

$$r = \overline{@}_2^3(\overline{@}_1^3(r)) = [[TwD_2, TwD_3, TwD_4], [TwR_1, TwR_5, TwR_6]]^T, \tag{27}$$

and the algorithm memorizes

$$ {}^{g}@_{1}(r) = [TwD,[TwR_1,TwR_5,TwR_6]]^T(r) = [Tw]^T, \\ {}^{g}@_{2}(r) = [[TwD_2,TwD_3,TwD_4],TwR]^T. \tag{28} $$

Step 3. The results are:

$$ r = r \oplus_{1,1} r_2 \oplus_{1,2} r_3 \oplus_{1,3} r_4 \oplus_{2,1} r_1 \oplus_{2,2} r_5 \oplus_{2,3} r_6 \\ = [[[TyD],[TyR,RaL],[TyR,RaL,TeC]],[[TyR], \\ [TyD,TeW],[TyD,TeW,SyO]]]^T,\ {}^{h_2\circ f}@_{1,1}(r) = [[TwD_2, \\ [TyR,RaL],[TyR,RaL,TeC]],[[TyR],[TyD,TeW], \\ [TyD,TeW,SyO]]]^T,\ {}^{h_3\circ f}@_{1,2}(r) = [[[TyD],TwD_3, \\ [TyR,RaL,TeC]],[[TyR],[TyD,TeW], \\ [TyD,TeW,SyO]]]^T,...,\ {}^{h_6\circ f}@_{2,3}(r) = [[[TyD], \\ [TyR,RaL],[TyR,RaL,TeC]],[[TyR],[TyD,TeW],TwR_6]]^T. \tag{29} $$

For the observations [Today is rain], [Rainfall is low], [Temperature is cold], [Sky is overcast] (TyR=1, RaL=1, TeC=1, SyO=1), the expert system output is computed in terms of:

$$ r = [[[TyD],[TyR,RaL],[TyR,RaL,TeC]],[[TyR],[TyD,TeW], \\ [TyD,TeW,SyO]]]^T = [[[0],[1,1],[1,1,1]],[[1],[0,0],[0,0,1]]]^T, \\ r = {}^{f}@_{1,1}({}^{f}@_{1,2}({}^{f}@_{1,3}({}^{f}@_{2,1}({}^{f}@_{2,2}({}^{f}@_{2,3}(r)))))) = [[[0],[1],[1]], \\ [[1],[0],[0]]]^T,\ r = {}^{h_2}@_{1,1}({}^{h_3}@_{1,2}({}^{h_4}@_{1,3}({}^{h_1}@_{2,1}({}^{h_5}@_{2,2}({}^{h_6}@_{2,3}(r)))))) \\ = [[0.285,0.909,0.6],[0.714,0.473,0.5]]^T, \\ r = {}^{g}@_{1}({}^{g}@_{2}(r)) = [0.909,0.714]^T. \tag{30} $$

This expert system response shows that the probability for “Tomorrow is Dray” is 0.909 and the probability for “Tomorrow is Rain” is 0.714. Prior to observations, the probabilities have been equal, i.e., 0.5 to 0.5.

Conclusions

This paper has given an application of signatures to expert systems modelling. Our modelling approach is backed up by a systematic modelling algorithm. The proposed approach has proved to be effective in accounting for certain observations, and the results have been generalized to uncertain observations.

The theoretical framework has been exemplified by a case study concerning a Bayesian model. We have shown how to elicit a rule base from the case study. The quality of the rules is not evaluated quantitatively, and this aspect will be considered as a future research direction of rule evaluation for confirming the effectiveness of the proposed modelling approach.

The future research will be focused on the reduction of the number of iterations which correspond to the step 2 of the algorithm. Since this case study is related to mechatronics applications but not to industrial problems, the proposed modelling results will be applied to other illustrative case studies such as those discussed in [37-46] to prove its effectiveness. More convincing mechatronics applications will be mapped onto our expert systems modelling approach.

Acknowledgement

This work was supported by a grant of the Romanian National Authority for Scientific Research, CNCS – UEFISCDI, project number PN-II-ID-PCE-2011-3-0109, and by the Romanian Ministry of Education, Research and Innovation through the PNII Idei project 842/2008.

References

[1] P. Angelov: Fuzzily Connected Multi-Model Systems Evolving Autonomously From Data Streams, IEEE Transactions on Systems, Man, and Cybernetics – Part B, Cybernetics, Vol. 41, No. 4, 2011, pp. 898-910

[2] P. Angelov, R. Yager: Density-Based averaging – A New Operator for Data Fusion, Information Sciences, Vol. 222, 2013, pp. 163-174

[3] Y. Dimopoulos, M. A. Hashmi, P. Moraitis: μ-SATPLAN: Multi-Agent Planning as Satisfiability, Knowledge-based Systems, Vol. 29, 2012, pp. 54-62

[4] H. H. L. C. Monte-Alto, A. B. Biasão, L. O. Teixeira, E. H. M. Huzita: Multi-Agent and Context-Aware Solutions for a Global Software Development Environment, International Journal of Artificial Intelligence, Vol. 11, No. A13, 2013, pp. 115-129

[5] D. Pal, K. M. Mandana, S. Pal, D. Sarkar, C. Chakraborty: Fuzzy Expert System Approach for Coronary Artery Disease Screening Using Clinical Parameters, Knowledge-Based Systems, Vol. 26, 2012, pp. 162-174

[6] D. Yazdani, B. Saman, A. Sepas-Moghaddam, F. Mohammad-Kazemi, M. Reza Meybodi: A New Algorithm Based on Improved Artificial Fish Swarm Algorithm for Data Clustering, International Journal of Artificial Intelligence, Vol. 11, No. A13, 2013, pp. 193-221

[7] A. Azarian, A. Siadat, P. Martin: A New Strategy for Automotive Off-Board Diagnosis Based on a Meta-Heuristic Engine, Engineering Applications of Artificial Intelligence, Vol. 24, No. 5, 2011, pp. 733-747

[8] W. Caicedo Torres, M. Quintana, H. Pinzón: Differential Diagnosis of Hemorrhagic Fevers Using ARTMAP and an Artificial Immune System, International Journal of Artificial Intelligence, Vol. 11, No. A13, 2013, pp. 150-169

[9] S. Szénási, Z. Vámossy: Evolutionary Algorithm for Optimizing Parameters of GPGPU-based Image Segmentation, Acta Polytechnica Hungarica, Vol. 10, No. 5, 2013, pp. 7-28

[10] C. Pozna, R.-E. Precup: Aspects Concerning the Observation Process Modeling in the Framework of Cognition Process, Acta Polytechnica Hungarica, Vol. 9, No. 1, 2012, pp. 203-223

[11] C. Pozna, R.-E. Precup: A General Formulation of Abduction Algorithms, Proceedings of IEEE 3rd International Conference on Cognitive Infocommunications (CogInfoCom 2012), Kosice, Slovakia, 2012, pp. 573-578

[12] P. Baranyi, A. Csapó: Cognitive Infocommunications: CogInfoCom, Proceedings of 11th IEEE International Symposium on Computational Intelligence and Informatics (CINTI 2010), Budapest, Hungary, 2010, pp. 141-146

[13] A. Csapó, P. Baranyi: A Unified Terminology for CogInfoCom Applications, Proceedings of 2nd International Conference on Cognitive Infocommunications (CogInfoCom 2011), Budapest, Hungary, 2011, 6 pp.

[14] A. Csapó, P. Baranyi: A Unified Terminology for the Structure and Semantics of CogInfoCom Channels, Acta Polytechnica Hungarica, Vol. 9, No. 1, 2012, pp. 85-105

[15] M. Nakano, Y. Hasegawa, K. Funakoshi, J. Takeuchi, T. Torii, K. Nakadai, N. Kanda, K. Komatani, H. G. Okuno, H. Tsujino: A Multi-Expert Model for Dialogue and Behavior Control of Conversational Robots and Agents, Knowledge-Based Systems, Vol. 24, No. 2, 2011, pp. 248-256

[16] C. A. Laurentys, C. H. M. Bomfim, B. R. Menezes, W. M. Caminhas: Design of a Pipeline Leakage Detection Using Expert System: A Novel Approach, Applied Soft Computing, Vol. 11, No. 1, 2011, pp. 1057-1066

[17] B. Jankowska, Using Semantic Data Integration to Create Reliable Rule-Based Systems With Uncertainty, Engineering Applications of Artificial Intelligence, Vol. 24, No. 8, 2011, pp. 1499-1509

[18] J. M. Gascueña, E. Navarro, A. Fernández-Caballero: Model-Driven Engineering Techniques for the Development of Multi-Agent Systems, Engineering Applications of Artificial Intelligence, Vol. 25, No. 1, 2012, pp. 159-173

[19] L. Chang, Y. Zhou, J. Jiang, M. Li, X. Zhang: Structure Learning for Belief Rule Base Expert System: A Comparative Study, Knowledge-Based Systems, Vol. 39, 2012, pp. 159-172

[20] C. Pozna, N. Minculete, R.-E. Precup, L. T. Kóczy, Á. Ballagi: Signatures: Definitions, Operators and Applications to Fuzzy Modeling, Fuzzy Sets and Systems, Vol. 201, 2012, pp. 86-104

[21] Á. Ballagi, L. T. Kóczy, T. D. Gedeon: Robot Cooperator Without Explicit Communication by Fuzzy Signatures and Decision trees, Proceedings of Joint 2009 International Fuzzy Systems Association World Congress and 2009 European Society of Fuzzy Logic and Technology Conference (IFSA-EUSFLAT 2009), Lisbon, Portugal, 2009, pp. 1468-1473

[22] L.T. Kóczy, T. Vámos, G. Biró: Fuzzy Signatures, Proceedings of 4th Meeting of the Euro Working Group on Fuzzy Sets and 2nd International Conference on Soft and Intelligent Computing (EUROPUSE-SIC'99), Budapest, Hungary, 1999, pp. 210-217

[23] C. Pozna, R.-E. Precup: Results Concerning a New Pattern of Human Knowledge, Proceedings of 2nd International Conference on Cognitive Infocommunications (CogInfoCom 2011), Budapest, Hungary, 2011, 18 pp.

[24] C. Pozna, R.-E. Precup: Novel Design of Cognitive System Strategies, Proceedings of 4th IEEE International Symposium on Logistics and Industrial Informatics (LINDI 2012), Smolenice, Slovakia, 2012, pp. 205-214

[25] M. Negnevitsky, Artificial Intelligence, Person Education Limited, Harlow, 2011

[26] N. Minculete: On Certain Inequalities About Arithmetic Functions Which Use Exponential Divisors, International Journal of Number Theory, Vol. 8, No. 6, 2012, pp. 1527-1535

[27] N. Minculete: A Note About Properties of Exponential Divisors, Applied Mathematics & Information Sciences, Vol. 7, No. 1, 2013, pp. 319-322

[28] F. C. Mitroi, N. Minculete: Mathematical Inequalities for Biparametric Extended Information Measures, Journal of Mathematical Inequalities, Vol. 7, No. 1, 2013, pp. 63-71

[29] S. Preitl, R.-E. Precup: On the Algorithmic Design of a Class of Control Systems Based on Providing the Symmetry of Open-Loop Bode Plots, Scientific Bulletin of "Politehnica" University of Timisoara, Romania, Transactions on Automatic Control and Computer Science, Vol. 41 (55), No. 1-2, 1996, pp. 47-55

[30] S. Preitl, R.-E. Precup: Introducere în conducerea fuzzy a proceselor, Editura Tehnica, Bucharest, 1997

[31] Z. C. Johanyák, S. Kovács: Polar-Cut Based Fuzzy Model for Petrophysical Properties Prediction, Scientific Bulletin of "Politehnica" University of Timisoara, Romania, Transactions on Automatic Control and Computer Science, Vol. 57 (67), No. 4, 2008, pp. 195-200

[32] R.-E. Precup, S. Doboli, S. Preitl: Stability Analysis and Development of a Class of Fuzzy Control Systems, Engineering Applications of Artificial Intelligence, Vol. 13, No. 3, 2000, pp. 237-247

[33] J. Vaščák, M. Rutrich: Path Planning in Dynamic Environment Using Fuzzy Cognitive Maps, Proceedings of 6th International Symposium on Applied Machine Intelligence and Informatics (SAMI 2008), Herľany, Slovakia, 2008, pp. 5-9

[34] E. S. Nicoară, F.-G. Filip, N. Paraschiv: Simulation-based Optimization Using Genetic Algorithms for Multi-objective Flexible JSSP, Studies in Informatics and Control, Vol. 20, No. 4, 2011, pp. 333-344

[35] A. Widyotriatmo, A. K. Pamosoaji, K.-S. Hong: Control Architecture of an Autonomous Material Handling Vehicle, International Journal of Artificial Intelligence, Vol. 10, No. S13, 2013, pp. 139-153

[36] P. Melin, L. Astudillo, O. Castillo, F. Valdez, M. Garcia: Optimal Design of Type-2 and Type-1 Fuzzy Tracking Controllers for Autonomous Mobile Robots Under Perturbed Torques Using a New Chemical Optimization Paradigm, Expert Systems with Applications, Vol. 40, No. 8, 2013, pp. 3185-3195

[37] R.-E. Precup, S. Preitl: Popov-Type Stability Analysis Method for Fuzzy Control Systems, Proceedings of Fifth EUFIT'97 European Congress, Aachen, Germany, 1997, pp. 1306-1310

[38] F. Previdi, T. Schauer, S. M. Savaresi, K. J. Hunt: Data-Driven Control Design for Neuroprotheses: A Virtual Reference Feedback Tuning (VRFT) Approach, IEEE Transactions on Control Systems Technology, Vol. 12, No. 1, 2004, pp. 176-182

[39] S. Preitl, R.-E. Precup, J. Fodor, B. Bede: Iterative Feedback Tuning in Fuzzy Control Systems. Theory and applications, Acta Polytechnica Hungarica, Vol. 3, No. 3, 2006, pp. 81-96

[40] R.-E. Precup, M. L. Tomescu, S. Preitl: Lorenz System Stabilization Using Fuzzy Controllers, International Journal of Computers, Communications & Control, Vol. II, No. 3, 2007, pp. 279-287

[41] G. Hermann, G. Benkó, A. Nagy: Retrofitting a Length Measuring Machine for Linear Scale Calibration, Proceedings of 13th International Conference on Intelligent Engineering Systems (INES 2009), Barbados, 2009, pp. 79-82

[42] G. Klančar, D. Matko, S. Blažič: A Control Strategy for Platoons of Differential-Drive Wheeled Mobile Robot, Robotics and Autonomous Systems, Vol. 59, No. 2, 2011, pp. 57-64

[43] M. Oprea: INTELLEnvQ-Air: An Intelligent System for Air Quality Analysis in Urban Regions, International Journal of Artificial Intelligence, Vol. 9, No. A12, 2012, pp. 106-122

[44] L. Horváth, I. J. Rudas: Active Knowledge for the Situation-driven Control of Product Definition, Acta Polytechnica Hungarica, Vol. 10, No. 2, 2013, pp. 217-234

[45] K. Lamár, J. Neszveda: Average Probability of Failure of Aperiodically Operated Devices, Acta Polytechnica Hungarica, Vol. 10. No. 8, 2013, pp. 153-167

[46] I. Mundi, M. M. E. Alemany, A. Boza, R. Poler: A Model-Driven Decision Support System for the Master Planning of Ceramic Supply Chains with Non-uniformity of Finished Goods, Studies in Informatics and Control, Vol. 22, No. 2, 2013, pp. 153-162

Analysis of Knowledge Base Units within Standardized Electrical Engineering Subfields

Živadin Micić, Momčilo Vujičić, Vera Lazarević

Faculty of Technical Sciences, University of Kragujevac
Svetog Save 32, Čačak, Serbia
micic@kg.ac.rs, momcilo.vujicic@ftn.kg.ac.rs, vera.lazarevic@ftn.kg.ac.rs

Abstract: The paper presents a comparative and multicriteria statistical analysis of knowledge base units as shown on the examples of standardized subfields of electrical engineering. A number of objectives have been realized on knowledge pathways, both globally and locally, and initial hypotheses have been confirmed. Ranking/clustering among standardized fields and subfields has been carried out based on time-frequent innovations presented by knowledge base unit quantities and values. Web applications for analysis and singling out of samples from the entire Web-population of local (SRPS) and global (ISO) standardization have been developed and implemented. These applications represent a significant contribution to authenticity, particularly the authenticity of statistical research methodology.

Original regression equations with accompanying indices of quantity, knowledge base unit values and ranking have been developed on the examples within the field of electrical engineering. Results of this study enable the realization of some higher practical objectives, i.e., (1) modification of teaching subjects and improving educational and financial resources for the quality of knowledge, (2) ranking of both narrow and broad standardized technologies and scientific fields.

Keywords: electrical engineering; knowledge; standardization; SRPS; ISO; subfield

1 Introduction

1.1 What Problem Does the Paper Look at?

The paper looks at the comparative and multicriteria analysis and evaluation of the international (ISO/IEC) and local (SRPS – designation of standard in Serbia, [1]) knowledge in the subfields of electrical engineering. In this particular field, knowledge pathways differ from other standardized fields of human endeavour. In this field more importance is given to local and individual rather than collective knowledge. According to the international classification of standards (ICS_1), electrical engineering is classified as field XXIX and it consists of 19 subfields (ICS_2 = 29.xyz, [2]): 29.020 – Electrical engineering in general, including voltages, general electrical terminology, electrical documentation, electrical tables, safety, fire hazard testing, etc. (electrical diagrams and charts, electrical power stations, electromagnetic compatibility), 29.030 – Magnetic materials, 29.035 – Insulating materials, 29.040 – Insulating fluids, 29.045 – Semiconducting materials, 29.050 – Superconductivity and conducting materials, 29.060 – Electrical wires and cables, 29.080 – Insulation, insulating fluids, insulating materials, insulating fluids, 29.100 – Components for electrical equipment, 29.120 – Electrical accessories, 29.130 – Switchgear and controlgear, 29.140 – Lamps and related equipment, 29.160 – Rotating machinery, 29.180 – Transformers, reactors, *instrument transformers,* 29.200 – Rectifiers, converters, stabilized power supply, *including semiconductor converters,* 29.220 – Galvanic cells and batteries, 29.240 – Power transmission and distribution networks, *electricity supply systems in buildings,* 29.260 – Electrical equipment for working in special conditions, 29.280 – Electric traction equipment, including railway electric fixed installations, electric road vehicles, non-electric railway tractive stock.

1.2 Why Has the 'Problem' Not Been Solved Before?

In the investigated field of electrical engineering, the 'problem' has been observed only in the past five years (2008 to 2012); e.g. in 2010 two local-scale standards (SRPS) were published daily! The same trend continued in 2011 and 2012. Currently, it is the field of electrical engineering that encompasses the greatest number of SRPS samples. Particularly *important* and *significant* are alternative solutions to the 'problem' of individual access to the knowledge base units. This is particularly important in various environments, institutions and corporations in developing countries. It is of great importance to provide organized help and support to teachers and assistants in order to achieve mass education based on the platform of standardization. Solving this 'problem' at the state level leads to a national model of e-learning excellence. Although, according to [3], the 'Quality is free' philosophy was prevalent in the previous century, nowadays, standards are extremely costly.

1.3 The Research Objectives

The initial hypotheses and research objectives in the field of *electrical engineering* provide opportunities for:

1 (P – Plan) resource *planning*, i.e. conducting comparative analyses and 'selection' of regression equations of the closest realities with explicit mathematical relations, with the aim of predicting the future resources and comparing them with the trends developing in other areas;

2 (D – Do) determining clear correlations in space and time, collectively and locally (ISO and SRPS), with annual trends of innovating the valued standardized knowledge base units;

3 (C – Check) *checking* and conducting multicriteria analyses of standardized knowledge base units by defining the comparative indices (quantity, value and ranking) of individual subfields, the field in its entirety and the future establishing of appropriate relations with scientific fields and the related teaching subjects at a higher education level;

4 (A – Act) deriving alternative solutions in order to replace discontinuous with continuous (cumulative) knowledge innovation, especially of teachers and the staff, but also of all the users in the related subfields.

2 Research Methodology and Framework

The PDCA (Plan, Do, Check, Act) statistical methodology of dynamic analyses and deductive-inductive reasoning methods has been used for predicting the future development and innovation of the pragmatic framework. Methodologically, statistical indices have been formed for the comparison of ISO–SRPS relations in the field of *electrical engineering* (ICS_1 = 29) with other fields of human endeavour, including: Quantity indices (Iq), value indices (Iv) and ranking index (Iqv). Quantity indices (Iq), defined and determined for both ISO and SRPS, refer to the following: Samples (Iqs), published standards (Iqp), standards under development (Iqu), standards withdrawn from use (Iqw), deleted projects (Iqd), innovations in various stages of development ($Iqi = Iqu_{/2012}$) – for the entire previous calendar year.

Value indices (Iv) are in correlation with quantity indices, whereby the results are graphically presented both cumulatively and by trends, within the majority of the 19 electrical engineering subfields, shown by figures consisting of two parts, (a) including time aspects for the entire research period – by years of publication, i.e., ∑Iv/year, as well as (b) including regression trend lines (exponential, linear, logarithmic and polynomial), based on data collected over the previous five years (2008 to 2012) and by the determined regression equations $y_{i/SRPS}$, Ivi/year.

For the complete 'game' (browsing, analysis, systematization and presentation of the results), original Web applications have been used for research purposes, as examples of IT innovations of software products on comparative samples in the fields of electrical engineering (for ICS_1 = 29), and on the ISO and SRPS levels. The first Web application gathers (and analyses) data on the local SRPS standardization level ($Iqs_{/SRPS/2012.01} \approx 34000$ local 'innovation', [1]) and sample of $Iqs_{/29/SRPS/2012.01} = 2111$ units). The second Web application browses (analyses) the ISO standardized unit base ($Iqs_{/ISO/2012.01} \approx 42000$ collective-global innovations on the examples of standards [2]) and the sample of $Iqs_{/29/ISO/2012.01} = 55$ units.

Multicriteria analysis is used to extract and evaluate the results achieved so far, both at the beginning and at the end of the year (01/2012, 11/2012 and 12/2012, Table 1). Indices of quantity, value and ranking allow comparisons within the examined field and the subfields, along with time-line based comparisons and external comparisons with other fields. This methodology enables the investigation and comparison of the results with those in other areas of work and human endeavour – as presented in [4].

3 Results

3.1 Comparative Analysis and Clustering

The results of the analysis of standardized knowledge base units within the field of electrical engineering are presented by comparative indices of quantity and value for each of the 19 subfields and the field as a whole (Table 1). Further, statistical samples (Iqs) with the accompanying comparative results and other indices of quantity and value are used to present the analysed knowledge parameters, as shown in Table 1.

Columns 7, 8 and 11 in Table 1 refer to SRPS, i.e., the current state of SRPS development, the sum of newly developed SRPS in 2012 and values based on the most appropriate regression equations singled out, respectively ((1.1) to (20.1)). In early 2012, SRPS standards (drafts) were developed faster than the ISO ones by some 300%. As for the value, a similar trend was recorded, as shown in Table 1, columns 12 and 13. CHF stands for the ISO-selling currency (in Switzerland).

Based on Table 1 (Iq, Iqs, Iqp, Iv, ∑Iv/year, yi/year+1), the field of electrical engineering, ICS_1 = 29, is classified in group VII (ranking index Iqv = 1, 2, 3 to 7) and, from the standpoint of higher education, it could/should be classified as an individual scientific field [5]. The quantitative criterion (Iqu/year) applies only to the assessment/evaluation of the innovation and clustering of sub/fields based on the development of SRPS, while regarding the ISO development, the criterion is replaced with Iqu.

The results of the analysis are presented in graphs, including only the subfields belonging to the second and higher groups. The graphs include two parts:

a) Aggregate analyses of all current standards (Std), corrections (Cor), amendments (Amd) and development projects,

b) Financial trend lines, only for Std, not including the period before 2008 (Figures 1 to 17).

In order to present the graphic results more clearly, only analyses by subfields are shown in detail, while the analyses by sub/subfields (62) are not included.

Table 1
Comparative elements of knowledge pathways ISO–SRPS
(For ICS_1 = 29, 2012 / January–December)

I	Sub/ field	Samples		Published		Developed		Withdrawn		Trend	∑Ivalues (CHF)	
		Iqs		**Iqp**		**Iqu**		**Iqw**		**yi/2013**	**∑Iv/2012**	
	ICS	ISO	SRPS	ISO	SRPS	SRPS	2012	ISO	SRPS	CHF	∑Iv/iso	∑Iv/srps
1)	2)	3)	4)	5)	6)	7)	8)	9)	10)	11)	12)	13)
1	29.020	-	194	-	152	12	44	-	30	1152	-	**5871**
2	29.030	-	19	-	19	-	10	-	-	0	-	462
3	29.035	-	299	-	266	–	40	-	33	1210	-	**6691**
4	29.040	-	60	-	53	-	7	-	7	97	-	1635
5	29.045	-	1	-	1	-	-	-	-	0	-	57
6	29.050	-	15	-	15	-	-	-	-	0	-	447
7	29.060	-	353	-	302	6	53	-	45	627	-	8930
8	29.080	-	88	-	73	1	4	-	14	53	-	2541
9	29.100	-	55	-	55	-	1	-	-	20	-	1291
10	29.120	-	257	-	215	8	26	-	34	402	-	8336
11	29.130	-	118	-	107	1	26	-	10	337	-	4743
12	29.140	-	448	-	327	1	64	-	120	1090	-	**12293**
13	29.160	35	76	17	53	12	3	17	11	149	1492	2336
14	29.180	7	109	4	96	3	8	2	10	96	348	2746
15	29.200	-	51	-	44	1	6	-	6	47	-	2041
16	29.220	6	48	2	46	-	-	4	2	0	200	1388
17	29.240	-	120	-	113	2	32	-	5	862	-	4227
18	29.260	5	113	2	95	3	35	-	15	1254	238	4079
19	29.280	-	62	-	57	-	8	-	5	170	-	1536
XII	∑ 29	53	2486	25	2089	50	367	23	347	6226	2278	71651
XI	∑ 29	53	2479	25	2060	88	345	23	326	-	2278	61906
I	∑29	55	2111	26	1837	187	0	23	87	7695	2352	51580

In Table 1, the sign Σ is not a mathematical "summation function", but a graphical symbol that indicates the cumulative values of some parameters, such as: ∑Ivalues, ∑Iv/2012, ∑Iv/iso, ∑Iv/srps and ∑29 for ∑29.xyz.

3.2 Knowledge Trends Analysis by Subfields

3.2.1 Subfield – Electrical Engineering in General (29.020)

Electrical engineering in general subfield belongs to cluster VI (Iqv = 6, as well as subfield 3, according to the criteria Iq and Iv, Table 1). The level of this subfield *includes: 1* voltages, general electrical terminology, electrical documentation, electrical tables, safety, fire hazard testing, 2* electrical diagrams and charts, 3* electrical power stations, 4* electromagnetic compatibility.*

In this subfield, out of the relevant 164 SRPS (Iqp + Iqu) standards, i.e., published + developed standards (value of $Iv_{(p+u)/29.020/SRPS}$ = 5,075.87 CHF), a significant number of standards were withdrawn ($Iqw_{/29.020/SRPS}$) and a number of projects were under development – drafts ($Iqu_{/29.020/SRPS}$, in value $Ivu_{/29.020/SRPS/2012.11}$ = 427.78 CHF).

According to the number of samples (Iqs), one of the most extensive and most representative of the five electrical engineering subfields is presented in Figure 1, along with the results of the analysis of local (SRPS) knowledge standardization.

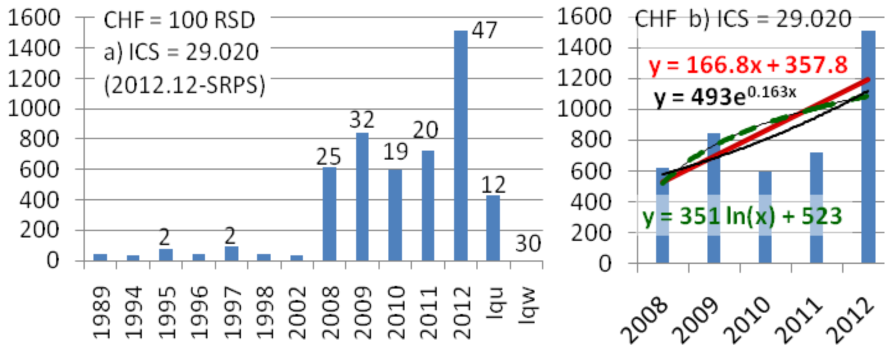


Figure 1

Analysis results for ICS_2 = 29.020 – Electrical engineering in general – **SRPS**

Figure 1 includes: a) aggregate analyses, over 1989-2012, and b) the trend of planned (future) annual needs according to the relation (1.1):

$$\mathbf{y_{29.020/SRPS/2008\text{-}2012.12} = 351\ ln(x) + 523}. \qquad (1.1)$$

3.2.2 Subfield – Magnetic Materials (29.030)

Magnetic materials subfield belongs to cluster I (Iqv = 1, Table 1). This subfield does not involve any documents under development – drafts (Iqu), or any documents that were withdrawn (Iqw) or deleted (Iqd). Other results can be represented: a) graphically, although the sample of all available $Iqs_{/29.030/SRPS}$ = 19 from the period of 2010-2011 is insufficient for a detailed statistical analysis; b) with the trend of planned future needs, according to the relation (2.1):

$$\mathbf{y_{29.030/SRPS/20010\text{-}2012} = 102.2\ x - 50.55}. \qquad (2.1)$$

3.2.3 Subfield – Insulating Materials (29.035)

Insulating materials subfield belongs to cluster VI (as well as subfield 1).

The samples available for analysis at this level involve 12 units, whereas the remaining 287 are distributed among the levels classified according to ICS_3: 29.035.01 – Insulating materials in general (121), 29.035.10 – Paper and board insulating materials (32), 29.035.20 – Plastics and rubber insulating materials, *including adhesive tapes* (114), 29.035.30 – Glass and ceramic insulating materials (6), 29.035.50 – Mica based materials (10), 29.035.60 – Varnished fabrics (2) and 29.035.99 – Other insulating materials (2).

SRPS standardization analysis is presented graphically in Figure 2:

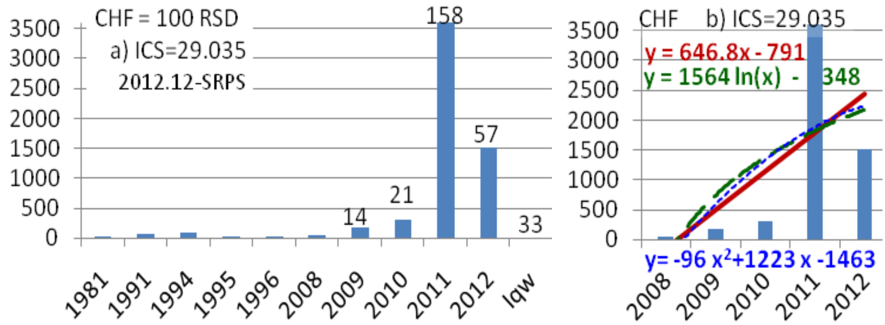


Figure 2

Analysis results for ICS_2 = 29.035 – Insulating materials – **SRPS**

Figure 2 includes:

a) the current standards published over 1981-2012, with a significant number of standards withdrawn from use (Iqw, Figure 2a),

b) a trend-line of planned (annual) future needs, i.e., the regression line, which is of a more theoretical nature, as the number of publications ($Iqp_{/2011}$ = 172) in 2011 shows a decreasing trend, whereby the planned values have been obtained using logarithmic or polynomial relation (3.1) (Figure 2b):

$$\mathbf{y_{29.035/SRPS/2008\text{-}2012.12} = -\,96\,x^2 + 1223\,x - 1463}. \tag{3.1}$$

3.2.4 Subfield – Insulating Fluids (29.040)

Insulating Fluids subfield belongs to cluster II (Iqv = 2, along with subfields 9, 16 and 19). This subfield level has one standard, whereas the other 59 are distributed among levels classified according to ICS_3: 29.040.01 – Insulating fluids in general (1), 29.040.10 – Insulating oils (24), 29.040.20 – Insulating gases (34) and 29.040.99 – Other insulating fluids (0).

The results of the standardization analysis in this subfield are presented in Figure 3, including:

a) all analysed samples developed over 1991-2012,

b) a trend of planned needs according to the relation (4.1):

$$\mathbf{y_{29.040/SRPS/2008\text{-}2012.12} = -\,43\,x^2 + 204\,x + 250}. \tag{4.1}$$

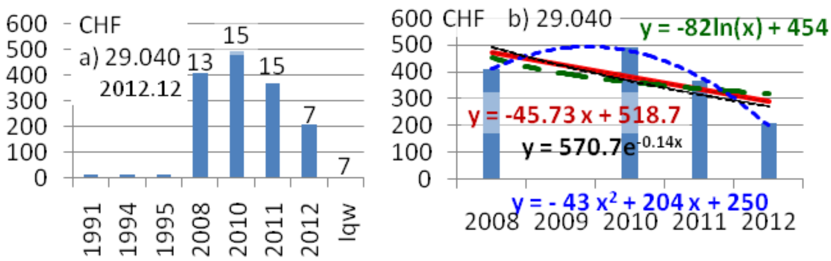


Figure 3
Analysis results for ICS_2 = 29.040 – **Insulating fluids – SRPS**

3.2.5 Subfield – Semiconducting Materials (29.045)

Semiconducting materials subfield belongs to cluster I (Iqv = 1). This subfield has the fewest samples (Iqs = 1). Other results are of no importance, because: a) the sample $Iqs_{/29.045/SRPS}$ = 1 was developed in 2011; b) the trend of planned future needs is not displayed.

3.2.6 Subfield – Superconductivity and Conducting Materials (29.050)

Superconductivity and conducting materials subfield belongs to cluster I. There are no sub/subfields for ICS_3. Other results are not significant for the field of electrical engineering: a) with the aggregate sample available $Iqs_{/29.050/SRPS}$ = 15 units from the period of 2004-2011; b) without the graph showing the trend of planned future needs according to (6.1):

$$\mathbf{y_{29.050/SRPS/2010\text{-}2011} = -\,371.8\,x + 773.4}. \qquad (6.1)$$

3.2.7 Subfield – Electrical Wires and Cables (29.060)

Electrical wires and cables subfield belongs to cluster V (Iqv = 5, along with subfield 10). Based on the levels classified according to ICS_3, some 350 samples are distributed as follows: 29.060.01 – Electrical wires and cables in general (1), 29.060.10 – Wires, *including electric rods, busbars* (141) and 29.060.20 – Cables (208).

The results of the analysis performed towards the end of 2012 are presented in Figure 4:

a) without currently valid standards dating back earlier than 1980, the majority of samples being developed only over the past four years (2009-2012),

b) with regression lines and the anticipated trend of planned future (annual) needs, according to (7.1):

$$\mathbf{y_{29.060/SRPS/2008\text{-}2012.12} = -\,390.7\,x^2 + 2605\,x - 2020}. \qquad (7.1)$$

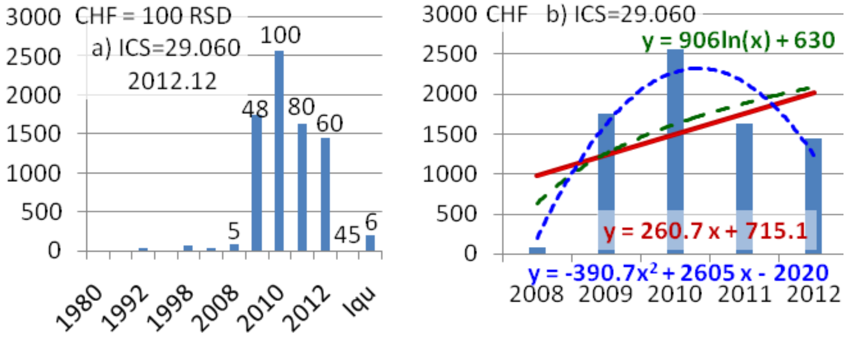


Figure 4

Analysis results for ICS_2 = 29.060 – Electrical wires and cables – **SRPS**

3.2.8 Subfield – Insulation, Insulating Fluids, Insulating Materials, Insulating Fluids (29.080)

Insulation, insulating fluids, insulating materials subfield belongs to cluster III (along with subfields 13, 14 and 15). The following subfields are related to this subfield: *Insulating materials, see 29.035 and insulating fluids, see 29.040.*

Samples available for analysis at this subfield level include 11 units, whereas the remaining 77 are distributed among levels classified according to ICS_3: 29.080.01 – Electrical insulation in general (15), 29.080.10 – Insulators, *including fittings and other components for insulators* (37), 29.080.20 – Bushings (10), 29.080.30 – Insulation systems (14), 29.080.99 – Other standards related to insulation (1). In this subfield, during early 2012, $Iqs_{/29.080/SRPS}$ = 80, two standards were under development, i.e. $Iqu_{29.080/SRPS}$ = 2 (Ivu = 39.45 CHF), 2 were withdrawn ($Iqw_{29.080/SRPS}$ = 2) and 76 were published, i.e. $Iqp_{29.080/SRPS}$ = 76.

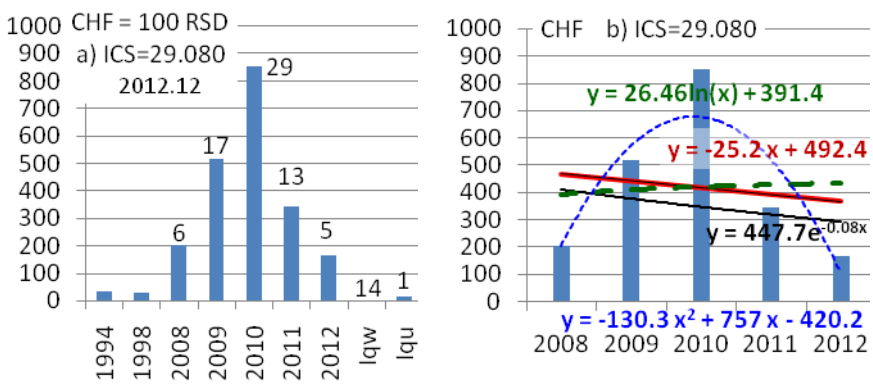


Figure 5

SRPS Analysis results for ICS_2 = 29.080 – Insulation – **SRPS**

Towards the end of 2012, the results of the analysis were slightly different, as shown in Figure 5, including: a) the aggregate analyses for the period 1994-2012; b) the trend of planned future (annual) needs according to (8.1) (Figure 5b):

$$\mathbf{y_{29.080/SRPS/2008\text{-}2012.12} = -\,130.3\,x^2 + 757\,x - 420.2}. \qquad (8.1)$$

3.2.9 Subfield – Components for Electrical Equipment (29.100)

Components for electrical equipment subfield belongs to cluster II (Iqv = 2, along with 4, 16 and 19 subfields), *including electronic components*.

Figure 6b shows the declining trend lines, while the anticipated trend line is (9.1):

$$\mathbf{y_{29.100/SRPS/2008\text{-}2012} = -\,249.9\,x^2 + 1633\,x - 2023}. \qquad (9.1)$$

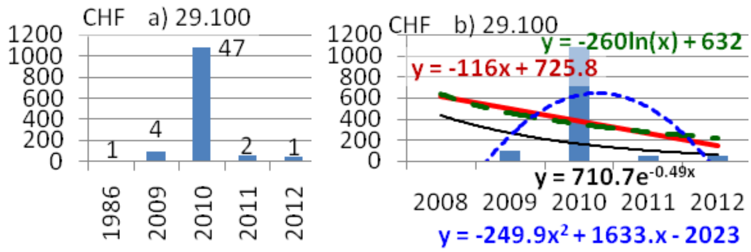


Figure 6

Analysis results for ICS_2 = 29.100 – Components for electrical equipment – **SRPS**

The samples from this subfield include two units, while the remaining 53 are distributed among levels classified according to ICS_3: 29.100.01 – Components for electrical equipment in general (0), 29.100.10 – Magnetic components (52), 29.100.20 – Electric and electromechanical components (1), 29.100.99 – Other components for electrical equipment (0).

3.2.10 Subfield – Electrical Accessories (29.120)

Electrical accessories subfield belongs to cluster V (along with subfield 7).

Out of the overall sample in this subfield ($Iqs_{/29.120}$), only five units were derived before 2008, which is only about 2% (Figure 7a). Out of the four descending trend lines, the lowest and most realistic planned value is given by the polynomial relation (10.1).

$$\mathbf{y_{29.120/SRPS/2008\text{-}2012.12} = -\,135\,x^2 + 660.5\,x + 876} \qquad (10.1)$$

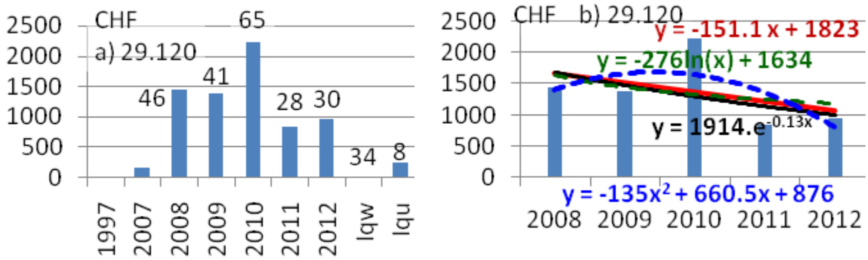


Figure 7

Analysis results for ICS_2 = 29.120 – Electrical accessories – **SRPS**

Some 279 knowledge base units available are distributed among the levels classified according to ICS_3: 29.120.01 – Electrical accessories in general (6), 29.120.10 – Conduits for electrical purposes (31), 29.120.20 – Connecting devices (27), 29.120.30 – Plugs, socket-outlets, couplers (26), 29.120.40 – Switches (39), 29.120.50 – Fuses and other over current protection devices (83), 29.120.70 – Relays (38), 29.120.99 – Other electrical accessories (18).

3.2.11 Subfield – Switchgear and Controlgear (29.130)

Switchgear and controlgear subfield belongs to cluster IV (Iqv = 4, along with subfields 17 and 18). The analysis samples in this subfield include 14 units, while the remaining 105 units are distributed among levels classified according to ICS_3: 29.130.01 Switchgear and controlgear in general (0), 29.130.10 High voltage switchgear and controlgear, *including enclosed switchgear and controlgear* (33), 29.130.20 Low voltage switchgear and controlgear, *including switchgear and controlgear assemblies* (65), 29.130.99 Other switchgear and controlgear (7).

The sample or the overall population in these subfields was formed in 2008 and later (Figure 8a). Out of the declining trend lines, the closest is (11.1), given the small Iqu number:

$$\mathbf{y_{29.130/SRPS/2008\text{-}2012.12} = -62\,x^2 + 346\,x + 493}. \qquad (11.1)$$

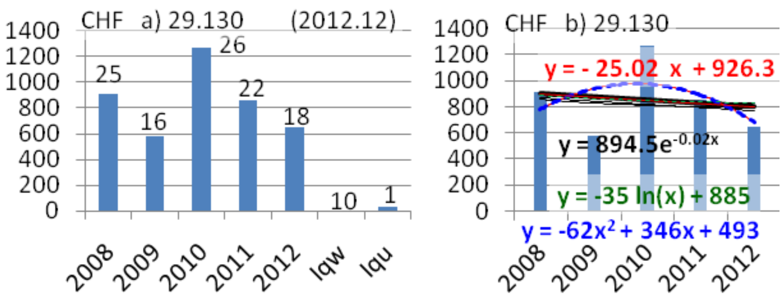


Figure 8

Analysis results for ICS_2 = 29.130 – Switchgear and controlgear – **SRPS**

3.2.12 Subfield – Lamps and Related Equipment (29.140)

Lamps and related equipment subfield belongs to cluster VII (Iqv = 7).

The results of the analysis are shown in Figure 9, including:

a) the aggregate analyses for the period 1984-2012, whereby only three units originated earlier than 2008, which is less than 1%,

b) the trend of the planned (annual) future needs based on the relation (12.1):

$$\mathbf{y_{29.140/SRPS/2008\text{-}2012.12} = -\,598.8x^{2} + 4257\,x - 4135}. \qquad (12.1)$$

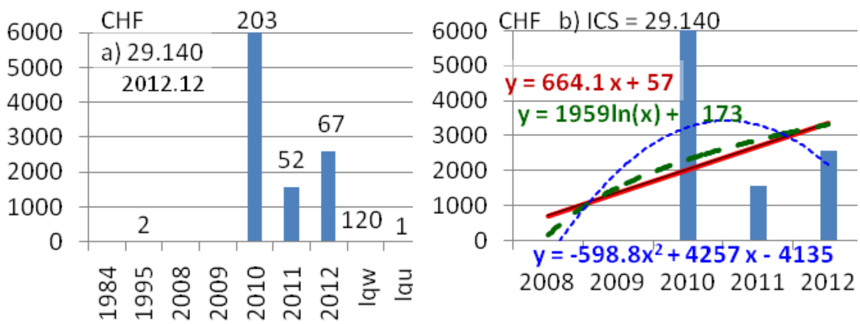


Figure 9

Analysis results for ICS_2 = 29.130 – Lamps and related equipment – **SRPS**

Samples available for analysis within the respective subfield include seven units, whereas the remaining 441 are distributed among levels classified according to ICS_3, viz. 29.140.01 – Lamps in general (3), 29.140.10 – Lamp caps and holders (213), 29.140.20 – Incandescent lamps (43), 29.140.30 – Fluorescent lamps, discharge lamps (73), 29.140.40 – Luminaires (38), 29.140.50 – Lighting installation systems, *including supply track systems* (21), 29.140.99 – Other standards related to lamps (50).

Compared with all the other 18 subfields, this subfield records the highest values of the quantity index (this includes Iqs samples as well as the published $Iqp_{29.140/SRPS}$ samples and the withdrawn $Iqw_{29.140/SRPS}$ samples), and the index value ($Iv_{/2011.11}$, $y_{29.140/SRPS/2013}$).

3.2.13 Subfield – Rotating Machinery (29.160)

Rotating machinery subfield belongs to cluster III (along with subfields 8, 14 and 15). This subfield has the greatest number of analysis samples (45 units), whereas the remaining 29 are distributed among levels classified according to ICS_3: 29.160.01 – Rotating machinery in general (7), 29.160.10 – Components for rotating machines (4), 29.160.20 – Generators (4), 29.160.30 – Motors (7), 29.160.40 – Generating sets (6), 29.160.99 – Other standards related to rotating machinery (1).

This subfield is the one with the highest quantity and value indices within the ISO standardization platform (see Table 1). The results of the analysis on the SRPS standardization platform are shown in Figure 10 with a) aggregate analyses for the 1992-2012 period, with only two units originating earlier than 2008, which is less than 3% (Figure 10a), and b) the clear downward trend of all the regression equations (Figure 10b), with the expected trend (13.1), given the significant development quantity of Iqu.

$$y_{29.160/SRPS/2008\text{-}2012.12} = \mathbf{16x^2 - 185x + 683}. \tag{13.1}$$

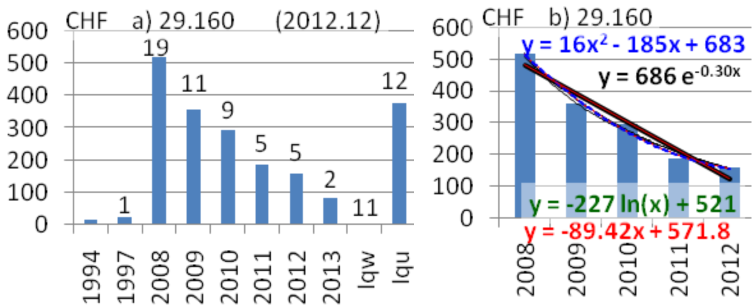


Figure 10

Analysis results for ICS_2 = 29.160 – Rotating machinery – **SRPS**

3.2.14 Subfield – Transformers, Reactors, Instrument Transformers (29.180)

Transformers, reactors, instrument transformers subfield belongs to cluster III (along with subfields 8, 13 and 15). The subfield is not subject to classification (in depth) at lower levels, and is related to *instrument transformers* (see 17.220.20). The results of the analysis on the SRPS standardization platform are shown in Figure 11 with:

a) the aggregate analyses for the period 1992-2012, whereby only one unit was developed before 2008, which is less than 1%,

b) the downward trend of all the regression equations and the most realistic one (14.1):

$$y_{29.180/SRPS/2008\text{-}2012} = \mathbf{-\,103.1\,x^2 + 577.3\,x - 116.6}. \tag{14.1}$$

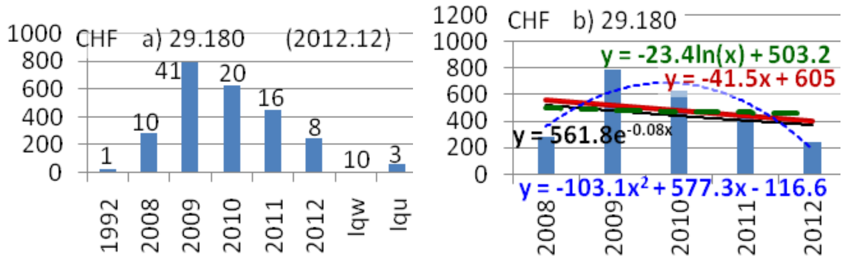


Figure 11

Analysis results for ICS_2 = 29.180 – Transformers, Reactors, *Instrument transformers* – **SRPS**

3.2.15 Subfield – Rectifiers, Converters, Stabilized Power Supply (29.200)

Rectifiers, converters, stabilized power supply subfield belongs to cluster III (along with subfields 8, 13 and 14). The subfield is not further classified (in depth) into sub/subfields, but it includes *semiconductor converters*.

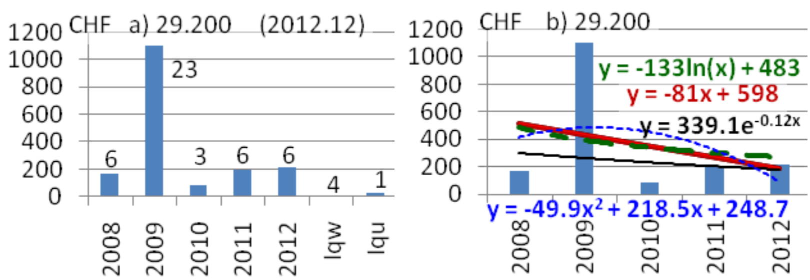


Figure 12

Analysis results for ICS_2 = 29.200 – Rectifiers, Converters, Stabilized power supply – **SRPS**

The results of the analysis on the SRPS standardization platform are shown in Figure 12, with: a) the aggregate analysis for the period 2008-2012, whereby no valid units derived before 2008 are recorded (Figure 12a); b) the downward trend of all the regression equations (Figure 12b), along with the ‘selection’ of (15.1):

$$\mathbf{y_{29.200/SRPS/2008\text{-}2012} = -\,49.9\,x^2 + 218.5\,x + 248.7}. \qquad (15.1)$$

3.2.16 Subfield – Galvanic Cells and Batteries (29.220)

Galvanic cells and batteries subfield belongs to cluster II (along with subfields 4, 9 and 19). Some 46 samples are classified among levels according to ICS_3, i.e., 29.220.01 – Galvanic cells and batteries in general (0), 29.220.10 – Primary cells and batteries (7), 29.220.20 – Acid secondary cells and batteries (26), 29.220.30 – Alkaline secondary cells and batteries (9), 29.220.99 – Other cells and batteries (4).

The results of the analysis are shown in Figure 13 with:

a) the aggregate analyses for the period 2000-2012, with only one unit produced earlier than 2008,

b) the downward trend of all the regression equations and the 'selection' of polynomial equation (16.1), which is the one closest to reality:

$$y_{29.220/SRPS/2008\text{-}2012.12} = -\mathbf{657\ x^2 + 3904\ x - 4919}. \tag{16.1}$$

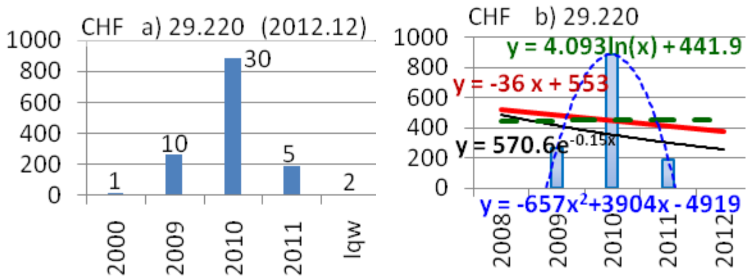


Figure 13

Analysis results for ICS_2 = 29.220 – Galvanic cells and batteries – **SRPS**

3.2.17 Subfield – Power Transmission and Distribution Networks (29.240)

Power transmission and distribution networks subfield belongs to cluster IV (as well as subfields 11 and 18), and is related to *electricity supply systems in buildings*. At this subfield level 18 samples are available *(for electricity supply systems in buildings),* while the remaining 103 are distributed among lower levels classified according to ICS_3: 29.240.01 – Power transmission and distribution networks in general, *power line telecommunications* (13), 29.240.10 – Substations, surge arresters (15), 29.240.20 – Power transmission and distribution lines (67), 29.240.30 – Control equipment for electric power systems, *measuring instruments for electrical quantities* (1), 29.240.99 – Other equipment related to power transmission and distribution networks, *including capacitors for power networks* (12).

The results of the analysis are shown in Figure 14 with:

a) the aggregate analyses for the period 1999-2012, whereby only two currently valid units were produced before 2008, which is less than 2%,

b) the growing trend of the regression equations and the 'selection' of the expected one (17.1):

$$\mathbf{y_{29.240/SRPS/2008\text{-}2012.12} = -64.6\ x^2 + 564.9\ x - 201.6}. \tag{17.1}$$

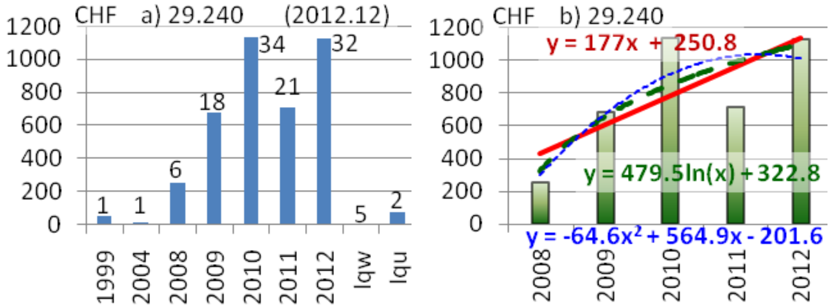


Figure 14

Analysis results for ICS_2 = 29.240 – Power transmission and distribution networks – **SRPS**

3.2.18 Subfield – Electrical Equipment for Working in Special Conditions (29.260)

Electrical equipment for working in special conditions subfield belongs to cluster IV (as well as subfields 11 and 17). In this subfield, six samples are available, while the remaining 107 are distributed among lower levels classified according to ICS_3: 29.260.01 – Electrical equipment for working in special conditions in general (0), 29.260.10 – Electrical installations for outdoor use (2), 29.260.20 – Electrical apparatus for explosive atmospheres (69), 29.260.99 – Other electrical equipment for working in special conditions (36).

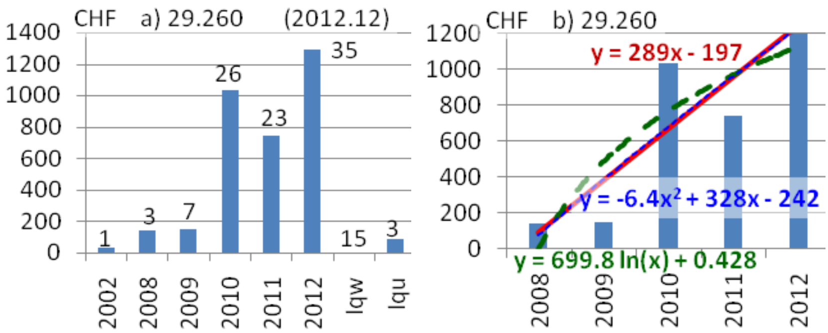


Figure 15

Analysis results for ICS_2 = 29.260 – Electrical equipment for working in special conditions – **SRPS**

The results of the analysis are shown in Figure 15 with:

a) the aggregate analyses for the period 1999-2012, whereby only one currently valid unit was developed before 2008, which is less than 1%,

b) the growing trend of all the regression equations, and the 'selection' of (18.1):

$$\mathbf{y_{29.260/SRPS/2008\text{-}2012.12} = 699.8\ ln(x) + 0.428}. \quad (18.1)$$

3.2.19 Subfield – Electric Traction Equipment (29.280)

Electric traction equipment subfield belongs to cluster II (along with subfields 4, 9 and 16), *including electric railway fixed installations, electric road vehicles, non-electric railway tractive stock.* This subfield was not subject to further classification (in depth) into levels based on ICS_3.

The results of the analysis are shown in Figure 16 with:

a) the aggregate analyses for the period 1999-2012, whereby only two currently valid units were developed before 2008, which is about 3%,

b) the growing trend of the majority regression equations, except for (19.1):

$$y_{29.280/SRPS/2008\text{-}2012} = -179\,x^2 + 1368\,x - 2025. \quad (19.1)$$

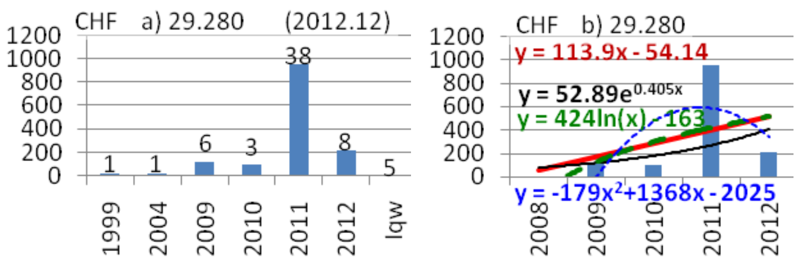


Figure 16
Analysis results for ICS_2 = 29.280 – Electric traction equipment – **SRPS**

4 Results and Discussion

4.1 Plan Phase Results (P)

Based on the results presented above, one can predict future resources and financial needs for each of the subfields as well as for the entire field. This is ensured by the established mathematical relations (1) to (20), represented theoretically by the trend lines and practically by the defined Iv/year value indices, as well as the given information on the stages of new projects development (along with Iqu).

The correctness of the trend line is discussed in *phase P*. Although the field of *electrical engineering* has been extensively innovated by SRPS standards in recent years (due to Serbia's preparations for EU accession), this does not mean that the same trend will continue in the years ahead. Therefore, the deviation of selected regressions is high: (7.1), (8.1), (9.1), (10.1), (11.1), (12.1) to (19.1).

4.2 Do Phase Results (D)

The results show significant 'index' issues in the analysed standardized knowledge base units among electrical engineering subfields. Figure 17a presents the specific results (Iv/year, Iqu) over 1980-2012 within the field analysed in its entirety (ICS_1 = 29). A negligible percentage of the currently valid knowledge base units were derived prior to 2008, the percentage being within the domain of statistical error.

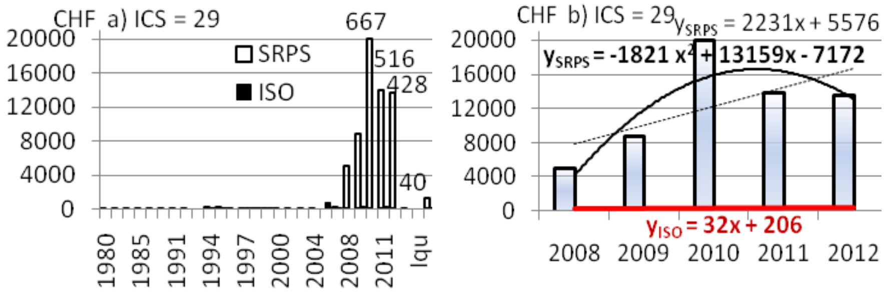


Figure 17
Examples of necessary resources and knowledge of all SRPS standards (ICS_1 = 29)

$$y_{29/SRSP/2008\text{-}2011/P0} = -2386\, x^2 + 15343\, x - 9370, \quad (20.1)$$

$$y_{29/ISO/2008\text{-}2011/P0} = y_{29/ISO/2008\text{-}2012/P1} = 32\, x + 206. \quad (20.2)$$

4.3 Check Phase Results (C)

Compared with some other areas of human endeavour, such as IT (ICS_1 = 35), or *IT applications in the industry* subfield (ICS_2 = 25.040), the opposite trend in the SRPS to ISO ratio of numbers of standardized innovations is evident. There is a greater number of SRPS units than ISO units. Some 367 SRPS standards ($Iqu_{/SRPS/29/2012}$) were developed in 2012, i.e., more than one standard each day. As for ISO, no standards were developed in 2012, with only a few (5) pathways being developed in 2012, i.e., $Iqu_{/ISO/29/2012} = 5$.

The overall results provide the means to develop an authentic methodology for establishing appropriate relations with scientific fields and related teaching subjects at a higher education level. This is achieved by defining the ranking index (Iqv = 1, 2, 3 to 7) as the criterion for grouping relevant subfields and/or fields of human endeavour into appropriate (teaching) groups [5].

Determining the index (or "degree") of innovation in the time dimension (Ity or It/year – per a year) is based on PDCA and defined values of quantity indices of development projects (Iqu), i.e., based on quantity indices of innovation (Iqi, phase Do). The period for checking innovations (*Check*) in certain fields or subfields is dependent on the time index of innovation, Ity. This index **Ity** is used to define groups or classes / clusters of innovation. The values of periodic checks (Check) on research for practice (1 – year, 2 – month, 3 – week or 4 – day) are assigned to this index. Ity = 1, for $1 \leq Iqu_{/SRPS/2012} \leq 10$ (planned annual audits/ checks), for example, ICS_2 = 29.030, 29.045, 29.050 etc (Iqv = 1, 2 and 3). Ity = 2, for $10 < Iqu_{/SRPS/2012} \leq 50$ (monthly Check, Table 1, column 8, Iqv = 4, 5 and 6). Ity = 3, for $50 < Iqu_{/SRPS/2012} \leq 250$ (weekly Check, Iqv = 6 and 7), of the third class, which requires continuous *weekly* monitoring of innovations. Ity = 4, for $Iqu_{/SRPS/2012} > 250$ (daily Checks), for example, for ICS_1 = 29 is an example of the field, which requires continuous *daily* monitoring of innovations.

4.4 Advancement Phase Results (A)

The results show some important details of the analyses of the field of electrical engineering and its subfields, with the possibility of comparison between the existing concept of "the gap between individual and collective knowledge" according to [4] and "Learning Strategies and Styles..." according to [6].

A new PDCA cycle (designated as $P_1D_1C_1A_1$) follows, with new plans for the next year (2013, Figure 17b) and the regression trend line based on (20.3):

$$\mathbf{y_{29/SRPS/2008\text{-}2012/P1} = -\,1821\,x^2 + 13159\,x - 7172}. \qquad (20.3)$$

Solving this issue at the state level results in a national model of excellence in e-learning, [7]. Future papers could deal with the incorporation of the high-ranking field of *electrical engineering* into the study curricula (concurrently with those in Europe and worldwide), and comparison of students' knowledge levels [8].

Conclusions

Based on the multicriteria analyses of knowledge units and the results presented through the examples of standardized electrical engineering subfields (ICS_1 = 29), and given also the confirmation of initial hypotheses, new conditions for future results/products have been provided.

1) According to the presented trends and dynamics of the development, original mathematical relations (1) to (20) have been derived and the trend lines have been presented. These provide an important theoretical aspect of the solution of the practical "problems".

2) Along with the analyses of knowledge base units in the field of electrical engineering, the correlation between local and international knowledge (SRPS to ISO) has been observed, whereas accessibility to individual teachers and associates in Serbia can be described as disputable.

3) The defined indices of quantity, value and ranking have enabled the comparisons above, as well as more efficient management of tasks and assignments related to respective scientific fields and related teaching subjects

at the higher education level. The ranking of *electrical engineering* is the basis for future work in conjunction with the new study curricula.

4) There is an obvious relationship in the index value between annual and cumulative knowledge innovation, as well as the need for alternative solutions to the knowledge innovation developed on the basis of standardized units.

Acknowledgement

The work presented here was supported by the Ministry of Education and Science of the Republic of Serbia – project III 44006, http://www.mi.sanu.ac.rs/projects/projects.htm#Interdisciplinary

References

[1] Institute for Standardization of Serbia: Advanced search: http://www.iss.rs/, http://www.iss.rs/standard/advance_search.php (accessed: 01. 01. 2013)

[2] ISO: ISO Store, http://www.iso.org/iso/home/store/catalogue_ics.htm, 29: Electrical engineering (accessed: 01. 01. 2013)

[3] P. B. Crosby, Quality is Free, McGraw-Hill (translated: Qualitass International, Beograd, 2010, p. 259)

[4] Ž. Micić, M. Demić, Knowledge Standardization in Road Vehicle Engineering, TTEM, Vol. 7, No. 3, 2012, 1281-1288

[5] Cluster Analysis, Ward's Method, © 2004 The Pennsylvania State University: http://sites.stat.psu.edu/~ajw13/stat505/fa06/19_cluster/09_cluster_wards.html

[6] P. Tóth, Learning Strategies and Styles in Vocational Education, Acta Polytechnica Hungarica, Vol. 9, No. 2, 2012, 195-216

[7] Ž. Micić, M. Tufegdžić, Knowledge Management Modeling to e-learning Excellence, TTEM, Vol. 6, No. 4, 2011, 1333-1344

[8] G. Kiss, The Concept to Measure and Compare Students Knowledge Level in Computer Science in Germany and in Hungary, Acta Polytechnica Hungarica, Vol. 5, No. 4, 2008, 145-158

Application of General Tool-life Function under Changing Cutting Conditions

János Kundrák, Zoltán Pálmai

Department of Production Engineering, University of Miskolc
Egyetemváros, H-3515 Miskolc, Hungary
janos.kundrak@uni-miskolc.hu; gepgyartas@uni-miskolc.hu

Abstract: The traditional Taylor formula $T=f(v_c)$ gives tool life of the cutting tool T as a function of the cutting speed at constant speeds in which the higher the speed, the lower the tool life. However, in a wider technological range, the function $T=f(v_c)$ has extreme values as well, which can be described by the general tool-life function. Frequently the tool is used until full wear of the edge at changing or alternating speeds, for which the general tool-life function can be extended. By using this extension, the general tool-life equation can be defined even under manufacturing conditions. Its applicability is demonstrated for examples of taper-turning and cutting optimisation.

Keywords: hard turning; tool life; economical cutting

1 Introduction

Several characteristics of machining are known, such as the surface quality of the workpiece, the shape of the chip, etc., and undoubtedly a great deal of attention is paid to tool life. Most often it is calculated by the well-known Taylor formula [1]

$$v_c \cdot T^{-\frac{1}{k}} = C_v \tag{1}$$

where v_c is cutting speed and T is tool life, while $-k$ and C_v are constants that are functions of the cutting conditions. Such conditions include the characteristics of the workpiece and the tool, the tool geometry and the cutting parameters (feed, depth of cut). On the basis of the Taylor formula several new research results have been obtained, such as the study by El Baradie [2] examining how the workpiece diameter influences the tool edge, or the examinations by Arsecuralante et al., who measured the volume of the material removed till the wear off of the tool edge [3], and who explained the tool wear by thematically activated processes [4]. Sahin [5], through a multiple regression analysis, attempted to apply a quadratic model that takes the hardness of the tool into account and also handles the product of the hardness and cutting speed, as well as the hardness and feed as separate technological variables. The cutting forces and chip morphology also were taken into account when examining coated tool materials as, for example, Chinchanikar and Choudhunry [6] did in their experimental study. Still, similarly to other cutting investigations, the common characteristic feature of these differently orientated studies applying different methods is that they summarise the results in Taylor-like formulas.

The tool life equation is just as important from the point of view of the theory and practice of cutting. When qualifying the surfaces of the workpieces, which is often complicated [7], the state of the tool also has to be defined accurately, in addition to the cutting parameters. In the case of complicated procedures like the 3D end ball milling of steels [8], the knowledge of tool wear is needed for the actual realisation of the theoretically expected surface roughness. In turning alloys which are difficult to machine, such as Inconel, the planning of tool wear is especially important [9], and when optimising the cutting technology it is essential [10].

However, there are two obstacles (limitations) to the general application of the Taylor formula. The first one is that v_c–T tool-life function is general, having extreme values but over a wide range of technological parameters rather than monotonously decreasing values, as is presented in Figure 1. The second obstacle is that the tool is used with alternating or changing cutting speed. A cutting test must be done where the cutting speed changes, both in the planning of practical technology and in the validation of the SEM simulation of tool wear [11]. In a case such as this not only the definition of tool life is questionable, but even its interpretation. Bridging these difficulties is the aim of this paper.

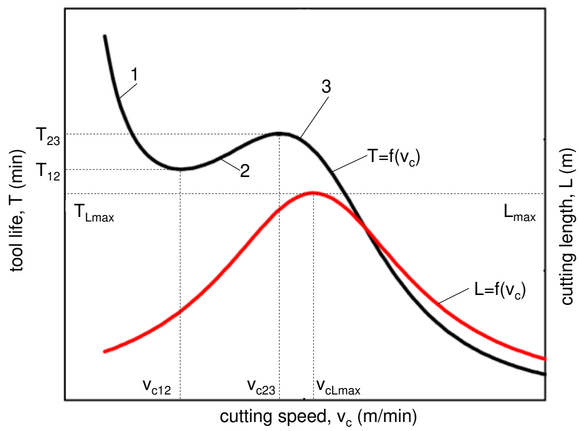


Figure 1

General tool-life function and the through cut length

2 General Tool-Life Function

Kundrák has demonstrated before that the general tool-life function worked out by him [12], which can be seen in Figure 1, describes the whole v_c–T function. Later other similar attempts became known [13] which rely on a rather complicated mathematical model, thus their practical application is laborious. The Kundrák formula that takes cutting speed into account can be written as:

$$T = \frac{C_{T1}}{v_c^3 + C_{T2} \cdot v_c^2 + C_{T3} \cdot v_c} \tag{2}$$

where C_{T1}, C_{T2} and C_{T3} are constants. At speeds v_{c12} and v_{c23} this function has an extreme value, which is naturally influenced by the technological parameters (feed f and depth of cut a_p):

$$v_{c12} = C_{v_{c12}} \cdot f^{x_{12}} \cdot a_p^{y_{12}} \cdot d_w^{q_{12}} \tag{3}$$

$$v_{c23} = C_{v_{c23}} \cdot f^{x_{23}} \cdot a_p^{y_{23}} \cdot d_w^{q_{23}} \tag{4}$$

$$T_{23} = C_{T23} \cdot f^{x_{T23}} \cdot a_p^{y_{T23}} \tag{5}$$

where d_w is the diameter of the workpiece, $C_{v_{c12}}$, $C_{v_{c23}}$, C_{T23} and x_{12}, y_{12}, q_{12}, x_{23}, y_{23}, q_{23}, x_{T23}, $y_{T_{23}}$ are constants that can be defined by technological examinations [12].

By their use already the constants of the general tool-life equation (1) can be calculated:

$$C_{T1} = T_{23} \cdot \left(v_{c23}^3 + C_{T2} \cdot v_{c23}^2 + C_{T3} \cdot v_{c23} \right) \tag{6}$$

$$C_{T2} = -\frac{3}{2} \cdot \left(v_{c12} + v_{c23} \right) \tag{7}$$

$$C_{T3} = 3 \cdot v_{c12} \cdot v_{c23} \tag{8}$$

It frequently occurs in technological planning practice that the length cut through the wear of the edge has to be examined or used, which, using (2), can be calculated by the formula:

$$L = T \cdot v_c = \frac{C_{T1}}{v_c^2 + C_{T2} \cdot v_c + C_{T3}} \tag{9}$$

This $L(v_c)$ function has only one extreme value (maximum) at speed

$$v_{L\max} = -\frac{C_{T2}}{2} \tag{10}$$

and its value can be calculated by

$$L_{\max} = \frac{C_{T1}}{0.75 \cdot C_{T2}^2 + C_{T3}} \qquad (11)$$

The tool-life function and cutting length defined in this way can be used in a wide range of technological parameters [14, 15] and thus one of the problems mentioned in the Introduction in connection with the Taylor formula can be eliminated. Later we provide a solution for the frequently occurring cases of cutting alternating in stages or changing continuously. Henceforth, tool life at fixed (unchanging) speeds is indicated with $T_{vc,fix}$, while it is indicated with $T_{vc,vari}$ at variable (changing) speeds.

3 Tool Life when Changing Cutting Speed

The solution is based on the principle of the proportional degradation of the tool. If speeds change in stages, which is the case presented in Figure 2, as cutting is done at different speeds the wear of the tool is naturally cumulative [16].

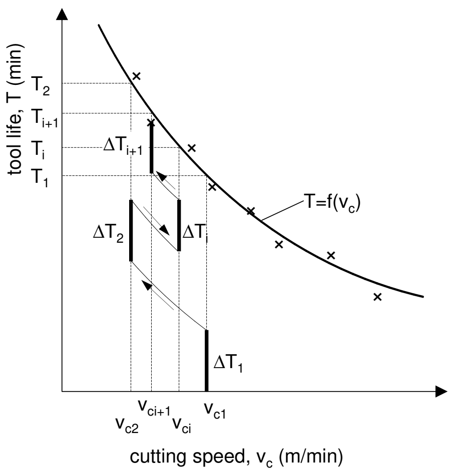


Figure 2

Tool life at changing cutting speed

It is a proved fact that if cutting is done for Δt_i time at cutting speed v_{ci}, the corresponding tool life is T_i, and the following equation is valid [16]:

$$\sum_{i=1}^{N} \frac{\Delta t_i(v_{ci})}{T_i(v_{ci})} \cong 1 \qquad (12)$$

where N is the number of the cutting stages [16].

Although it can easily be proved that by means of the Zorev power function

$$VB=C_{VB}\cdot t^{u} \tag{13}$$

(u=0.5-1) [17] the sum of Equation (12) can be described exactly as 1 in wear functions, this is valid only until inflection is reached on the wear curve. In wear curves in which there is inflection, too, Equation (12) can be considered as an approximation before reaching the wear defined as tool-life criterion. As is known, the degradation of the tool is the consequence of the accumulated influence of several processes, where occasionally a sudden disorder occurs such as chipping away or pitting [18]. Such cases are not covered by this investigation. If cutting speed changes continuously, as in cross turning or taper turning, the equation takes the form

$$\int_{t_1}^{t_2} \frac{dt}{T(v_c)} \cong 1 \tag{14}$$

where $v_c=v_c(t)$ is a function of time [19]. By connecting Equations (2), (12) and (14), a general tool-life equation can be gained which is valid for a wide range of cutting speeds, in the case of both constant and changing cutting speed. For cutting speed changing in stages

$$\sum_{i=1}^{N} \Delta t_i \cdot (v_{ci}^3 + C_{T2} \cdot v_{ci}^2 + C_{T3} \cdot v_{ci}) \cong C_{T1} \tag{15}$$

however, for continuously changing cutting speed (e.g. taper turning or cross turning)

$$\int_{t_1}^{t_2} \left[v_c^3(t) + C_{T2} \cdot v_c^2(t) + C_{T3} \cdot v_c(t) \right] dt \cong C_{T1} \tag{16}$$

Using these equations and the general tool-life equation (12) valid for the cutting speed changing in stages,

$$T = \frac{\sum_{1}^{N} \Delta t_i (v_{ci}^3 + C_{T2} \cdot v_{ci}^2 + C_{T3} \cdot v_{ci})}{v_c^3 + C_{T2} \cdot v_c^2 + C_{T3} \cdot v_c} \tag{17}$$

and using Equation (14) at continuously changing cutting speed

$$T = \frac{\int_{t_1}^{t_2} \left[v_c^3(t) + C_{T2} \cdot v_c^2(t) + C_{T3} \cdot v_c(t) \right] dt}{v_c^3 + C_{T2} \cdot v_c^2 + C_{T3} \cdot v_c} \tag{18}$$

Equations (17) and (18), once we know constants C_{T2} and C_{T3}, can be handled numerically.

These functions are valid if all conditions with the exception of cutting speed v_c remain unchanged throughout the tool life. T calculated in this way cannot be considered as tool life in the strict sense, since tool life, according to the definition, means the time of cutting during which at a given v_c cutting speed a predetermined amount of wear – the tool-life criterion – develops on the tool. However, here the cutting speed is not unchanged till the edge wears off, rather it changes either periodically or continuously. The usual v_c–T data pair may occur if, besides the cutting time calculated by Equations (17) or (18), using Equation (2) $v_{c,eq}$ is also calculated which is considered constant, and $T=T_{eq}$ can really be considered as tool life. This has significance in the economic or technological optimisation of productivity.

Essentially there are two basic cases of the practical utilisation of Equations (17) and (18) in the tool-life function (1):

a) C_{Ti} constants of Equations (17) and (18) are known and identical tools are used in operations carried out at either periodically or continuously changing speeds. Now, to define the number of workpieces that can be produced before the tool wears off N_p, using the data of the operation plan and also Equations (12, 14),

$$N_p = \frac{C_{T1}}{\sum_1^N \Delta t_i \cdot (v_{ci}^3 + C_{T2} \cdot v_{ci}^2 + C_{T3} \cdot v_{ci})} \tag{19}$$

and

$$N_p = \frac{\int_0^{t_p} \left[v_c^3(t) + C_{T2} \cdot v_c^2(t) + C_{T3} \cdot v_c(t) \right] dt}{k \cdot (v_c^2 + C_{T2} \cdot v_c + C_{T3})} \tag{20}$$

can be applied, where

$$k = \frac{t_p}{v_c} = \frac{L_p \cdot D_w \cdot \pi}{1000 \cdot v_c \cdot f}, \tag{21}$$

where D_w (mm) is the diameter of the workpiece, f (mm/rev) is feed, L_p (mm) is the length that can be cut by the piece, t_p is the operation time. These parameters can be used for data collection during production as well if the approach is reversed: the number of the different machined parts is registered for the tool used until it is worn off. The results can be used for the refinement of Equations (15) and (16) that is the C_{Ti} constants, as well.

b) C_{Ti} constants of the equation are not known. Then data $N_{p,i}$ can be collected for the values of Δt_i and v_{ci}. There will be $N_{p,i}$ data, too, for all the worn out tools, for which Equations (12) and (14) are true. Many equations (15), (16) offer themselves for the unknown C_{T1} and other constants as the worn off edges; that is, a linear equation system with a minimum of three equations will be dealt with, from which these constants can be calculated.

4 Cutting Model Experiment

Hard turning was completed by cutting an Ø45 mm inner cylindrical surface of 100Cr6 hardened steel (HRC=62±2) on an E400-1000 centre lathe. The length of the machined bore hole was L_p=30 mm. The characteristics of the tool: CBN K10, $\gamma=-5^o$, $\alpha=\alpha'=15^o$, $\lambda=0^o$, $\kappa_r=45^o$, $\kappa_r'=15^o$, the applied feed f=0.075 mm/rev., depth of cut a_p=0.1 mm. The constants of the tool-life formula given in Equation (2) related to the boring of this material are summed up in Table 1, and the v_c-$T_{vc,fix}$ tool life curve defined this way for cutting by fixed speed can be seen in Figure 3.

For the comparison of cutting at a fixed and at periodically changing speed, Table 2, for 5 chosen revolution values, displays the $N_{p,i}$ number of workpieces completed before a tool wears out; that is, tool life is indicated by the number of the turned bore holes. Table 3 summarises the model experiment series in which the number of pieces $N_{p,i}$ is measured that can be produced by turning at two speeds each, chosen from the 5 different speeds. Index I refers to the cutting speeds listed in Table 2, and $\Delta t_i=N_{p,i}\times t_{pi}$. These T_{meas} tool lives can also be seen in Figure 3.

Table 1

Constants of functions in Equations (3)–(5) with K10 tool [2]*

$C_{v_{c12}}$	$x_{v_{12}}$	$y_{v_{12}}$	$q_{v_{12}}$
1.06	-0.41	-0.40	0.19
$C_{v_{c23}}$	$x_{v_{23}}$	$y_{v_{23}}$	$q_{v_{23}}$
3.32	-0.24	-0.49	0.16
C_{T23}	$x_{T_{23}}$	$y_{T_{23}}$	$q_{T_{23}}$
13.99	-0.19	-0.09	0.48

*$[v_c]$=m/min, [f]=mm/rev., $[a_p]$=mm.

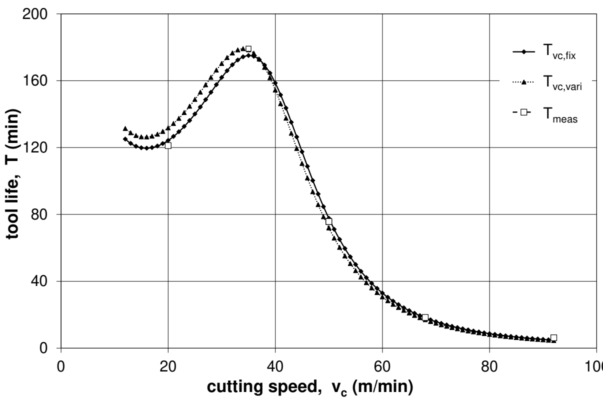


Figure 3

T-v_c curves defined by fixed speed ($T_{vc,fix}$), changing speed ($T_{vc,vari}$) and measured (T_{meas}) values

As Table 3 shows, the model experiment series consist of 3 variations each. The first part of the series covers the range of cutting speed where there is the maximum of v_c–T curve, at cutting speed v_{c23}.

Table 2

Technological parameters for fixed and periodically changing turning

Designation	Symbol	i=1	i=2	i=3	i=4	i=5
Number of revolutions	$n_{spindle}$ [1/min]	142	250	354	480	650
Cutting speed	$v_{c,i}$ [m/min]	20.06	35.33	50.02	67.82	91.85
Operation time	$t_{p,i}$ [min]	2.8169	1.6000	1.299	0.8333	0.6154
Piece number	$N_{p,i}$ [piece]	43	112	67	22	10
Tool life	$T_{vc,fix}$ [min]	121.1	179.2	75.7	18.3	6.15
	$T_{vc,vari}$ [min]	131.2	178	71.8	17.2	4.6
$T_{vc,fix}$: constant v_c, continuous cutting; $T_{vc,vari}$: variable v_c, periodical cutting						

Table 3

Number of bore holes turned with two speeds each until the wear off of the cutting edge

Number of experiment / Speed	1		2		3		4		5		6		$\Sigma\Delta t_i/T_i$
	$N_{p,1}$ [piece]	Δt_1 [min]	$N_{p,2}$ [piece]	Δt_2 [min]	$N_{p,3}$ [piece]	Δt_3 [min]	$N_{p,4}$ [piece]	Δt_4 [min]	$N_{p,5}$ [piece]	Δt_5 [min]	$N_{p,6}$ [piece]	Δt_6 [min]	
v_{c1} and v_{c3}	24	67.61	—	—	31	35.03	—	—	—	—	—	—	1.021
v_{c1} and v_{c2}	30	84.51	40	64	—	—	—	—	—	—	—	—	1.055
v_{c2} and v_{c3}	—	—	36	57.6	43	48.59	—	—	—	—	—	—	0.963
v_{c3} and v_{c4}	—	—	—	—	—	—	28	31.64	11	9.17	—	—	0.918
v_{c3} and v_{c5}	—	—	—	—	—	—	40	45.20	—	—	5	3.08	1.097
v_{c4} and v_{c5}	—	—	—	—	—	—	—	—	15	12.5	3	1.85	0.982
Numerical data of speeds from Table 2													

By means of Equation (15), in any part of the series 3 equations can be set up from which the constants C_{T1}, C_{T2}, and C_{T3} can be calculated. Table 4 summarises the constants defined from the results of the first part of the series.

The last column of Table 3 also contains the sums calculated by Equation (12), whose dispersion is indicated in Figure 4 on a so-called Gauss paper. Here on the vertical axis (ordinate) the scale was made on the basis of the error function; P_i, however, is the estimation of the error function calculated from the results obtained by the formula given in Figure 4. The dispersion of the data $\Sigma\Delta t_i/\Delta T_i$ is typical in cutting examinations; their average, however, closely approaches the value 1 given in Equation (9).

By means of the constants summarised in Table 4, the values can be calculated of $T_{vc,vari}$ in the v_c–T curve which can be defined by turning at periodically changing speeds. The bottom line of Table 2 contains this data, while the close relationship of values $T_{vc,fix}$ and $T_{vc,vari}$ is presented in Figure 5.

The number of the parts produced with the tool edge until reaching the allowed value of wear (e.g. VB) can also be calculated from the data above. For the variations indicated in Table 3, the number of the parts is shown in Figure 6.

Table 4

Constants of function (2) at speeds changing periodically*

Constant C_{Ti}	Value
C_{T1}	1334540
C_{T2}	-74.51304
C_{T3}	1596.56

*Calculated from v_{c1}-v_{c3} data

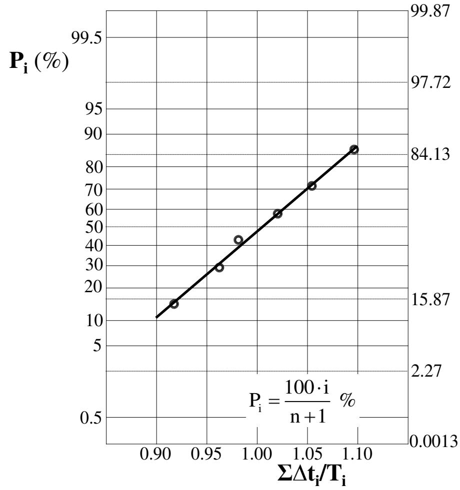


Figure 4

Dispersion of $\Sigma\Delta t_i/T_i$ values

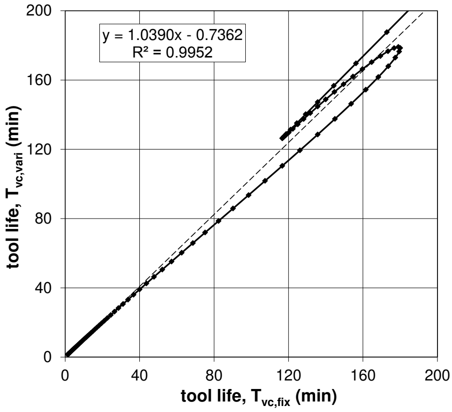


Figure 5
The relationship of tool lives stated by $T_{vc,fix}$ fixed speed and $T_{vc,vari}$ changing speed cutting

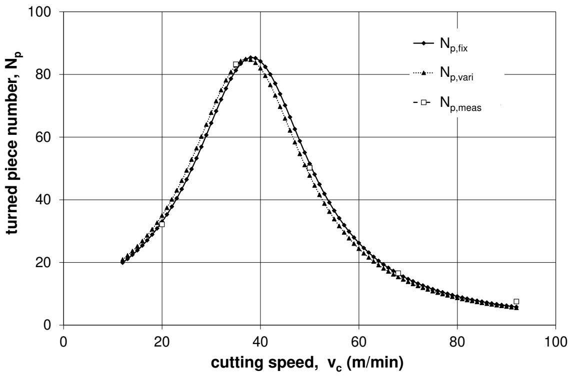


Figure 6
Number of parts as a function of cutting speed

5 Applications

The applicability of the general tool life function extended to the cutting with changing or alternate speed can be illustrated well in the examples of face turning and taper turning and also by the economic examination of tool performance.

5.1 Turning of Conical Surfaces

In the machining of internal cylindrical surfaces it has been found that constants C_{Ti} also depend on diameter of the workpiece D_w besides feed f and depth of cut a_p [12]. In these cases Equation (12) takes the form of

$$\sum_{i=1}^{N}\Delta t_i \cdot \frac{v_c^3 + C_{T2i}\cdot v_c^2 + C_{T3i}v_c}{C_{T1i}} \cong 1 \tag{22}$$

with continuously changing cutting speed; however, in face turning or taper turning, for example, it takes the form of

$$\int_0^{t_p}\frac{v_c^3 + C_{T2}\cdot v_c^2 + C_{T3}\cdot v_c}{C_{T1}}\,dt \cong 1 \tag{23}$$

In the frequent case when N_p parts are produced in sequence until the edge is worn off, on the right side of Equations (21) and (22) there is $1/N_p$ instead of 1.

For an example of the continuous change of cutting speed, the task is the hard turning of an internal cylindrical surface with a 1:5 taper made of 100Cr6 (HRC60±2) hardened steel. In this example the length of the cone is ℓ=40 mm, the tool geometry, the feed f and depth of cut a_p are the same, however, diameter D changes:

$$D(t) = D_0 - \frac{n\cdot f}{5}t \tag{24}$$

Therefore here the values v_{c12}, v_{c23} and T_{23} defining C_{Ti} constants also depend on time. Putting technological data into Equations (3)-(5) we obtain

$$v_{c_{12}}=7.7007\cdot D^{0.19}$$
$$v_{c_{23}}=24.1192\cdot D^{0.16} \tag{25}$$
$$T_{23}=101.6348\cdot D^{0.48}$$

with which the constants C_{T1}, C_{T2}, C_{T3} can be calculated. The number of parts N_p that can be produced until the tool edge has been worn off can be simply calculated with a numerical method by means of Equation (23)

$$\int_0^{t_p}\frac{v_c^3 + C_{T2}\cdot v_c^2 + C_{T3}\cdot v_c}{C_{T1}}\,dt \cong \frac{1}{N_p}$$

The cutting time before the tool wears off

$$T_{cut}=N_p\times t_p, \tag{26}$$

where $t_p=\ell/(f\times n)$. The results of the calculation are shown in Figure 7. As can be seen, the character of the v_c-T_{cut} function can be compared with tool life T according to Figure 3; however, because of the continuous change of speed in taper turning, the locations of the extreme values are somewhat modified: they are shifted towards the higher values, however, their maximums are almost the same.

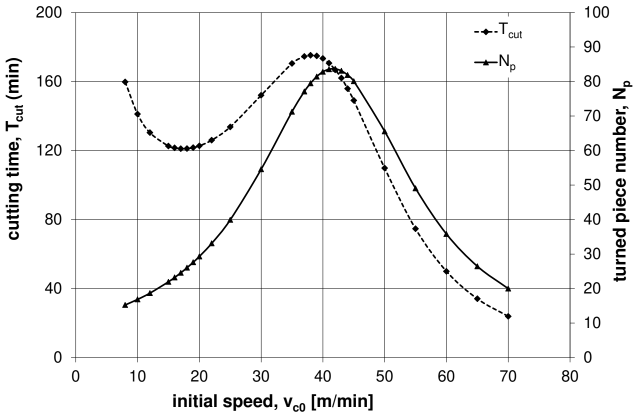


Figure 7

Number of parts and cutting time machined by taper turning until the wear off of the edge

5.2 Economics of Cutting

Several points of view have to be considered in economic examinations of cutting; however, in the present case the aim the economic analysis is to explore how the tool life influences the economy and productivity. So at first approach, the other cost factors of production beyond the technological parameters and the related different technological limits can be ignored.

Productivity can be characterised in the simplest way by the material volume cut during a time unit, that is, it can be calculated by

$$Q = \frac{a_p \cdot f \cdot v_c}{t_{ch} + T} \cdot T \ \text{(cm}^3\text{/min)} \tag{27}$$

where t_{ch} is the time needed to change the tool.

To define the cutting speed economically from the point of view of tool life, it is expedient to calculate the cost of cutting of the material quantity by unit of volume:

$$K = \frac{1}{a_p \cdot f \cdot v_c} \left[k_m + \frac{t_{ch} \cdot k_m + K_{tool}}{T} \right] \ \text{(€/cm}^3\text{)}, \tag{28}$$

and here k_m is the so-called minute cost of the machine tool, however K_{tool} is the cost of one edge of the tool.

As is known, these specific economic indexes have extreme values depending on the cutting speed, the position of which can simply be expressed with formulas; in the case of the Taylor formula, by calculating the tool life that is optimal from the point of view of productivity and economy. In the case of the general v_c-T tool life function the calculation of this optimal tool life would need the solution of a cubic equation originating from Equations (2) and (17), which can be handled by the Cardano method – a program can be created for it, but that would be rather labor intensive. However, by means of a direct numerical method, e.g. an ordinary spreadsheet program, all the necessary information on productivity and economy can be gained.

Functions (27) and (28) can be seen in Figure 8 with the experimental data of turning discussed in Section 4, where the time of the tool change is t_{ch}=5 min, the cost of the edge is K_{tool}=4 €, the minute cost of the machine tool is k_m=0.25 €/min and $a_p \times f$=0.0075 mm^2. So that the characteristic curve of $\mathbf{v_c}$ cutting speed and *K* specific cost of cutting as well as *Q* productivity can be illustrated in a diagram: Figure 8 shows a 20-fold *K* specific cost and a 120-fold *Q* specific chip output.

Here tool life *T* is an independent variable of the calculation; in reality, however, the aim is to define the cutting speed that is optimal from the point of view of productivity or expenditure. The characteristic curve of economic indices can be evaluated on the basis of Figure 8, then from the related spreadsheet table the $v_{c,opt}$ value, which is in Table 5, can be read with optional accuracy. It can be seen that the specific economic characteristics defined for turning with fixed speed and for turning with periodically changing speed are almost the same.

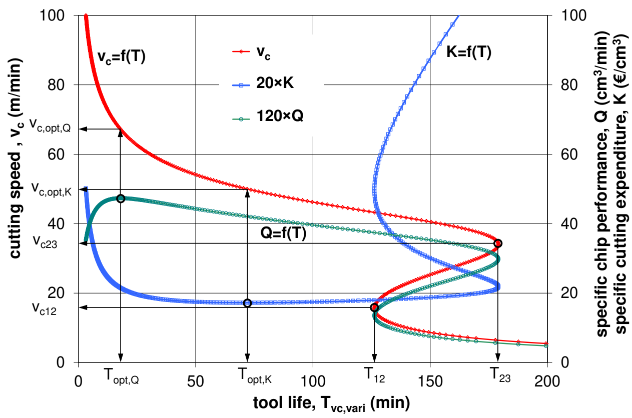


Figure 8

Specific cutting expenditure (K, €/cm^3) and chip performance (Q, cm^3/min) in hard boring

It follows from Equation (28) that the increase of tool expenditure K_{tool} increases the tool life optimal for economy. In an extreme case it may occur that $T_{opt,K}$ slides in between the borders of T_{12}–T_{23}, where – as can be seen in the T–v_c curve – up to three values offer themselves for the speed and consequently for machining cost *K*. Obviously, the most economically beneficial one is to be chosen from them if the other circumstances make that possible. When the application of optimal cutting speed defined in that way encounters some barriers, it can easily be decided what technology may closest approach the optimum.

Table 5

Comparison of specific characteristics of cutting in continuous cutting and cutting with changing speed

Cutting method	$v_{c,opt,Q}$ [m/min]	$T_{opt,Q}$ [min]	Q_{opt} [cm^3/min]	$v_{opt,K}$ [m/min]	$T_{opt,Q}$ [min]	K_{opt} [€/cm^3]
Cutting with fixed speed	67.5	18.8	0.400	51.0	71.1	0.847
Cutting with changing speed	67.8	17.3	0.400	49.3	76.9	0.881

Summary

The most important point of view of cutting ability is tool life. Usually it is described by the well-known Taylor formula, which, in the basic case gives the tool life as function of the cutting speed, but this is only valid for a monotonously decreasing $T–v_c$ curve. To describe a $T–v_c$ curve having extreme values as well, we introduced a general tool life function proved by cutting experiments. A frequent problem is, however, that cutting goes on sequentially with the same tool at several different speeds, and then this formula can be used only with difficulty. With periodically changing cutting speed the equation $\Sigma\Delta t_i/T_i \cong 1$ is valid theoretically and experiments have proven this. From this a general form of $T–v_c$ curve can be deduced which can also be used with cutting at periodically changing cutting speeds. In the case of cutting at different speeds, an equivalent speed and the related tool life can be defined which can already been handled by the general Kundrák tool-life function. By the calculation of the equivalent speed it is possible to define the $T–v_c$ curve under manufacturing conditions, even in cutting with changing speed, and the general tool life function can be defined by operational measurements. The applicability of this function is demonstrated by examples of taper turning and cutting optimisation.

Acknowledgements

This research was (partially) carried out in the framework of the Centre of Excellence of Innovative Engineering Design and Technologies at the University of Miskolc. The work was presented with the support of the Hungarian Scientific Research Fund (Number of Agreement: OTKA K 78482), which the authors greatly appreciate.

References

[1] ISO Standard 3685 Tool-Life testing with Single-Point Turning Tools

[2] El Baradie M. A.: The Effect of Varying the Workpiece Diameter on the Cutting Tool Clearance Angle in Tool-Life Testing. Wear 195 (1996) pp. 201-205

[3] Arsecularatne J. A., Zhang L. C., Montross C., Mathew P.: On Machining of Hardened AISI D2 Steel with PCBN Tools, Journal of Material Processing Technology 171 (2006) pp. 244-252

[4] Arsecularatne J. A., Zhang L. C., Montross C.: Wear and Tool Life of Tungsten Carbide, PCBN and PCD Cutting Tools, International Journal of Machine Tools and Manufacture 46 (2006) pp. 482-491

[5] Sahin Y.: Comparison of Tool Life between Ceramic and Cubic Boron Nitride (CBN) Cutting Tools when Machining Hardened Steel, Journal of Material Processing Technology 209 (2009) pp. 3478-3489

[6] Chinchanikar S., Choudhury S. K.: Investigations on Machinability Aspect of Hardened AISI 4340 Steel Different Levels of Hardness Using Coated Carbide Tools, Int. Journal of Refractory Metals and Hard Materials 38 (2013) pp. 124-133

[7] Fekete G., Horváth S., Czifra Á.: Microgeometry Tests of 'Contradictory' Surfaces with Various Evaluation Techniques, Acta Polytechnica Hungarica, Vol. 4, No. 2, 2007, pp. 87-97

[8] Mikó B., Beňo J., Maňková I.: Experimental Verification of Cups Heights when 3D Milling Rounded Surfaces, Acta Polytechnica Hungarica, Vol. 9, No. 6, 2012, pp. 101-116

[9] Khan S. A., Soo S. L., Aspinwall D. K., Sage C., Harden P., Fleming M., White A., M'Saoubi R.: Tool Wear/Life Evaluation when Finish Turning Inconel 718 Using PCBN Tooling, 5th CIRP Conference on High Performance Cutting, Procedia CIRP Vol. 1, 2012, pp. 283-288

[10] Aykut S., Kentli A., Gülmez S., Yazıcıoğlu O.: Robust Multiobjective Optimization of Cutting Parameters in Face Milling, Acta Polytechnica Hungarica, Vol. 9, No. 4, 2012, pp. 85-100

[11] Salvatore F., Saad S., Hamdi H.: Modelling and Simulation of Tool Wear During the Cutting Process, 14th CIRP Conference on Modelling of Machining Operations (CIRP CMMO) Procedia CIRP 8 (2013) 305-310

[12] Kundrák J.: The Scientific Principles of Increasing the Effectiveness of Inner Surfaces Cutting with CBN Tools, Kharkov, 1996, p. 368

[13] Pluta Z., Hrynieviez T.: Advanced Model of the Tool Edge Blunting under Machining, Int. J. Manuf. Technol. 51 (2010) pp. 35-43

[14] Mamalis A. G., Kundrák J., Horváth M.: Wear and Tool Life of CBN Cutting Tools, Int. J. Adv. Manuf. Technol. 20 (2002) pp. 475-479

[15] Kundrák J., Pálmai Z.: Investigation of Machinability in Boring of Inner Cylindrical Surfaces of Hardened Bearing Steels, Proc. ICT-2012, 13th International Conference on Tools, 2012 Miskolc, Hungary, pp. 165-170

[16] Pálmai Z.: Anwendung der Taylorschen Gleichung auf die Zerspanung mit wechselnd sich Änderungen Schnittgeschwindigkeiten. Archiv für das Eisenhüttenwesen 49 Nr. 2 (1979) pp. 89-93

[17] Zorev N. N., Granovskij G. I., Loladze T. N., Tretyakov I. P.: Razvitie nauki o rezanii metallov, Mashinostroenije, Moscow, 1967, p. 413

[18] Olovsjö S., Nyborg L.: Influence of Microstructure on Wear Behaviour of Uncoated WC Tools in Turning of Alloy 718 and Waspaloy, Wear 282-283 (2012) pp. 12-21

[19] Pálmai Z.: Standzeit des Werkzeuges beim Plandrehen. Archiv für das Eisenhüttenwesen 49 Nr. 10 (1984) pp. 499-502

Monetary Policy in the Face of the Eurocrisis How does the European Central Bank Absorb Excess Liquidity?[1]

Marc Piazolo

International and Monetary Economics
University of Applied Sciences Kaiserslautern
Graduate School of Management
Amerikastr. 1, D-66482 Zweibrücken, Germany
marc.piazolo@fh-kl.de

Abstract: The objective of this paper is to demonstrate the mechanism by which the European Central Bank (ECB) has been successfully absorbing excess liquidity – albeit to a small extent. These Fine Tuning Operation (FTO) were introduced to neutralize the liquidity effects of the Securities Markets Programme initiated in May 2010. With an out-of classroom experiment, students simulated investment decisions for financial institutions which needed to park excess liquidity with the ECB through an auction process that replicates the Fine Tuning Operations in place. The auctions showed to be very competitive, thereby enhancing the students understanding of the real word behaviour of major European financial institutions. A thorough analysis of the bidding results and the various determinants of total profits is presented. We start with describing as well as analyzing the evolution of the ECBs' monetary policy instruments.

Keywords: Financial crisis; monetary policy instruments; auction design; experimental economics

Introduction

The financial crisis of 2007 started out in the US subprime market (real estate finance) - a market segment of relative limited scope. As US housing prices started to come down, several subprime originators failed. When Lehman Brothers had to file for Chapter 11 bankruptcy protection on 15 September 2008, the previous market turmoil escalated into a full-fledged global crisis. As a growing number of major financial institutions were facing the risk of default, a crisis of

[1] I thank two anonymous referees for their valuable comments, though I am fully responsible for the remaining shortcoming of the article.

confidence resulted - quickly spreading across markets and countries.[2] The run on money market mutual funds triggered "shareholders redemptions that resembled a bank run" (BIS 2009 p. 26). The crisis of confidence hit most of the financial intermediaries and quickly spread via the credit channel to the real sector of the economy (corporates and private households). Only when six major central banks engaged in unprecedented interventions and coordinated joint reductions in their policy rates did the financial markets stabilize at a much more fragile level.

The worst global crisis since the Great Depression led world output drop for the first time since the late 1940ies – as real GDP contracted by 0.6 % (2009). At the same time, the breakdown of international trade was even more pronounced – trade in goods and services contracted by almost 11 % in 2009. Therefore, open and trade-oriented countries were hit hard by the contraction of international import and export flows. Germany and Japan experienced sharper drops in output than the United States (-3.1 %) even though their financial sector remained relatively sound.[3] Crisis countries with fragile banking systems, large current account deficits and subsequent rising public debt ratios had to go through a much more pronounced recession. For some of them, the end of their plight is still not in sight; e.g. Greece, Portugal or Spain.

The major advanced countries quickly coordinated their fiscal stimuli packages.[4] They undertook measures to stabilize the financial sector through recapitalization, liability guarantees and asset purchases. Importantly, they did not resort to damaging protective policy measures as during the Great Depression. The major central bank's coordination was even more extensive – they sharply reduced interest rates and provide extensive liquidity support.[5] Rapid growth in Asia as well as in other developing country regions led to a sharp revival of world trade (+12.5 %) in 2010. Combined with domestic expansionary fiscal and monetary measures, real GDP growth for most advanced crisis countries jumped forward again. Within the Euro Area, Germany's rebound was the strongest, while advanced countries as a group experienced a growth rate of 3.0 % (2010). [6] Due to the short economic contraction in world output, the financial crisis 2007-2013 is often also labeled as the Great Recession (2008-2009).[7]

2 For an overview, Gorton; Metrick (2012) and Mishkin (2011) make a good read. In Table 1 of Gorton; Metrick (2012), a financial crisis major events timeline is presented. Another one is to be found in BIS (2009 Table II.1).

3 Germany contracted by 5.1 %, though small open economies like Hungary (-6.7 %) or Finland (-8.5%) were hit even harder (IMF 2013 Table A2 and A4).

4 Fiscal stimuli seemed to be of less importance than the stabilization of the financial sector (Mishkin 2011 p. 64).

5 See Gorton; Metrick (2012 Table 2) for a summary of supportive measures by government and central banks.

6 Germany's real GDP expanded 4.0 % in 2010 and 3.1 % in 2011 and was mainly export and investment driven (IMF 2013 Table A2).

7 See Schularick; Taylor (2012) p. 1042.

The quick recovery of the real sector in 2010-11 was in no way paralleled in the financial sector, and it did not cover all countries. In Europe, a major sovereign debt crisis in the countries of the periphery emerged in late 2009. The debt levels of Greece, Portugal and Ireland had become unsustainable.[8] In May 2010 Greece was the first country in need of an international financial aid package (three year lending facility, \$110 bn). The crisis management of the Euro Area countries was in the hands of a Troika - European Commission, ECB and International Monetary Fund (IMF). Especially the German government had been very keen on for an active role of the IMF. Besides participating with its own share of \$30 bn, the IMF was seen as more experienced and independent. Due to the threat of contagion, further stabilization measures were adopted. The EU finance ministers established the European Financial Stability Mechanism (EFSM €60 bn) and the Euro Area states agreed on the European Financial Stability Facility (EFSF €440 bn). On a case-by-case basis, the IMF could top-up the financial aid coming from the two European funds by 50 %.[9] These decisions were made end of May 2010, in hope to calm down the turmoil on international financial markets. Ireland was the first country to slip under this new "European financial umbrella" in November 2011 (€67.5 bn). Portugal, Greece, Cyprus and Spain were to follow by mid 2013. As it became apparent that temporary measures were not successful enough, the European Council decided to establish a permanent crisis management mechanism. The European Stability Mechanism (ESM €700 bn) enter into force on July 1 2013, thereby replacing the measures mentioned before.[10]

After the US subprime crisis (2007) and the bankruptcy of Lehman Brothers (2008), the European sovereign debt crisis (2010) emerged as the main challenge for policy makers in Europe. The paper starts to look into the monetary policy reaction of the European Central Bank (ch. 1), followed by a discussion on the economic effects of various policy instruments (ch. 2). In chapter 3, the monetary policy experiment is presented. It includes the set-up as well as the analysis of the findings. Finally, the learning outcomes of the classroom experiments as well as broad implications on the effectiveness of monetary policy instruments are discussed. A brief outlook on future monetary policy developments is given.

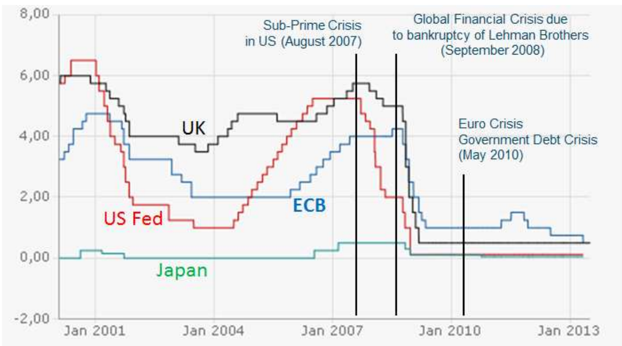
8 Lane (2012 p. 51f.) provides a overview on the accumulation of European countries sovereign debt and its link to financial as well as external imbalances.

9 More specifics to these measures can be found in Bundesbank (2011a) p. 90 and in ECB (2011b) p. 75.

10 To limit the potential liability of the Euro Area states until the EFSF has run out, total consolidated ESM and EFSF lending may not exceed €500 bn (ECB 2011b p. 74).

1 Financial Crisis and ECB's Monetary Policy Reactions

Based on the experience of the Great Depression, John M. Keynes recommended for central banks to initiate anti-cyclical policy actions by providing liquidity to the economy in order to dampen the detrimental effects of an imminent recession in the real sector. In the economic aftermath of the bust of the dot.com bubble (2000), the US Federal Open Market Committee (FOMC) cut the federal funds rate aggressively in 2001 to cushion the fall in demand and employment. After the 9/11 terror attack, all major central banks accommodated the rise in uncertainty by providing liquidity to the financial system in a coordinated move. The sharp expansionary policy stance prevented a worldwide recession to emerge. As expectations were well anchored, there was no significant hike in inflation.



U.S. federal funds target rate 0-0.25 %; Bank of Japan 0-0.1 %; UK base rate 0.5 %; ECB main refinancing rate 0.5 %. www.finanzen.net (05 August 2013)

Figure 1
Leading Interest Rates of Major Central Banks (2000-2013)

Figure 1 presents the development of the leading interest rates of major central banks: the US Federal Reserve, Bank of England, Bank of Japan and the European Central Bank. During the last 15 years the variation in interest rates has been a very close one. Only the degree of variation in national rates differs significantly. Traditionally, decisions of the US FOMC set the pace for the other central banks to follow.[11] Essentially, central banks aim to determine short-term market interest rates – money market rates - by influencing the cost of interbank lending. Through various channels of transmission, they want to target real economic activity, and eventually consumer price inflation.[12]

11 One exception was the temporary rise of the ECBs' main refinancing rate in 2011, before a sharpening of the debt crisis led to a reconsideration of its interest rate policy.

12 Transmission mechanisms from rates to prices are given in ECB (2011c p. 59).

Following the bankruptcy of Lehman Brothers and the subsequent crisis of confidence all major central banks cut down their leading interest rates fast and substantially – until they reached the zero-bound. They started off in a concerted action, when six central banks reduced their rates by 50 basis points on October 8 2008.[13] These standard monetary policy actions are often asymmetric in nature – slowly raising interest rates, when the economy is in danger of overheating, and rapidly reducing the rates, when the economy is over the cliff (*Figure 1*). This is called the "mop-up" approach in monetary policy.[14]

Growing uncertainty about the financial health of major commercial banks (loss of confidence) led to a collapse of the money-market and other financial markets. Interbank lending virtually came to a halt. Central banks quickly provided massive liquidity and introduced various kinds of non-standard measures to reduce systemic risks as well as to stabilize financial markets. In addition, the channels through which interest rate decisions were normally transmitted to the economy came under severe strain.[15] As the key interest rates reached the zero lower bound, the monetary authorities' ability to further reduce real interest rates is rather limited since a reduction in nominal rates is not possible anymore. Due to the above reasons, the central banks had to undertake other, more innovative measures.

The size of non-standard measures undertaken by the monetary authorities is reflected in the "explosion" of central bank balance sheets. Changes in the minimum reserve requirements and the purchase of government bonds also led to an expansion of balance sheets. In percent of GDP, the balance sheet of the ECB almost tripled over the course of the global financial crisis, peaking at 27 % of GDP in 2012. The increase in liquidity provision to a country's economy falls under the term *Quantitative Easing*. At the same time, the balance sheet underwent significant structural changes (*Qualitative Easing*). Parts of the structural changes were the considerable extension of maturities (refinancing operations) as well as the reduction in quality standards for collateral central banks would accept.

Table 1 presents the various non-standard measures implemented by the ECB during the course of the financial crisis. They either fall into money market-based measures – with a focus on general liquidity provision - or securities market-based measures – focusing on providing liquidity to specific, dysfunctional market

[13] Canada, European Central Bank, Sweden, Switzerland, United Kingdom, United States (ECB 2010b p. 65).

[14] See Blinder (2010) for a thorough discussion of how the "mop-up approach" evolved, and when it should be implemented.

[15] The financial crisis impaired the pass-through from official to short- and long-term market rates, banks found it hard to access funding, bank's borrowers experienced a fall in assets, and finally creditworthiness as well as risk perceptions changed (ECB 2011a p. 55; Woodford 2010 p. 37 and Figure 5).

segments. ECB studies found the (money market-based) non-standard measures to be quite successful in dampening the recession as bank lending rates to households and non-financial corporations declined with short delay, and the level of lending contracted only moderately (ECB 2011a p. 61). Though, the differences in development between periphery and core countries remained substantial.

Due to the continued crisis of interbank confidence (counterparty risk), many financial institutions asked the ECB for much more liquidity than they actually worked with. Excess liquidity ballooned to €800 bn, when the two 36-month refinancing operations took place (12/2011 and 03/2012). The banks reduced their excess liquidity especially when the overnight deposit rate at the ECB dropped to zero and (unlimited) outright monetary transactions were announced.[16]

Table 1

Non-standard Measures of the ECB (2008-2013)

	Money Market-Based	**Securities** Market-Based
Fixed-rate full allotment in		
- main refinancing operations	2008	
- longer-term refinancing operations	2008	
Special maintenance-period operations	2008	
Supplementary LTROs	2007	
6 / 12 / 36-month LTROs	2008 / 2011	
USD / SFR providing operations	2007 / 2008	
Covered Bond Purchase Programme (CBPP)		2009
Securities Markets Programme (SMP)		2010-2012
Reduction of Reserve Requirements (to 1%)	2012	
List of eligible collateral extended	2008	
Outright Monetary Transactions (announced)		2012

LTROs = Longer-Term Refinancing Operations (year of introduction – or phase-out); based on ECB (2011a) p. 57 and Deutsche Bundesbank (2013) pp. 47-48

As the global financial crisis calmed down in early 2010, the Governing Council of the ECB planned to return to normal ("standard times") by reducing the scale of longer-term refinancing operations as well as changing the auction design of the refinancing operations to variable rates and limited allotment (pre-crisis design). But in May 2010 the European sovereign debt crisis hit, unsettling international financial markets further. So, the ECB prolonged the non-standard measures and expanded them in scope as well as in scale by fall 2012 (*Table 1*). To support dysfunctional financial markets within the Euro Area the Governing Council decided on May 10 2010 to implement the Securities Markets Programme (SMP). It targeted illiquid public and private debt securities markets – mainly government bond markets of countries on the European periphery like Portugal, Italy, Ireland,

[16] Deposit facility: 0 % (11.07.2012); Governing Council decided on OMTs (06.09.2012). In March 2013, excess liquidity of the Eurosystem stood at €400 bn (Deutsche Bundesbank 2013a p. 46).

Greece and Spain. The Eurosystem purchases are strictly limited to secondary markets – therefore being "kind of" in line with no direct financing of fiscal deficits. Direct purchases of securities do not provide additional liquidity to the Euro Area. They have to be fully sterilized by conducting liquidity-absorbing operations.[17] The SMP was heavily disputed in the German public as well as within the Governing Council – especially by its two German members. The friction on this fundamental policy issue led to major personal consequences and the election of Mario Draghi as the current president of the ECB.[18]

2 How do ECB‘s Open Market Operations Work?

Open market operations are central in implementing the monetary policy of the Eurosystem. They are used to target interest rates and the liquidity of Euro Area's financial markets. In addition, these operations provide crucial information to market participants on future monetary policy actions of the ECB. Their function of signalling medium-term policy developments has become much more important lately.[19] Up to fall 2007 its main refinancing operations – with a maturity of one week – had been dominating the liquidity management. Between 60-75 % of all liquidity provided to the economy were short-term in nature. After the outbreak of the global financial crisis in October 2008, the focus shifted towards longer-term refinancing operations. Herewith, reducing the uncertainty of financial intermediaries in tapping liquidity they needed. Instead of using interbank lending, the Central Banks provided them with direct access to liquidity over a longer term.

Since the start of the Eurosystem in January 1999 up until the outbreak of the global financial crisis, the ECB had already undertaken several important changes in implementing its main refinancing operations. *Table 2* gives an overview of the various structural changes for liquidity providing operations within the whole

17 ECB (2010b) S. 74.

18 Mario Draghi was the unanimous candidate for the ECB only, after Axel Weber dropped out as the prefered candidate unexpectably in the spring of 2011. Weber stepped down as president of the Bundesbank due to the monetary policy decisions the ECB had taken. In September 2011, the chief economist and German board member of the ECB, Jürgen Stark, announced to end his career with the ECB before his term of service was over.

19 The former member of the Board of Governors of the Federal Reserve, F.S. Mishkin (2011 p. 62), stresses the importance of monetary policy to focus on the management of market participants' expectations. In his July 4 2013 statement to the press, the president of the ECB (M. Draghi), for the first time, gave explicit *forward guidance* towards the interest rate policy of the Governing Council: "Looking ahead, our monetary policy stance will remain accommodative for as long as necessary. The Governing Council expects the key ECB interest rates to remain at present or lower levels for an extended period of time." (Draghi 2013).

period of 1999-2013. The main refinancing operations are liquidity-providing reverse operations that are executed regularly by the Eurosystems National Central Banks (NCBs) through standard tenders.[20] In the case of *fixed rate (volume) tenders*, the central bank specifies the interest rate (i_{fix}) in advance and the participating counterparties (e.g. commercial banks) bid the amount of money they want to transact at the specified rate. During the period 1999-2000 the bid of the counterparties usually exceeded the volume of money the ECB provided the banking sector with. Therefore, the submitted bids were satisfied pro rata – according to each bids share of the aggregate amount. The allotment ratio reflects the amount to be allotted to the aggregate amount bid. The competition between the counterparties led to totally unrealistic and therefore non-informative bids ("moon bids"). On June 7 2000, counterparties bid for €8,491.2 bn while the ECB allotted only €75 bn - the allotment ratio being a meagre 0.9 %.[21]

Table 2

Tender Procedures for Main Refinancing Operations (1999-2013)

Start of Period	**Fixed Rate** Tender (ltd Allotment)	**Fixed Rate** Tender (Full Allotment)	**Variable Rate** Tender (American Auction)	Maturity (# of days)
07.01.1999	X*			14 **
28.06.2000			X	14 **
10.03.2004			X	7
15.10.2008		X		7

*ltd – limited; * with extremely small allotment ratios; ** overlapping – with each week a new tender*
Deutsche Bundesbank (2011c); ECB (2011c)

Due to continued overbidding the Governing Council decided to switch to a more market-oriented procedure by the end of June 2000: *variable rate (interest) tenders*. The ECB specifies (usually) a minimum interest rate (i_{min}). With this rate in mind, the counterparties bid the amounts of money and the interest rates at which they want to enter into transaction with the Eurosystems NCBs. Counterparties are not limited to submit only a single bid. They can submit up to ten bids that differ in the amount of money and the interest rate.[22] The ECB ranks all incoming bids in descending order (liquidity-providing operations) based on the offered interest rate. In general, the bids with higher interest rates offered are satisfied first and to full extent. Bids with successively lower rates are accepted until the total liquidity to be allotted is exhausted. If the aggregate amount of bids at the lowest acceptable rate – marginal interest rate ($i_{marginal}$) – exceeds the

20 A detailed presentation of the different forms of monetary policy instruments and procedures can be found in ECB (2011d).

21 Deutsche Bundesbank (2011c). The maximum allotment ratio is 100 %. Frenkel et al. (2003) analyze the process of overbidding in more depth.

22 The interest rates bid must be expressed as multiples of 0.01 percentage point and the minimum bid amount is €1 mn (ECB 2011d L331/23).

amount of liquidity still to be allotted, this remaining amount is allocated pro rata.[23] All bids with interest rates below the marginal interest rate do not receive any allotment. As these variable rate tenders proceed with an *American auction* design, the bidding counterparties have to pay their individually interest rates offered. By switching the procedure in June 2000, overbidding was sharply reduced. The allotment ratio rose substantially, whereby the information content of the bids for the ECB also increased significantly.[24]

On October 15 2008, the European Central Bank switched back to *fixed rate (volume) tender*, though this time all of the counterparties bids are fully allotted. It covers the main refinancing operations as well as the longer-term refinancing operations. As the financial institutions knew that they would receive any amount they were bidding for, the problem of extreme overbidding did not arise. There are still two reasons that account for the excess liquidity: firstly, the liquidity management of each individual counterparty wants to be on the safe side – therefore, they will bid for larger amounts than necessary. Secondly, due to the crisis in confidence the banks reduced their usage of interbank lending, which increased their demand for ECB's liquidity provision. Since liquidity was provided for a longer-term (up to 3 years) excess liquidity should prevail, even if the confidence crisis subsides and interbank lending picked up.

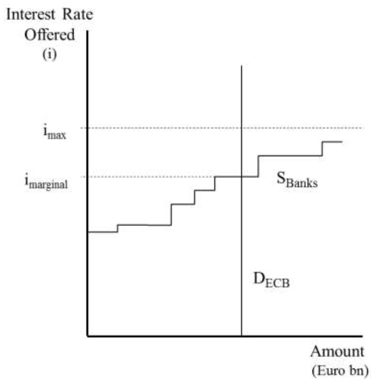
Direct purchase of securities by the ECB through the Covered Bond Purchase Programmes, the Securities Markets Programme and the Outright Monetary Transactions should under no circumstances provide additional central bank liquidity to the financial system. Special Fine Tuning Operations (FTOs) were initiated to absorb the provided liquidity by exactly the same amount – the purchase programmes were fully sterilized. Practically, this was (still is) done through a *variable rate tender* with an *American auction*.

In *Figure 2* such a procedure is presented graphically. The ECB announces a tender operation and specifies a maximum interest rate. Then, the counterparties offer excess liquidity to the ECB; e.g. the latter collects fixed-term deposits for one week. As the interest rates offered represent costs for the ECB, the central bank tries to minimize these costs by accepting the bids with the lowest interest rates first. Subsequently, bids with higher interest rates are accepted until the total liquidity to be absorbed is exhausted.

---

23 Additional information can be found with ECB (2011d L331/24).

24 The various tender procedures are discussed in Frenkel et al. (2003 p. 112f). Numerical examples are given in ECB (2011d) Appendix 1.



Collection of Fixed Deposits

Tender-ID: 20130188	
Quick Tender (variable rate tender, American auction)	
Submission date for bids:	27.08.2013 at 12:05
Maturity:	7 days
Maximum interest rate bid:	0.50 %
Minimum amount being allotted:	-
Minimum amount of bid:	€1.00 mn
Targeted allotment amount:	-
Aggregate amount bid:	€287.539 bn
Number of counterparties:	123
Total amount of allotment:	€190.500 bn
Marginal interest rate:	0.13 %
Ratio of allotment at marignal interest rate:	64.47 %
Average interest rate:	0.11 %
Bids spread:	0.00 % - 0.50 %

Piazolo (2011) Figure 5; www.bundesbank.de / FTO 2013-08-27 (in German)

Figure 2
Liquidity-absorbing Tenders

At the highest interest rate just being accepted, (marginal interest rate) an allotment pro rata takes place, if the aggregate amount of bids exceeds the remaining amount to still be allotted.[25] In August 2013 the ECB absorbed €190.5 bn at an average interest rate of 0.11 % (p.a.).

3 Experiment on Tender Procedures

3.1 Setup of the Monetary Out-of Classroom Experiment

To enhance the students understanding of the bidding behaviour of major financial institutions – as reflected by the results of the fine tuning operation in the right hand column of *Figure 2*, we asked them to participate as investment officers in two different kinds of liquidity-absorbing operations. The banks, they are working for, are in excess of €4 mn of liquidity. The participants are supposed to engage with €2 mn each in both investment opportunities – thereby maximizing their own profits.

The first kind of investment is a Fine Tuning Operation, which almost exactly resembles the liquidity-absorbing *collection of fixed-term deposits* through the ECB presented above. Though, the following adjustments are made for reasons of simplification:

25 For more details see ECB (2011d L331/25).

- Maturity is lengthened to one year; therefore, the interest rate is easily captured as per annum.
- €2 mn is to be invested in two tranches of €1 mn each, or alternatively in one tranche only. Instead of being able to submit bids with up to 10 different interest rate levels,[26] we restrict them to a maximum of two different bids.
- As the aggregate amount of bids is unknown to us in advance, we announce that only 2/3 of this aggregate will be allotted by the central bank. Therefore, all participants know that they are part of a competitive auction process.
- An alternative investment vehicle is introduced: German Federal Treasury Notes ("Bundesschatzbriefe") with comparatively low interest rates. They essentially serve as a lower bound of an interest rate corridor for the bidding process.

The tender procedure is a *variable rate tender*, whereas the participating counterparties will have to submit bids stating the interest rate(s) offered as well as the amount of liquidity to be deposited. In *Table 3*, we mimicked an official ECB announcement for a tender operation. It was presented to the students via email and the university's learning management system. The announcement was made 8-10 days in advance of the submission date.

Table 3
ECB Announcement of Fine Tuning Operation (June 13 2013)

Liquidity-Absorbing Collection of Fixed-Term Deposits	
Standard Tender (variable rate tender, American auction)	
Submission date for bids:	21.06.2013 at 12:00
Allotment date:	21.06.2013
Maturity:	360 days
Maximum interest rate:	0.50 %
Minimum amount being allotted:	---
Minimum amount of bid:	€1 mn
Maximum amount of bid:	€2 mn
Targeted allotment amount:	2/3 of aggregate amount of bids

This announcement was part of general information material on the various auction designs for tenders in ECBs main refinancing operations. In the participation sheet the students had to submit (*appendix 1*), they were made aware that the amount (excess liquidity) of the unsuccessful bids would automatically be invested in an alternative government security. In the 2013 experiment, these German Federal Treasury Notes carried an interest rate of 0.25 % (p.a.).

---

26 See for comparison ECB (2011d L331/23).

In the second part of the experiment,[27] we made use of the fact that Deutsche Bundesbank acts, besides being an integral part of the Eurosystem, as fiscal agent for the German government. Deutsche Bundesbank is given the mandate to offer Federal bonds, Federal Treasury notes and Treasury discount paper for sale by auction, and it also is in charge for providing liquidity to the market by directly engaging for the account of the Federal Government in trading securities. Technically, the bids are to be transmitted electronically through the Bund Bidding System run by Deutsche Bundesbank. Even though there are no restrictions on the range of potential buyers, only members of the "Bund Issues Auction Group" can directly participate in the auctions. In August 2013, 37 major international financial institutions – mainly commercial banks – were members of the auction group.[28] In 2012, 70 auctions took place - with a total allotment of €214 bn. In 2008, fewer auctions were conducted with less securities being allotted to (37 auctions and €170 bn respectively). Though, the peak of allotment had been reached in 2010 with a total volume of €273 bn - due to the additional financing needs of the Federal Government to stabilize financial markets. For 2013, a volume of €250 bn is scheduled – almost 70 % of which are capital market securities, while 30 % are money market securities like Treasury discount papers.[29] The latter are so called *Bubills* with maturities varying between 3-12 months.

We mimic an auction for Treasury discount papers (*Bubills*) with a maturity of 12 months. For discount papers, the interest or effective rate of return for the investor results from the difference between the nominal value (100.00) and the purchase price. The design is a *variable rate tender* run as an American auction. The minimum amount to bid for is €1 mn. Several bids can be submitted – varying in price and volume. Though, in our out of classroom experiment only two bids can be made, and price bids must be expressed as 0.01 percentage points.

As the issuer aims to keep interest rate payments as low as possible, higher price offers are served first and fully. Bids which are above the lowest price accepted (marginal price) will be allotted to full extent. Bids below the marginal price will not be considered. Deutsche Bundesbank ranks all incoming bids in descending order based on the prices offered. The marginal price is manifested when the total volume of *Bubills* to be allotted is exhausted. If the aggregate amount of bids at the marginal price exceeds the amount of *Bubills* still to be allotted, this remaining amount is allocated pro rata.

---

27 This part of the experiment resembles the T-bills auction by Saros (2009).

28 Financial institutions have to fullfill specific criteria to become and remain a member (Deutsche Bundesbank 2012 p. 1 and 2013).

29 For more detailled information see Finanzagentur (2012 p. 1). Small amounts of inflation-indexed bonds as well as foreign currency bonds are also auctioned off.

In addition, it is also possible to submit non-competitive bids. These bids are allotted at the weighted average price of the competitive bids accepted.[30] The introduction of non-competitive bids adds a third investment option for our participants – besides competitive bids and German Federal Treasury Notes. To be eligible to participate, the students represented a member institution of the Bund Issues Auction Group. In the participation sheet the students had to submit (*appendix 1*), they were made aware that the amount (excess liquidity) of the unsuccessful bids would automatically be invested in German Federal Treasury Notes at an interest rate of 0.25 % (p.a.) in 2013.

The experiment was offered in 2010-2013 to students of Financial Services (B.A.) at the University of Applied Sciences Kaiserslautern (Germany). We presented the invitation to participate in their second year compulsory course in *Monetary Economics*. The experiment took place in the latter part of the course weeks after the topics on main refinancing operations with different kinds of tender procedures had been discussed in class. In 2010, we also invited MBA students at an international seminar in Budapest, Hungary, to participate in that years' experiment.[31] The coursework consisted of one final exam – two hours, written and open answers. A total of 120 points were given on the exam. To pass the course, a minimum of 60 points was required. During the semester we offered minor assignments on a voluntary basis: participating in two knowledge surveys and one or two classroom experiments as well as giving a ten-minute presentation on a current topic in monetary policy. For our monetary tender experiment 2 to 5 bonus points were at stake: 2 points for participation only, and additional 2-3 points when reaching the top 20% of all students profit-wise. Between 31% and 76% of each cohort of students participated. The bonus points served as incentive for participation as well as for submitting competitive bids.[32]

3.2 Analysis of Bidding Behaviour

In total, 126 students participated over the course of four years. The largest number of participants was reached in 2013 (44). In 2010, we started with the smallest group – in absolute terms (26) as well as in relation to all signups to the course (31 %). For each cohort we ran a separate experiment. As monetary conditions changed over time – e.g. the leading interest rate for the Euro Area varied, the experiments were adjusted accordingly. The analysis consists of two parts: First, we present the summary results for the largest cohort (2013) in a

30 Deutsche Bundesbank (2012 p. 2).

31 ECB's monetary policy - including its instruments and procedures - were part of the course work of the International Seminar at Obudá University. The students had to submit a poster presentation, on which their grade was based on.

32 Based on previous, more basic auction experiments in the introductionary *Economics* course, we expected strong competition and little collusion between the participants.

descriptive manner. It is similar to the feedback the participants were provided with. Thereafter, we analyze the aggregate bidding behaviour of all four years.

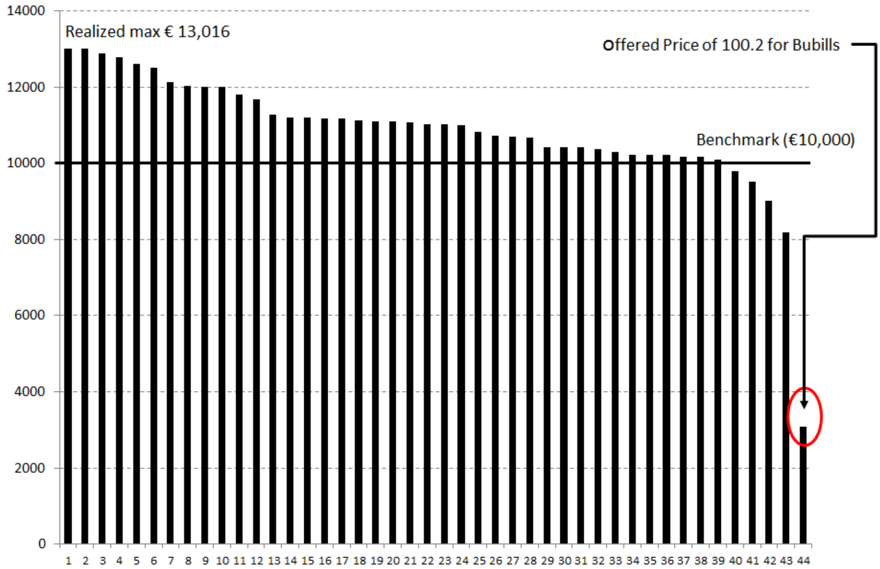


Figure 3
Total Interest Rate Profits 2013 (Excess Liquidity of €4 mn)

3.2.1 Descriptive Analysis of the Class of 2013

The benchmark of profits for the 44 participating students, when investing €4 mn in excess liquidity, depends on the interest rate for the alternative investment opportunity: German Federal Treasury Notes. As their rate was set at 0.25 % (p.a.), participants should at the least end up with €10,000 in profits. This represents a return on a fully risk free investment.[33] It serves as the lower bound for profits to all rational decision makers. This is also true for students, whose bids were not successful in the auction process.[34] The upper bound for profits is twice the amount (€20,000), as the ECB and the German Government are willing to accept a maximum interest rate of 0.50 %. In *Figure 3*, total interest rate earnings are shown for all participants. Two main findings can be generalized; they hold for the previous cohorts as well.

33 “Risk free” in the sense of not being dependent on the other participant’s bids.

34 As noted in the participation (decision) sheet, bids that did not receive an allotment were automatically shifted to be invested in German Federal Treasury Notes (*appendix 1*).

- None of the participants gets close to the upper limit of possible profits. The opposite is true, average profits remain pretty close to the lower bound. In 2013, average profits are €10,848 (median at €11,014) with the coefficient of variation standing at 14.5 % only.
- There are always a few participants, who do not behave rational in a sense that they do not even reach the risk free rate of return (€10,000).

These findings seem to reflect a tense competition among the participants. The drive for competitiveness led a significant number of students (11 %) to be carried away during their bidding process by dropping beneath the risk-free benchmark. Either the relative small incentive for bonus points works well, or the drive for "winning" against their peers dominates. In *Table 4* we present the results of the bids of the two tender procedures.

Table 4
Tender Procedures - Summary of Bids (2013)

Collection of Fixed-Term Deposits		***Bubills* - Federal Treasury Discount Paper**	
Bids (amount)	€ 88 mn	Competitive Bids (amount)	€ 68 mn
Offered Interest Rates	0.10 % - 0.50 %	Offered Prices	97.60 - 100.20
Allotment (amount)	€ 58.7 mn	Allotment (amount)	€ 45.3 mn
Marginal Interest Rate	0.39 % € 11 mn p.r.	Marginal Price	99.65 € 5 mn p.r.
		Non-competitive Bids (amount)	€ 20 mn
		Average Price of Competitive Bids (Weighted)	99.74
Alternative Investment - Federal Treasury Notes (0.25 %)			
Amount	€ -- mn	Amount	€ -- mn

p.r. = pro rata; n = 44 participants.

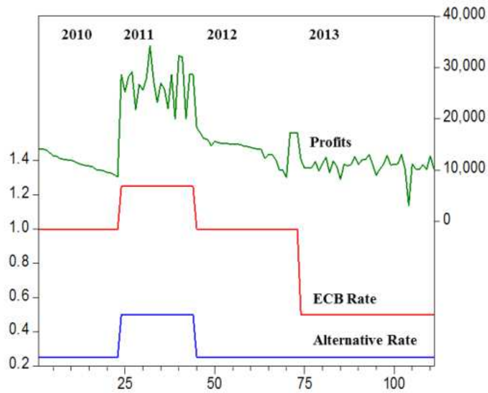
Overall, none of the participants opted for the alternative investment. As the Federal Treasury Notes represent their fallback option, this is a fully rational approach.[35] For the collection of fixed-term deposits, five bids were submitted at 0.25 % or less the interest rate of the Federal Treasury Notes. In addition, eleven competitive bids for *Bubills* were set at prices of 99.75 and above. These 16 bids - 9 % of the total - were not rational as they contained prices and interest rates that were the "below" the ones for Federal Treasury Notes.

For the *Bubills* auction, almost a quarter of bids were non-competitive in nature. In reality, financial institutions often submit an even higher share of non-competitive bids.[36] The non-competitive bids were allotted to full extent at the

[35] In 2010-2012 each, a few participants chose to invest in the risk-free alternative.

[36] For example, the Bundesbank's tender for 12 month *Bubills* on August 26 2013 resulted in 43 % of the total amount bid being non-competitive (Deutsche Bundesbank 2013c).

weighted average price offered of 99.74 % - just marginally above the interest rate of Federal Treasury Notes. The tight concentration of bids in interest rates as well as in prices offered is reflected in the amount of bids (€11 mn & €5 mn) being allotted pro rata at the marginal rate & price respectively. Overall, the strong competition between participants led to relative minor profits - quite similar to the real world (e.g. right hand side of *Figure 2*). Though, one would not expect professional investment agents to overbid.



Alternative Rate: Federal Treasury Notes; ECB Rate: Main Refinancing Rate in percent p.a. (left y-axis); Total Profits in € (right y-axis).

Figure 4
ECB Conditions and Total Interest Rate Profits in 2010-2013

3.2.2 Analysis of Aggregate Tender Procedures in 2010-2013

Of the 126 students participating, we could only use of 111 sheets for our statistical analysis - due to missing information on several of the participation sheets submitted.[37] The monetary conditions were varied according to the ECB framework in each year. *Figure 4* reflects these developments in the change of the main refinancing rate as well as in the interest rate of the Federal Treasury Notes. The main refinancing rate varied from 0.50 % (2013) to 1.00 % (2010, 2012) and to 1.25 % (2011). While the interest rate of the alternative investment - representing the lower bound - remained at 0.25 % except for 2011 (0.50 %). The benchmarks of the risk-free investment alternative were €10,000 and €20,000 respectively. On the opposite side, maximum profits for the €4 mn investment of excess liquidity could have reached €20,000 (2013), €40,000 (2010, 2012) and €50,000 (2011). *Figure 4* shows that the actual realization of total interest rate profits remained far below these maximum rates. One observation seems to be obvious: In 2011, as the interest rate for Federal Treasury Notes doubled profits

[37] In 2013, data for 38 students out of a total of 44 students was applied.

rose accordingly. Though, the higher profit level could also be driven by the rise of the main refinancing rate (upper bound). Additional comparative statistics are presented in *appendix 2*. Total profits (mean) differ significantly between 2011 and the other years, but the relative variation in profits remains small. Age-wise, 2010 is the cohort with the oldest participants. The difference in age is solely explained by the participating post-graduate MBA students - this group of nine students was 39 years old (mean) whereas the remaining second-year B.A. students of 2010 were in line with their mates of the other cohorts (23.5 years). Of the 102 bachelor students that enrolled in the course *Monetary Economics*, exactly half of them had been placed in *Group A* and *B*. To provide a better learning environment it is departmental policy to split a course into two parallel groups if total enrollment exceeds 40 students. The assignment to *Group A* or *B* is not random. For *Group B*, students with higher final marks in their High School Diploma ("Abitur") are selected. Total profits of *Group A* students are on average €14,798, whereas the mean of total profits of *Group B* stands at €15,868.

Our next step is to sort out, which factors determine *total profits*. Based on the above descriptive statistics, we can formulate the following hypotheses:

- The rates of return are expected to influence total profits positively. As Federal Treasury Notes rates (lower bound) and ECBs main refinancing rate (upper bound) go up, profits on average are expected to increase. (*Hypothesis # 1a, 1b*) Whereby the lower bound should have a greater impact on profits, as competition drove down the mean rates of profits closer to this boundary.
- The number of participants might influence the depth of the competition process - more students (e.g. in 2013) increase the competition and therefore drive down total profits (*Hypothesis # 2*).
- The age of students does not differ too much, but minor age difference (25 vs. 22) could reflect that the student took the more demanding *Monetary Economics* course at a later stage of his/her study programme. This is especially the case, when he/she had to struggle to pass the basic courses of the first year already. As struggling students would be older age-wise, we can expect age to exert a negative influence on profits (*Hypothesis # 3*).
- MBA students - many of them with a Bachelor's degree in Engineering - have less of an academic background in economics. Therefore, we expect them to do not as well in the auction process (*Hypothesis # 4*).
- Due to the selection process for the two groups of *Monetary Economics* students, we expect the better students - based on their secondary education - to understand the auction design more thoroughly. *Group B* students should be rewarded with higher total profits (*Hypothesis # 5*).

- Only a quarter to a third of the participants is female (*appendix 2*). Women are underrepresented since their share in *Monetary Economics* is on average 45 %. Due to the competitive nature of the experiment, it seems to attract more of the male students. This self-selection bias might result in an outperformance of male vs. female students (*Hypothesis # 6*).

Table 5
Determinants of Total Interest Rates Profits (2010-2013)

Total Profits [1)] n = 111	**Model I**	**Model II**	**Model III**	**Model IV**	**Model V**
Constant	-2,484.1 (978.7)***	5,482.8 (4,129.5)	2,456.7 (4,102.2)	-1,971,336 (652,377)***	-19,334.5 (5,403.0)***
Alternative Rate	58,555.7 (3,466.0)***	48,887.9 (3,969.8)***	46,146.4 (3,882.9)***	43,386.8 (3,942.2)***	45,708.8 (3,910.2)***
ECB Rate [2)]		5,026.3 (903.0)***	6,251.3 (950.0)***	9,578.1 (1,484.4)***	12,829.7 (2,671.5)***
log(Age)		-2,975.8 (1,328.1)**	-2,362.9 (1,305.2)*		
Monetary Economics Group B			1,776.0 (501.4)***	1,201.2 (542.7)**	1,400.8 (545.2)**
Year of Experiment				949.7 (323.9)***	
MBA Course				-1,449.1 (634.4)**	-1,554.5 (625.9)**
Male					780.8 (496.2)°
# of Participants					285.1 (111.1)**
R²	*0,812*	*0.844*	*0.861*	*0.875*	*0.878*
Adj.R²	*0,810*	*0.840*	*0.856*	*0.869*	*0.871*
Prob.(F-Statistic)	*0.000****	*0.000****	*0.000****	*0.000****	*0.000****

*[1)] Method: OLS white heteroscedasticity-consistent standard errors & covariance (with EViews). Significance level: ° 15%, * 10%, ** 5%, *** 1%. [2)] ECB's main refinancing rate.*

Both - the interest rate of Federal Treasury Notes as well as the ECB's main refinancing rate - have a major and significant impact on total profits. The rate of the alternative investment in itself explains 81 % of the variation in overall profits. But, profit levels still increase as the ECB's leading interest rate (upper bound) rises (*hypothesis # 1a, 1b* √). As participants get older their profits decrease. One reason being that less inclined students tackle *Monetary Economics* at a later stage than proposed by the official study program (in their 3rd or even 4th year). Secondly, *MBA* students, who are on average 16 years older than the B.A. students, do less well. The moment we include *MBA* as an independent variable, *age* drops out - so, the *MBA* criteria seems to override *age* (*hypothesis # 4* √). Pre-selected students of *Group B* in *Monetary Economics* do significantly better by €1,201 up to €1,776 (*hypothesis # 5* √). *Male* students improve profits compared to their female counterparts. Though, the support for *hypothesis # 6* is weak and its effect relatively meager - being only half in size compared to the pre-selection effect of *Group B*. When comparing model IV with model V, the *year of the experiment* is significant in the former, while the *number of participants* is in the latter. Model V is the one we consider to be our best fit. A possible transfer of know-how filtering through the cohorts could explain the positive impact of the year the experiment took place. Though, the moment the number of participants is included, years drops out. The correlation between the two independent variables is strong as the number of participants increase over time (*appendix 2*), but the variation in participants seems to dominate time. On the contrary to *hypothesis # 2*, more participants are driving up profits significantly. This is a stable as well as an intriguing result. In 2013, some cooperative behavior between the participants combined with better knowledge on the bidding process might have enhanced profit levels.

In addition, we also wanted to find out, if the bidding behavior of students affected their overall points in the course - and therefore their final grade. Besides the obvious impact that bonus points have in ratcheting up total points on a 1:1 base, *total profits* as well as the *alternative rate* lead to a significant positive impact on overall points of the course on their own.[38]

4 Insights Gained on Monetary Policy Implementation

The build-up of the different phases of the global financial crisis setting in the summer of 2007, led major central banks to react aggressively first, by using standard monetary policy instruments. After the bankruptcy of Lehman Brothers, they applied non-standard measures to tackle the crisis of confidence within the financial sector as well as to revive the real economy through the credit channel. In addition, the ECB still has to cope with the crisis of debt ridden countries in the periphery of the Euro Area. Early on, Deutsche Bundesbank as well as the leading economic research institutes in Germany had been very critical of the course the ECB had taken. Especially when it implemented the different programs of direct securities markets interventions.[39] The strict separation of fiscal versus monetary policy seemed to dissolve as central banks increased their balance sheet risks when taking in 2nd or 3rd rated (non-investment grade) government and corporate bonds. At least concerning the liquidity effect of such purchases and its long-term inflation risk, the ECB made use of Fine Tuning Operations to sterilize the liquidity enhancing purchases.

---

38 On request, the author provides interested readers with the econometric findings.

39 See Deutsche Bundesbank (2011b p. 70). In June 2013, Deutsche Bundesbank and ECB met on behalf of Outright Monetary Transactions and the European Stability Mechanism on opposed sides infront of Germany's Federal Constitutional Court.

For the better understanding of ECBs main refinancing operations, we invited students to take part in an out of classroom experiment. As financial institutions in the Euro Area have soaked up long-term central bank liquidity in excessive amounts, they have been bidding to invest their excess liquidity with the ECB in the short-term. To make use of the competition between the banks, the ECB chose to implement a variable rate tender procedure (American auction) - effectively reducing its interest rate costs. This is the framework in which our experiments settled in. Of their excess liquidity, the participants submitted competitive bids in the amount of €2 mn for *Fixed-Term Deposits* with the ECB (interest rates), as well as an additional €2 mn to the Bundesbank acting as fiscal agent of the German government for *Bubills* (prices). The auction design for *Bubills* had the additional twist that non-competitive bids were also possible to submit. The experiment ran with four different cohorts of students with varying conditions on ECB's main refinancing rate (upper bound) as well as on the interest rate of the alternative investment vehicle - German Federal Treasury Notes (lower bound). The main insights based on descriptive statistics as well as on the econometric analysis of the participants' behavior can be summarized as follows:

- very tense *competition* drove down profits - closely reflecting real world behavior of financial institutions. Though, some participants were carried away by submitting irrational bids.
- unsurprisingly, *Total Profits* were driven by financial market conditions - whereas the *Alternative Rate* exerted a greater influence than the *ECB Rate*. *Male* students and academically more inclined *Group B* students outperformed, while older participants (*Age*) and *MBA* students did less well. The former, because they delayed enrolling in the demanding course of *Monetary Economics*, the latter, because their academic background had been in engineering mainly (*MBA*).

Overall, the experiment deepened the students' understanding of the behavior of financial sector market agents when participating in competitive tender procedures. Based on personal feedback to the lecturer, the students seemed to have enjoyed this interactive teaching approach thoroughly. For monetary economics, the experiment gave additional empirical support for central banks to return to *variable rate tenders* (*American auctions*) for their main refinancing operations when providing liquidity to the economy. It is the most market-oriented tender procedure. Unfortunately due to the enduring problems and challenges of the international financial system, the return to - monetary speaking - standard times is still somewhere in the far future (2015).

Appendix 1 – Participation (Decision) Sheet of Individual Institution

Name ______________________________ Age ______ Sex male / female

I. ECB Fine Tuning Operation: Liquidity-Absorbing Collection of Fixed-Term Deposits - 12 Months

In the name of your employer (financial institution), you invest €2 mn of excess liquidity.
Standard Tender (variable rate, American auction)
Offered interest rates must be expressed as multiples of 0.01 percentage points;
the maximum rate is 0.50 %. Two bids can be submitted.
2/3 of aggregate amount of bids will be allotted by ECB.

Bid 1: amount ___ (€ mn); interest rate offered _____ (% p.a.)
Bid 2: amount ___ (€ mn); interest rate offered _____ (% p.a.)
Investment alternative – Federal Treasury Notes at 0.25% p.a. Amount ____ (€ mn)

II. Bundesbank Tender: German Treasury Discount Papers (*Bubills*) - 12 Months

In the name of your employer (financial institution), you invest €2 mn of excess liquidity.
Standard Tender (variable rate, American auction)
Offered prices must be expressed as multiples of 0.01 percentage points; the minimum price is 99.50 %. Two competitive bids can be submitted. 2/3 of aggregate amount of bids will be allotted by ECB. Non-competitive bids are fully allotted at the (weighted) average price of the competitive bids being successful.

Bid 1: amount ___ (€ mn); price offered ______ (%)
Bid 2: amount ___ (€ mn); price offered ______ (%)
Bid 3: amount ___ (€ mn); non-competitive bid
Investment alternative – Federal Treasury Notes at 0.25% p.a. Amount ____ (€ mn)

Unsuccessful competitive bids are automatically invested in Federal Treasury Notes.
Date of submission: June 21 2013 at 12:00 (electronically).
General aim: maximize total profits by investing €4 mn of excess liquidity.

Appendix 2 - Comparative Summary Statistics for 2010-2013

n = 111	2010	2011	2012	2013
Total profits (TP) - mean	€11,422	€26,814**	€14,514	€10,845
Total profits (TP) - median	€11,300	€27,467	€15,000	€11,087
TP - Standard deviation	€1,583	€3,750	€2,239	€1,676
TP - Coefficient of variation	14.0 %	13.7 %	14.9 %	15.1 %
Possible Range of Total Profits*	€10,'-40,000	€20,'-50,000	€10,'-40,000	€10,'-20,000
# of participants	23	21	29	38
Age of participants - mean	29.2	23.6	23.1	23.4
Age of participants - CV	31.6 %	9.6 %	9.0 %	9.0 %
Female Participants	30.4 %	33.3 %	24.1 %	26.3 %
Final points - mean	79.0	75.2	63.6	72.6
Failure rate of participants	8.7 %	14.3 %	41.4 %	23.7 %
Bonus points - mean	4.0	5.0	5.9	8.6

(*) Possible Range of Total Profits: Profits of Federal Treasury Notes to ECB main refinancing rate.
CV - Coefficient of Variation. In 2013, the data differs slightly compared to the quotes in the text as the number of participants analyzed dropped from 44 to 38.
(**) Total Profits in 2013 differ at the 1 % level significantly with the ones in 2010, 2011 and 2012 (Difference in Means Test).

References

[1] *BIS* 79th Annual Report (1 April 2008 – 31 March 2009) Bank for International Settlements, Basel 29 June 2009 www.bis.org

[2] *Blinder, A.* How Central Should the Central Bank be? Journal of Economic Literature Vol. 48:1 (2010) 123-133

[3] *Deutsche Bundesbank* Annual Report 2010, Frankfurt (March 2011a)

[4] *Deutsche Bundesbank* Konsequenzen für die Geldpolitik aus der Finanzkrise, Monatsbericht (März 2011b) 55-71

[5] *Deutsche Bundesbank* Geldpolitische Geschäfte des Eurosystems (Tenderverfahren) www.bundesbank.de (26. April 2011c)

[6] *Deutsche Bundesbank* Auction rules for the issue of Federal bonds, five-year Federal notes, Federal Treasury notes and Treasury discount paper (2012) www.bundesbank.de Bundeswertpapiere / Tenderverfahren

[7] *Deutsche Bundesbank* Annual Report 2012 (March 2013a)

[8] *Deutsche Bundesbank* Mitglieder der Bietergruppe Bundesemissionen, www.bundesbank.de (05. August 2013b)

[9] *Deutsche Bundesbank* Tenderergebnis Unverzinsliche Schatzanweisungen der Bundesrepublik Deutschland („Bubills"), Pressenotiz (26. August 2013c)

[10] *Draghi, M.* Introductory Statement to the Press Conference (with Q&A), President of the ECB, Frankfurt (4 July 2013)

[11] *ECB* The ECB's non-standard Measures – Impact and Phasing-out, Monthly Bulletin, European Central Bank (July 2011a) 55-69

[12] *ECB* The European Stability Mechanism, Monthly Bulletin, European Central Bank (July 2011b) 71-84

[13] *ECB* The Monetary Policy of the ECB, European Central Bank, Frankfurt (2011c)

[14] *ECB* The Implementation of Monetary Policy in the Euro Area, General Documentation on Eurosystem Monetary Policy Instruments and Procedures, European Central Bank, Frankfurt (2011d)

[15] *EZB* Jahresbericht 2009, Europäische Zentralbank, Frankfurt (April 2010a)

[16] *ECB* The ECB's Response to the Financial Crisis, Monthly Bulletin, European Central Bank (October 2010b) 59-74

[17] *Finanzagentur* Emissionsplanung des Bundes, Jahresvorausschau und erstes Quartal 2013, Pressemitteilung Nr. 25/12, Frankfurt (20. Dezember 2012)

[18] *Finanzen* Aktueller Leitzins, www.finanzen.net (05. August 2013)

[19] *Frenkel, M., Pierdzioch, C., Stadtmann, G.* Die optimale Ausgestaltung der EZB-Offenmarktpolitik, Volkswirtschaftslehre, WISU Heft 1 (2003) 111-118

[20] *Gorton, G.; Metrick, A.* Getting up to Speed on the Financial Crisis: A One-Weekend-Reader's Guide, Journal of Economic Literature Vol. 50:1 (2012) 128-150

[21] *IMF* Hopes, Realities, Risks, World Economic Outlook, International Monetary Fund, Washington D.C. (April 2013)

[22] *Lane, P. R.* The European Sovereign Debt Crisis, Journal of Economic Perspectives Vol. 26:3 (2012) 49-68

[23] *Mishkin, F. S.* Over the cliff: From the Subprime to the Global Financial Crisis, Journal of Economic Perspectives Vol. 25:1 (2011) 49-70

[24] *Piazolo, M.* Geldpolitik in Zeiten der Eurokrise, Wie funktioniert die Liquiditätsabschöpfung der Europäischen Zentralbank in der Praxis? 121-141 in: Proceedings, 9th International Conference on Management, Enterprise and Benchmarking MEB 2011, Budapest (2011)

[25] *Saros, D. E.* Teaching Undergraduate Money and Banking: T-Bill Auctions and Stock Market Models, Economic Instruction, Journal of Economic Education Vol. 40 (Summer 2009) 315-330

[26] *Schularick, M.; Taylor, A. M.* Credit Booms Gone Bust: Monetary Policy, Leverage Cycles, and Financial Crisis, 1870-2008, American Economic Review Vol. 102:2 (2012) 1029-1061

[27] *Woodford, M.* Financial Intermediation and Macroeconomic Analysis, Journal of Economic Perspectives Vol. 24:4 (2010) 21-44

Different Mathematical Solutions on Gas Oscillation

Ferenc Szlivka

Institute of Technology, College of Dunaújváros
Táncsics Mihály u. 1, 2401 Dunaújváros, Hungary
e-mail: szlivkaf@mail.duf.hu

Abstract: The mathematical description of gas-pressure oscillations excited by a piston at the end of a straight duct has been investigated in an analytical way and it has been compared to numerical solutions and measured results. Two different analytical solutions have been found in the literature in series form which have been transformed into closed form solutions showing their identity. This evaluation can significantly be simplified and the comparison of the results of the analytical and the numerical solutions are presented in this study accordingly. In the results are presented the frequency of piston is close to the resonant frequency.

Keywords: gas pressure oscillation; analytical solution; infinite series; closed form

1 Introduction

The studied geometrical model with piston movement of constant frequency at one end of a pipe and with a free stream outflow to the atmosphere at the other end is shown in Figure 1. Such unsteady oscillations inside the pipe are calculated with different mathematical approaches. E.g. if amplitude of the oscillation is considered to be sufficiently small; the mathematical model of the current physical problem is the one-dimensional wave equation with the appropriate boundary conditions. In case of large amplitudes, a substantially different mathematical model must be used because the propagation of pressure waves cannot be described correctly with the solution of the linear wave equation. In the following discussion different wave equation solutions will be shown.

Jimenez [2] investigated the numerical solution of a similar gas pressure oscillation problem. An analytical investigation for the same problem is presented by Hoffmann and Fényes [1], where the problem is solved using the same mathematical model of a partial differential equation system as well as initial and boundary conditions.

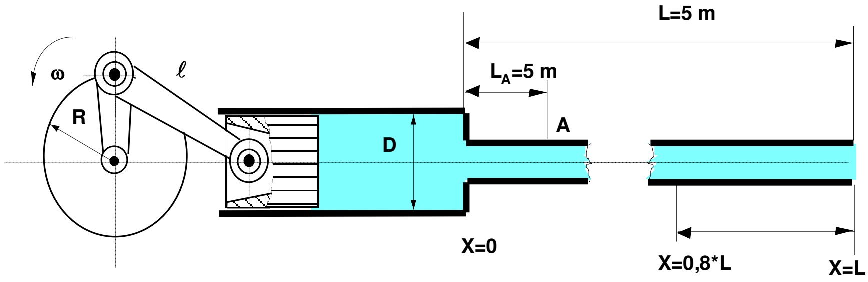


Figure 1
Schematic diagram of the gas-pressure oscillation problem

Different analytical methods are used by Hoffmann and Fényes [1] to solve the mathematical model of the pressure wave propagation. The developed analytical solutions give the same result although they are in very different forms. The aim of this study is to present a solution of simplified form for the mathematical model of the pressure wave propagation. A closed form formula is developed from the different infinite series solutions worked out by Hoffmann and Fényes [1]. If both solutions are brought into a closed form, they can be formed also formally perfectly immediately. (That out of the momentum and mass transport equations corresponding to the physical process and continuity equation diverted mathematical model that on the velocity c(x,t) and the pressure is p(x,t) related partial differential equation system). Some parts of the solutions was published by Szlivka [4] but in German language and only in Hungary, nevertheless it was not complete. It is shown some calculated results and comparison with other theories and measured results.

The mathematical model of the pressure wave oscillation stems from the momentum and mass transport equations of the continuous fluid. It is called in the literature as acoustic model see [5], [6].

$$\frac{\partial c}{\partial t} = -\frac{1}{\rho} \cdot \frac{\partial p}{\partial x}; \qquad \frac{\partial c}{\partial x} = -\frac{1}{\rho \cdot a^2} \cdot \frac{\partial p}{\partial t} \tag{1}$$

With the appropriate initial and boundary conditions

$$c(x,0) = 0 \qquad p(x,0) = p_0 \tag{2}$$

$$c(0,t) = R\omega \cdot \sin\omega \cdot t \qquad p(L,t) = p_0$$

where the meaning of the symbols are: "ρ" the density of the medium; "$a=347,8$ m/s" the speed of sound (constant); "p" the pressure; "R " the radius of the crankshaft; "ω " drive its circle frequency; "L" the pipe length. In the first case “l”, the rod length is infinite. In chapter 6 there will be used finite rod lengths [1] following solution is based on the Bernoulli-Fourier-method:

$$c(x,t) = R\omega \cdot \left[\frac{sin\,\omega \cdot \left(t - \frac{x}{a} - \frac{L}{a} \right)}{2 \cdot cos\,\frac{\omega \cdot L}{a}} - \frac{\omega \cdot a}{L} \cdot \sum_{n=0}^{\infty} \frac{cos\,\omega_n \cdot \left(t - \frac{x}{a} \right)}{\omega_n^2 - \omega^2} \right] + \\ + R\omega \cdot \left[\frac{sin\,\omega \cdot \left(t - \frac{x}{a} - \frac{L}{a} \right)}{2 \cdot cos\,\frac{\omega \cdot L}{a}} + \frac{\omega \cdot a}{L} \cdot \sum_{n=0}^{\infty} \frac{cos\,\omega_n \cdot \left(t + \frac{x}{a} \right)}{\omega_n^2 - \omega^2} \right] \tag{3}$$

$$p(x,t) = p_0 + \rho a \cdot R\omega \cdot \left[\frac{sin\,\omega \cdot \left(t - \frac{x}{a} - \frac{L}{a} \right)}{2 \cdot cos\,\frac{\omega \cdot L}{a}} - \frac{\omega \cdot a}{L} \cdot \sum_{n=0}^{\infty} \frac{sin\,\omega_n \cdot \left(t - \frac{x}{a} \right)}{\omega_n^2 - \omega^2} \right] - \\ - \rho a \cdot R\omega \cdot \left[\frac{sin\,\omega \cdot \left(t - \frac{x}{a} - \frac{L}{a} \right)}{2 \cdot cos\,\frac{\omega \cdot L}{a}} - \frac{\omega \cdot a}{L} \cdot \sum_{n=0}^{\infty} \frac{cos\,\omega_n \cdot \left(t + \frac{x}{a} \right)}{\omega_n^2 - \omega^2} \right] \tag{4}$$

With the Laplace-transformation, the

$$c(x,t) = R\omega \cdot \sum_{n=0}^{\infty} (-1)^n \cdot \left[sin\,\omega \cdot \left(t - \frac{x + 2 \cdot n \cdot L}{a} \right) + sin\,\omega \cdot \left(t - \frac{2 \cdot (n+1) \cdot L - x}{a} \right) \right] \tag{5}$$

$$p(x,t) = p_0 + \rho a R\omega \sum_{n=0}^{\infty} (-1)^n \left[sin\,\omega \left(t - \frac{x + 2nL}{a} \right) - sin\,\omega \left(t - \frac{2(n+1)L - x}{a} \right) \right] \tag{6}$$

Where $sin\,\omega(t-\gamma) = 0$ if $t \le \gamma$, as well as $t > 0$, $\gamma > 0$.

We can see two different solutions above. One with Fourier method (3), (4), the other with Laplace-transformation (5), (6). Both can be brought into closed form in order to reach the result, which will be presented in Chapter 3 and 4.

It is to be recognized that the only difference between formula (3) and (4) is constant 'p_0', multiplicator 'ρa' and sign of the second term. Therefore simplification is shown only for formula (3), which can be used for formula (4) analogically.

3 Developing the Closed Form of the Infinite Series form Derived from the Bernoulli-Fourier-Method

Formula (3) contains two separate infinite series of cosine functions which can be reformulated into closed form what can be seen in expression (15) and (16). For simpler handling, the following symbols are introduced:

$$T=\frac{a\cdot t}{L};\quad V=\frac{c}{a};\quad P=\frac{p}{p_0};\quad X=\frac{x}{L};\quad \Omega=\frac{\omega\cdot L}{a}$$

$$\Omega_n=(2n+1)\cdot\frac{\pi}{2},\quad B=\frac{R\omega}{a};\ D=\frac{\rho a\cdot R\omega}{p_0} \tag{7}$$

and these symbols applied in expression (3)

$$V(X,T)=\frac{B}{2}\cdot\left[\frac{\sin\Omega\cdot(T-X+1)}{\cos\Omega}-2\Omega\cdot\sum_{n=0}^{\infty}\frac{\cos\Omega_n\cdot(T-X)}{\Omega_n^2-\Omega^2}\right]+$$
$$+\frac{B}{2}\cdot\left[\frac{\sin\Omega\cdot(T+X-1)}{\cos\Omega}+2\Omega\cdot\sum_{n=0}^{\infty}\frac{\cos\ \Omega_n\cdot(T+X)}{\Omega_n^2-\Omega^2}\right] \tag{8}$$

The infinite series expression in equation (8) can be divided into the sum of two other infinite cosine series, which comes from Ryzhik, I. M.-Gradstein, I. S.: Tafeln [3].

$$2\Omega\sum_{n=0}^{\infty}\frac{\cos(2n+1)\frac{\pi}{2}y}{\left[(2n+1)\cdot\frac{\pi}{2}\right]^2-\Omega^2}=2\Omega\frac{4}{\pi^2}\sum_{i=1}^{\infty}\frac{\cos\ i\frac{\pi}{2}y}{i^2-\left(\frac{2\Omega}{\pi}\right)^2}-\frac{2\Omega}{\pi^2}\sum_{j=1}^{\infty}\frac{\cos\ j\pi y}{j^2-\left(\frac{\Omega}{\pi}\right)^2} \tag{9}$$

where y can be both $T+X$ or T-X, and $\Omega\neq k\cdot\frac{\pi}{2}$ *and* $i;j;k$ *and* n are integer numbers. Multiplicator *(2n+1)* on the left side is an odd number. The right side consists of infinite series in eq. (9)

The following notations are introduced:

$$\frac{y}{2}=N_1+z_1;\ where\ N_1=int\left(\frac{y}{2}\right)\ and\ 0\le z_1<1 \tag{10}$$

$$y=N_3+z_3;\ where\ N_3=int(y)\ and\ 0\le z_3<1 \tag{11}$$

where *int()* function is the largest integer part of real value.

$2N_1=N_3$ if N_3 an even number and (12)

$2N_1=N_3-1$ if N_3 is an odd number.

$$2\Omega\cdot\frac{4}{\pi^2}\sum_{i=1}^{\infty}\frac{\cos\left(i\cdot\pi\cdot N_1+i\cdot\pi\cdot z_1\right)}{i^2-\left(\frac{2\Omega}{\pi}\right)^2}=\begin{cases}=2\Omega\cdot\frac{4}{\pi^2}\sum_{i=1}^{\infty}\frac{\cos\left(i\cdot\pi\cdot z_1\right)}{i^2-\left(\frac{2\Omega}{\pi}\right)^2} & N_1\ \ even\\ =2\Omega\cdot\frac{4}{\pi^2}\sum_{i=1}^{\infty}\left(-1\right)^i\frac{\cos\left(i\cdot\pi\cdot z_1\right)}{i^2-\left(\frac{2\Omega}{\pi}\right)^2} & N_1\ \ odd\end{cases}\tag{13}$$

Second term at the right side of the equation (9) is substituted with expression (11) and the following formula has been received

$$\frac{2\Omega}{\pi^2}\cdot\sum_{j=1}^{\infty}\frac{\cos\left(j\cdot\pi\cdot N_3+j\cdot\pi\cdot z_3\right)}{j^2-\left(\frac{2\Omega}{\pi}\right)^2}=\begin{cases}=\frac{2\Omega}{\pi^2}\cdot\sum_{j=1}^{\infty}\frac{\cos\left(j\cdot\pi\cdot z_3\right)}{j^2-\left(\frac{2\Omega}{\pi}\right)^2} & N_3\ \ even\\ =\frac{2\Omega}{\pi^2}\cdot\sum_{j=1}^{\infty}\left(-1\right)^j\frac{\cos\left(j\cdot\pi\cdot z_3\right)}{j^2-\left(\frac{2\Omega}{\pi}\right)^2} & N_3\ \ odd\end{cases}\tag{14}$$

Let us use the following equations from Tafeln [3]:

$$\sum_{k=0}^{\infty}\frac{\cos k\cdot z}{k^2-\alpha^2}=\frac{1}{2\alpha^2}-\frac{\pi}{2}\cdot\frac{\cos\alpha\cdot\left(\pi-z\right)}{\alpha\cdot\sin\alpha\cdot\pi}\quad\text{with}\ \ 0\le z\le\pi\tag{15}$$

and

$$\sum_{k=0}^{\infty}\left(-1\right)^k\cdot\frac{\cos k\cdot z}{k^2-\alpha^2}=\frac{1}{2\alpha^2}-\frac{\pi}{2}\cdot\frac{\cos\alpha\cdot\left(\pi-z\right)}{\alpha\cdot\sin\alpha\cdot\pi}\quad\text{with}\ \ -\pi\le z\le\pi\tag{16}$$

supposing that "*k*" is an integer number, " α " is not an integer number, (13) and (14) can be written in a simpler form. If N_1 and N_3 are even numbers, we get expression (15) and if N_1 and N_3 are odd numbers we get expression (16), if

"*z*" value fulfils the above indicated conditions. If N_1 is an even number inequality (17) must be valid by expression (10).

$$0\le\pi\cdot z_1\le\pi\tag{17}$$

This condition is almost equivalent to expression (10). It is easy to realize that after the definition by z_1 and z_3 conditions of the equivalences (15) and (16) are always fulfilled.

Under the application of the previous executions, we simplify the equation (9) in case if N_1 and N_3 are even numbers. Expression (9) can be formulated in a substantially simplified form with the application of expression (15) and (16) with the corresponding conditions for N_1 and N_3.

$$2\Omega \sum_{n=0}^{\infty} \frac{\cos\left(2\cdot n+1\right)\cdot\frac{\pi}{2}\cdot y}{\left[\left(2\cdot n+1\right)\cdot\frac{\pi}{2}\cdot\right]^2-\Omega^2} = 2\Omega\frac{4}{\pi^2}\sum_{i=1}^{\infty}\frac{\cos\ i\cdot n\cdot z_1}{i^2-\left(\frac{2\Omega}{\pi}\right)^2}-\frac{2\Omega}{\pi^2}\sum_{j=1}^{\infty}\frac{\cos\ j\cdot n\cdot z_1}{j^2-\left(\frac{\Omega}{\pi}\right)^2}=$$

$$=\frac{\cos\Omega\left(2z_1-2\right)}{\sin 2\Omega}-\frac{\cos\ \Omega\left(z_3-2\right)}{\sin\Omega}=$$

$$=\frac{2\cos\ \Omega\left[y-2\left(N_1+1\right)\right]-2\cos\Omega\left[y-\left(N_3+1\right)\right]\cos\Omega}{\sin 2\Omega}=$$

This modification was carried out if N_1 and N_3 are even numbers, but rather also can be solved in all other cases, as well. After application of the relations (12), one receives the expressions:

$$\begin{cases} =\dfrac{\sin\Omega\left[y-\left(2N_1+1\right)\right]}{\cos\Omega} & \textit{If } N_1 \textit{ even, } N_3 \textit{ odd, or } N_1 \textit{ and } N_3 \textit{ even numbers are.} \\ =-\dfrac{\sin\Omega\left[y-\left(2N_1+1\right)\right]}{\cos\Omega} & \textit{If } N_1 \textit{ odd, } N_3 \textit{ even, or } N_1 \textit{ and } N_3 \textit{ odd numbers are.} \end{cases}$$

Introduced the expression $\varepsilon=\Omega-\frac{\pi}{2}$ and use in the last two expressions we get the same result for all cases of odd and even N_1 and N_3 numbers.

$$=\frac{\sin\ \Omega\cdot\left[y-\left(2N_1+1\right)\right]}{\cos\Omega}=\frac{\sin\cdot\left[\Omega\cdot y-\left(\varepsilon+\frac{\pi}{2}\right)\left(2N_1+1\right)\right]}{\cos\Omega}=\frac{\sin\left[\Omega\cdot y-\left(2N_1+1\right)\varepsilon-N_1\cdot\pi-\frac{\pi}{2}\right]}{\cos\Omega}=*$$

$$=-\frac{\sin\ \Omega\cdot\left[y-\left(2N_1+1\right)\right]}{\cos\Omega}=-\frac{\sin\left[\Omega\cdot y-\left(\varepsilon+\frac{\pi}{2}\right)\left(2N_1+1\right)\right]}{\cos\Omega}=-\frac{\sin\left[\Omega\cdot y-\left(2N_1+1\right)\varepsilon-N_1\cdot\pi-\frac{\pi}{2}\right]}{\cos\Omega}=*$$

Using the $sin\left[\alpha \pm \frac{\pi}{2}\right] = \pm cos\,\alpha$; $cos\left[\alpha \pm \frac{\pi}{2}\right] = \mp sin\alpha$ and $sin\left[\alpha \pm \pi\right] = -sin\alpha$; $cos\left[\alpha \pm \pi\right] = -cos\,\alpha$ and also $sin\left[\alpha \pm 2\pi\right] = sin\alpha$; $cos\left[\alpha \pm 2\pi\right] = cos\,\alpha$ expressions, we can get one formula from the two:

$$* = -\frac{cos\left[\Omega \cdot y - \left(2 \cdot N_1 + 1\right) \cdot \varepsilon\right]}{cos\,\Omega}$$

At the end of this simplification we can write in the form of the infinite series in a closed form:

$$2\Omega \cdot \sum_{n=0}^{\infty} \frac{cos\left(2 \cdot n + 1\right) \cdot \frac{\pi}{2} \cdot y}{\left[\left(2 \cdot n + 1\right) \cdot \frac{\pi}{2} \cdot\right]^2 - \Omega^2} = -\frac{cos\left[\Omega \cdot y - \left(2 \cdot N_1 + 1\right) \cdot \varepsilon\right]}{cos\,\Omega}$$

This formula can be used in the equation (8) if y eider $T + X$ or $T - X$

$$V(X,T) = \frac{B}{2} \cdot \left[\frac{sin\,\Omega \cdot \left(T - X + 1\right)}{cos\,\Omega} + \frac{cos\left[\Omega \cdot \left(T - X\right) - \left(2N_1 + 1\right) \cdot \varepsilon\right]}{cos\,\Omega}\right] + \\ + \frac{B}{2} \cdot \left[\frac{sin\,\Omega \cdot \left(T + X - 1\right)}{cos\,\Omega} - \frac{cos\left[\Omega \cdot \left(T + X\right) - \left(2 \cdot N_4 + 1\right) \cdot \varepsilon\right]}{cos\,\Omega}\right] \tag{18}$$

where

$$N_1 = int\left(\frac{T - X}{2}\right) \qquad N_4 = int\left(\frac{T + X}{2}\right) \tag{19}$$

Using the expression in the equation (18) we can get the final form of the velocity in equation (20).

$$V(X,T) = B \cdot \frac{sin\left[\Omega \cdot \left(T - X - N_1\right) + N_1 \cdot \frac{\pi}{2}\right] \cdot cos\left[(N_1 + 1) \cdot \Omega - N_1 \cdot \frac{\pi}{2}\right]}{cos\,\Omega} + \\ + B \cdot \frac{cos\left[\Omega \cdot \left(T + X - N_4 - 1\right) + N_4 \cdot \frac{\pi}{2}\right] \cdot sin\left[N_4 \cdot \left(\Omega - \frac{\pi}{2}\right)\right]}{cos\,\Omega} \tag{20}$$

und the pressure in equation (21).

$$P(X,T)=1+D\cdot\frac{sin\left[\Omega\cdot\left(T-X-N_1\right)+N_1\cdot\frac{\pi}{2}\right]\cdot cos\left[(N_1+1)\cdot\Omega-N_1\cdot\frac{\pi}{2}\right]}{cos\,\Omega}-$$
$$-D\cdot\frac{cos\left[\Omega\cdot\left(T+X-N_4-1\right)+N_4\cdot\frac{\pi}{2}\right]\cdot sin\left[N_4\cdot\left(\Omega-\frac{\pi}{2}\right)\right]}{cos\,\Omega} \tag{21}$$

$cos\,\Omega\neq 0$.

It is important to review whether the result fulfils the initial and boundary conditions or not (2). Next investigation is to substitute these conditions to expressions (20) and (21). We circumscribe the conditions (2) with the labels (7) into non-dimensional form to this, therefore:

$T=0 \qquad V(X,0)=0, \qquad P(X,0)=1$

$X=0 \qquad V(0,T)=B\cdot sin\,\Omega T$

$X=1 \qquad P(1,T)=1$

At $T=0$ is out of (19) it is to recognize that $N_1=-1$ and $N_4=0$. In this case value of the factors $sin(N_1+1)\cdot\varepsilon$ and $sin\,N_4\cdot\varepsilon$ in (18) and (19) is zero, and so directly that $V(X,0)=0$ and $P(X,0)=1$.

In $X=0$ arises out of (19) that $N_1=N_4$ used; this in (18) and with trigonometric identities transformed receives one the prescribed boundary condition:

$$V(0,T)=-B\left[\frac{sin\left[\Omega T+N_1\cdot\varepsilon\right]\cdot sin\left[(N_1+1)\cdot\varepsilon\right]-sin\left[\Omega T-\left(N_1+1\right)\cdot\varepsilon\right]\cdot sin\,N_1\cdot\varepsilon}{cos\,\Omega}\right]=$$

$$=-B\left[\frac{-cos\left(\Omega T+\varepsilon\right)\cdot cos\left(\Omega T-\varepsilon\right)}{2\cdot cos\,\Omega}\right]=-B\frac{sin\,\Omega T\cdot sin\,\varepsilon}{cos\,\Omega}=B\frac{sin\,\Omega T\cdot cos\,\Omega}{cos\,\Omega}=B\,sin\,\Omega T$$

In $X=1$ is out of (19) to see, that $N_1=N_4-1$; this in (21) set and moved together, receives one (using ε):

$$P(1,T)=1-D\frac{\left\langle sin\left[\Omega T-\Omega-\left(N_4-1\right)\cdot\varepsilon\right]+sin\left[\Omega T+\Omega-\left(N_4+1\right)\cdot\varepsilon\right]\right\rangle\cdot sin\,N_4\varepsilon}{cos\,\Omega}$$

If the identity is used regarding the sum of the sinus both angle, one receives the expression:

$$P(1,T)=1-D\cdot\frac{sin\,N_4\cdot\varepsilon}{cos\,\Omega}\cdot sin\left(\Omega T-N_4\cdot\varepsilon\right)\cdot cos\left(\Omega-\varepsilon\right)$$

Out of (19) $\Omega - \varepsilon = \frac{\pi}{2}$, therefore the cosine of this expression is also zero. $P(1,T) = 1$ constant, that with the desired boundary condition agrees. Expressions (20) and (21) satisfy the non-dimensional forms of the equation system (1) with which exception of the points are not where $P(X,T)$ and $V(X,T)$ derivable. At these places, also the original expressions (3) and (4) kink points have however. Latter is maintained only based on the record of the functions.

4 Developing the Closed Form Calculated by the Laplace-Transformation Method

Only the expression (5) of the speed is reduced - like under A - because the expression of the pressure in more by more analogy manner can be transformed. With the labels (7), becomes (5) written:

$$\begin{aligned} V(X,T) = {} & B \cdot sin\Omega \cdot (T - X) \cdot \sum_{n=0}^{N_1} (-1)^n \cdot cos\, 2n \cdot \Omega - \\ & - B \cdot cos\, \Omega \cdot (T - X) \cdot \sum_{n=0}^{N_1} (-1)^n \cdot sin 2n \cdot \Omega + \\ & + B \cdot sin\Omega \cdot (T + X - 2) \cdot \sum_{n=0}^{N_1} (-1)^n \cdot cos\, 2n \cdot \Omega - \\ & - B \cdot cos\, \Omega \cdot (T + X - 2) \cdot \sum_{n=0}^{N_1} (-1)^n \cdot sin 2n \cdot \Omega \end{aligned} \tag{22}$$

where under application of the next to conditions $sin\omega(T - \gamma) = 0$ if $T < \gamma$ and $T > 0$; $\gamma > 0$ the equations (5) and (6), and the labels

$$N_1 = Max\left[0; int\left(\frac{T - X}{2}\right)\right]; \; N_2 = Max\left[0; int\left(\frac{T + X}{2} - 1\right)\right] \tag{23}$$

The meaning of that is, if $T - X < 0$; $sin\, \Omega(T - X) = cos\, \Omega(T - X) = 0$ and if

$T + X - 2 < 0$; $sin\, \Omega(T + X - 2) = cos\, \Omega(T + X - 2) = 0$ must be.

The expressions with summation in (22) can be written in same manner in a simpler form. One from that, for example

$$\sum_{n=0}^{N} (-1)^n \cdot sin 2n \cdot \Omega \tag{24}$$

The label introduced $\varepsilon = \Omega - \frac{\pi}{2}$ and used into the context (22), receives one after the same conversions

$$\sum_{n=0}^{N}(-1)^n \cdot sin(n\pi + 2n\cdot\varepsilon) = \sum_{n=0}^{N}(-1)^n \cdot sin\, n\pi \cdot cos\, 2n\varepsilon + \\ + \sum_{n=0}^{N}(-1)^n \cdot cos\; n\pi \cdot cos\, 2n\varepsilon \tag{25}$$

The first term on the right side is always same zero, that factor $(-1)^n \; cos\, n\pi$ in the second term yields for all “n” one, therefore Eq. (24) can write further similar:

$$\sum_{n=0}^{N} sin\, 2n\varepsilon = \frac{cos\,(N+1)\varepsilon \cdot sin\, N\varepsilon}{sin\,\varepsilon}$$

The equivalence was taken from Tafeln [3]. If the expression

$$\sum_{n=0}^{N}(-1)^n\, cos\, 2n\Omega$$

in the same manner is simplified, one receives the simplified form:

$$\frac{sin\,(N+1)\varepsilon \cdot cos\; N\varepsilon}{cos\;\varepsilon}$$

The results on the expression (22) for V(X, T) used, receives one the simplified context

$$\begin{aligned} V(X,T) &= B\cdot sin\,\Omega\cdot(T-X)\cdot\frac{sin(N_1+1)\cdot\varepsilon\cdot cos\, N_1\varepsilon}{sin\,\varepsilon} - \\ &- B\cdot cos\,\Omega\cdot(T-X)\cdot\frac{sin(N_1+1)\cdot\varepsilon\cdot cos\, N_1\varepsilon}{sin\,\varepsilon} + \\ &+ B\cdot sin\,\Omega\cdot(T+X-2)\cdot\frac{sin(N_2+1)\varepsilon\cdot cos\, N_2\varepsilon}{sin\,\varepsilon} - \\ &- B\cdot cos\,\Omega\cdot(T+X-2)\cdot\frac{sin(N_2+1)\varepsilon\cdot cos\, N_2\varepsilon}{sin\,\varepsilon} \end{aligned} \tag{26}$$

Let's introduce the label $N_4 = N_2 + 1$ and simplify like (26) under application of the well known trigonometric equivalences and $\varepsilon = \Omega - \frac{\pi}{2}$, therefore:

$$V(X,T) = B \cdot \frac{sin\left[\Omega \cdot (T - X - N_1) + N_1 \cdot \frac{\pi}{2}\right] \cdot cos\left[(N_1 + 1) \cdot \Omega - N_1 \cdot \frac{\pi}{2}\right]}{cos\,\Omega} + $$
$$+ B \cdot \frac{cos\left[\Omega \cdot (T + X - N_4 - 1) + N_4 \cdot \frac{\pi}{2}\right] \cdot sin\left[N_4 \cdot \left(\Omega - \frac{\pi}{2}\right)\right]}{cos\,\Omega} \tag{27}$$

After similar simplifications, the pressure arises to

$$P(X,T) = 1 + D \cdot \frac{sin\left[\Omega \cdot (T - X - N_1) + N_1 \cdot \frac{\pi}{2}\right] \cdot cos\left[(N_1 + 1) \cdot \Omega - N_1 \cdot \frac{\pi}{2}\right]}{cos\,\Omega} - $$
$$- D \cdot \frac{cos\left[\Omega \cdot (T + X - N_4 - 1) + N_4 \cdot \frac{\pi}{2}\right] \cdot sin\left[N_4 \cdot \left(\Omega - \frac{\pi}{2}\right)\right]}{cos\,\Omega} \tag{28}$$

$cos\,\Omega \neq 0$ and $T > 0$ where $N_1 = int\left(\frac{T - X}{2}\right)$ $N_4 = int\left(\frac{T + X}{2}\right)$.

In (23) a condition was made for N_1 and N_2 ($N_4 = N_2 + 1$) that they cannot be negative. Condition that T<0 must not be made because it is out of the initial conditions. The values $\varepsilon = k \cdot \pi$ do not belong to the domain of the velocity formulas (27) and pressure formulas and (28), while in the formulas (5) and (6) it belong to the definition area, however the corresponding expressions have the same limit. (When $\varepsilon = k \cdot \pi$ it is the own resonance frequency of the pipe.)

It is to be recognized that velocity relation and pressure relation (20) and (21) simplified with the Fourier method will match relations (27) and (28) simplified with Laplace-transformation.

The boundary conditions of the physical process were indicated through the equations (2). A fully sinusoidal excite expression was supposed on the closed end of the pipe. As equations (1) and the boundary conditions are homogeneous linear, the resulting solution can be produced as a sum of divided solutions of several sinusoidal expressions with different amplitudes and frequencies. The simplified forms have especially large advantages in these cases as in [1] indicated end formulas. The advantages arise in the simpler calculation and in the less required time duration. The next chapter is presenting a practical calculation of these results.

5 Results of Analytical and Numerical Solutions

Let us look through the gas dynamic solution of this mathematical problem. The gas dynamic model was solved with the help of method of characteristics. The differential equation system has a different form from the equation (1).

$$\frac{\partial c}{\partial t}+c\cdot\frac{\partial c}{\partial t}+\frac{1}{\rho}\cdot\frac{\partial p}{\partial x}=0;\qquad \frac{\partial \rho}{\partial t}+c\cdot\frac{\partial \rho}{\partial t}+\rho\cdot\frac{\partial c}{\partial x}=0 \tag{29}$$

The main difference between the two models is that the "a" value, the speed of sound is not constant in the gas dynamic model. The wave equation of gas dynamics uses the basic relations without neglect, although it is more expedient to modify the equation system in such a manner that the dependent variables in it should be the velocity c = c (x,t) and the speed of sound a = a (x, t). Hence

$$a\cdot\frac{\partial c}{\partial x}+\frac{2}{\kappa-1}c\cdot\frac{\partial a}{\partial x}+\frac{2}{\kappa-1}\cdot\frac{\partial a}{\partial t}=0;\quad \frac{\partial c}{\partial t}+c\cdot\frac{\partial c}{\partial t}+\frac{2}{\kappa-1}a\cdot\frac{\partial a}{\partial x}=0 \tag{30}$$

The quasi-linear partial differential equation system is hyperbolic. The applied numerical solution is called the method of characteristics. The details are in the [2]. And there is no analytical solution for this differential equation system. At the physical plane [x, t,] along the projection of the characteristic curve with a given tangent, the correlations of the independent variables can be written in the non-dimensional form, in the well-known manner, as

$$\frac{\kappa-1}{2}\cdot V+A=\alpha;\quad \frac{\kappa-1}{2}\cdot V-A=\beta;\qquad \cdot V=\frac{c}{a_0}$$

Where α *and* β are the Riemann-variables and $V=c/a_0$: the non-dimensional velocity and $A=a/a_0$: the non-dimensional speed of sound. The isentropic change of state has been used in each case to determine pressure. The boundary condition is a little bit different, the pressure on the open end of the pipe is not constant. The details can be seen in the [5]. In Figure 2 you can see different model solutions and measured results from [7]. The calculated gas oscillation frequency $\omega=317,291/s$ is very close to the resonant frequency $\omega_r=327,191/s$.

According to Figure 2 the solution of the wave equation of acoustics has shown a difference of about l0 % less in the amplitude, related to that one of gas dynamics. On the basis of the examinations it can be stated that the pressure patterns forming in the system are equal regarding the frequency, depending on the calculation method, but from the point of view of the pressure values, considerable differences appear, according to the preceding valuations. The analytical solution can be used to predict the gas velocity and pressure fluctuation inside a straight pipe. It is simpler to use than the gas dynamic model.

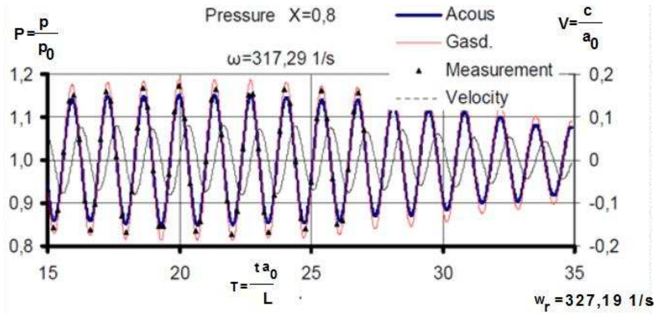


Figure 2
Comparison of the different solutions

We have some measurement results, too. The measurement data are between the acoustic and the gas dynamic model results. In the real case the friction force decreases the velocity and the pressure amplitude. For example, in the [8] J. J. Patel and U. V. JoshI show CFD calculation results of a gas oscillation where there is friction in the pipe.

6 Some Calculated Results

The gas oscillation problem can be solved analytically not only in a clear sinusoidal excitation. It is an opportunity for the analytical solution if the piston velocity is more complicated function of time, for example when the piston rod is not infinite. The differential equation system is the same (1)

$$\frac{\partial c}{\partial t} = -\frac{1}{\rho} \cdot \frac{\partial p}{\partial x}; \qquad \frac{\partial c}{\partial x} = -\frac{1}{\rho \cdot a^2} \cdot \frac{\partial p}{\partial t} \tag{31}$$

The appropriate initial and boundary conditions are the next:

$$c(x,0) = 0 \qquad p(x,0) = p_0$$

$$c(0,t) = R\omega \cdot \left[sin(\omega \cdot t) + \frac{\lambda}{2} \cdot sin(2\omega \cdot t) \right] \quad p(L,t) = p_0 \tag{32}$$

where the meaning of the symbols are: " ρ " the density of the medium; "a_0=347,8 m/s" the speed of sound (constant is); "p" the pressure; "$R = 0{,}011 m$" the radius of the crankshaft; “$\ell = 0{,}044 m$“ the piston rod length; “$\lambda = \frac{\ell}{R} = 0{,}25$“ the piston rod ratio; "ω" drive its circle frequency; "$L = 5m$" the pipe length. The velocity $c(0,t)$ function is an approximation: the first two terms of a Taylor’s series for a finite rod length. The first term is at infinite piston rod length in the previous section.

The (31) equation system is homogenious linear so the solution of a sum of two functions is the sum of the two solutions. In this case pressure and velocity oscillation were calculated. Figures 3 and 4 show the results. Non-dimensional velocity, $V = \frac{c}{a_0}$ and non-dimensional pressure, $P = \frac{p}{p_0}$)are used in the diagrams. Figure 3 shows the pressure functions at infinite and finite piston rod while Figure 4 shows the velocity functions. Pressure and velocity functions were calculated at $X = 0{,}2$ and $X = 0{,}8$. The function is zero until the wave from the piston reaches the observation point at $X = 0{,}2$ or $X = 0{,}8$. The difference between the finite and infinite functions is bigger until two or three oscillations. After some periods the difference can be neglected. You can see that the functions of pressure and velocity have some broken points where the functions are not derivable. It is an interesting property of the solution.

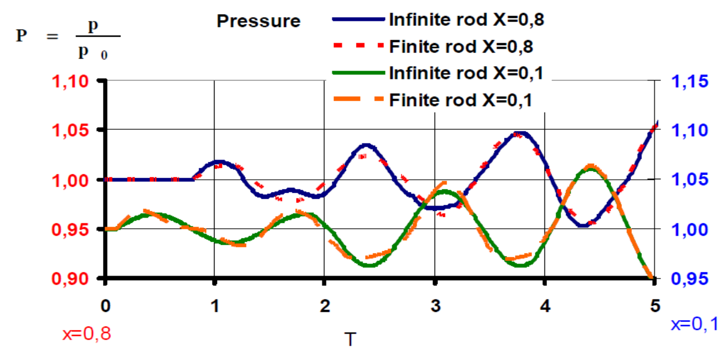


Figure 3
Comparison the pressure functions of the “infinite” and “finite” piston rod solutions
(The vertical scale is shifted)

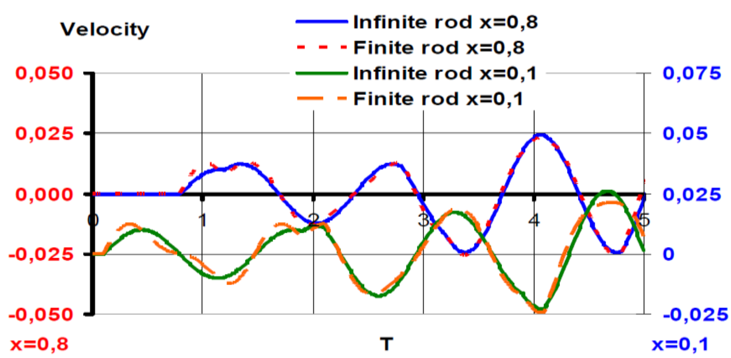


Figure 4
Comparison the velocity functions of the “infinite” and “finite” piston rod solutions (The vertical scale is shifted)

Conclusion

Two different mathematical infinite series solutions exist in the literature. In this article they were transformed out of an endless row into closed form. Both simplifying methods reached the same result as a closed form. It is simpler to use these solutions than the gas dynamic numerical simulation. The comparison with other numerical solutions and measured results shows a good agreement.

References

[1] Hoffmann, A - Fényes, T. (1982) 'Rechenferfharen zur Unterzuchungen von Wellenerseinungen in Rohrleitungen' *Periodika Polytechnica Mech. Eng.* 26/1 (1982) U.T. Budapest

[2] Jimenez, I. (1973) 'Fluid Mechanics', Great Britain, 23

[3] Ryzhik, I. M. - Gradtein, I. S. (1957) 'Tafeln', Berlin Deutscher Verlag der Wissenschaften

[4] Szlivka, F. (1983) 'Herstellung der geschlossenen Lősung des Differentialgleichungssystems von Gasschwingungen', *Periodica Polytechnica Mech. Eng.* 27/1 (1983)

[5] Bencze, F. - Hoffmann. A. - Szlivka, F. (1983) 'Untersuchung der Wellenerscheinungen in Rohrleitungen', *ZAMM* 63. pp. 228-231

[6] Engelhardt, J. (2012) 'Schwingungen von Rohrleitungen aktiv mindern, Artikel aus der Zeitschrift' *3R Fachzeitschrift für sichere und effiziente Rohrleitungssysteme* ISSN: 2191-9798. Nr. 5, 2012, Seite 396-400, Abb.

[7] Bencze, F. - Hoffmann. A. - Szlivka, F (1983) 'Examination of Forced Gas Oscillations by the Aid of Computer*', The Seventh Conference on Fluid Machinery*, Budapest 1983. pp. 228-231

[8] J. J. Patel and U. V. JoshI (2012) 'Study of the Oscillating Flow in the Circular Tube Using CFD', *World Journal of Science and Technology* 2012, 2(4) pp. 46-49, ISSN: 2231-2587

Symbolic Analysis as Universal Tool for Deriving Properties of Non-linear Algorithms – Case study of EM Algorithm

Vladimir Mladenović, Miroslav Lutovac, Dana Porrat

The Higher Technical School of Professional Studies
Nemanjina 2, 12000 Požarevac, Serbia

The University Singidunum
Danijelova 32, 11000 Belgrade, Serbia

The School of Computer Science and Engineering
Hebrew University of Jerusalem
Jerusalem 91904, Israel

Email: vlada@open.telekom.rs, mlutovac@singidunum.ac.rs, dana.porrat@huji.ac.il

Abstract: Many researchers start their work by studying theory in order to get better insight into measured phenomena. Sometimes they cannot get the numeric values of parameters used in published equations. This is even more complicated when the theory is statistical and there are no closed form deterministic solutions. In this paper we introduce an original approach and method to analyzing a popular and frequently cited tutorial paper on Expectation-Maximization (EM) algorithm. The original paper has sufficient information to understand the algorithm. Using tools of Computer Algebra System and methods of Symbolic Processing (SP), the typewriting errors are discovered and the re-derived equations are used for automatic generation of algorithmic code. The examples are evaluated using automatically derived code and the final numeric values agree with the values from the original paper. The derived results are used for further optimization, such as deriving the computational error in the closed form. From the closed form solutions, the precision can be derived in terms of number of iterations, or the minimal number of iterations can be expressed in terms of the required precision. This helps to optimize the algorithm parameters so that the algorithm becomes more efficient.

Keywords: Symbolic processing; EM algorithm; iteration; convergence; ML estimation; Computer algebra system

1 Introduction

Computer algebra is a field of computer science concerned with the development, implementation, and application of algorithms that manipulate and analyze expressions. Some of the issues are developing algorithms (such as algebraic algorithms, symbolic-numeric algorithms), studying how to build computer algebra systems (memory management, higher-order type systems, optimizing compilers), and mathematical knowledge management (representation of mathematical objects). Computer algebra is based on well-defined semantics, compose constructions, and algebraic algorithms. Symbolic computation is concerned with alternative forms, partially-specified domains, and symbolical derivations in general. Computer algebra systems is used to simplify rational functions, factor polynomials, find the solutions to a system of equations, give general formulas as answers, model industrial mathematical problems, and various other manipulations. Symbolic processing can be treated as a transformation of expression trees, such as symbols for operations ("+", "sin"), variables, constants, and simplification (for example, expression equivalence).

The EM (Expectation-Maximization) algorithm is analysed with a new type of processing, called symbolic processing. We illustrate the main drawbacks of numerical tools that have led to erroneous conclusions in the application of the algorithm through a practical example presented in [1].

Using symbolic processing, we derive the error function in closed form as a function of the number of iterations. Next, we determine the minimum number of iterations required to obtain a correct result, and as well as the required number of iterations as a function of the number of accurate bits of the final result of an algorithm. The main aim of this paper is to show how Computer Algebra (CA) and symbolic processing can be used in the design of numeric algorithms, and in discovering the properties of EM algorithm.

2 The Method of Symbolic Processing

Modern research in the field of engineering science usually starts with theoretical analysis. The next step can be simulation in order to prove the theoretical assumptions. The final step is practical implementation and measurements. Usually, the theoretical derivations are made on ideal parameter values, such in infinite number of samples or known signal values for time from $-\infty$ to $+\infty$. Simulation is used for simulation of systems in order to gain insight into their functioning. For a finite number of measured values, one can expect that the result agrees well with the simulation.

Probably, the first critical step is to set the simulation parameters and write a code that fits the theory. The second critical step is to choose the number of iterations or the number of input samples.

Symbolic processing can help to error-free derive simulation code, and to find typewriting errors in published results. This paper presents such an analysis. Next, symbolic processing can be used for finding the processing errors as the closed-form expression for computing the number of required iteration steps; or the error function due to finite word-length [2, 3]. Therefore, symbolic processing can help gain insight into how a system works, which is preferred to experimenting with numeric simulations.

Symbolic processing plays an important role in education. Even more it can be used in practice because the numerical tools may have problems with accuracy and a significant increase in the complexity. Primarily, the analysis area is essentially continuous and infinite [4]. This paper aims to highlight the role of symbolic processing applications, especially where the numerical approaches fail to produce the right algorithms that are important for many areas of engineering research, especially in communications.

Symbolic processing repairs certain disagreement between theoretical performance and numerical simulations. Numerical processing may show unsatisfactory results [5]. It can drive the researchers to erroneous conclusions. Using the symbolic tools, it is easy to find and prove mistakes and gain a better insight into the whole process of analysis, simulation and modeling.

Symbolic processing works with symbols. Any symbol can be replaced by a number when symbolic expressions become impractical, for example, for plotting a response to a specific excitation.

Therefore, writing programs using symbolic processing can be seen as a set of instructions that manipulate the symbols and can be used to perform a much wider range of activities [6]. The efficiency of symbolic processing becomes more important if systems and signals are more complex.

There is no algebraic solution for all forms of processing, and symbolic processing will not be able to solve all problems. Even then, the final result can be expressed in a form of special function that can be computed for known numerical values of system parameters.

In the next section, we will demonstrate the derivation of the EM algorithm using symbolic processing.

3 The Expectation-Maximization Algorithm

The Expectation-Maximization (EM) algorithm is ideally suited to estimate the parameters of a probability function, especially when direct access to all of the data is impossible [5]. A brief description of the algorithm can be found in literature [5] that is suitable for signal processing practitioners. An illustrative example is also presented there. Unfortunately, the algorithmic steps there disagree with the theoretical statements, and do not provide the numeric results given in Table of [5].

Those who want to test the knowledge learned from the example may be confused about the reason for such disagreements. The error can be in the initial statements, in the derived algorithmic steps, or it can be typewriting error.

A computer algebra system (specifically Mathematica) and symbolic processing are used in this paper to start analysis from the theoretical statements, derive the algorithmic code, and calculate the final results.

It is important to emphasize the role of the new analysis of the EM algorithm because EM is the backbone for algorithms that are used in significant analyse, measurement data processing and simulations [1, 7, 8].

The main goal of EM is approximation Maximum Likelihood from incomplete data. The most general form starts from the pdf of the incomplete data:

$$g(y|\theta) = \int_{\chi(y)} f(x|\theta)dx \tag{1}$$

where $f(x|\theta)$ is the probability density function of complete data, and θ is set of parameters.

An observation y determines a subset of χ that is denoted as $\chi(y)$.

The EM algorithm consists of two steps: the E-step and the M-step. The E-step (Expectation step) takes an expectation with respect to the unknown variables using the current estimate of the parameter and conditioned upon the observation. The M-step (Maximization step) then provides a new estimation of the parameters, in that it produces a maximum-likelihood (ML) estimates of parameters when there is many-to-one mapping. These two steps are iterated until convergence [5].

For the E-step compute:

$$Q(\theta|\theta^{[k]}) = E[\log f(x|\theta)|y, \theta^{[k]}] \tag{2}$$

where $Q(\cdot)$ is $Q-$ function.

For the M-step compute:

$$\theta^{[k+1]} = \arg\max_{\theta} Q(\theta \mid \theta^{[k]}) \tag{3}$$

4 An Example Problem

The example problem is to find the unknown value of *p* using the EM algorithm. In this example the initial value of p is defined as a symbol, and we know that its true value is 1/2 (as specified in the example in [9]). The application of EM to this example is calculated symbolically (using Mathematica), and we show mathematical description.

The expressions which are derived in this paper, were not derived in the paper [5] and cannot be obtained by hand. On the other side, the estimated parametar *p* is obtained with the value of 0.52 in the paper [5], but nowhere it explains the deviation from the true value of 0.5. We obtain the exact value of 0.5 and explain the way how we get it.

This section repeats Example from [5]. The application of EM to this example is calculated symbolically (using Mathematica), and we show mathematical description. Let a set of random variables is $\{X_1, X_2, \ldots, X_m\}$. For example X_1 represents the number of round dark objects, X_2 represents the number of square dark objects, and X_m represents the number of light objects. Let x_1, x_2, …, x_m be a sequence of outcomes of the random variables X_1, X_2, …, X_m that have been observed. It is assumed that X_i is independent of X_j for $i \neq j$. If the number of outcomes (total number of observed objects) is n, then x_i are nonnegative integers such that

$$\sum_{i=1}^{m} x_i = n \tag{4}$$

If X_1, X_2, …, X_m are mutually exclusive events with $P(X_1 = x_1) = p_1$, ..., $P(X_m = x_m) = p_m$, then the probability X_1 that occurs x_1 times, ..., X_m occurs x_m times is given by

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m) = n! \prod_{i=1}^{m} \frac{p_i^{x_i}}{x_i!} \tag{5}$$

where p_i are constants with $p_i > 0$ and

$$\sum_{i=1}^{m} p_i = 1 \tag{6}$$

The joint distribution of X_1, X_2, …, X_m is a multinomial distribution and $P(X_1 = x_1, X_2 = x_2, \ldots, X_m = x_m)$ is given by the corresponding coefficient of the multinomial series

$$\left(p_1 + p_2 + \cdots + p_m\right)^n \tag{7}$$

Assume further that probabilities are also known As an example, assume that the objects are known to be trinomial distributed, that is $m = 3$, as in [5]. Assume further that probabilities are also known

$$p_1 = \frac{1}{4} \tag{8}$$

$$p_2 = \frac{1}{4} + \frac{p}{4} \tag{9}$$

Then, from (6) follows

$$p_3 = 1 - \left(p_1 + p_2\right) = \frac{1}{2} - \frac{p}{4} \tag{10}$$

We can use a notation for probability density function

$$f(x_1, x_2, x_3 \mid p) = P(X_1 = x_1, X_2 = x_2, X_3 = x_3) \tag{11}$$

For the unknown parameter p, the function becomes

$$f(x_1, x_2, x_3 \mid p) = \frac{n!}{x_1! x_2! x_3!} \left(\frac{1}{4}\right)^{x1} \left(\frac{1}{4} + \frac{p}{4}\right)^{x2} \left(\frac{1}{2} - \frac{p}{4}\right)^{x3} \tag{12}$$

Suppose that a feature extractor is employed that can distinguish the color of the objects, but cannot distinguish the shape. From the measurement, the number of dark objects is known, $y_1 = x_1 + x_2$, but the number of x_1 or x_2 is unknown. The basic idea of EM algorithm is to find expressions that can be used for iterated computing of x_1 or x_2 based on known $y_1 = x_1 + x_2$ and distribution function.

4.1 Expectation Step

In the E-step, the expected values of x_1, x_2, and x_3, have to be computed using initial estimate and measured data. The algorithm is derived using the knowledge of the multinomial distribution.

Starting with measured y_1 (where $y_1 = x_1 + x_2$) we have to derive expressions that can be used in iterated algorithm in order to find the number of objects x_1 and x_2, and the parameter p. In other words, we have to derive expected values of x_1 and x_2 for measured y_1 and estimate of p in the k th iteration

$$x_1^{[k+1]} = E[x_1 \mid y_1, p^{[k]}] \tag{13}$$

$$x_2^{[k+1]} = E[x_2 \mid y_1, p^{[k]}] \tag{14}$$

The expected values can be derived using the multinomial distribution (assuming that $x_1 + x_2 + x_3 = n$)

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = n! \frac{p_1^{x1}}{x_1!} \frac{p_2^{x2}}{x_2!} \frac{p_3^{x3}}{x_3!} \tag{15}$$

Notice that $(X_1 + X_2, X_3)$ is binomial with class probability $(p_1 + p_2, p_3)$, then (assuming $p_1 + p_2 + p_3 = 1$, $y_1 + x_3 = n$) the probability $P(Y_1 = y_1, X_3 = x_3)$ becomes

$$P(Y_1 = y_1, X_3 = x_3) = n! \frac{(p_1 + p_2)^{y1}}{y_1!} \frac{p_3^{x3}}{x_3!} \tag{16}$$

Assume that the probability of some event x is known, say $P(x)$, and the number of events can be from 0 to n, the expectation value of x is defined by

$$E[x] = \sum_{x=0}^{n} x\, P(x) \tag{17}$$

Noticing that $y_1 = x_1 + x_2$ and $0 \le x_1 \le y_1$, expectation of x_1 requires to know conditional probability $P(x_1)$,

$$E[x_1] = \sum_{x=0}^{y_1} x_1\, P(x_1) \tag{18}$$

Assuming that y_1 has occurred, the conditional probability $P(x_1)$ equals

$$P(x_1) = \frac{P(X_1 = x_1, Y_1 = y_1)}{P(Y_1 = y_1)} \tag{19}$$

Applying those definitions on $x_1^{[k+1]}$, the iterated algorithm can be derived. Unfortunately, the presented relations in [5] disagree with expected.

4.2 Maximization Step

The maximum of some function, say $f(x_1, x_2, x_3 \mid p)$, can be obtained by taking the derivative of the function f with respect to parameter p, equating to zero, and solving for p. Since the logarithm is monotonically increasing, maximizing $\log(f)$ is equivalent to maximizing f.

The function f is known in terms of the parameter p, and we can try to solve for p the following equation

$$\frac{\mathrm{d}}{\mathrm{d}p}\log\left(f(x_1, x_2, x_3 \mid p)\right) = 0 \tag{20}$$

If solution exists, it becomes the second part of EM algorithm.

5 Implementation of Computer Algebra System Code into EM Algorithm

In this paper, we re-derive the expressions obtained in [5] because they do not produce the expected results in the numeric example presented in [5].

Computer algebra systems (CAS), such as Mathematica [10], can be used for derivation of all equations required for algorithm. The main motive is to avoid manual derivation and thus derive all relations without error.

All derivations are in a document called notebook that looks like technical paper in electronic form. There are cells with textual description, such as title, section head, text with formula, but also cells named input and output. In the input cells we write expressions that can be evaluated.

5.1 Entering Knowledge in CAS

The simplest way to re-derive equations is to start with definitions. The first line of the first input cell can be the definition of the trinomial distribution

$$\texttt{f[x1\_, x2\_, x3\_, p1\_, p2\_, p3\_] := (x1 + x2 + x3) !}\ \frac{\texttt{p1}^{\texttt{x1}}}{\texttt{x1 !}}\ \frac{\texttt{p2}^{\texttt{x2}}}{\texttt{x2 !}}\ \frac{\texttt{p3}^{\texttt{x3}}}{\texttt{x3 !}}$$

By executing the input cell there is no visible output cell. As a matter of fact, we define new knowledge that becomes a part of the CAS. The meaning of the code is that we define a function called **`f`**, a function with 6 arguments, (**`x1`**, **`x2`**, **`x3`**, **`p1`**, **`p2`**, **`p3`**), where instead of any argument with **_** we can use any other symbol or number (instead of **x1_** we can use x_1). The symbol **:=** tells to CAS to remember the expression that follows, but without execution, and so there is no output cell.

Once the knowledge is inputted to CAS, we can use for presentation or evaluation. The mathematical expressions may look different than a code. Let us replace symbolic arguments with symbols that appear in textbooks (use x_1 instead of **x1**), and present in the traditional form (transform the expression using command **//TraditionalForm**).

This expression still does not look same as equation (15). Therefore, we evaluate the expression again for the same symbolic arguments, and in the next line use the expression (% stays for previous expression) and apply substitution as in equation (4), and finally, display the result in the traditional form

```
In[2]:= f[x1, x2, x3, p1, p2, p3] // TraditionalForm
Out[2]//TraditionalForm=
```

$$\frac{(x_1 + x_2 + x_3)!\, p_1^{x_1}\, p_2^{x_2}\, p_3^{x_3}}{x_1!\, x_2!\, x_3!}$$

Now, we obtain the same result as equation (15). This means that inputted knowledge is visually identical as in textbook. Notice that the last output is a result of the line that follows input In[3], that is actually input In[4], and the displayed output is Out[4].

The binomial distribution can be defined in a similar way

```
In[3]:= f[x1, x2, x3, p1, p2, p3];
        % /. x1 + x2 + x3 → n // TraditionalForm
Out[4]//TraditionalForm=
```

$$\frac{n!\, p_1^{x_1}\, p_2^{x_2}\, p_3^{x_3}}{x_1!\, x_2!\, x_3!}$$

This defines a function called f, a function with 5 arguments, (**y1**, **y2**, **p1**, **p2**, **n**). The head of the function is the same as for the trinomial distribution, but the difference is in the number of arguments. Instead of any argument with _ we can use any other symbol or number (instead of **y2_** we can use x_3, and instead of **p1_** we can use $p_1 + p_2$). Replacing arguments with appropriate symbols, and using the command for displaying expression in the traditional form, equation (16) is derived

```
In[6]:= f[y1, x3, p1 + p2, p3, n] // TraditionalForm
Out[6]//TraditionalForm=
```

$$\frac{n!\, p_3^{x_3}\, (p_1 + p_2)^{y_1}}{x_3!\, y_1!}$$

5.2 Derivation of The Expectation Step

In order to find out expectation of x_1, the conditional probability $P(x_1)$ should be computed using equation (19)

```
In[7]:= cp1 = f[x1, x2, x3, p1, p2, p3] / f[x1 + x2, x3, p1 + p2, p3, x1 + x2 + x3] ;
        Px1 = cp1 /. x2 → y1 - x1 // Simplify;
        % /. {p1 → p₁, p2 → p₂, x1 → x₁, y1 → y₁};
        % // TraditionalForm
```

Out[10]//TraditionalForm=

$$\frac{y_1!\, p_1^{x_1} (p_1 + p_2)^{-y_1} p_2^{y_1 - x_1}}{x_1!\,(y_1 - x_1)!}$$

In the nominator the trinomial distribution is used, while in the denominator the binomial distribution is applied. The command **`Simplify`** is use to find the simplest form of the expression. Next, the derived expression is used in equation (18), and the final result is not the same as in [5]

```
In[11]:= ex1 = Sum[x1 Px1, {x1, 0, y1}];
         % /. {p1 → p₁, p2 → p₂, y1 → y₁}
         % // TraditionalForm
```

Out[12]= $\frac{p_1\, y_1}{p_1 + p_2}$

Out[13]//TraditionalForm=

$$\frac{p_1\, y_1}{p_1 + p_2}$$

The same procedure is used for deriving expected value of x_2, but this time the same result is as in [5]

```
In[14]:= cp2 = f[x2, x1, x3, p2, p1, p3] / f[x1 + x2, x3, p1 + p2, p3, x1 + x2 + x3];
         Px2 = cp2 /. x1 → y1 - x2 // Simplify;

In[16]:= ex2 = Sum[x2 Px2, {x2, 0, y1}];
         % /. {p1 → p₁, p2 → p₂, y1 → y₁};
         % // TraditionalForm
```

Out[18]//TraditionalForm=

$$\frac{p_2\, y_1}{p_1 + p_2}$$

Sum of expected values of x_1 and x_2 should be y_1, that is obvious from the derived expressions using CAS, but also proves that one of two results in [5] is not correct. Now, it is obviously that the sign – in the denominator of the expectation of x_1 is not correct in [5, Box 2, equation (37)]. Also, in the original text [5], it used y and y_1 for the same variable, that may be also confusing for some readers.

The presented derivation in this paper shows that all results can be found out automatically, and intermediate results and defined functions can be visually identified to be the same with those from textbooks.

Once the general expressions are derived, a particular solution follows as a special case. For example, if the probabilities are known, they can be denoted by subscript **e**

```
In[19]:= p1e = 1/4;
         p2e = 1/4 + p/4;
```

The expected value of x_1 for this specific case can be derived by substituting parameters in the general solution for x_1

```
In[21]:= ex1 /. {p1 → p1e, p2 → p2e, y1 → y₁};
         % // TraditionalForm
```

Out[22]//TraditionalForm=

$$\frac{y_1}{4\left(\frac{p}{4} + \frac{1}{2}\right)}$$

Following the same procedure, the expected value of x_2 is derived

```
In[23]:= ex2 /. {p1 → p1e, p2 → p2e, y1 → y1};
         % // TraditionalForm
```

Out[24]//TraditionalForm=

$$\frac{\left(\frac{p}{4}+\frac{1}{4}\right) y_1}{\frac{p}{4}+\frac{1}{2}}$$

The derived results differ from that in [5, equations (5) and (6)], and instead of p, the automatically derived expressions have $p/2$.

5.3 Derivation of The Maximization Step

For known probabilities of a specific case, the trinomial distribution can be expressed in terms of x_1, x_2, x_3, and the parameter p

```
In[25]:= f[x1, x2, x3, p1, p2, p3];
         % /. {p1 → p1e, p2 → p2e, p3 → (1 - p1e - p2e)};
         fe = % /. {x1 → x1, x2 → x2, x3 → x3};
         % /. x1 + x2 + x3 → n;
         % // TraditionalForm
```

Out[29]//TraditionalForm=

$$\frac{4^{-x_1}\, n!\left(\frac{1}{2}-\frac{p}{4}\right)^{x_3}\left(\frac{p}{4}+\frac{1}{4}\right)^{x_2}}{x_1!\, x_2!\, x_3!}$$

The same expression is presented in [5]. Since this is a continuous function in p, the first derivative can be computed

```
In[30]:= d = D[Log[fe], p] // Simplify
```

Out[30]= $\frac{(-2+p)\, x_2+(1+p)\, x_3}{(-2+p)\,(1+p)}$

The parameter p can be expressed in terms of for the extreme (and again, the same expression is presented in [5])

```
In[31]:= sol1 = Solve[d == 0, p] // Flatten
```

Out[31]= $\left\{p \to \frac{2\, x_2 - x_3}{x_2 + x_3}\right\}$

But, the value of x_2 is unknown, Instead of x_2, we can use the expected value of x_2, that was derived in the expectation step. In order to obtain the same form of the final results as in [5], some rearrangements can be evaluated (to collect all terms in numerator and denominator that are multiplied by p)

```
In[32]:= sol1 /. x2 → ex2;
         s = % /. {p1 → p1e, p2 → p2e, y1 → y1} // Simplify
                   Collect[Numerator[p /. s], p]
         pnew = ─────────────────────────────────
                 Collect[Denominator[p /. s], p]
```

Out[33]= $\left\{p \to \frac{-(2+p)\, x_3 + 2\,(1+p)\, y_1}{(2+p)\, x_3 + (1+p)\, y_1}\right\}$

Out[34]= $\frac{-2\, x_3 + 2\, y_1 + p\,(-x_3 + 2\, y_1)}{2\, x_3 + y_1 + p\,(x_3 + y_1)}$

The final result is not the same as in [5]. It seems that $2p$ is used instead of p in the recurrence relation in [5].

The presented derivation in this section shows that the final result can be found out automatically following the mathematical notation from textbooks or published papers.

5.4 Derivation of The Code

Once the expressions are derived from inputting the knowledge into CAS, the algorithm in the form of code can be simply transformed.

The code can be derived for some input parameters, such as the total number of objects, number of dark or light objects, initial guess for p, required accuracy, or probabilities

```
In[35]:= Clear[k, p]
         n = 100;
         y1 = 63;
         x3 = n - y1;
         p[0] = 0;
         a = 10^-7;
         p1e = 1/4;
         p2e = 1/4 + p/4;
```

The expressions used in the code is a special case of the general expressions in terms of the known parameters or the previous value of p

```
In[43]:= pnew = pnew /. {x3 → x3, y1 → y1} /. p → p[k - 1];
         x1new = ex1 /. {p1 → p1e, p2 → p2e} /. p → p[k - 1];
         x2new = ex2 /. {p1 → p1e, p2 → p2e} /. p → p[k - 1];
```

Iterated algorithm can be a while loop, in that new number ob objects or new p is calculated

```
In[46]:= e = 1; k = 1;
   While[e > a, {
     p[k] = p_new;
     x1[k] = x1_new;
     x2[k] = x2_new;
     e = p[k] - p[k - 1];
     k = k + 1}]
```

Finally, the results of iterations can be viewed as table of values

```
Table[N[{i, x1[i], x2[i], p[i]}, 8], {i, 1, k - 1}] // TableForm
```

Number of displayed digits can be also specified

Out[48]//TableForm=

1.0000000	31.500000	31.500000	0.37956204
2.0000000	26.475460	36.524540	0.49029997
3.0000000	25.298157	37.701843	0.51409288
4.0000000	25.058740	37.941260	0.51883995
5.0000000	25.011514	37.988486	0.51977276
6.0000000	25.002255	37.997745	0.51995551
7.0000000	25.000441	37.999559	0.51999129
8.0000000	25.000086	37.999914	0.51999829
9.0000000	25.000017	37.999983	0.51999967
10.000000	25.000003	37.999997	0.51999993
11.000000	25.000001	37.999999	0.51999999

The values from table are the same as in [5]. Therefore, using CAS, we derive all expressions and code for EM algorithm. Also, we avoid typewriting errors in the original text, prove all of the steps, and compute the numerical results for a specific case. For example, it is obvious that y_1 cannot be 100 [5, page 50 after equation (8)], because $n = 100$ and $y_1 < n$. A typewriting error is in the derivation of expectation step [5, page 53, Box 2] where index in the summation formula is from 10. Another typewriting error is in the derivation of expectation step [5, page 49, equation (6)] where x_2 should be used instead of x_1.

6 Analytical Solution of The EM Algorithm

The EM algorithm can be further analyzed by employing the symbolic expressions derived using CAS. Instead of the numeric analysis, a closed form expressions can be derived. This can help us to gain insight into how some system works and understand the properties of the algorithm.

6.1 Finding Closed Form Solution

Firstly, we generate an equation that describes iterated computing of parameter p, knowing the initial value

```
In[49]:= Clear[k, p]

In[50]:= eq1 = {p[k + 1] == (pnew /. p → p[k]), p[0] == 0}
```

Out[50]= $\left\{p[1+k] == \dfrac{-2\,x_3 + 2\,y_1 + p[k]\,(-x_3 + 2\,y_1)}{2\,x_3 + y_1 + p[k]\,(x_3 + y_1)},\ p[0] == 0\right\}$

Command **RSolve** can be used to derive a closed form expression for p in terms of the iteration step k and other parameters kept as symbols

```
In[51]:= sol2 = RSolve[eq1, p[k], k] // Flatten // FullSimplify
```

Out[51]= $\left\{p[k] \to -\dfrac{2\,(x_3 - y_1)\left(\left(-1-\frac{x_3}{y_1}\right)^k - \left(-3-\frac{3\,y_1}{x_3}\right)^k\right)}{2\left(-1-\frac{x_3}{y_1}\right)^k (x_3 - y_1) - (x_3 + y_1)\left(-3-\frac{3\,y_1}{x_3}\right)^k}\right\}$

In the specific case, when numeric values of some parameters are known, the closed form expression can be in terms of k, only

```
In[52]:= sol3 = sol2 /. {x₃ → x3, y₁ → y1} // Simplify
```

Out[52]= $\left\{p[k] \to -\dfrac{13\left(\left(\frac{100}{63}\right)^k - \left(\frac{300}{37}\right)^k\right)}{13\left(\frac{100}{63}\right)^k + \left(\frac{12}{37}\right)^k 25^{1+k}}\right\}$

The numeric values of p in each iteration step can be displayed in table format

```
In[53]:= Table[N[p[k] /. sol3, 8], {k, 1, 11}] // TableForm

Out[53]//TableForm=
        0.37956204
        0.49029997
        0.51409288
        0.51883995
        0.51977276
        0.51995551
        0.51999129
        0.51999829
        0.51999967
        0.51999993
        0.51999999
```

Closed form expression can be used to find out the exact value after infinite number of iteration

```
In[54]:= Limit[p[k] /. sol3, k → +Infinity]
         % // N

          13
Out[54]= ──
          25

Out[55]= 0.52
```

Any of the parameters from the general solution can be changed to compute another set of values, such as for any value of n and specified exact value for p (named $\mathbf{p_{True}}$)

```
In[56]:= Clear[n]
                  1
         pTrue = ─ ;
                  2
                1
         p1e = ─ /. p → pTrue;
                4
               1   p
         p2e = ─ + ─ /. p → pTrue;
               4   4
         y1 = Round[p1e n] + Round[p2e n];
         x3 = n - y1;
         Limit[p[k] /. sol2 /. n → 100, k → +Infinity]
         % // N

          13
Out[62]= ──
          25

Out[63]= 0.52
```

Since we already have a limit value as closed form solution in terms of n, the exact values can be computed for a range of different n

```
nrange = Range[50, 150];
pLimit = Limit[p[k] /. sol2 /. n → nrange, k → Infinity];

ListPlot[Transpose[{nrange, pLimit}],
 Filling → Axis, AxesOrigin → {50, 0.45}, AxesLabel → {n, p}]
```

Those results can be illustrated as it is shown in Fig. 1. The probability can be from the range 0.45 to 0.55, depending on n. Regardless of the fact that initial value of p is 1/2, named $\mathbf{p_{True}}$ in the code, the computed value of is from the range 0.45 to 0.55 as exact result of the algorithm after infinite number of iterations. The exact value $p = 0.5$ can be obtained for a large number of objects, or as a mean value for several n.

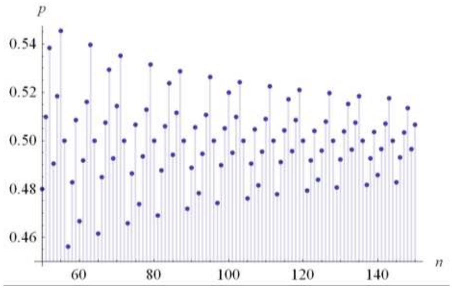


Figure 1
Estimate value of p in terms of the number of observed objects n

6.2 Finding Number of Iterations k

Usually, the errors in iterated algorithms are estimated. Since we already have the exact value after infinite number of iterations, and the current value at any iteration, the error can be expressed in closed form

```
error = (Limit[p[k] /. sol2 /. n → 100, k → +Infinity] -
       p[k] /. sol2 /. n → 100) // Simplify
```

$$\frac{494}{325 + 625 \left(\frac{189}{37}\right)^k}$$

```
a = 1/2^b
```

$$2^{-b}$$

In addition, we specify the acceptable error as $a = 1/2^b$, where b is the number of accurate bits. Solving the equation when error is equal to required number of bits, we can determine b in terms of the iteration index k

```
sol4 = Solve[error == a, b] // Flatten // Simplify
```

$$\left\{b \to -\frac{\text{Log}\left[\frac{494}{325+625\left(\frac{189}{37}\right)^k}\right]}{\text{Log}[2]}\right\}$$

For tree different number of iteration, it follows that 4 iterations are sufficient for 8 bits,

```
b /. sol4 /. k → {3, 6, 10} // N
b /. sol4 /. k → {4, 7, 11} // N
```

```
{7.40333, 14.4561, 23.8672}

{9.7516, 16.8089, 26.22}
```

Figure 2 shows the number of accurate bits is liner function of number of iterations

```
Plot[b /. sol4, {k, 1, 11},
 PlotStyle → Thick, AxesLabel → {k, b},
 GridLines → {{4, 7, 11}, {8, 16, 24}},
 Ticks → {{0, 2, 4, 7, 9, 11}, {0, 8, 16, 24}}]
```

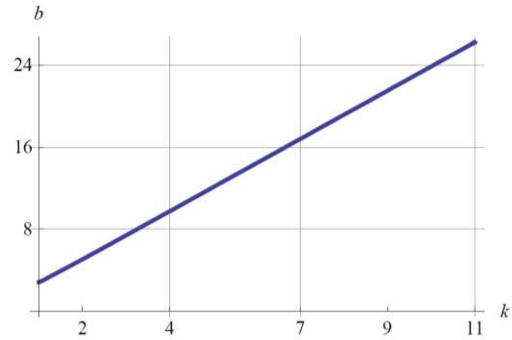


Figure 2

Number of exact bits of estimated *p* vs. number iterations *k*

Instead of number of accurate bits, we can set number of accurate decimal digits

$$a = \frac{1}{10^{digits}}$$

$$10^{-digits}$$

Solving the equation when error is equal to required number of accurate decimal digits, we can determine minimal number of iterations

```
sol5 = Solve[error == a, k] // Flatten // Simplify
```

$$\left\{k \to \text{Log}\left[-\frac{13}{25} + 247 \times 2^{1+digits}\, 5^{-4+digits}\right] \Big/ \text{Log}\left[\frac{189}{37}\right]\right\}$$

The minimal number of iterations is a function of digits

```
k /. sol5 /. digits → Range[0, 15, 3] // N
```

```
{-0.801956, 4.0911, 8.32723, 12.563, 16.7987, 21.0344}
```

For a 6-digit precision, the minimal number of iterations is 9

```
Ceiling[k /. sol5 /. digits → 6 // N]
```

```
9
```

Recalling the table with results, we can see that the 9^{th} iteration has 6-digit precision.

This result can be used for fixed number of iterations instead of while loop. This is more suitable for numeric algorithms, especially when cash instructions can improve the speed of computing (no branch at the end of set of instructions).

In a similar way, by increasing n, the maximal number can be determined for the required precision. For n from 20 to 10000, 9^{th} iteration are sufficient for 6-digit precision.

Conclusions

In this paper, we analyze EM (Expectation-Maximization) algorithm using a new approach and method named symbolic processing. We present a straightforward systematic procedure for automatically derivation of expressions starting from the basic knowledge of the phenomena, automatic derivation of the code, processing with the code and reevaluate the numeric results from published papers. Even more, a closed form solution can be derived, and symbolic optimization may be performed. The presented procedure can be used as template for symbolic processing instead of the classic numeric processing and manual derivations of new advanced challenging algorithms. We introduce an original approach to analyzing a popular and frequently cited tutorial paper on Expectation-Maximization (EM) algorithm, but with a number of typewriting errors. We prove that a more elegant and accurate solution can be obtained by using symbolic processing. Also, we provide better solutions for the exact values of the unknown parameters of probability.

References

[1] J. A. Fessler and A. O. Hero, “Space-Alternating Generalized Expectation Maximization Algorithm”, IEEE Trans. Signal Processing, Vol. 42, pp. 2664-2677, Oct. 1994

[2] M. D. Lutovac and D. V. Tošić, “Symbolic Analysis and Design of Control Systems using Mathematica”, International Journal of Control, Special Issue on Symbolic Computing in Control, Vol. 79, No. 11, 2006, pp. 1368-1381

[3] M. Lutovac, J. Ćertić, Lj. Milić, “Digital Filter Design Using Computer Algebra Systems,” Circuits Syst. Signal Process., Vol. 29, No. 1, 2010, pp. 51-64

[4] M. Lutovac, D. Tošić, “Symbolic Signal Processing and System Analysis”, Facta Universitatis, Electronics and Energetics, Vol. 16, No. 3, 2003, pp. 423-431

[5] T. Moon, “The Expectation-Maximization Algorithm”, IEEE Signal Processing Magazine, 1997, pp. 47-60

[6] M. D. Lutovac, V. M. Mladenović, “Development of Aeronautical Communication System using Computer Algebra Systems”, Proc. XXV Simposium of a New Technologies in Post Office and Telecommunication Traffic - PosTel 2007, Belgrade, December 2007

[7] B. H. Fleury, D. Dahlhaus, R. Heddergott, and M. Tschudin, “Wideband Angle of Arrival Estimation using the SAGE Algorithm”, Proc. IEEE Fourth Int. Symp. Spread Spectrum Techniques and Applications (ISSSTA ’96), Mainz, Germany, Sept. 1996, pp. 79-85

[8] D. Porrat, “The Diffuse Multipath Component and Multipath Component Visibility”, The 5th European Conference on Antennas and Propagation, EUR Congressi in Rome, Italy, EuCAP 2011, 11-15 April 2011

[9] M. D. Lutovac, V. M. Mladenović, “Post-Processing with Advanced Symbolic Simulation and Design using SchematicSolver and Mathematica”, 9th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services- TELSIKS – IEEE societies: MTT, AP, Communications, and Region 8, October 7-9, 2009, Nis, Serbia&Montenegro

[10] S. Wolfram, The Mathematica Book, Cambridge: Cambridge University Press, 2003

Application of RSM Method for the Examination of Diamond Tools

Richárd Horváth[1], Ágota Drégelyi-Kiss[1], Gyula Mátyási[2]

[1] Bánki Donát Faculty of Mechanical and Safety Engineering, Óbuda University, Népszínház u. 8, H-1081 Budapest, Hungary, E-mail: horvath.richard@bgk.uni-obuda.hu, dregelyi.agota@bgk.uni-obuda.hu

[2] Budapest University of Technology and Economics, Egry József u. 1, H-1111 Budapest, Hungary, matyasi@manuf.bme.hu

Abstract: The aluminium alloys are used by the automotive, aerospace industries increasingly because of their numerous advantageous mechanical and chemical properties. The cutting of such aluminium alloys is performed with almost exclusively diamond tools. The most often used work piece materials are the so-called Al-Si and Al-Mg-Si alloys, of which the most widespread are reinforced aluminium-silicon alloys. In the practice of technology planning, the average value of surface roughness and surface height of machined work pieces are as important criteria as geometrical sizes (e.g.: length, diameter tolerance). The authors in this article examine the cutting performance of diamond tools having different edge materials with the help of design of experiment. The correlation between the examined surface roughness and the cutting performance was evaluated; additionally the optimal cutting parameters were determined.

Keywords: fine turning; design of experiment; response surface method; aluminium alloy; surface roughness measurement; optimal cutting parameter

1 Introduction

Surface roughness measurements are essential in the characterization of the features of a machined surface. To examine the effects of cutting parameters on surface roughness thoroughly, a huge number of experiments are needed, depending on the number of input parameters. By utilizing the method of design of experiments (DoE), the number of experiments can be reduced in such a way that the effects of parameters could be assessed appropriately. If the linear effects (correspondences) of cutting parameters are considered, then fractional factorial design is sufficient, but to examine the quadratic terms, RSM method has to be applied [1].

DoEs are often applied in cutting research. Aouchi et al. [2] and Noordin et al. [3] examined hard turning with a CBN and hard metal tool and evaluated surface roughness with the help of DoE. Asiltürk et al. [4] examined stainless steel turning with coated hard metal tools. Dry, wet, and MQL turning was examined with the help of DoE by Young Kug Hwang [5]. Harničárová et al. studied the topography of laser-cut surfaces [6]. Lazarevic et al. examined the surface roughness of engineering polymers by using Taguchi method [7]. In this study, cutting parameters and surface roughness parameters were correlated to determine the relationship between them in case of aluminium fine turning. Our goal was to establish a mathematical model that can be easily used in technology planning to estimate the expected values of surface roughness. In this paper, using a statistical approach, the dependence of the *Ra* and *Rz* parameters on cutting parameters will be analysed, and additionally, the optimal cutting parameters will predicted.

2 Materials and Methods

2.1 Workpiece and Tool Materials

Turning experiments were performed in dry conditions using CNC lathe type NCT EUROTURN 12B, with 7 kW spindle power and 6000 1/min rotation speed. The workpiece material was AS17, frequently used in automotive, aerospace and defence industries. The chemical composition (in wt.%) is: 74.35% Al; 20.03% Si, 4.57% Cu; and 1.06% Fe. The hardness of the workpiece was 114 $HB_{2.5/62.5/30}$. The examined part was a cylinder with a diameter of 110 mm. The experimental runs were made every 10 mm.

The standard designation of the tools selected was DCGW 11T304 FN with ISO geometry (MDC, PCD, CVD-D) and they were manufactured by TiroTool (PCD, CVD-D) and WNT (MDC). The holder of the tool was codified as SDJCR 1616H 11. The average surface roughness values (*Ra*) and roughness heights (*Rz*) were measured by a Mitutoyo SJ-301 surface roughness tester. Parameters related to surface roughness measurements were: l=4 mm, λc=0.8, N=5. The measurements were repeated three times at three reference lines equally positioned at 120° and the results presented were the average of the measured values.

2.2 Design of Experiments

Response surface methodology (RSM) is a procedure which is able to determine a relationship between independent input process parameters (e.g. cutting parameters) and output data (process response, *Y*, in our case *Ra, Rz*). In current study the phenomenological relationship between the input, called the cutting parameters (cutting speed (v_c, m/min), feed rate (*f*, mm), depth of cut (*a*, mm), and the output, machinability parameters (*Y=Ra, Rz*) was given in the form

$$Y = \Omega(v_c, f, a) \tag{1}$$

where Ω is the so-called response function. For the estimation of ouput parameter Y, we used the following equation including the linear and quadratic terms of input parameters, and their interactions:

$$\begin{aligned} Y &= b_0 + b_1 \cdot v_c + b_2 \cdot f + b_3 \cdot a + b_{11} \cdot v_c^2 + b_{22} \cdot f^2 + \\ &+ b_{33} \cdot a^2 + b_{12} \cdot v_c \cdot f + b_{13} \cdot v_c \cdot a + b_{23} \cdot f \cdot a + \varepsilon \end{aligned} \tag{2}$$

In the above formula, b_0, b_i and b_{ij} are the calculated coefficients, v_c, f and a are the input parameters, and ε is the experimental error. In the course of design of experiments the response surface method was chosen, namely a central composite design (CCD). The CCD was performed by taking into account three controllable factors, like cutting speed (v_c), feed rate (f) and depth of cut (a). Each factor had 5 different levels. The number of experimental runs was 16, in which two trials were examined in the centre of the design (15(C) and 16(C) points). The design layout is given in Table 1.

Table 1

Design of experiments with the levels of factors

2^3 central composite design			
Runs	v_c	f	a
1	-1	-1	-1
2	-1	-1	1
3	-1	1	-1
4	-1	1	1
5	1	-1	-1
6	1	-1	1
7	1	1	-1
8	1	1	1
9	-1.28719	0	0
10	1.28719	0	0
11	0	-1.28719	0
12	0	1.28719	0
13	0	0	-1.28719
14	0	0	1.28719
15 (C)	0	0	0
16 (C)	0	0	0

It is advisable to set the studied cutting parameters in order to represent the values used by industries currently, and so that they should meet the requirement of high speed cutting (HSC) applications, as well. Based on these considerations the selected setting values of the cutting parameters are summarized in Table 2.

Table 2
The setting values of the cutting parameters

Factor level	v_c, m/min	f, mm	a, mm
-1.28719	500	0.05	0.2
-1	667	0.058	0.267
0	1250	0.085	0.5
1	1833	0.112	0.733
1.28719	2000	0.12	0.8

3 Results and Discussion

The results of the series of experiments (*Ra* and *Rz* individual values) for different tools are demonstrated in Fig. 1 and Fig. 2. As can be seen, that the scatter *Ra* and *Rz* among work pieces machined by different tools is smaller that the difference between each tools. It is clear that there are such experimental runs (e.g.: 4, 8, 11), when the different surface roughness values of work pieces machined by different tools are clearly separated from each other.

Analysis of variance (ANOVA) was applied to determine the effects of input parameters on the cutting process. First of all the behaviour of individual tools was examined one by one, and the validity of the mathematical model (represented by Eq. 2) was analysed by the response surface method (RSM).

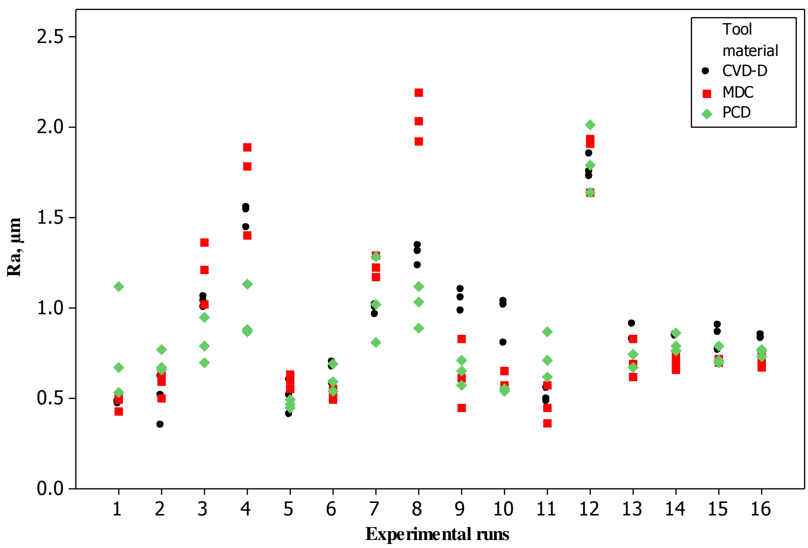


Figure 1
The values of average surface roughness (*Ra*) obtained by the examined tools

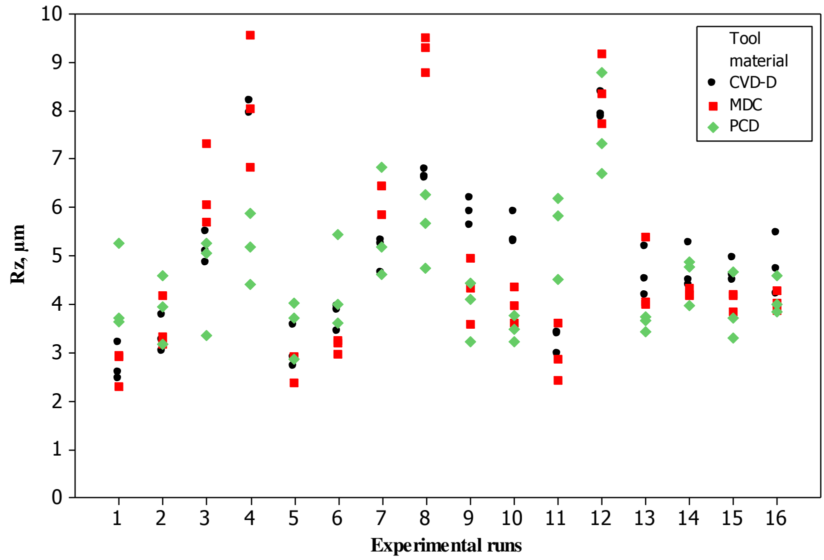


Figure 2

The values of roughness heights (*Rz*) obtained by the examined tools

As a result it was found that several factors and interactions are significant and some parameters could be neglected. The results of significance test are summarized in Table 3.

Table 3

Table of significant coefficients (x - significant, o – no significant)

	v_c	f	a	v_c^2	f^2	a^2	$v_c \cdot f$	$v_c \cdot a$	$f \cdot a$
Ra_{MDC}	x	x	x	o	x	o	o	o	x
Rz_{MDC}	x	x	x	o	x	o	o	o	x
Ra_{PCD}	x	x	x	x	x	o	x	o	o
Rz_{PCD}	x	x	x	x	x	o	o	o	o
Ra_{CVD}	x	x	x	o	x	x	x	o	x
Rz_{CVD}	x	x	x	o	o	x	x	o	x

As can be seen, during the whole examination performed within the specified range of input parameters, the $v_c \cdot a$ interaction is never significant. In order to pool the various influences of non-significant factors, reduced mathematical models were developed.

The reduced models corresponding to the different tools and surface roughness parameters (together with the R^2 values characterizing the goodness of fit) are represented by the following equations:

$$Ra_{MDC} = 2.943 + 6.644 \cdot 10^{-5} \cdot v_c - 66.599 \cdot f - \\ -1.688 \cdot a + 426.86 \cdot f^2 + 25.698 \cdot f \cdot a \quad (3)$$

$R^2 = 92.71\%$

$$Ra_{PCD} = 2.182 + 7.449 \cdot 10^{-4} \cdot v_c - 53.716 \cdot f + \\ +0.113 \cdot a - 4.946 \cdot 10^{-7} \cdot v_c^2 + 327.926 \cdot f^2 + 0.005 \cdot v_c \cdot f \quad (4)$$

$R^2 = 73.7\%$

$$Ra_{CVD} = 0.012 + 2.298 \cdot 10^{-4} \cdot v_c - 5.023 \cdot f + 1.166 \cdot a + \\ +102.206 \cdot f^2 - 1.862 \cdot a^2 - 0.0035 \cdot v_c \cdot f + 12.385 \cdot f \cdot a \quad (5)$$

$R^2 = 89.84\%$

$$Rz_{MDC} = 8.85 + 2.293 \cdot 10^{-5} \cdot v_c - 181.3 \cdot f - \\ -3.74 \cdot a + 1339.18 \cdot f^2 + 69.35 \cdot f \cdot a \quad (6)$$

$R^2 = 91.02\%$

$$Rz_{PCD} = 10.87 + 3.935 \cdot 10^{-3} \cdot v_c - 253.08 \cdot f + \\ +0.88 \cdot a - 1.537 \cdot 10^{-6} \cdot v_c^2 + 1633.02 \cdot f^2 \quad (7)$$

$R^2 = 63.43\%$

$$Rz_{CVD} = -2.105 + 0.0013 \cdot v_c + 49.229 \cdot f + 6.1536 \cdot a - \\ -9.414 \cdot a^2 - 0.0177 \cdot v_c \cdot f + 65.106 \cdot f \cdot a \quad (8)$$

$R^2 = 88.66\%$

With the above equations (from Eq. 3 to Eq. 8) the surface roughness values (*Ra* and *Rz*) obtained by individual tools can be easily predicted from the cutting parameters to be set (within the range of the input parameters examined).

To calculate the surface roughness produced by tools having different edge materials a united (combined) statistical model was developed, by which simpler, well-working equations can be obtained. (See Eq. 9, and Eq. 10). In these equations as a novel variable, the type of edge materials (*TM)* was also inserted. The values *TM* to be substituted are either 1, 2 or 3 depending on the type of edge materials selected, namely: 1 – PCD, 2 – CVD-D and 3 – MDC. As a result of computations, the following formulas are obtained:

$$Ra = 2.739 - 0.476 \cdot TM + 4.32 \cdot 10^{-4} \cdot v_c - 51.363 \cdot f - 1.174 \cdot a - \\ -1.723 \cdot 10^{-7} \cdot v_c^2 + 285.93 \cdot f^2 + 5.05 \cdot TM \cdot f + 0.191 \cdot TM \cdot a + 13.092 \cdot f \cdot a \quad (9)$$

$R^2 = 81.43\%$

$$Rz = 12.622 - 2.241 \cdot TM + 2.838 \cdot 10^{-5} \cdot v_c - 199.566 \cdot f - \\ -2.088 \cdot a + 1032.12 \cdot f^2 + 28.252 \cdot TM \cdot f + 45.391 \cdot f \cdot a \quad (10)$$

$R^2 = 78.37\%$

Equations 9 and 10 are demonstrated in Fig. 4 and Fig. 5. For the 3 different edge materials (denoted by 1,2, and 3), these diagrams show the change of surface roughness (Ra and Rz) as a function of cutting speed (v_c) and feed (f) with constant depth of cut (a = 0,5 mm). From the figures it can be concluded that the surface roughness values obtained by the tool denoted by 1 (PCD) are found in a thinner range (it is less sensitive to changes of cutting parameters), than that of the other two tools examined.

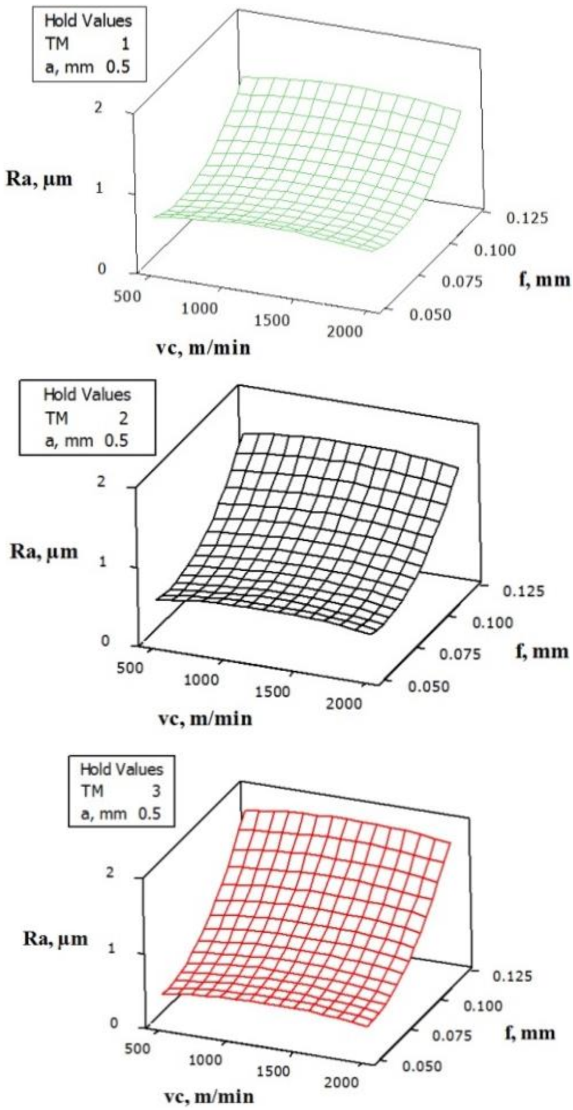


Figure 4

The expected average surface roughness (*Ra*) plotted against cutting speed (v_c) and feed (f) rate

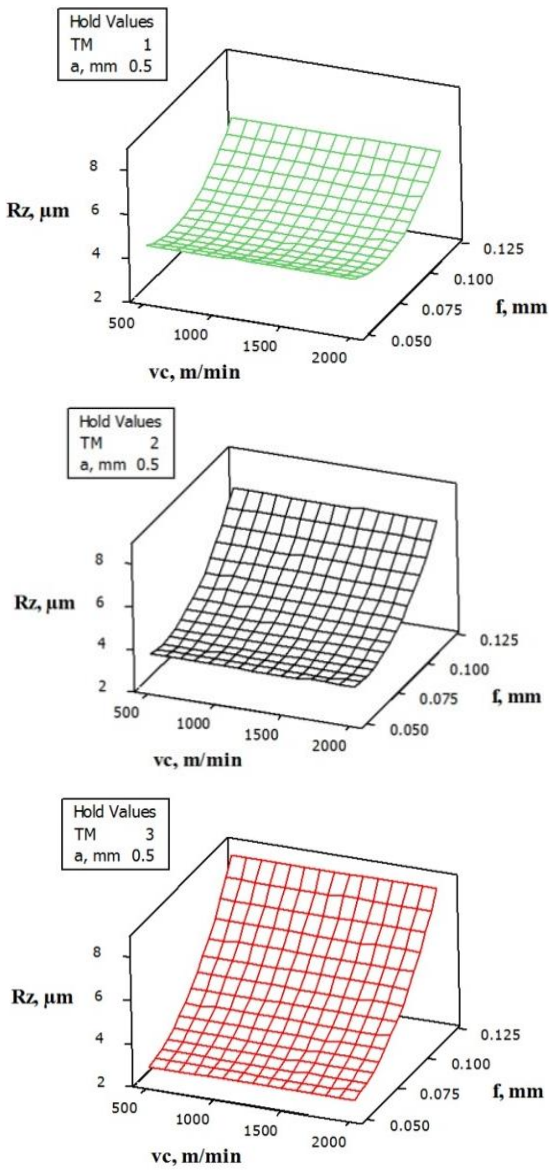


Figure 5
The expected surface height (*Rz*) plotted against cutting speed (v_c) and feed rate (f)

4 Estimating Optimal Cutting Parameters

To calculate optimal cutting parameters it is important to choose an proper target function (objective function). The optimal point (i.e. a maximal or a minimum value) can be obtained at different sets of independent factors, however, the requirements (restrictions) occasionally contradict one another. In this case the aim is to minimize the machined surface roughness parameters (*Ra, Rz*) and to maximize the productivity factor (*Pf*). To fulfil the requirements mentioned the 3 target functions were formulated as follows:

$$Ra \Rightarrow Min \tag{11}$$

$$Rz \Rightarrow Min \tag{12}$$

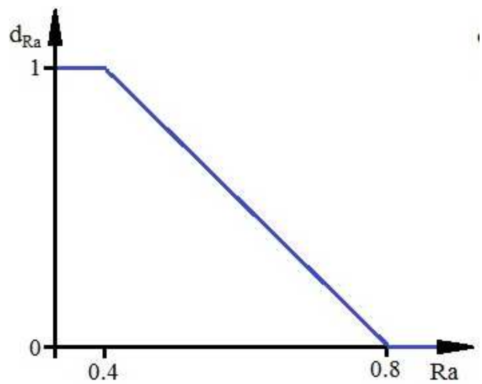
$$Pf = v_c \cdot f \Rightarrow Max \tag{13}$$

In our case we used the desirability function proposed by Harrington [8], for the combination of features expressed in different measures and to be considered with different weights. As it is known, the equations representing the desirability functions take values from the interval (0, 1). In our investigation, the selected desirability functions denoted by d_{Ra}, d_{Rz} and d_{Pf} are shown in Fig. 6. The limits of surface roughness were determined, to be the same as the minimum values (*Ra, Rz*) expected during grinding. In order to fulfil the requirements represented by Eqs. (11, 12 and 13), a composite desirability function was constructed.

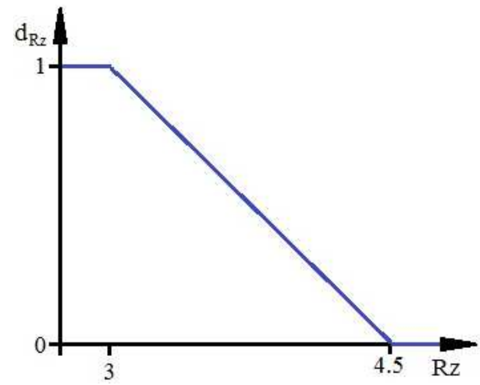
The composite desirability function (*D*) which is designated to look for the optimal point, is obtained by computing the geometrical average of the 3 individual desirability functions:

$$D = \sqrt[3]{d_{Ra} \cdot d_{Rz} \cdot d_{Pf}} \tag{14}$$

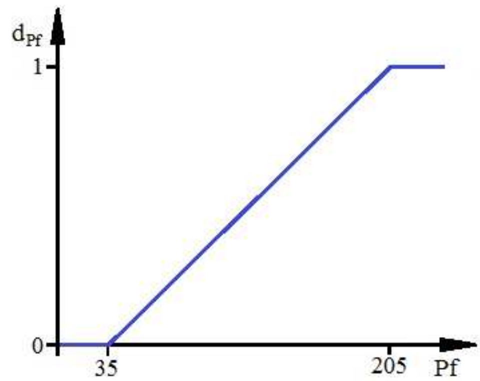
The value of the composite desirability function is 1 (i.e. *D*=1) if the values of selected desirability functions (d_{Ra}, d_{Rz} and d_{Pf}) are 1 as well.



a) desirability function of average surface (*Ra*)



b) desirability function of surface height (*Rz*)



c) desirability function of productivity factor (*Pf*)

Figure 6
Desirability functions

The results obtained by optimization with 0.83527 composite desirability value (*D*) are: TM = 3; v_c = 2000 m/min; f = 0.0677 mm; a = 0.2 mm. In this case the optimum *Ra* value is 0.4001 µm with d_{Ra}=0.99965; the optimum *Rz* value is 2.9957 with d_{Rz}=1.000 and the productivity factor value is 135.3535 which has d_{Pf} = 0.58296.

Conclusions

The effects of cutting parameters (v_c, f, a) were examined in case of fine turning of a cylindrical work pieces of AS17 alloy with a MDC, PCD, CVD-D tools. To calculate the average surface roughness (*Ra*) and the surface height (*Rz*) for different edge materials, statistical methods and empirical equations based on the use of design of experiments were applied. The performed investigation shows which factors (cutting parameters) have significant influences on average surface roughness (*Ra*) and surface height (*Rz*). Based on experimental data, combined models were established by which the surface roughness obtained with different edge materials can be efficiently predicted (the edge materials as a changing input factor were included in the model equations). By using an appropriately defined combined objective function, this phenomenological model enables to estimate the optimal cutting parameters.

Acknowledgement

The project was realised through the assistance of the European Union, with the co-financing of the European Social Fund, TÁMOP-4.2.1.B-11/2/KMR-2011-0001 Research on Critical Infrastructure Protection.

References

[1] Á. Drégelyi-Kiss, R. Horváth, B. Mikó.: Design of Experiments (DOE) in Investigation of Cutting Technologies, *Development in Machining Technology/Scientific-Research Reports* 3 (2013) (book chapter, in press)

[2] H. Aouici, M. A. Yallese, K. Chaoui, T. Mabrouki, J. F. Rigal: Analysis of Surface Roughness and Cutting Force Components in Hard Turning with CBN Tool, Prediction Model and Cutting Conditions Optimization, *Measurement* 45 (2012) 344-353

[3] M. Y. Noordin, D. Kurniawan, Y. C. Tang, K. Muniswaran: Feasibility of Mild Hard Turning of Stainless Steel Using Coated Carbide Tool, *The International Journal of Advanced Manufacturing Technology* 60 (2012) 853-863

[4] I. Asiltürk, S. Neseli: Multi Response Optimisation of CNC Turning Parameters via Taguchi Method-based Response Surface Analysis, *Measurement* 45 (2012) 785-794

[5] Y. K. Hwang, C. M. Lee: Surface Roughness and Cutting Force Prediction in MQL and Wet Turning Process of AISI 1045 Using Design of Experiments, *Journal of Mechanical Science and Technology* 24 (2010) 1669-1677

[6] M. Harničárová, J. Valíček, M. Kušnerová, R. Grznárik, J. Petrů, L. Čepová: A New Method for the Prediction of Laser Cut Surface Topography, *Measurement Science Review* 12 (2012) 195-204

[7] D. Lazarević, M. Madić, P. Janković, A. Lazarević: Surface Roughness Minimization of Polyamide Pa-6 Turning by Taguchi Method, *Journal of Production Engineering* 15 (2011) 29-32

[8] E. C. jr. Harrington: The Desirability Function, *Industrial Quality Control* 21 (1965) 494-498

Utilizing Experiments Designed Results during Error Identification and Improvement of Business Processes

Zuzana Hajdu [(1)], Marek Andrejkovič [(2)], Ladislav Mura [(3)]

[(1),(2)] Faculty of Business Economics of the University of Economics in Bratislava with seat in Košice, Department of Mathematics and Informatics, Tajovského 13, 041 30 Košice, Slovakia; e-mail: zuzana.hajduova@euke.sk

[(3)] Faculty of Social Sciences, University of Ss. Cyril and Methodius in Trnava, Nám. J. Herdu 2, 917 01 Trnava, Slovakia; e-mail: ladislav.mura@ucm.sk

Abstract: The improvement of business processes is a necessary part of innovations in business, aimed at improving of customer satisfaction and achieving more reliable productions. The above mentione of steps are inevitable in today's world of high competitiveness and therefore the need for improvement of competitive abilities of an enterprise and its products grows steadily. Excessive error rate creates additional costs that are reflected in the price of the product. This is the reason for less effective competitive ability of the given product. Therefore, this contribution shows designed experiments as a method of decreasing error rate in production processes.

Keywords: Quality; business processes; DOE; process improvementt

1 Introduction

Design of experiments represents a very important step towards the improvement of processes and towards the optimization of production. Optimization in this sense must be understood as a whole, not only as a minimization of errors. In this case, minimum of errors may also be achieved by not producing at all, which, of course, is not the aim of any enterprise. For this reason, this contribution points out the optimization of selected business processes with the aim of improvement of quality within the given enterprise in the sense of improvement of production processes. The statistical software SAS version 9.2 and SAS Enterprise Guide 4.2 will be used for supporting statistical calculations. The aim of this contribution is to improve the existing processes in the selected enterprise by means of optimization methods. With this aim, various methods wil be used, such as observation, analysis, synthesis, methods of inductive statistics, optimization methods, mathematical modelling and many more.

2 Design of Experiments

Design of experiments (DOE) is in literary works also called statistically designed experiments (*statistically designed experiments*). (El-Haik - Yang, 2003)

Experiment is a process that is prepared and organized in order to better understand the measured subject. Experiment consists of a set of trials which are diferentiated from each other by individual settings of factor levels. A factor is any definite cause that may influence the response. Factors are further divided according to various points of view. The experiment and the subsequent analysis of data obtained aims at finding the causal-consequential relationships between inputs and experimented factors in the process. (Anderson – Lorenzen, 1993)

Researches of many authors have come to a conclusion that DOE is currently not utilized enough (Gremyr et al., 2003 and Bergquist and Albing, 2006). Tanco et al. (2007) has found out that as many as 94% branches of industry use experiments, however, only in connection with the OFAT (one factor at a time) method.

While carrying out experiments our aim is to establish a mathematical functional relation:

$$y = f(x_1, x_2, ..., x_n) + \varepsilon \qquad (1),$$

where ε is the error of the experiment (experiment deviation), y means the response and the $x_1, x_2, ..., x_n$ mean factors which should be influenced. This experiment deviation expresses the difference between expected and real values of the response, which means that there is probably no deterministic functional relationship between y and $(x_1, x_2, ..., x_n)$. This can be caused by: (Bailey, 2008)

1. Uncontrollable factors $(z_1, z_2, ..., z_p)$ influencing the response of y which are not contained in the equation (1).
2. The fact that both experiment deviations and measurement deviations are mutually related in a functional relationship (y and $x_1, x_2, ..., x_n$).

Design of experiments represents a reliable tool. We use it for the observation and study of factors and/or parameters influencing the process in question. Such conduct enables to divide factors into significant and less significant with regard to the conditions of the environment. It also helps to gain information on mutual influence of individual factors $(x_1, ..., x_n)$. (Anderson – Whitcomb, 2007)

The aim of DOE is: (Phadke, 1989)

- Submit the necessary information with minimum effort and minimum costs,
- Specify whether it is possible to answer the questions by means of the experiment or not,

- Specify whether a series of experiments or rather a single experiment is desirous,
- Clearly define individual stages of the experiment,
- Utilize previous knowledge and experience to state hypotheses especially when specifying input factors and their levels.

The theory mentions theoretical models, such as ”hard“ models. These are usually based upon the specified relations stated by theory. They must be based upon the idea of natural sciences and their principles. On the other hand, there are also empirical models, the so called ”soft“ models. In these we describe relationships between factors and their response on a local interval. (Montgomery, 2005)

The aim of this stage of DOE is, above all, to identify all relevant factors that will further appear in the analysis and also to describe their mutual interaction. The identification of these factors may be performed by means of brainstorming, especially by directed brainstorming. There is also the need to specify the levels of individual factors which will be later used in the experiments. In general, two levels of each factor are sufficient – a minimum and a maximum one. (Pyzdek – Keller, 2009)

Design of an experiment must also include any and all restrictions (drawbacks) that may occur during the course of individual stages of the experiment: (Phadke, 1989)

- *economic restrictions* – if we observe a great number of factors
- *time restrictions* – if there must be multiple experiments due to a great number of factors
- *process restrictions* – following from the fundamental principle of the process, where it is impossible to set individual levels of factors
- *factor restrictions* – in case no level can be specified at all
- *diversity of results* – if we get different result while having the same setting of factor levels – due to this, individual stages of the experiment must be repeated

The importance of designing experiments lies also in the fact that only necessary experiments are performed. Also, there is a need to state the aim of such experiments. Such formulation clearly shows how to achieve improvement.

3 Methodology of Research

Improvement of quality and processes in businesses is the fact that helps businesses become more competitive. The customer pressure to have get better quality as well as internal pressure to decrease the number of defect products and the effort to produce less waste creates predispositions for the implementation of mathematical and statistical methods, the aim of which is to optimize the above production processes. Such methods may result in lower enterpreneurial costs either directly, in the form of lower production costs, or indirectly, by lower costs for repairs and/or customer claims. This method may also be used to increase profits.

This contribution shows how to improve existing processes in a selected enterprise by means of optimization methods. This main aim is fulfilled through statistical and mathematical methods as well as through managerial methods.

The process of production has been identified in the enterprise. We analyzed the production process by means of observation and analysis of individual stages and sub-processes. We created process maps. The identification of these subprocesses was vital in order to identify further possibilities of quality improvement. Wrong understanding of production processes may cause the failure of optimization. After the analysis of processes we identified critical points, in our case we measured the number of faults and analysed their occurence in processes. By means of a Pareto diagram we identified the most common failures and concenrated our attention upon them.

During the effort to improve a selected failure we used the Ishikawa diagram in which we identified possible causes of this fault that may have happened in the given production process. We need to emphasize that we fully cooperated with the employees of the enterprise who provided many opinions and tips as to the possible origination of this key fault. Subsequently, after the identification of causes we analysed the production process again, this time with a special emphasis on the observed defect.

During this minute observation of the production process we discovered two key factors that may have influence the origination of the above faults. These factors were subsequently used to design an experiment further conducted in this enterprise. This designed experiment was conducted in the form of a 2^k full factor experiment. I.e., each factor was observed on two levels. These levels were specified according to recommendations of the Chief Production Manager and our observation of production processes.

After having conducted the experiment and measuring the response, we started to analyse results and tried to find the optimal setting of the factors in order to minimize the number of defects. Apart from this, we also focused our attention to finding other possible improvements that could optimize the overall production process in the given business company while taking into account real procurement costs of such investments.

During this solution we used methods such as observation, analysis, synthesis, deduction and induction, and also mathematical and statistical optimization methods, method of designed experiments, regress and correlation analysis and testing of statistical hypotheses.

The improvement of processes has taken part in an enterprise in Hungary in a relative vicinity of Budapest. In the year 2012 we were addressed by the representatives of this enterprise and asked for help with the improvement of their production processes.

4 Production Process and Experiments

This part will show the results we analysed and measured in the company. Also analyses that we performed.

4.1 Description of Production Process

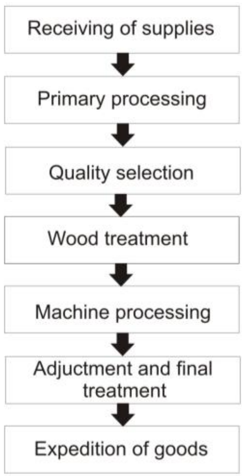


Figure 1
Production process

Production process of wood parquet consists of several stages that will be described in this part. The production process consists of seven main production stages, each of these has also its internal stages. We will describe them one by one.

The first stage starts by receiving supplies to the factory dealing with the production of wood parquet. This stage consists of the delivery and unloading of wood material from trucks.

Second stage consists of processing and cutting of raw wood and its preparation for further processing within the production process.

Stage three represents the selection of suitable wood material that has been cut for further processing according to its quality. Three groups are formed in this stage, namely, first quality, second quality and waste wood that fails to meet the quality requirements.

Subsequently stage four consists of the treatment of wood. In this stage the wood is dried and further treated which is followed by the machine processing consisting of final cutting to required width and length and planing.

After machine processing, there is again the selection of quality products. Afterwards, groove and tongue are cut in the wood parquet and the surface is ground. After such processing and subsequent quality control there are final treatments that shall provide for the removal of possible faults.

In the end, the wood parquet is packed and placed onto euro-pallets in larger bulks containing appr. 130 m^2 and weighing appr. 800-1100 kilograms. Then these pallets are placed into the store and further dispatched upon individual orders.

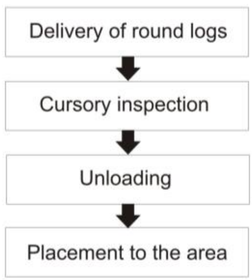


Figure 2
First sub-process – Receiving of supplies

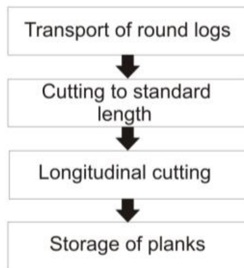


Figure 3
Second sub-process – Primary processing

The first stage consists of the delivery of round logs. These are usually delivered by trucks. However, there is a railroad next to the factory, although not leading directly to the factory. The logs may be delivered also by means of rail transport. Round logs are delivered from various areas, mostly from Slovakia. After the delivery, the logs are visually checked by the employee in charge. He confirms the receipt of the goods and subsequently orders to unload the trucks.

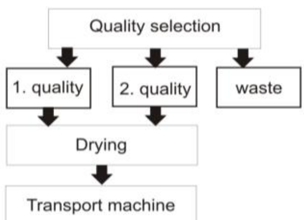


Figure 4
Third sub-process – Quality selection

After having received the round logs in their storage space on the premises of the factory, these logs are further processed and cut to standard length of 4 meters. Such 4-meter round logs are further transported to longitudinal cutting during which the logs are cut into wood planks 4 meters long. The thickness of such planks is appr. 20 mm. Afterwards, the planks are stored in the stock.

After having cut the wood into planks, these are subdivided according to their quality into two qualitative levels, namely:

- 1. Quality class,
- 2. Quality class,
- waste.

This sorting is performed according to the bark on these planks. In case the bark is damaged, this may create predisposition for bad internal conditions of the wood, e.g. the existence of bark beetle or other worms, and/or dry-rot inside wooden planks.

Subsequently, the wood is dried to achieve the humidity of 8 – 12 %. Such drying takes approximately 48 hours, i.e. 2 days. After two days, wood planks are further transported for processing by means of carts pushed by the employees.

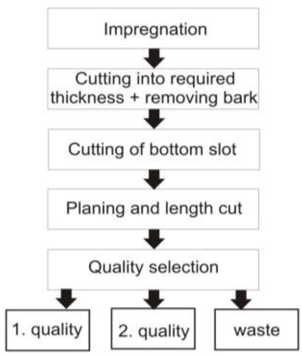


Figure 5
Fourth sub-process – Wood treatment

In this stage, only products that have been marked as either first or second quality follow to another stages. The dried wood is impregnated against possible future woodworm exposure. After impregnation, the planks are cut into required width and the bark is removed from the edges of the planks. After this step, the bottom slot may be cut out. In the end, the wood is planed so that the surface is relatively smooth without coarse structures. After this treatment the planks are cut into final length. Subsequently, they are again assessed as to the quality and they are subdivided into three groups as listed in the previous stage.

After the adjustment of planks into their final length and width in the previous stage, the planks undergo final treatment consisting of manual sealing. Then, a machine cuts out the groove and tongue at the edges of the parquet. After a succesful cutting, the surface is machine ground so that ists surface is smooth. After this treatment, the products are assessed according to their qaulity. First quality products are faultless. Second quality represents products with slight faults that may be repaired. Recognized faults are manually sealed and subsequently, the plank surface is machine ground again.

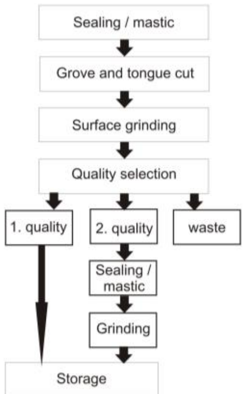


Figure 6
Fifth sub-process – Machine processing

After final treatment, the parquet is transported into the stock. Here they are packed into boxes, in one box there are five layers of parquet placed on top of each other. The pallet is wrapped and tied so that the parquet is not harmed during transport. After having tied and wrapped the pallet, this is transported to the stock of final products and prepared for dispatching for a customer. This factory only delivers parquet in large bulks and works upon individual orders. Therefore, it is not necessary to prepare small consumer packages and the storage through palletts is suitable and sufficient for the needs of the customers.

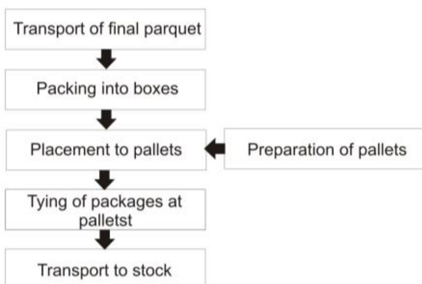


Figure 7
Sixth sub-process – Adjuctment and final treatment

During the production process, waste material occurs. Tihis waste is not a standard communal waste, but these are products of insufficient quality. Such faulty products are spearated in the production process.

Technically, the origination of faulty products is as follows:

1. primary processing, in which they refuse unsuitable pieces of wooden logs and cut thick planks that are unsuitable for further processing,
2. after machine processing, where some faults can be discovered after planning and cutting of the wood into required length and thickness due to the following reasons:
    a. nodes,
    b. colour,
    c. dry rot,
    d. wrong length,
    e. other,
3. last stage, in which faulty products may occur is before packaging, when the parquet may be damaged during the cutting out of groove/tongue or during grinding. Several of these faults are impossible to be repaired.

The assessment of ratio of depreciation and the fact whether the product may be repaired into a suitable shape or no is done by a technician who assesses every item indivually. This may cause delay in many cases. Any and all products and parts marked as unsuitable are held on a separate place. Waste consists also of sawdust and wood material that was left after having processed the parquet, namely cuttings. These are processed in the following way.

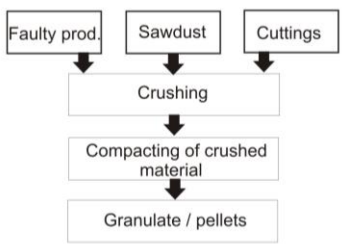


Figure 8

Waste processing in company

Subsequently, these wood remains are processed into crushed material that is compacted by means of a high-pressure compact press. The result is in the form of granulate or pellets.

4.2 Analysis of Errors in the Process

During production process we identified the origination of faults. With regard to the long-term production plan in this enterprise, we identified an individual set of errors. In the communication with the factory workers we found out that our assesment of errors and measured number of faults is really close to a real long-term condition in the enterprise.

We found out that the following errors are most likely to occur:

- Black longitudinal lines,
- Black vertical lines,
- Inaccurately cut grooves,
- Jagged edges,
- Faults in wood material.

Most frequently, the fault of longitudinal black marks occurs during production. In most of the cases, these marks cause the product to be marked as faulty because it is impossible to remove them from the surface. Therefore, we decided to find reasons for these black marks and created an Ishikawa diagram showing these reasons.

During research of individual possibilities the most probable cause of such marks was identified as black supporting wheels that prevent the parquet from elevation during machine processing. The parquet moves on a conveyor belt between individual machines during the processing. Due to this reason, we decided to further study only these objects and tried to remove this fault, or, at least, minimize its occurence.

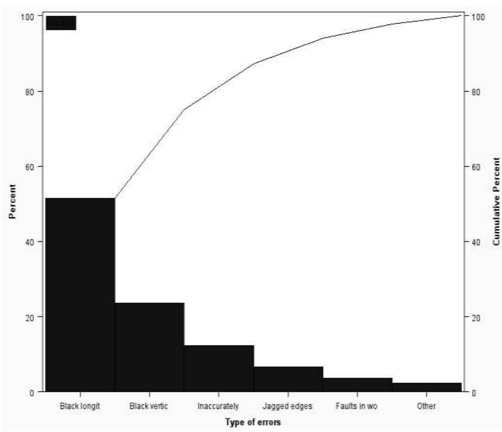


Figure 9

Pareto diagram of errors in company

4.3 Design of Experiment

With this purpose, we concentrated or attention to the problem of supporting wheels the surface of which is made of rubber. During the assessment production we found out that these wheels get stuck sometimes and subsequently the surface of the produced parquet is damaged.

With regard to the above, we decided to explore the influence of two variables:

- Velocity of conveyor belt and, therefore, of parquet,
- Pressure of wheels to the conveyor belt, therefore, to the parquet itself.

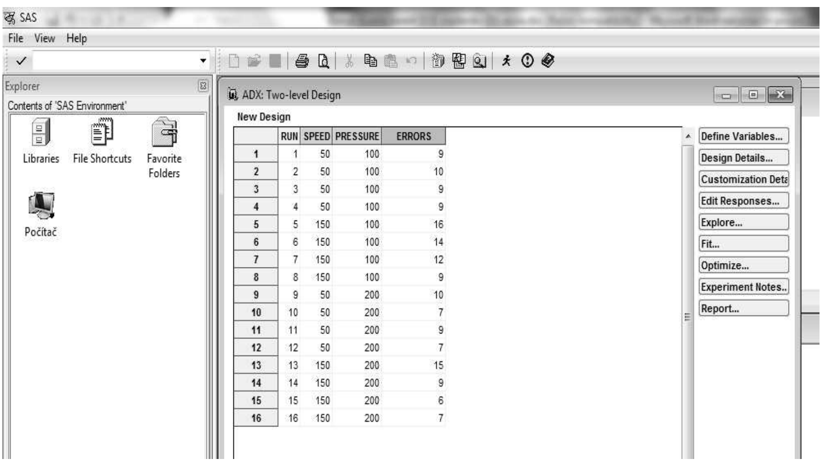
That is why we defined an experiment in which we shall take into account the above parameters. We decided to perform a full factor experiment. We set two levels of each factor as follows:

Chart 1
Levels of factors used in experiment

Factor	Minimum level	Maximum level
Pressure	100	200
Speed (Velocity of movement)	50	150

Subsequently, we performed measurements for all combinations, whereas we decided to perform four replications of the given experiment. This means that we measured individual settings one by one, each of the settings was measured four times.

Input values can be seen in the following figure from software SAS 9.2 that was used for planning and calcualtion of individual values of the experiment.



	RUN	SPEED	PRESSURE	ERRORS
1	1	50	100	9
2	2	50	100	10
3	3	50	100	9
4	4	50	100	9
5	5	150	100	16
6	6	150	100	14
7	7	150	100	12
8	8	150	100	9
9	9	50	200	10
10	10	50	200	7
11	11	50	200	9
12	12	50	200	7
13	13	150	200	15
14	14	150	200	9
15	15	150	200	6
16	16	150	200	7

Figure 10
Input measurements of Factors and Response

5 Results and Discussions

5.1 Discussion

Now we may show the output that points out the number of errors we measured. The graph describes the number of faults discovered on produced parquet in the same part of the production batch. We always changed the settings after having produced the same production batch which was, in our case, 200 pieces of parquet produced continually.

It is clear that the largest number of errors was detected in Batch 5 where we obtained as many as 16 faulty pieces of parquet.

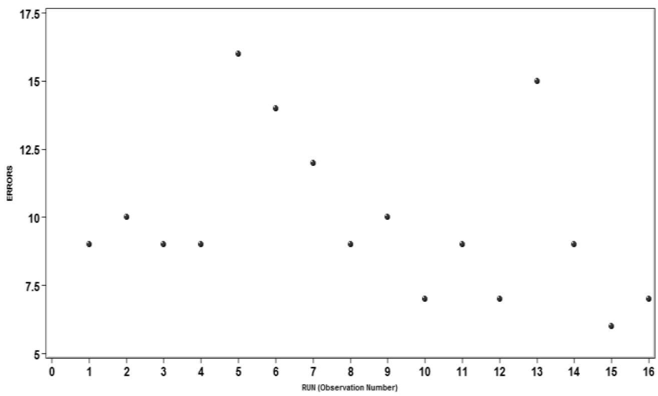


Figure 11

Chronological measurements of Result (Errors)

Upon these measurements we calculated the influence of individual factors. We can see that the higher velocity of movement results in higher occurence of faults. On the other hand, the analysis of pressure showed opposite results. The higher pressure was applied to the cylinder (wheel), the lower number of faults occured.

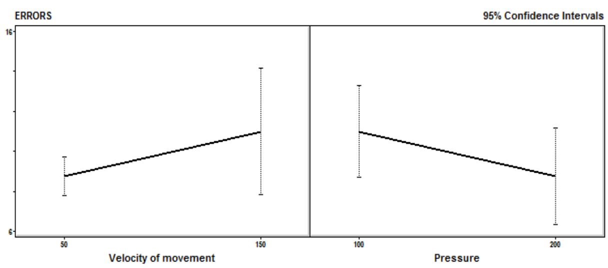


Figure 12

Graph of influence of Level of factor on Response

Apart from this, we also decided to explore mutual interactions between these two variables. We found out that the interaction between these variables is not significant because the curves do not intersect, what would have been a sign of important interaction. From this reason it is not necessary to explore the development of interaction because it means that the influence of both factors at the same time is not significantly higher than the sum of their influences analysed individually.

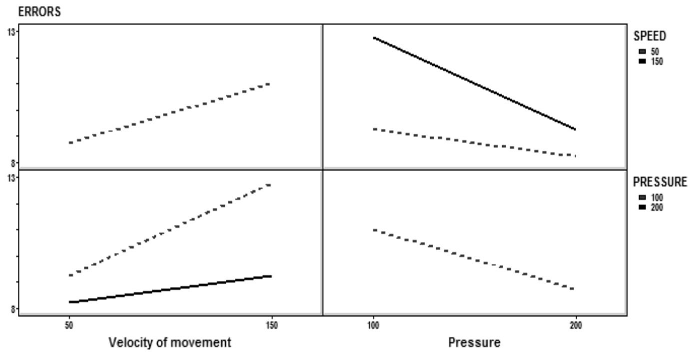


Figure 13

Interaction of factors measured on Response

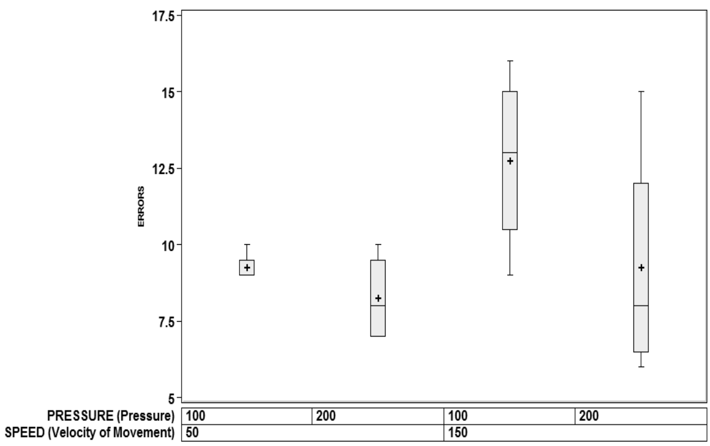


Figure 14

Divisions of individual subfiles based on levels of factors

Subsequently, all researched groups were shown in box plots that represent results of measurements. However, we need to emphasize that these box plots consist of only four data and therefore cannot be understood in an exact way. What occurs as a very important phenomenon is the fact that in case of low speed the number of faults is smaller than in case of higher speed.

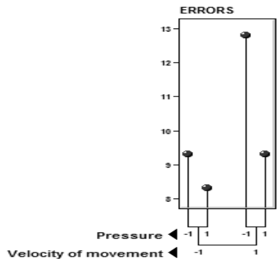


Figure 15
Group averages

The above fact may be described by means of a point graph that shows average values of individual groups.

Apart from this, we also calculated 95% Prediction intervals for the curves of influence of individual factors to the number of faults.

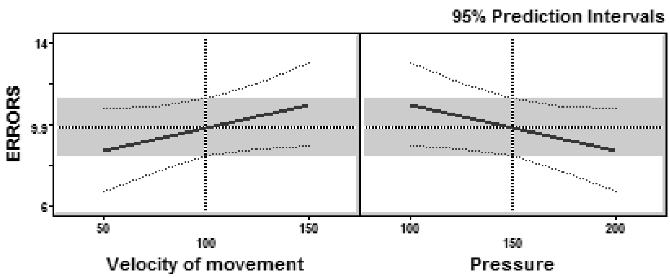


Figure 16
Reliability intervals of individual groups

Upon the above measurement we further defined a predictive model that serves the purpose of calculating the expected number of faults at a given speed and presuure.

Predictive Model for FAULTS				
ERRORS	=	9.875 + 1.125***SPEED** - 1.125***PRESSURE** - 0.625***SPEED*PRESSURE**		
ERRORS	=	7.25 + 0.06* **SPEED** + 0.0025***PRESSURE** - 0.00025***SPEED*PRESSURE**		

Figure 17
Predictive model for the number of faults

The above conclusions may be represented by means of a so-called „contour graph“ that shows the number of faults through increase of surface in a two-dimensional area with the help of the so-called „contour lines“. This graph shows that the smallest number of faults occurs at a higher pressure and slower speed of movement.

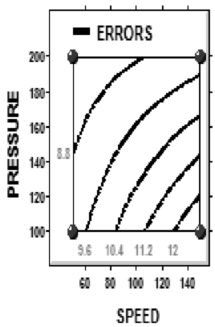


Figure 18
Contour plot for the number of faults

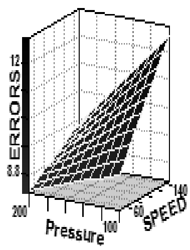


Figure 19
3D surface graph of Response

The above conclusion may also be shown by means of a 3D graph in which we can see that the lowest number of expected faults is likely to be expected at higher pressure and lower speed of the conveyor, i.e. parquet on the production line.

Upon these analyses and results we may subsequently define recommendations that form the content of the next chapter.

5.2 Recommendations

Upon the above analyses we can define the following recommendations for the enterprise. We found out that the most optimum setting of the given factors is as follows:

Prediction Profile Settings					
SPEED	Velocity of movement	50	**ERRORS**		8.25 [6.817455,9.682545]
PRESSURE	Pressure	200	**Desirability**		48.44%

Figure 20
Optimum factor setting

The factory should therefore set the speed of the conveyor, and, thus the movement of the parquet to level 500 and at the same time, increase the pressure of the cylinder to the level of 200 Newton, which is appr. 20 kg/force to a given parquet. As the load-bearing capacity of individual parquet is much higher, these are safe from harm.

Apart from this, there is the need to regularly perform lubrication of bearings of these wheels and cylinders that help to press the parquets to the conveyor during their movement on a production line. It is important to ensure that the excessive oil or vaseline does not spray from these bearings during the movement of wheels or cylinders. For this reason, there is a need to use suitable lubrication oils that are solid enough but have sufficient characteristics. Another problem that requires constant attention is the correct shape of wheels and cylinders. In case they are repeatedly stuck, they are often misshapen and such small damages may cause the wheel and/or cylinder to get stuck again and the damage may become greater. For this reason, there is a need to regularly replace these wheels and cylinders and check them in order to prevent such stucking and damaging of the material.

Apart from this, there are some other improvements that may be implemented in this enterprise and that we encountered during our work in the factory through solving the previous task.

We found out that the factory uses a lot of pushcarts for the manipulation with the material. It would be useful to think of their replacement by conveyor belts as these would transport the parquet more carefully, because many times there was the damage to the material due to bad handling and/or inattentiveness. The transfer takes place between three production halls which are interconnected by means of roofed corridors. These corridors may be used for the construction of conveyor belts that would automatically transfer the semiproducts between individual construction stages.

Apart from this, another problem is the failure to perform the steeping of the wooden parquet. These wooden parquet are used in the interior. Steeping currently takes place only at the application in the interior and three layers of steeping must be performed. Each new layer may be applied only after complete drying of the previous one. Therefore, the primary steeping at the factory would save a lot of time. Moreover, such steeping may be performed automatically and as such saves costs also in the form of a labour force.

Conclusions

We applied designed experiments in a concrete enterprise that deals with the production of wood parquet. By means of this method we achieved the improvement of production through less errors, i.e. less faulty products causing additional costs for the factory. The use of planned experiments is very effective in company activities. Such experiments represent the way of improving business processes at relatively low costs, and describe the method of effectively spent improvement costs with the aim of achieving aimed improvement of business proccesses. Upon the above we came to conclusions described in the last paragraph of this contribution. Listed improvements helped the factory to gain a better place on the market due to lower costs caused by faulty products and subsequent possibility to offer lower sale price to the customers. In such way, we achieved a higher competitiveness of products of this company.

Acknowledgement

This paper was supported by the research project VEGA No. 1/0822/13 „Energy efficiency and economic support of regional energy policies”.

References

[1] Anderson, M. J. – Whitcomb, P. J. 2007. DOE Simplified: Practical Tools for Effective Experimentation. New York : Productivity Press, 2007. ISBN 978-1563273445

[2] Anderson, V. – Lorenzen, T. 1993. Design of Experiments. New Jersey : CRC Press, 1993. ISBN 978-0824790776

[3] Bailey, R. A. 2008. Design of Comparative Experiments. Cambridge : Cambridge University Press, 2008. ISBN 978-0521865067

[4] BERGQUIST, B. – ALBING, M. 2006. Statistical methods—Does anyone really use them? In: Total Quality Management & Business Excellence, 17(8):961–972

[5] Box, G. E. P. - Hunter, J. S. W. - Hunter, G. 2005. Statistics for Experimenters—Design, Innovation and Discovery. 2nd ed. New Jersey: John Wiley & Sons

[6] El-Haik, B. – Yang., K. 2003. Design for Six Sigma: A Roadmap For Product Development. New York : McGraw – Hill, 2003. ISBN 0071412085

[7] Gremyr, I. – Arvidsson, M. - Johansson, P. 2003. Robust design methodology: A status in the Swedish manufacturing industry. In: Quality and Reliability Engineering International. 19:285–293

[8] JMP. 2005. Design of Experiments. Cary : SAS Institute, 2005. ISBN 1-59047-816-9

[9] Lewis, J. P. 2007. Fundamentals of Project Management. 3rd ed. AMACOM, United States

[10] Lye, L. M. 2005. Tools and toys for teaching design of experiments methodology. In: 33rd Annual General Conference of the Canadian Society for Civil Engineering, 2–4 June 2005, Toronto, Ontario, Canada

[11] Montgomery, D. C. 2005. Design and Analysis of Experiments. New Jersey: John Wiley & Sons

[12] Phadke, M. 1989. Quality Engineering Using Robust Design. New Jersey : Prentice Hall, Upper. 1989

[13] Prokeinová, R. 2009. The impact of ecological footprint in enterprise decision-making. In: Scientia Iuventa 2009: proceedings of reviewed papers from international scientific conference of PhD. Banská Bystrica, 2009

[14] Pyzdek, T. – Keller, P. 2009. The Six Sigma Handbook. New York : McGraw Hill Professional, 2009. ISBN 978-007162338-4

[15] SAS. 2005. Concepts of Experimental Design. Cary : SAS Institute, 2005

[16] Stehlíková, B. Statistical Analysis with SAS System. Nitra: SUA, 2003, 127p.

[17] Tanco, M. et al. 2007. Is Design of Experiments really used? A survey of Basque Industries. In: Journal of Engineering Design. 19(5):447–460

[18] Viles, E. et al. 2009. Planning Experiments, the First Real Task in Reaching a Goal. In: Quality Engineering, 21: pp. 44–51, 2009 ISSN: 0898-2112

Strategy Support of Residential Energy Saving Investments in Hungary with the Method of Technology Roadmapping

Szilvia Bíró-Szigeti

Department of Management and Corporate Economics, Budapest University of Technology and Economics, Magyar tudósok krt. 2, Building Q, A-309, H-1117 Budapest, Hungary, e-mail: szigetisz@mvt.bme.hu

Abstract: This paper discusses residential energy saving market in Hungary, applying a relatively new method. The study is a comprehensive presentation of the results of an extensive research, exploration of relationship systems and system-based processing, in the framework of which the design, the implementation and the operation of the energy investments of buildings takes place. The market is examined with strategic approach and the reseach findings are systematized in intuitive ways with the method of technology roadmapping (TRM). These aspects of analysis give new and original contribution to the Hungarian building energetics sector and energy policy. The results are obtained from interviews and questionnaire survey with representatives of stakeholder groups and experts. The establishment of residential energy saving investments is highly influenced by environmental factors, technological development of building energy market and stakeholders interests and the interrelation of all these. The problems originate mostly in the long-standing lack of the comprehensive strategic plans and marketing tools of building energetics enterprises and concrete objectives and state's guidelines for the residential market of building energetics.The aim of the study is to serve as a practical instrument for the decision-makers, it can promote more efficient planning, support and implementation of residential energy saving investments.

Keywords: energy saving; strategy; technology roadmapping; TRM

1 Introduction

Residential energy saving approach has become significant in recent decades. As much as 28% of household energy saving potential [8] can be identified in new or replacement building energy investments in the future. Complex macro- and market level conditions and sustainability requirements effect the realisation of the energy saving investments (both on the demand and supply side) which requires certain marketing features from the enterprises. Building energetics micro and small enterprises[1] have to operate continuously under difficult circumstances due to uncertainty in energy prices, in subsidy systems, in economic condition and social attitude. Competition is increasing, which requires strategic thinking, marketing approach and the use of marketing tools in the corporate functions.

The main objectives of this research are to reveal the building energetics market and the stakeholders of the micro and small enterprises that provide energy saving products and services with a strategic analytical and marketing approach; to systematically organise the available information with the help of TRM methods; and to draw conclusions regarding practical feasibility. These aspects of analysis of the building energetics sector can be considered new and original in Hungary. The necessary information sources were collected and used after the recommendation of the TRM method. The frame of the strategic technology roadmap of the Hungarian residential energy saving is new, the paper can support the development process of this method [43].

In the course of the feasibility studies of Hungarian residential energy saving investments[2] – when analysing the conditions and the market of energy saving solutions – a knowledge gap occurs between the necessary and available basic information. Studying the traditionally examined factors is vitally important but not enough in itself. There is a lack of indispensable information that should be examined and used but is disregarded by Hungarian research and literature. The state and market level energy policy has not been worked out with processing all the necessary information in a systemic way, using management methods of high standard like TRM so far. In this paper I would like to contribute to the filling of this methodological gap.

At first the TRM method will be presented here (in Chapter 2) briefly, and after that a method will be adapted (in Chapter 3) on the basis of the information available from different sources on the Hungarian residential energy saving investments. Alphabetical symbols and figures help the reader to make the examined background consistent with the related results.

---

1 *Micro-sized enterprise:* it is defined as an enterprise which employs fewer than 10 persons and whose annual turnover and/or annual balance sheet total does not exceed EUR 2 million (~HUF 550 million). *Small-sized enterprise:* it is defined as an enterprise which employs fewer than 50 persons and whose annual turnover and/or annual balance sheet total does not exceed EUR 10 million (~HUF 2,75 billion) [20].

2 The *definition of energy saving or building energetics investment* is used in this research area in a way that it refers to the new or replacement energy saving investments carried out by micro and small enterprises in the residential and corporate market segment. The term, *building energetics* includes all building industry products and services (electrical, mechanical, etc.) that actively or passively influence the energy consumption of the building for long-term use.

2 The Research Topic and Processing Methods

2.1 Brief Description of the Technology Roadmapping Method

TRM is a widely used method all over the world for developing economic and social policies, making environmental decisions, introducing legislations and laying down R&D guidelines that contribute to the national level energy policy [37, 40]. Phaal [42] has listed 159 publications that support the application of TRM in the field of energetics. TRM became a popular topic in management literature in the 2000s, the method is still evolving, and has to be further researched and developed.The TRM is used not only at company-level, but industry-level investigations and decision-making too [28]. Industry-level aspect of the method will be discussed in this study, however company-level applications of the method will not be mentioned.

The authors, Phaal, Farrukh, Mitchell and Probert [39] decided that it would be more useful to call the method technology-product, product-, business or strategic mapping. These names express that the method is suitable for the process, product and market prospects of integrating and synchronizing. According to Moehrle, Insenmann and Phaal [31] TRM is ordered directly to strategy and innovation. The roadmap name in this paper contains a strategic approach which is based on research information. Hronszky and Várkonyi [19] distinguish between two main kinds of industry-level roadmapping. In the case of forecasting future trends are more or less predictable, however in the case of foresighting future events are very uncertain. The energy saving strategic technology roadmap, described in this paper, is an example of the first type supporting the forecast.

TRM has no graphing standards like e.g. project management. The standard versions of e.g. network diagrams can be learned from several project management books and other teaching media. These graphic project management tools are more or less standardized, like the symbols of electronic circuit diagrams. However there are several examples for graphing TRM standards but there are no standardized diagramming techniques.

2.2 The Dimensions of the Roadmap Analysis

The structural principle of roadmap presented in the study (Figure 1) is suitable for first general-level analysis for the industry, but indirectly, is able to support processes and business level strategies as well. The top layer of TRM is usually called the "market". Experiences show [38] that considering the market in itself is not enough, as sooner or later we always end up analysing the broader environment. The assessment of the company's environment is essential to reveal the strategic situation. The traditional marketing approach divides the environment of the company into macro- and micro level [24]. The micro environment covers those elements that have stronger relationship with the company. The model presented in this paper will be constructed by following this principle. In this study Porter's five forces model [44] is used to analyse the micro environment. Furthermore the widely known STEP [1] model is used to analyse the macro environment.

The literature recommends converting maps and graphics structure for the actual specific purposes univocally. The frame of the model is constructed by merging two structural principles [38, 40]. The information is derived on the basis of presented methods and models in section 2.4. Results of the earlier studies are in Chapter 3.

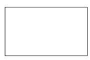
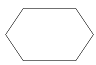
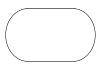
2.3 Legend

Time: Three periods are drawn altogether (current, gaps and migration paths, vision).

Categories: micro level enviromental factors (market, suppliers, product and services) and macro level environmental factors (technology, energy policy, economic contraints, social influence and nature).

Subcategories: the environmental factors are divided into subcategories. (The sustainability factors are merged into macro environmental subcategories.)

Key to the symbols of the periods:

- Current phenomena: 
- Things to be done (agendas): 
- Ideal future trends: 

2.4 Identifying the Knowledge Gaps

It is important to clarify the knowledge gaps [40] during every step of the TRM process. The necessary information for putting together the strategic technology roadmap of the Hungarian residential-level energetics is partly missing in the literature. I summarize this fragmented information using the insights and recommendations of Duinker and Greig [12], Brummel and MacGillivray [6], Mahmoud et al. [27], Shoemaker [47], Postma and Liebl [45], and Swartz et al. [48] about producing a proper roadmap. Industry-level energy generation is widely covered by the literature but there are only a few cross-references to the residential-level, so the adaptation of the roadmap to the circumstances in Hungary must be limited. That is – among others – why the expected technological information and other numbers on the roadmap are taken from a document published by the Japanese Ministry of Economy, Trade & Industry [21], corrected slightly when it was customized to the situation in Hungary.

The following knowledge gaps are identified as (A) (B) (C) and (D). (A) New energy policy strategic programs, directives, elements of fiscal policy. (B) Need for radical innovations or such new technological trajectories that are not yet visible or are still in the critical experimental phase. (C) Concrete priorities (the roadmap contain the different technological and energy policy areas in general). (D) Future trends and counter-trends, and potential but unexpected events (mainly in the case of the environmental, economic and social factors) that can influence the foresight and cause a shift in any field; the roadmap contains the estimations based on the previously known consumption habits and behaviours.

The majority of the knowledge gaps listed above originate in the lack of the detailed energy strategy in Hungary. The recent map contains changes in attitudes, motivations and resources required from the stakeholders, and it contains their collective and investment risks, furthermore different customer types, that are listed in knowledge gaps of Bíró-Szigeti – Pataki [2] work too.

The background information and specification needed are presented to draw the strategic technology roadmap of residential energy saving investments and interpret the data shown on roadmap (see Figure 1) in the next sections.

2.5 Information Sources of Macro and Micro Environmental Analysis

Both primary and secondary information sources were used to the description of the market structure and its dispersion and to the analysis of main participants. Later I will refer to the examined background of the related results in Chapter 3 according to the same classification (alphabetical symbols and figures) below.

(I) Information sources of micro environmental factors and processing methods of the model are detailed below. These were identified as (1), (2) and (3).

(1) The method of in-depth interviews was used to analyse the strategies of the micro and small enterprises (2008-2010). 25 personal (qualitative) interviews were carried out with experts and executives of this subtopic. The obtained information was organised according to *(a)* Kotler's [25] five factors of competition conditions, *(b)* Doyle [11] differential advantages model *(c)* Doyle [11] customer needs, *(d)* consumer decision-making processes [13, 34], *(e)* decision-making unit and consumer behaviour models [50], as well as *(f)* other corporate and marketing strategy describing standard methods and tools.

(2) Detailed analyses were carried out concerning building energetics micro and small enterprises and the energy saving investments, based on the online

(quantitative) questionnaires. The objective of the research was to reveal the market position, strategies, management knowledge, customer value and success factors of the building energetics micro and small enterprises, and also to compare the results in different segments. The online data survey took place in 2010. The enterprises in HBI 2010/1 company database were considered as population, that was classified on energy, energy building technology, building and electrical Hungarian NACE (TEÁOR) numbers (4110, 4120, 7111, 7112 and 7120). There were 2549 companies altogether. Each of them was asked to complete the questionnaire. Thus this survey can be considered complete. 212 questionnaires were analysed altogether. The response rate was around 8.3% for the main population and also for the total sample. The respondents are representative for the range of energy saving investment projects of the enterprises. The representation was examined on the basis of company size property: sales and number of employees. The results from the survey include information of the subcategories of the Figure, 'Market' and 'Goods and services'.

(3) Both secondary (literature) and primary (in-depth interviews) sources were used to reveal the stakeholders' interests and relationships, and the gained information was organized, according to the "stakeholders of environmental investments" method [3] developed by Overseas Development Administration (ODA) [35]. The results and the conclusions of the ODA analysis can help to increase the number and the rate of the successful residential energy saving investments [4]. On this subtopic, 17 in-depth (qualitative) interviews were conducted with experts and executives according to stakeholders' analysis method. The main results of the analysis of the models 'Suppliers' were built in category 'Public energy company', 'Financial institution', 'University, Research institute' subcategories of the Figure. The 'Energy Policy' category of the Figure contains information on state, local governments, environmental and other influencers.

(II) Macro environmental factors of the theme are based on the literature of secondary sources, which is detailed in section 3.2. These are the well known macro environmental factors of the STEP model that was extended in several ways. The roadmap is made up of layers which represent the usual structure of STEP model of the following environmental factors: 'social influence', 'technological', 'economic constraints', '(energy) political', and – in this context – 'natural (ecological)'. In particular, the technologies were previously analysed in the form of a so-called "Technology roadmap on the Hungarian residential energy saving" [2], showing each technological development tendencies. It is important to note that energy policy can be connected to mostly every other layers because residential-level building energetics is a tightly state controlled industry. The particular entities on the roadmap could be assigned for more than one environmental factor, so this layout is only one of the possible graphing structures.

ENVIRONMENT – Micro level

Time →

		CURRENT	GAPS, MIGRATION PATHS	VISION
Market	Enterprises	- Lack of cooperation, strategy building, differentiation. - Features: short-term benefits, primitive marketing tools, word of mouth advertising, mass marketing, one-step segmentation, pessimism, strong interdependence relationship - a commission system.	- Rapid response for macro-environmental changes and for steps of competitors. - Consumer research on basic necessities. - Strengthening management and marketing skills : ~ high quality products are offered in medium-high price class, ~ consumer loyalty, relationship management, brand design and strengthening, distinguishing advantage, ~ using concentration or differentiation strategy, ~ developing traditional marketing tools, ~ emotional (subjective) effects. - Paying attention to decision maker persons. - Buying decision process should be supported by application of the Nicosia model.	- Using basic strategy and marketing tools of enterprises that appropriate for the area of their size and activity. - Increasing the price margin successfully. - Appling the target market based on wants approach. - Applicating the marketing orientation for consumers' needs and customized solutions. - Strengthening the communication between the state and the enterprises: ~ Preventing or correcting the aversions and the misperceptions.
Market	Consumers	- Types of consumers: ~ consumers belonging in high-income category: a sense of prestige, secondary return on investment. ~ consumers belonging to medium-income category: feel good deed for future generations. - Motivation: energy savings. - Appearance the environmentally conscious consumer behavior. - High expectations on high price. - Customized products and construction at the lowest price as possible. - Buying decisions procces based on some product characteristics. - The price sensitivity is less typical, quality is more important. - Directly supported stakeholders (KEOP tenders, subsidies of social policy, real estate loan payments). - Facilities: creating financial conditions. - Risk factors: the companies 'maximum profit' aspect.	- Shaping energy-conscious attitude and lifestyle. - Acquiring financial and savings culture.	- Traditional consumer behavior should be replaced by the environmentally conscious consumer behavior in higher proportion. - Involving more lower-income categories of consumers.
Suppliers	Energy companies	- Opposing interest with stakeholders: reduction in energy consumption results decreasing income of the energy companies.	- Promoting the spread of residential energy saving approach. - Being more flexibility. - Making up for lost income with specialized professional knowledge.	- Implementation of the corporate social responsibility. - Profitable enterprises.
Suppliers	Financial institutions	- Credit and social support is available only with limited rate and strict conditions. - Risk aversion: return on investment slowly. - None state guarantee.	- Developing calculation guides: the household investment are returnable under what conditions. - Increasing the risk-taking of financial institutions.	- Supporting the credit products helps the spreading of investments. - High level of culture in finances and savings.
Suppliers	Uni.&Re. institutions	- Tool-providing stakeholder: intellectual resources.	- Keeping up with market requirements and continuous renewal of approach. - Increasing both influence and importance (enforcement of interest) is important.	- Ensuring continuos technology development. - Testing different solutions. - Training experts. - Exercising an opinion-forming role.
Products and services		- Wide portfolio of products and services. - Complexity.	- Continuously enlarging product and service portfolio.	- Complex investments are realized on finance funds. - Prevailing sustainability criteria.

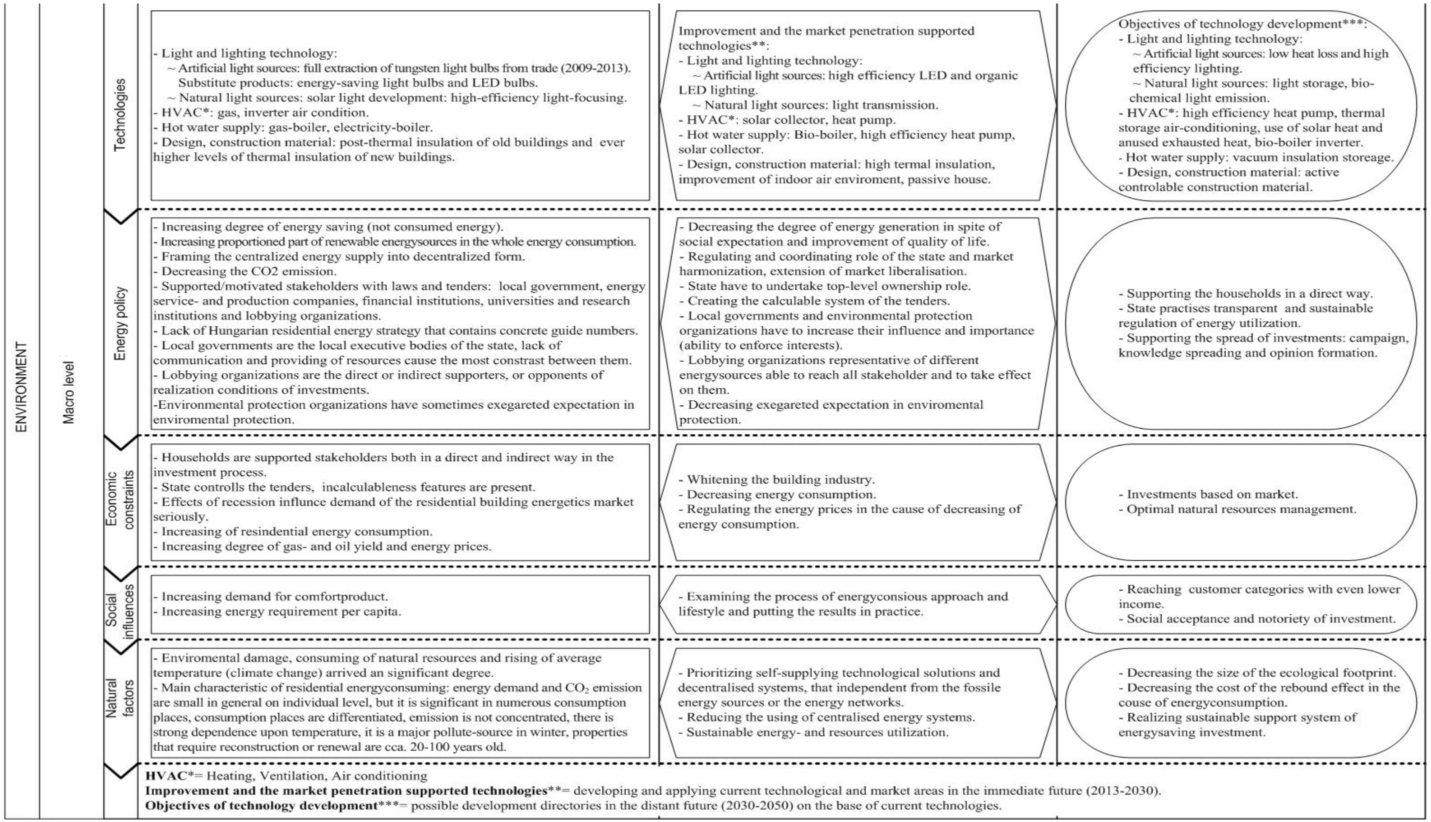


Figure 1: Strategic technology roadmap of the Hungarian residential energy saving.

Source: own figure. Using data from: [4, 2, 5, 7, 9, 10, 14, 15, 16, 17, 18, 21, 26, 29, 30, 33, 36, 38, 41, 44, 49, 51]

3 Presentation of Results and Discussion

3.1 Results from the Micro Environmental Analysis (I)

(1) One of the main directions of this research is to identify the strategy of building energetics enterprises, its product- and service concepts and, the main characteristics and requirements of the customers in an indirect way. In the following those research results are detailed that were obtained on the basis of the models already described in paragraph (1) of the section 2.2. These were identified as *(a), (b), (c), (d), (e)* and *(f)*.

(a) There is no cooperation among the competitors on the residential market. Aggressive position-taking can be seen region-wide to deter the competitors when a new technology or a product is launched. The enterprises are continuously making price comparisons and also they have different missions and objectives for their operation fields. Their main business objective is the same (to win orders and to generate profit).

(b) Enterprises do not have conscious corporate strategy, however, they can adapt successfully to the changing environment. They mainly use basic marketing tools. The word-of-mouth advertisement has a significant role when they are doing business. They do not know or realize the added-value of their products or services. There is a tendency in the building energetics sector that the profit of the businesses will be attributed to the selling of the physical products at a decreasing rate. Currently, most of the micro and small enterprises are willing to do business with anyone who can pay for their products and services. In this way, the enterprises will have a large number of customers with low profit margin. Furthermore, they lay more emphasis on the short-term benefits than long-term profits. The factors of the customer value analysis for determining the value offer can be divided into two groups. The sense of prestige dominates among the high-income customers, and the return on investment is a secondary criterion, whereas among the middle-income customers, the sense of a positive step for the sake of the future generations dominates. However, high expectations that are attached to high prices can create substantial customer value in both categories. Even if the enterprises are able to name some of their target groups, they usually do not use different communication strategies for addressing them or also to attract them with different marketing tools. They communicate and behave differently with consumers depending on the personality of the salesman and/or the contractor based mainly on their knowledge of human nature. Their strategic tools are quite similar to mass marketing.

(c) The enterprises do not analyse the customer needs due to time and resource limitations. Moreover, it is difficult for them to formulate a need that they could not possibly satisfy, because it would cause a competitive disadvantage for the enterprises.

(d) The classical model of consumer buying decision process [13] was changed to a 0+5-step model, that is very similar to the Nicosia-model [34]: 0. Interest, 1. Problem awareness 2. Information search 3. Evaluation of alternatives 4. Purchase 5. Post-purchase evaluation.

(e) Most of the time, the decision maker is a person with technological and economic background, but other family members, and sometimes even friends, acquaintances and colleagues can influence the decision making process. It can be assumed that the buying decision is based on some product characteristics. The traditional consumer behaviour has to be replaced by the environmentally conscious consumer behaviour to a greater extent in order to increase the number of successful investments.

(f) The results presented that only a small part of the analysed building energetics enterprises use conscious segmentation in the sphere of the residential market. Most of the time, the segmentation is built on a certain technology and rarely on the information obtained from the analysis of the consumer behaviour, or the values that the customers expect. Segmentation is usually done in one step, i.e. the enterprises categorise the customers on the based of only one criterion.

(2) Results of quantitative market research in paragraph (2) of section 2.2 will be presented in the following. The following factors were analysed in the empirical research in order to reveal the market success and marketing management characteristics of the micro and small enterprises undertaking building energetics investments: identifying and evaluating the marketing operation (tools and strategies); the quality and price of the product and the income situation of the customers; the number and characteristics of the products and business operations; mission; consumer loyalty; subjective and objective factor of success; price sensitivity; and the importance of energy saving. The more an enterprise undertakes higher-level and extended marketing operations, the more successful it can become. Based on the variables of the analysis included, it can be stated that this increases the importance of the differential strategy. Positioning for product quality plays a crucial role in increasing success among the marketing operations. A connection was discovered between the product quality and the price level. The most successful strategy is to offer high quality products for high or medium price. Successful building energetics enterprises are optimistic about their future prospects. They consider the market to be continuously expanding and their increasing margin reflects a rising demand. Their customers are loyal, which indicates adequate sales and communication activities and assumes branding. The subjective factors (they are not easily measurable and quantifiable, or they build on emotions) proved to be the most successful marketing tools among the factors

of success. Indirectly, this can relate to the conscious and wide-ranging marketing activities of the enterprises. The recession has exerted a more dramatic impact on the demand of the residential market. Due to this fact the enterprises see their future more pessimistically on the residential market. Based on the opinion of the survey respondents, the energy saving motivation is more dominant in this market segment that can be considered as an important factor when developing a strategy that builds on the costumers' (stakeholders) motivation. In the residential market, marketing tools play a more significant role in sales and can be used much more efficiently. Moreover, enterprises have developed different communication and sales strategies for their target groups. The primary stakeholder groups identified in the residential customer segment are less price sensitive than the corporate segment. This partly contradicts the "standard" marketing concept that consumers are price sensitive, whereas corporations – mainly business corporations – are not, or less, since they add their costs to their products and make their customers pay for them. Moreover, it can be stated on the base of the research results that the supply of the enterprises operating on the residential market includes wider range of products and services and these are more complex.

(3) Third main direction of our primary market research is the stakeholder analysis. The results will be described in the following. The methodology was already presented in paragraph (3) of section 2.2 formerly. The lack of interest appears in case of two primary stakeholders *(building energetics enterprises and household customers)* in connection with the financial terms and targets. "Maximum profit" approach of *enterprises* could mean risk factor for the customers, if the quality of professional implementation and the return of residential investment are compromised. The customer base is narrowed down primarily according to their income due to the typically high investment cost. The four groups of the enterprises can be identified in scope of activities: manufacturers, designers, distributors and constructors. These enterprises ensure the necessary tools for mental and physical implementation. There are strong interdependence relationships between these enterprises. Their investment interests are mostly derived from the recommendations / mediation fees received directly from manufacturers and these are proportional with commission rate. The primary stakeholders have financial (savings, loans, grants, equity), mental (expertise, innovation), total assets (technology, labour, real estate) resources. The problems usually occur in financial resources and pose a barrier to the realization of investments. Among the primary stakeholders, the household customers are supported in the process both directly and indirectly. However, I met a number of references in the qualitative research that indirect forms of support (mainly aimed at whitening the building industry) have not been successful. Any organizations that can serve valuable / useful information for the creation of the investments were considered *suppliers* - in addition to the traditional interpretation. In this context, the following statements for suppliers were made: *energy service- and production companies* should equip themselves with great flexibility and advanced technical knowledge in the future that is characteristic of small

enterprises, which attracts the modification of the legal background as well. The taking risk of *financial institutions* should be increased for spreading of loan products. Increasing both influence and importance would be important for *universities and research institutions* because they can fall out the scope of secondary stakeholders due to the improper lobbying.

3.2 Results from the Macro Environmental Analysis (II)

The macro environmental band of the model describes the current trends and what changes are necessary that influence the market and the implementation conditions of the investments. Resources of the discussed secondary stakeholders (state, local governments, environmental organizations and lobbying) are extremely complex, vary according to each stakeholder subcategory of the Figure. Consumption of resources has significant impact on household energy savings investment plans and their feasibility. The economic, the social and the environmental factors are the same with the basic pillars of sustainable development. Ideally, states should take all these three elements into account to shape their energy policy that, as the model also shows, has impact on the developing technology areas. Macro environmental and sustainability aspects of this study have already been consulted by Bíró-Szigeti [4]. The gathered information is summarized in the form of the rough strategic technology roadmap.

The main *technological possibilities* appearing in residential energy saving areas are following [21, 29, 51]: to develop new lighting technologies: conservation and using natural light, different models of LED, with low heat loss and high efficiency luminous bodies; to develop HVAC (heating, ventilating, air conditioning) and hot water supply: to replace air-conditioning equipment, to equip heat pump and solar collector, to use unused energy/heat; to develop energy saving planning and construction materials for housing and building: developing materials for thermal insulation and applying to building materials. Further possibilites are: spreading high quality housing by convenient and highly accurate housing performance design and assessment technologies.

Energy policy. The biggest knowledge gap is in the field of the Hungarian household energy saving strategy. Hungary currently has no comprehensive, detailed and coherent energy strategy that contains concrete (with data plans and feasibility) goals. The available guidelines and concepts are not sufficient. Control of the state coordinating role and market harmonization are essential for the next steps [33]. State has to undertake top-level ownership role over the industry roadmap, because serious results can not be expected without lack of top-level ownership. The state has a major role and attempts to provide support policies (laws and regulations) and incentives (tenders) for the local authorities, energy suppliers and energy companies, financial institutions, universities and lobby organizations as top-down pressure. Organizations representing the fossil fuel lobby are able to influence almost all the stakeholders. It would be necessary to reduce the excessive environment protection requirements of the environmental protection organizations. It would be important to improve the influence and the significance of the environmental organizations and the local authorities, because they could fall out of the scope of secondary stakeholders due to inadequate lobbying.

Social influence. Energy demand decreases if for example the population decreases [45], or the energy conscious way of thinking and lifestyle is spreading (i.e. the demand for air conditioning is reduced), and if new buildings are designed in a more enviromental friendly way as a result of complex energy saving housing (houses with low energy demand, passive houses) [21, 29]. By the spreading of environment-conscious thinking and technology applications broader and broader consumer range and consumer categories with lower and lower income levels can be reached and involved into saving energy. Fábián et al. [16] separated the main groups of Hungarian society in the course of their research according to three ranking criteria: flat, material and cultural consumption. The first group is the elite characterized by outstanding material and cultural consumption, the second is the wealthy group, the third is the group that accumulate savings, the fourth is the free time oriented group, the fifth is the deprived group with good flat, and the sixth is the deprived-poor. Building energetics investments can be accomplished in the first place by the members of the first three groups (elite, wealthy and the group that accumulate savings) because of their motivation, values and financial situation. The conclusion of their research is that the exploration of the possibilities of crossing the boarders of the separate consumption models could contribute to the motivation of some particular consumption patterns and to the understanding of the tools of influencing the formation of consumption patterns.

Economic constraints. The most important technological requirement in the residential building energetics sector is that the energy consumption per household must be reduced by 80% till 2100, in spite of the fact that in the same time the share of energy consumption within the GDP is expected to increase continually [7]. Increasing proportion of energy consumption within the GDP appears to be unavoidable because every government tries to raise the living standard but it involves more demand for energy [7, 15]. The share of electric energy within all types of energy should reach 100% by 2100 [21]. Within electric energy the proportion of fossil energy sources should be continually decreased. This requirement involves that the energy demand of buildings and household appliances, the usage of long distance transmission line networks, the magnitude of centralized energy generation, and CO2 emission must be reduced. The proportion of renewable energy sources and operations by using electricity must be increased while the users' comfort level and other requirements must be maintained [10, 15]. It can be stated that the spread of investments cannot be expected to be based on only the market, so cooperation of the secondary stakeholders, strengthening and harmonization of their interest with the investment might get to the foreground. The responsibilities and influences of the secondary stakeholders impact the establishing of the residential energy saving investment to greater extent.

Natural factors. The residential sector is typically characterized by physical factors [9, 21]. These should be considered when the substitution or the energy saving replacement of the currently most widely used technologies with alternative technological solutions in the future is planed. Climate change and population growth determine the size of the ecological footprint. The running out of our energy resources, the decreasing natural ground area, and the increasing population are continuously reducing the size of the ecological footprint per capita while the inhabitants of the developed and developing countries have bigger and bigger ecological footprints [36]. The less is the magnitude of the rebound effect [22] and the more efficient is the operation of the new energy saving devices, the higher living standard can be achieved with the same ecological footprint size [18]. At last but not least one of the most serious natural factors urging energy saving is the rise of the average temperature, resulting in climate change. Signs of climate change appear in temperature and precipitation intensity and frequency contexts in Hungary [33]. Expert opinions diverge about of the magnitude of this future change, but this problem must be taken seriously and must be dealt with.

Conclusions

The study aims to contribute to description of the current state of the Hungarian building energetics market and the residential energy-saving investments with a specific approach. The approach provides system and market (strategic) aspects for the variety stakeholders in the energy saving market. Complex framework conditions of the residential building energetics investments will be summarized on the base of the literature and the expedient collected primary information. I assumed if I developed a strategic technology roadmap than the necessary information would be currently available, on the base of which the proposed model can help to define the main guidelines of consistent energy policy and strategy for the long-term and successful marketing strategy and tools for building energetics enterprises.

I collated and summarized the most important information from the result of the enviromental analyses (in section 3.1 and 3.2) and from the strategic technology roadmap. Note that these results contain complex information from macro and micro level analyses which are necessary together for the most important actions to achieve the 'Vision' future state proposals (see Figure 1). These results (in conclusion form) were identified as (i), (ii), (iii) and (iv). The issues could been parallel between some results from other SME’s researchwork [23].

(i) To fulfil the required *technological, economical, social and natural expectations* in the residential sector the energy consumption must be reduced in the first place. At this point those household appliences and pieces of equipment should be taken into consideration that will come out only in the future. In the second place such energy generation methods must be found that utilize continuously available renewable resources like solar and geothermal energy. Those technological solutions and decentralised systems are prioritized which are self-supplying, independent from the fossil energy sources or the energy networks. In this way less energy produced by the existing (centralised) energy system will be transmitted for long distances, consequently efficiency can be increased and energy generation demand can decrease.

(ii) *State* and industrial coordination of implementation of the residential energy saving investments – in relation with sustainability criteria – is burdened with numerous problems. There is a lack of strategic plan and state guidelines that contain concrete goals and comprehensive to the building energetics industries. The available guidelines and concepts are not adequate. The regulating and coordinating role of the state, as well as the market liberalisation are essential for the next steps. The work of a multidisciplinary expert team would be necessary to develop a complete energy strategy plan for Hungarian that is acceptable for all stakeholders.

(iii) It is difficult for *consumers* to take energy-saving efforts themselves due to existing misconceptions, lack of information, and any bad experiences, lack of financial resources and short-term thinking, etc., while the energy prices are continuously rising, the demands of the comfort level are getting even higher and the residential energy saving has increasing impact on the national economy. It would be necessary to spread energy-conscious attitudes and lifestyle for the environmentally conscious consumer behaviour. Higher level financial and savings culture could support the market size and the number of segments efficiently in the future.

(iv) *Micro and small enterprises* in the building energetics sector present strategic behaviour, the modern marketing concept and many elements of marketing tools in their practice, but at low levels, due to the skill gaps, resulting mostly non-conscious and non-systematic marketing management application. The success factors of marketing strategies and tools of the building energetics micro and small enterprises are summarized in the following. These could serve as examples to establish future strategies. These are based on the differential advantage concept, follow a focus strategy, and offer high quality products in the medium or high price category. Consumer loyalty and relationships, brand name and subjective (not easily measurable and quantifiable) factors dominate among the marketing tools of the successful building energetics enterprises. Moreover, their product portfolio and field of operations will be characterised by increasing market demand and margin.

The strategic technology roadmap is not complete in its present form. In order to improve it the most important elements of the knowledge gap were collected. Further studies and research in the following areas are needed especially to fill the knowledge gaps. That work can be continued exclusively by a multidisciplinary expert team: to examine customer specifications; to survey social misconceptions; to adapt success enterprise strategies in different business segments; to research applicable technologies and future development trends; to complete the Hungarian state energy strategy. The roadmap is applicable to develop the corporate-level roadmap for building energetics enterprises. So the enterprises can customize it into the field of their activity and market in their strategy.

In addition, the results of the study can serve with practical information for the concerned enterprises, and may also be used for the state or other institutional stakeholders to design the promoting energy savings policy decisions (regulations, subsidies) in residential building energetics and to evaluate their effectiveness. If the energy efficient building energetics investments spread on market based, it would have a positive impact on the economic, social and natural environment of Hungary, and so it could support sustainable development in long-term.

Acknowledgement

Thank for methodologycal advices and helping my work for Mária Vágási, PhD, associate professor; Béla Pataki, PhD, associate professor and Györgyi Danó, PhD student, market researcher.

References

[1] Aguilar, F. J.: Scanning the Business Environment, Macmillan, New York, 1967

[2] Bíró-Szigeti Sz., Pataki B.: Applying Technology Roadmapping for the Energy Saving Industry, Periodica Polytechnica – Social and Management Sciences 20:2, 2012, pp. 105-113

[3] Bíró-Szigeti Sz., Vágási M.: A háztartási energiamegtakarítási beruházások érintettjeinek rendszerelvű megközelítése - egy kutatás tapasztalatai alapján (Systemic approach of the residental energy saving investments stakeholders - Based on the experience of a research) (in Hungarian), in: Topár J. (ed.): A műszaki menedzsment aktuális kérdései, Műszaki Kiadó, Budapest, 2012, pp. 323-342

[4] Bíró-Szigeti Sz.: Mikro- és kisvállalkozások marketingfeltételeinek vizsgálata az energiamegtakarítás lakossági piacán (Analysing the marketing conditions of the micro and small enterprises on the residential energy saving market) (in Hungarian), PhD dissertation, Budapest University of Technology and Economics, Budapest, 2011

[5] BPA (Bonneville Power Administration): Energy Efficiency Technology Road Map, USA BPA Technology Innovation Office, 2006 http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=4077537&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D4077537

[6] Brummel, A., MacGillivray, G.: Introduction to Scenarios

http://scenarios2strategy.com/pdf/Introduction%20to%20Scenarios%20and%20Scenario%20Planning.pdf

[7] CEC (Commission of the European Communities): A Technology Roadmap for the Communication on Investing in the Development of Low Carbon Technoilogies (SET-Plan). Commission staff working document. SEC(2009) 1295, Brussels, 2009 http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0519:FIN:EN:PDF

[8] CEC (Communication from the Commission): Action Plan for Energy Efficiency: Realising the Potential. COM(2006)545 final SEC(2006)1173-1175, Brussels, 2006 http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:52006DC0545:HU:NOT

[9] Csete J., Horánszky B., Szunyog I.: A lakossági energiafelhasználás környezetterhelése (Enviromental load of residential energy consumption) (in Hungarian), Energiagazdálkodás, 46:5, 2005, pp. 19-25

[10] Dinya L., Domán Sz., Fodor M., Tamus A.: Ökoenergetikai marketing kihívások (Eco-energetics marketing challenges) (in Hungarian), Marketing Oktatók Klubjának 12. Országos Konferenciája, Conference proceedings on CD, Budapesti Műszaki és Gazdaságtudományi Egyetem, Budapest, 2006

[11] Doyle, P.: Value-based Marketing: Marketing Strategies for Corporate Growth and Shareholder Value. Wiley, Pennsylvania, 2000

[12] Duinker, P. N., Greig, L. A.: Scenario Analysis in Environmental Impact Assessment: Improving Explorations of the Future, Environmental Impact Assessment Review, 27:3, 2007, pp. 206-219., cited in: Moen, 2009

[13] Engel, J., Blackwell, R., Kollat, D.: Consumer Behavior, Rinehart and Wiston, New York, 1973

[14] EPT (Európai Parlament és Tanács): 2010/31/EU irányelve az épületek energiahatékonyságáról – átdolgozás (2010/31 EU policy on the energy efficiency of buildings – revised version) (in Hungarian), 2010 http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2010:153:0013:0035:HU:PDF

[15] EREC (European Renewable Energy Council): Renewable Energy Roadmap up to 2020, Renewable Energy House, 2007 http://www.erec.org/fileadmin/erec_docs/Documents/Publications/EREC-Technoloy_Roadmap_def1.pdf

[16] Fábián Z., Kolosi T., Róbert P.: Fogyasztás és életstílus (Consumption and lifestyle) (in Hungarian), in: Kolosi T, Tóth I, Vukovich Gy, (ed.): Társadalmi riport 2000 (Social report 2000), TÁRKI, Budapest, 2000, pp. 346-376

[17] GKM (Gazdasági és Közlekedési Minisztérium): Magyarország megújuló energiaforrás felhasználás növelésének stratégiája 2007-2020 (Hungary's strategy for increasing the usage of renewable energy sources 2007-2020) (in Hungarian), GKM, Budapest, 2007

[18] Harangozó G.: A javuló energiahatékonyság szerepe az energiafelhasználás csökkentésében: lehetőségek és buktatók (The role of improving energy efficiency in the reduction of energy consumption: opportunities and obstacles) (in Hungarian), Budapest, BCE, 2009 http://www.uni-corvinus.hu/index.php?id=26651

[19] Hronszky I., Várkonyi L.: Radikális innovációk menedzselése (Managing radical innovations) (in Hungarian), Harvard Businessmanager, 8:10, 2006, pp. 28-41

[20] HUP (Hungarian Parliament): 2004. évi XXXIV. törvény a kis- és középvállalkozásokról, fejlődésük támogatásáról (Act XXXIV of 2004: Small and Medium-sized Enterprises and their Development Act) (in Hungarian), 2004 http://net.jogtar.hu/jr/gen/hjegy_doc.cgi?docid=A0400034.TV

[21] IAE (Institute of Applied Energy: Japanese Ministry of Economy, Trade & Industry): Strategic Technology Roadmap (energy sector) – energy technology vision 2100, 2005 http://www.iae.or.jp/2100/main.pdf

[22] Jevons, W. S.: The Coal Question. An Inquiry Concerning the Process of the Nation, and the Probable Exhaustion of our Coal-Mines, MacMillan, London, 1865

[23] Kadocsa, Gy., Francsovics A.: Macro and Micro Economic Factors of Small, Enterprise Competitiveness. Acta Polytechnica Hungarica, 8:2, 2011, pp. 23-40

[24] Kotler, P.: Marketing Management, 11th ed., Prentice Hall, Englewood Cliffs, 2003

[25] Kotler, P.: Marketing menedzsment, Budapest, KJK-KERSZÖV, 2004

[26] KvVM (Környezetvédelmi és Vízügyi Minisztérium): Nemzeti Éghajlatváltozási Stratégia (National climate change strategy) (in Hungarian), PRUDENCE project, 2007 klima.kvvm.hu/documents/14/NES_6.4c.pdf

[27] Mahmoud, M., Liu, Y., Hartmann, H., Stewart, S., Wagener, T., Semmens, D., Stewart, R., Gupta, H., Dominguez, D., Dominguez, F., Hulse, D., Letcher, R., Rashleigh, B., Smith, C., Street, R., Ticehurst, J., Twery, M., van Delden, H., Waldick, R., White, D., Winter, L.: A Formal Framework for Scenario Development in Support of Environmental Decision-Making, Environmental Modelling & Software, 24:7, 2009, pp. 798-808, cited in: Moen, 2009

[28] McCarthy, R. C.: Linking Technological Change to Business Needs, IEEE Engineering Management Review, 31:3, 2003, pp. 49-53

[29] Medgyasszay P.: Fűtési energia megtakarítás épület léptékben (Saving heating energy on the scale of buildings) (in Hungarian), 2007a http://fenntarthato.hu/epites/leirasok/nes/futes-epulet-leptekben

[30] Medgyasszay P.: Várható éghajlatváltozás hatása az épített környezetre (The effect of the expectable climate change to the built environment) (in Hungarian), 2007b http://fenntarthato.hu/epites/leirasok/nes/eghajlatvaltozas-hatasa

[31] Moehrle, M. G., Isenmann, R., Phaal R.: Technology Roadmapping for Strategy and Innovation, Springer, New York, 2013

[32] Moen, J., Jonsson, R.: Scenarios in the context of Future Forests, Future Forests Working Report, 2009, p. 11 http://www.futureforests.se/download/18.5686ae2012c08a47fb5800023704/Scenarios+in+the+context+of+Future+Forests.+PDF.pdf

[33] NeFMi (Nemzeti Fejlesztési Minisztérium): Nemzeti Energiastratégia 2030 (National Energy Strategy 2030) (in Hungarian), 2011 http://www.kormany.hu/download/5/39/40000/Energiastrategia.pdf

[34] Nicosia, F.: Consumer Decision Process, Prenntice Hall, Englewood Cliffs, New York, 1966

[35] ODA (Overseas Development Administration Social Development Department): Guidance note on how to do stakeholder analysis for aid projekts and programmes, London, 1995

[36] Pappné Vancsó J.: Az ökológiai lábnyom számítási módszerének bemutatása Magyarország példáján keresztül (Introducing the calculation method of ecological footprint through the example of Hungary) (in Hungarian), Táj, tér, tervezés, Geográfus Doktoranduszok VIII. Országos Konferenciája, Szeged, 2004 http://geography.hu/mfk2004/mfk2004/phd_cikkek/pappne_vancso_judit.pdf

[37] Pataki B., Bányai A.: Egy új stratégiai eszköz: a technológiai úttérképezés. (A new strategic tool: technology roadmapping) (in Hungarian), CEO 11:1 2010, pp. 8-12

[38] Pataki B., Bíró-Szigeti Sz., Szalkai Zs.: Company-level Technology Roadmapping, in: Ran, B. (ed.): The Dark Side of Technological Innovation (Contemporary Perspectives on Technological Innovation, Management, and Policy) Information Age Publishing, Charlotte, 2013, pp. 83-109

[39] Phaal, R, Farrukh, C, Mitchell, R, Probert, D, Starting-up Roadmapping Fast, IEEE Engineering Management Review, 31:3, 2003, pp. 54-60

[40] Phaal, R., Farrukh, C., Probert, D.: Customizing Roadmapping, Research-Technology Management, 47:2, 2004, pp. 26-37

[41] Phaal, R., Farrukh, C., Probert, D.: Technology Roadmapping – A Planning Framework for Evolution and Revolution. Technological Forecasting and Social Change, 71, 2003, pp. 5-26

[42] Phaal, R.: Public-Domain Roadmaps, Centre for Technology Management University of Cambridge, 2009 http://www.ifm.eng.cam.ac.uk/uploads/Research/CTM/Roadmapping/public_domain_roadmaps.pdf

[43] Phaal, R.: Roadmapping Bibliography, University of Cambridge, 2013 http://www.ifm.eng.cam.ac.uk/uploads/Research/CTM/Roadmapping/Roadmapping_Bibliography_Phaal.pdf

[44] Porter, M. E.: How Competitive Forces Shape Strategy. Harvard Business Review 57:2, 1979, pp. 86-93

[45] Postma, T. J. B. M., Liebl, F.: How to Improve Scenario Analysis as a Strategic Management Tool?, Technological Forecasting and Social Change, 72:2, 2005, pp. 161-173

[46] PRB (Population Reference Bureau): 2010 World Population Data Sheet, Washington, 2010 http://www.prb.org/pdf10/10wpds_eng.pdf

[47] Shoemaker, P. J. H.: Twenty Common Pitfalls in Scenario Planning, in: Fahey, L., Randall, R.M. (eds.): Learning from the Future, Competitive Foresight Scenarios, Wiley, New York, 1998, pp. 422-431, cited in: Moen, 2009

[48] Swartz, P.,: The Art of the Long View, Doubleday, New York, 1991, cited in: Moen, 2009

[49] TNM (a regionális fejlesztésért és felzárkóztatásért felelős tárca nélküli miniszter): 7/2006. (V.24.) TNM rendelet az épületek energetikai jellemzőinek meghatározásáról (minister without portfolio regulation of determining the energetics characteristics of buildings) (in Hungarian), 2006 http://net.jogtar.hu/jr/gen/hjegy_doc.cgi?docid=A0600007.TNM

[50] Törőcsik, M.: Fogyasztói magatartás trendek – új fogyasztói csoportok. (Consumer Behaviour Trends – New Customer Groups) (in Hungarian) KJK-KERSZÖV, Budapest, 2003

[51] Ürge-Vorsatz D., Füle M.: Economics of Greenhaus Gas Mitigation, Hungary Country Study, Prepared for UNEP, 1998

Software Environment for Learning Continuous System Simulation

Marijana Despotović-Zrakić[1], Dušan Barać[1], Zorica Bogdanović[1], Branislav Jovanić[2], Božidar Radenković[1]

[1]Faculty of Organizational Sciences, University of Belgrade, Jove Ilića 154, 11000 Belgrade, Serbia, maja@fon.bg.ac.rs, barac.dusan@fon.bg.ac.rs, zoricab@fon.bg.ac.rs, boza@fon.bg.ac.rs

[2]Institute of Physics, University of Belgrade, Pregrevica 118, 11000 Belgrade, Serbia, brana@ipb.ac.rs

Abstract: In this paper, we describe a new graphical environment for learning continuous simulation. This open source software solution enables engineering students to learn and understand various real system models in the CSMP (Continuous Simulation Modelling Programme) language easily, and at the same time learn about the implementation aspects of simulation languages. The software provides efficient tools for creating, storing, and executing continuous system simulation models that are described with differential equations. The graphical block-diagram interface allows students to drag-and-drop predefined modelling blocks, connect them together and create different models of continuous systems. Further, we present an example of using the developed application in solving a typical continuous simulation problem. Finally, we present the results of the evaluation performed on a testing sample of 160 students of undergraduate course Simulation and simulation languages at the Faculty of Organizational Sciences, University of Belgrade.

Keywords: Continuous system simulation; teaching simulation and simulation languages; CSMP software

1 Introduction

In this study, the authors set out to develop a software tool that would effectively support teaching and learning continuous simulation and simulation languages. The focus is on creating an open source tool that would provide an environment for learning continuous systems modelling and simulation as effectively as using commercial software such as Matlab/Simulink, but, unlike commercial tools, provide a testbed for learning simulation languages and the concepts of their implementation.

Simulation is used to describe the behaviour of a dynamic system by means of a model and using this validated model in a series of experiments designed to provide an insight into the future behaviour of the system under specific conditions [1]. The area of computer simulation has been successfully applied to the study and modelling of processes, applications, and real-world objects. Simulation tools provide an environment for the analysis of various features of the system: feasibility, behaviour, stability, performance, etc. [2], [3], [4]. In addition to its many practical applications in industry, commerce, research, and other areas, computer simulation can also be used as a powerful teaching aid [5], [3]. Computer simulation gives the freedom and flexibility to adjust various parameters of the system before design [6]. In some cases, simulation models can substitute real systems that are too expensive or unsafe to install in teaching laboratories [1]. In addition, using simulation is sometimes a requirement, e.g. when designing a manufacturing prototype [7].

All physical systems exist in the time-space continuum. The type of model one selects depends on the degree of aggregation of individual phenomena. One may choose either a continuous or a discrete approach to model development. Continuous models are useful when the behaviour of the system depends more on the aggregate flow of events than upon the occurrence of individual events [5]. A discrete system is one in which the state variables change only at a discrete set of points in time, while in a continuous system the state variables change continuously over time.

1.1 Continuous Simulation e-Learning Course

Simulations play a vital role in engineering education, especially in laboratory exercises [8]. Computer simulation enables engineering students to perform the analysis of various types of systems, investigate relations and interactions among a system's components, project design, and implementation of a system, verify different solutions and environments, test capabilities and technical characteristics [9]. The knowledge gained from simulation courses help students understand, analyze and design systems, develop software components, test software solutions, and solve different real life problems [10], [11].

There are lot of specialized languages for continuous simulation, such as: CSMP (Continous Simulation Modelling Programme) [12], [13], ESL (European Simulation Language) [14], ACSL (Advanced Continuous Simulation Language) [15], CSSL4 (Continuous System Simulation Language 4) [16], Simulink [17], [18], Matlab [19], Modelica [20] and others that have been developed to simplify modelling, and to minimize problems related to programming continuous systems [21]. In the past years, there has been much research dealing with developing tools, languages, and methods for continuous simulation [6], [15], [22], [23], [24], [25]. Using Matlab/SIMULINK for designing simulation models is one of the most frequent scenarios [6]. Matlab is one of the most powerful simulation platforms as it provides a plethora of features [26]. However, a majority of simulation tools have limitations such as high costs, platform dependence, maintenance difficulties, and limited reusability.

Computer simulation has been studied for many years at the Faculty of Organizational Sciences, University of Belgrade, in the scope of the Simulation and simulation languages course. The course is organized at the fourth year of undergraduate studies in the area of information systems and Internet technologies. Before attending this course, students are obliged to take several exams in the field of programming and are familiar with programming concepts and several programming languages (Java, C). The main goal of the course is to introduce the basic concepts and applications of computer simulation and simulation languages.

The course is realized using the combination of traditional classroom-based teaching and e-learning technologies, i.e. blended learning [27]. As a support to the teaching process, we use a learning management system Moodle [28]. The course lasts for three months. The course syllabus is divided into three main blocks: continuous simulation, discrete event simulation and 3D simulation. All the topics are covered by practical exercises using an appropriate simulation language: CSMP/FON (Faculty of Organizational Sciences - in the Serbian language "*Fakultet Organizacionih Nauka*"), GPSS (General Purpose Simulation System), X3D (XML-based file format for representing 3D objects), respectively. Students are obliged to implement their own models and solutions to different problems in the simulation of complex, real systems. In the part of the course that deals with continuous simulation, students model systems from the area of: motion dynamics, electrical engineering, mechanics, fluids, social phenomena.

Solving real cases by means of specialized software helps students to be highly motivated and contributes to overcoming possible deficiencies in their mathematical background. Many of the simulation programmes based on mathematical models have to be used by experts. These programmes cannot be used for providing basic knowledge or gaining experience to engineers or users who are not experts in this subject. On the other hand, an important part of the course curriculum is related to design and implementation of the simulation languages.

Therefore, the main goal of this research was to develop an efficient software tool for creating, storing, and executing continuous system simulation models. This software tool will provide students with a cheap environment for practice, particularly suited for students in the areas of information technologies and software engineering. The idea was to use all the existing features of CSMP and enrich it with interactive graphic environment. The main requirements in the design and implementation of the software environment for learning continuous system simulation were the following:

Create a testbed for learning and realization of continuous simulation languages – the main idea behind the CSMP/FON is to provide students with an environment for both using the simulation language features and learning how to implement key continuous simulation concepts within a simulation language. These main simulation concepts should include emulation of parallelism, solving problems related to algebraic loops, blocks, implicit functions, etc. [4]. Using the CSMP/FON students are enabled to learn more about the structure of the CSMP language and how the CSMP implements all the above-mentioned concepts. As an open source and free solution the CSMP/FON can be extended and modified according to any kind of request. For instance, each student can implement an integration method of their own choosing and perform simulation using the implemented method.

Matlab/Simulink, as the most widely used solution on the market, provides numerous features, but the source code is not open. For its users, blocks are provided as "black boxes", so users cannot change or adapt anything. Accordingly, users can use advanced features related to simulation, but they cannot learn anything about how the simulation language was realized. Therefore, Matlab/Simulink is not entirely adequate for a course like this.

Support for scientific research – scientific research within our laboratory often requires creating complex models that include both continuous and discrete event simulation. In order to combine different simulation models effectively, we need simulation environments that are interoperable, opened for modifications, and easily adjustable for specific problems.

Saving costs - One of the main reasons for implementing our own version of the CSMP was related to the cost of licenses for commercial CSMP and Simulink implementations. Licenses would have to be provided for more than 100 computers at the Faculty of Organizational Sciences. Compared to Matlab/SIMULINK, the CSMP/FON has almost the same features. However, we made the CSMP/FON free and open source; therefore, the application of this software in teaching and learning is cheap and every student can easily learn about the simulation of complex systems on a personal computer.

Simple and rich user interface - In the CSMP/FON we implemented all the newest concepts that are used in the modern software design: simplicity, clarity, responsiveness, efficiency, richness of colours and images, etc. The user interface of the CSMP/FON can be adapted and changed according to the users' needs. In addition, the architecture of the application is designed in order to enable an easy integration of CSMP/FON functionalities in the web-based simulation solutions. Although Matlab provides quite clear interface, it is much less user-friendly, its design is old-fashioned and does not follow any of the abovementioned concepts.

Suitable and easy to use for educational purpose - while Matlab/Simulink provides students with various features and possibilities, a research showed [26] that it is not suitable for beginners in the area of simulation. Using

Matlab/Simulink requires good comprehension of numerical analysis, linear algebra and many mathematical functions, which makes this solution quite complex and inappropriate for our students. The CSMP/FON abstracts mathematical issues and enables students to learn basic simulation concepts without the need to have advanced knowledge in the area of mathematics. Considering suitability of other solutions for learning simulation, for instance, in [29] the authors introduce a comprehensive object-oriented, distributed, and extendable research solution for business simulation, named DSOL. Numerous features and simulation concepts are provided. Further, this software focuses on linking simulations to business information systems, such as ERP systems and databases. The main constraint of this solution is that it requires advanced knowledge of the Java programming language. Further, GUI features are not rich enough for learning basic simulation concepts. Accordingly, the DSOL is not entirely suitable for teaching and learning simulation languages.

Minimize required hardware resources and improve speed of execution – Matlab/Simulink provides a wide range of advanced features and services, but consequently requires powerful hardware infrastructure. Further, our experience with simulation in Matlab/Simulink showed that it could be time consuming. The CSMP/FON is a lightweight software solution that minimizes hardware consumption and optimizes the time of simulation execution.

2 CSMP/FON Application

The CSMP simulation language can be classified into the group of block-oriented languages for solving systems of differential equations. Each block is specified by a set of inputs and parameters and a graphic symbol. The graphic display of elements in the general form is presented in Fig. 1. Each available element specifies the relation of three input variables e_1, e_2, e_3 and three parameters p_1, p_2, p_3. The output is a scalar whose value depends on the concrete relation f for that element $e_0=f(e_1, e_2, e_3, p_1, p_2, p_3)$.

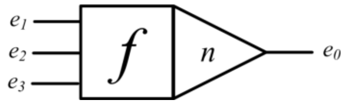


Figure 1

Graphic display of elements in the general form

The CSMP was developed by IBM in the early 1960s and it represented a real analogue simulator [15]. The language development of the CSMP II and later the CSMP III was followed by hardware development in IBM. These programmes could not be used interactively nor simultaneously from different platforms. These shortages were eliminated in an interactive implementation of simulation language CSMP/FON. The CSMP/FON implements the simulation concepts introduced by IBM's CSMP, but all the libraries and classes in our solution were implemented from the scratch. GUI elements are completely new and adapted for teaching and learning simulation. In this implementation of the CSMP/FON a user-friendly graphical user interface was developed, and the efficiency of the programme improved. The CSMP/FON is an open source solution available to the students and can be downloaded from the web site http://www.elab.rs/laboratorija-za-simulaciju/. The CSMP/FON simulation tool allows users to develop complete, interactive simulations in three steps: describing the mathematical model, building the user interface using off-the-shelf graphical elements and connecting certain properties of these elements to the variables of the model. Operating principle of the CSMP/FON is shown in Fig. 2.

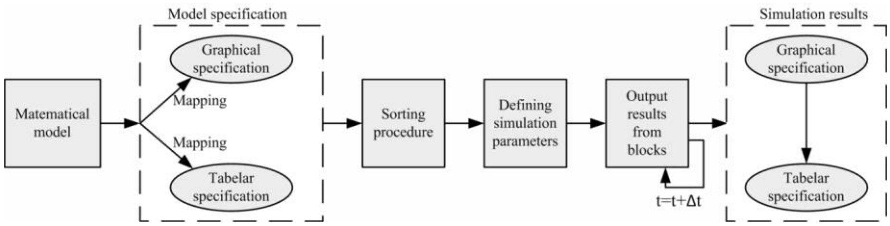


Figure 2
How CSMP/FON works

The programme's engine is based on an algorithm for sorting ordinal numeral of blocks. It gives an order needed to calculate the output values from every block and after each integration interval the output values for every block are known. We use the Runge Kutta IV method for integration.

The CSMP/FON provides an integrated and user-friendly environment for creating, testing, and analyzing continuous system models. It enables students to configure simulation models, execute simulation, and analyze results. Particular benefits from using the CSMP/FON include an improved performance of teaching activities, integration of activities in teaching and learning processes, learning simulation on a variety of real models, learning about the implementation of a continuous simulation language. An interactive environment with resident editor, processor and result analyzer, a fast and easy model debugging, different views and analyses of simulation results are some of the most important features of the application. Users are allowed to interact with a simulation, monitor values of variables, change parameter values, and rerun the simulation, redefine output requirements, etc.

2.1 CSMP/FON Architecture

The architecture and key components of the CSMP/FON application are presented in Fig. 3. The components are explained in the following text.

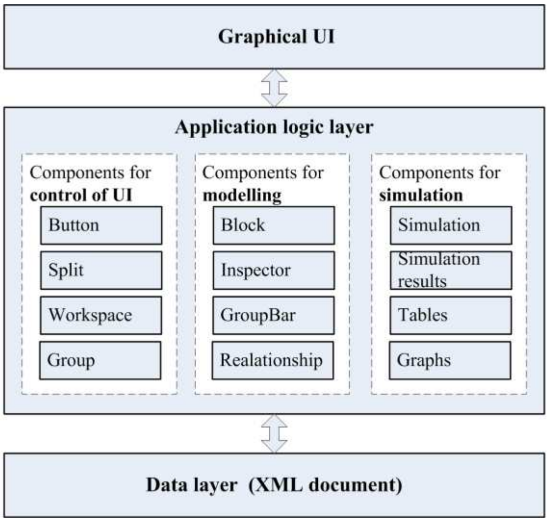


Figure 3
Conceptual model of CSMP/FON software solution

The architecture of the CSMP/FON solution is based on the MVC pattern (Model View Controller) [30]. Each component in the application logic layer in Figure 3 is implemented through appropriate MVC classes.

2.1.1 User Interface

A software application for learning simulation, in addition to the model, should have an interface for displaying the simulation in a proper way, and a facility to change the parameters while the model is running to provide dynamic control. The assigning of values to parameters, the specification of initial conditions, and control during the execution need to be achieved in ways that directly correspond to the concepts or procedures that are to be learned [4]. The user interface shown later in Fig. 5 is one of the key advantages of the CSMP/FON in comparison with similar solutions. In essence, the user interface of the application is a collection of graphical windows whose components are active, dynamic, and clickable. Students can change any active element in the graphical interface, the recalculation and dynamic presentation are immediate. In this way students instantly perceive how their modifications affect the model.

The graphical interface contains three basic menus: File, Edit, Help; and a toolbox with eight buttons for model manipulation. On the left side of the main screen there is a list of available blocks. Blocks are categorized in groups: Mathematical elements, Trigonometry, Generators, Limiters and Other. The user creates a model by dragging and dropping blocks from the list to the workspace, and then connects blocks using options from the toolbox. After the block is dropped, a form for input parameters of the specific block is shown. Each block has a predefined number of parameters (min 1, max 3). When the model is created, the user chooses the option for starting the simulation from the toolbox, and sets simulation parameters: the length of the simulation, the integration interval, and the interval for printing the results. The simulation results are shown in the form of tables and graphs.

2.1.2 Application Logic

The source code of the application is grouped into logical parts: components for control of user interface, components for modelling, and components for simulation (Fig. 3).

Components for control of user interface implement functionalities of the standard Windows forms buttons with some handy visual add-ons, functionalities for visual split of controls and space saving. There is a container for any type of a control that enables an easy hiding and showing of controls. A visual representation of controls can be changed during the run time. “Workspace” is a component that represents the canvas on which the model is drawn. This component handles most of the work related to visual representation and editing of the model. We use workspace as a canvas to draw CSMP blocks and relationships.

Components for modelling implement functionalities for creating and saving block diagrams. “Block” is the main component of this programme and it handles the graphical representation of any element of the CSMP language. This component is abstract and extensible, therefore programming new blocks with new properties and parameters is as simple as changing the elements of an XML file. The XML file Manifest.xml contains the properties of each block in the model and keeps the information about the visual grouping of blocks. Using this approach we achieved extensibility of the application. For each CSMP block in the model an object (named TCSMPBlock) is instanced. This object holds the properties and parameters of the block, information about all of the input and output connections, as well as the values calculated during the simulation. We can move blocks across the workspace to achieve a better representation of the model. When moving blocks, the position of all input and output connections is preserved. “Connections” is a component for managing the graphical display of connections between blocks. To achieve a “breaking” of the line, we introduced helper objects called “holders”. We pin the holder on a line and create a polyline. These holders are also used to create junctions (a connection between a relationship and a block). Junctions are used when there already is an output link from a block and we want to create another one. The CSMPInspector is a component for displaying and editing properties of the selected CSMP block.

Part of the application that handles the simulation resides in a component titled "**Simulation**". This part of the programme prepares the model for simulation, checks the model for consistency, and represents the "brain" responsible for the simulation itself. The program is implemented in Delphi on Win32 platform.

2.1.3 Data Layer

Models created using the CSMP/FON are stored in XML files. XML is widely used and it is de facto a standard in modern software architectures. The structure of this XML file (with the extension .CSMP) is shown in Fig. 4:

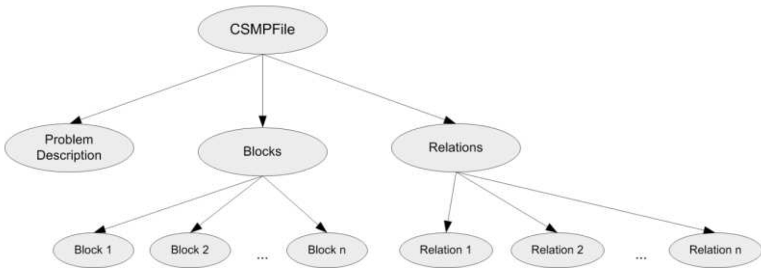


Figure 4
Structure of the CSMP file

Each CSMP model is described using the <block> XML tag. Each block is described through the following tags: <block_id>, <name>, <parameters>, <inputs>. Using the XML technology for storing models enables platform independence, reading and editing the models in any software tool that supports XML, providing compatibility of models developed using old or future versions of the application. Further, XML notation enables an integration of simulation models in web-based solutions.

3 Using CSMP/FON in Simulation of a Spacecraft Landing

3.1 Problem Description

In this problem, we discuss a simplified model of vertical landing of a spacecraft on the surface of the Moon. The total mass of the spacecraft is labelled as M and consists of two elements: the fixed mass of spacecraft (m_s) and the variable mass of fuel in the spacecraft (m_f). The acceleration of the spacecraft can be represented with the well-known formula shown in Equation (1).

$$M \cdot a = -\frac{g\left(m_s + m_f\right)}{\left(1 + s/r\right)^2} + g_p * g_f \tag{1}$$

where: g – gravitational acceleration on the Moon 1.62ms^{-2}; r – Moon radius (1 738000 m); m_s – fixed mass of the spacecraft (15000 kg); m_f – variable mass of fuel in the spacecraft (initial value 1000 kg); s – distance of the spacecraft from the initial position; g_p – fuel consumption (kg s^{-1}); g_f – factor for conversion of fuel consumption in units of force (4 000 Ns kg^{-1}).

The change in the mass of fuel in the spacecraft is equal to the negative value of variable g_p that depends on time t and can be calculated using the formula (2).

$$g_p = 2 + 1{,}5t \tag{2}$$

If the mass of the fuel drops to zero, fuel consumption must also be equal to zero. The spacecraft stops when it lands on the Moon surface. In the beginning of the simulation, the distance from the Moon surface is 5000 m.

3.2 Solution

The total mass of the spacecraft M is the sum of fixed mass of the spacecraft m_s and variable mass of the fuel in the spacecraft m_f. Since the acceleration of the spacecraft can be presented as the second derivative of the distance of the spacecraft, we can transform (1) into (3). Equation (3) is suitable for creating a block diagram.

$$\ddot{s} = -\frac{g}{\left(1 + s/r\right)^2} + \frac{g_p * g_f}{m_s + m_f} \tag{3}$$

Using (3) we created a block diagram in CSMP/FON, shown in Fig. 5.

After creating the block diagram, we set up simulation parameters (integration interval and time of simulation) and perform the simulation. The simulation results are shown in graphical format (Fig. 6).

Fig. 6 shows how the distance of the spacecraft from the land changes in time. The initial position of the spacecraft is set to 4000m above the ground. It takes about 38 time units to land. After 38 time units, the simulation is over, because the quit block stops the simulation when the spacecraft touches the ground.

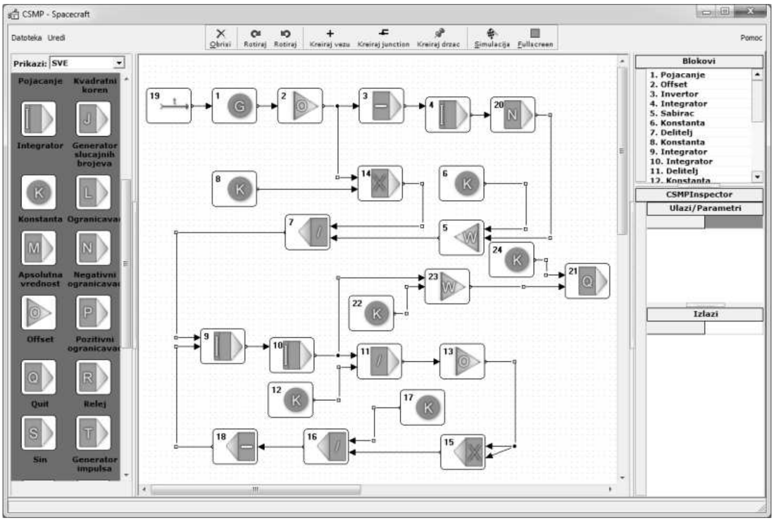


Figure 5
CSMP/FON block diagram

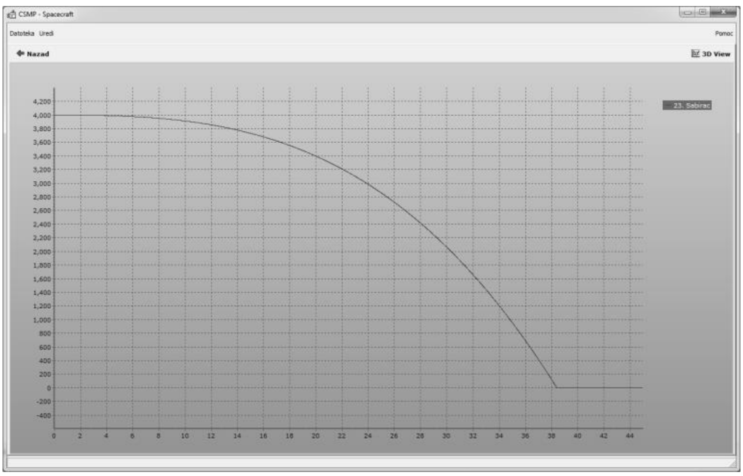


Figure 6
Simulation results showing the distance of the spacecraft

4 Evaluation

This study aims to investigate if the CSMP/FON application can be effectively used to test students’ knowledge in the area of continuous simulation and simulation languages. The experiment was conducted on a sample of 160 undergraduate students of the Faculty of Organizational Sciences, University of Belgrade. Students who attended the Simulation and simulation languages course were randomly divided into experimental (80 students) and control groups (80 students). All students, both experimental and control groups, had a block of lectures presented in the traditional way. Then they attended lab classes where they used the CSMP/FON application to simulate continuous systems. In order to measure the research results, we used the knowledge test that students take at the end of semester. In the knowledge test, students solve problems from the area of continuous systems modelling and simulation. A typical task in the test includes the following requirements: 1. Mathematical modelling of the continuous system described through a verbal model; 2. Creating a block diagram; 3. Creating a table of configuration; 4. Editing a simulation model and adjusting it to specific requirements. The test we applied in the experiment was a standard test used for testing students for more than five years. The students in the experimental group took the final test using the CSMP/FON applications, while the control group students took the final test in the standard form, i.e. on paper. A descriptive comparative statistics of results achieved on the knowledge test is presented in Table1.

Table 1

Descriptive comparative statistics of results achieved on knowledge test

	N	**Mean**	**Std. Deviation**
Experimental group	80	9.26	1.088
Control group	80	8.82	1.271

The distribution of grades in the Experimental and Control groups is shown in Table 2. Grades range from 6 (equivalent to E in ECTS grading scale) to 10 (equivalent to A in ECTS grading scale). We can see that the number of students who achieved high grades is higher in the experimental group. The results of further statistical analysis show that a larger number of students from experimental group that achieved high marks in comparison with the number of students from the control group is statistically significant, $F(1,159)=5.473$ ($p<0.05$).

Table 2

Distribution of grades in Experimental and Control groups

	10	**9**	**8**	**7**	**6**
Experimental group	49	13	9	8	1
Control group	34	18	12	12	4

The described quantitative analysis shows that usage of CSMP/FON within the exam contributed to students' results. This may be explained by the fact that students in the experimental group had a feedback during the exam, but on the other hand, this confirms that this software needs to be included as an integral part of the course. Also, a qualitative analysis is required in order to make better conclusions.

Therefore, after the test, students in the experimental group were asked to fill in a questionnaire and evaluate the CSMP/FON application. Table 3 shows the students' responses.

Table 3

Questionnaire about CSMP application quality results

(Strongly agree – 5; Agree – 4; Neutral – 3; Do not agree – 2; Strongly disagree - 1)

Question	5	4	3	2	1	Mean score	Std. dev
Using CSMP is simple and clear	70	27.5	1.25	1.25	0	4.66	0.57
User interface for developing simulation model is well designed	45	41.25	10	2.5	1.25	4.26	0.84
User interface for simulation results representation is well designed	35	45	13.75	5	1.25	4.08	0.90
The application contains appropriate information for creating a simulation model	55	37.5	7.5	0	0	4.48	0.64
Information and dialogues during testing and executing simulation models are appropriate	38.75	46.25	12.5	1.25	0	4.24	0.72
Simulation executing speed is satisfactory	70	25	2.50	1.25	1.25	4.61	0.72
I think that user interface suits me	48.75	40	7.5	1.25	1.25	4.35	0.79
The application has enhanced my interest in the area studied within the course	13.75	28.75	33.75	13.75	8.75	3.25	1.14
I have understood the course contents better after using the application	42.50	41.25	11.25	3.75	0	4.24	0.80
While using the application, I have discussed with my colleagues how to create a proper simulation model	28.75	31.25	12.50	10	16.25	3.47	1.43

Mean scores are mainly over 4, so it can be concluded that students are generally satisfied with the CSMP application. However, lower mean values indicate that the application needs to be improved in several aspects: a) the user interface for simulation results should be enhanced. This can be done by implementing better methods for zooming across results, a function for searching results, and a better design. b) The application does not enhance students' interest in the area of continuous simulation. In order to improve this aspect, we should implement some features that support game-based learning and edutainment concepts. c) The application does not encourage collaboration. In further improvements, new features for collaboration need to be implemented.

One of the most important questions was related to the role of the CSMP/FON in understanding the course contents. Since the mean score for this question was

4.24, we can conclude that the usage of this application contributes to students' understanding of continuous systems simulation and simulation languages. This can be explained by the openness and ease of use of the CSMP/FON.

One of the drawbacks of the evaluation presented in this paper is related to the lack of comparison between the results that students achieve with the CSMP/FON and the results achieved with Matlab/Simulink. Additional research is needed to compare how these two software tools affect students' knowledge. On the other hand, the impact of Matlab/Simulink and CSMP/FON on students' knowledge in simulation languages is incomparable, since the former cannot be applied for this purpose.

Conclusion

The main contribution of this work is the development of a new, efficient, and open source environment for learning continuous simulation. The CSMP/FON is a free simulation tool that improves students' understanding of the theoretical concepts of computer simulation and simulation languages. The proposed solution keeps all the existing characteristics of the CSMP language and adds new graphical user interface. The CSMP/FON supports creating interoperable simulation models. This flexible and extensible solution enables developers to implement new simulation algorithms or integration methods easily. The results of the experiment realized in order to evaluate the developed application show that the CSMP/FON application can be used effectively to learn the simulation and simulation languages concepts. This is also in agreement with the fact that the CSMP/FON has already been adopted by other faculties that have a computer simulation in their curriculum, i.e. Faculty of Transport and Traffic Engineering, University of Belgrade.

Finally, we acknowledge some limitations of this study. The CSMP/FON is a desktop application not fully integrated in the e-learning process. Therefore, future researches are directed towards the development of a web CSMP/FON application with the same features and integration of the CSMP/FON into a learning management system. A full coordination of all processes in e-learning will lead to an improved quality of the e-learning system. In addition, we plan to create a library of the most important models and sub models from the area of continuous simulation that can be simply integrated as model's components. Simulation models created in the CSMP/FON should be fully interoperable with analogue models created in other simulation languages.

Acknowledgement

This work was supported by the Ministry of Education, Science, and Technological Development of the Republic of Serbia, grant number 174031.

References

[1] J. Banks: Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice, John Wiley & Sons, New Jersey, 1998

[2] J. Banks, J. S. Carson, B. L. Nelson, D. M. Nicol: Discrete-Event System Simulation, 5th Ed. Prentice Hall, New Jersey, 2010

[3] J. Calvo, M. Boada, V. Díaz, E. Olmeda: SIMPERF: SIMULINK-based educational software for vehicle's performance estimation, Computer Applications in Engineering Education, 17(2), pp. 139-147, 2009

[4] R. Granlund, E. Berglund, H. Eriksson: Designing web-based simulation for learning, Future Generation Computer Systems, 17 (2), pp. 171-185, 2000

[5] C. François, K. Ernesto: Continuous System Simulation, Springer-Verlag, New York, 2006

[6] M. Sadiku, M. Tofighi: A tutorial on simulation of queueing models, International Journal of Electrical Engineering Education, 36 (2), pp.102-120, 1999

[7] F. Kentli, H. Çalik: Matlab-Simulink Modelling of 6/4 SRM with Static Data Produced Using Finite Element Method, Acta Polytechnica Hungarica, 8(6), pp. 23-42, 2011

[8] C. Blake, E. Scanlon: Reconsidering simulations in science education at a distance: features of effective use, Journal of Computer Assisted Learning, 23 (6), pp. 491-502, 2007

[9] M. Despotović-Zrakić, D. Barać, Z. Bogdanović, B. Jovanić, B. Radenković: Integration of web based environment for learning discrete simulation in e-learning system, Simulation Modelling Practice and Theory, 27, pp. 17-30, 2012

[10] A. Rodić, G. Mester: The Modeling and Simulation of an Autonomous Quad-Rotor Microcopter in a Virtual Outdoor Scenario, Acta Polytechnica Hungarica, 8(4), pp. 107-122, 2011

[11] M. Badida, R. Králiková, E. Lumnitzer: Modeling and the Use of Simulation Methods for the Design of Lighting Systems, Acta Polytechnica Hungarica, 8(2), pp. 91-102, 2011

[12] F.H. Speckhart, W. L. Green: A guide to using CSMP-the Continuous system modeling program: a program for simulating physical systems, Prentice Hall, New Jersey, 1976

[13] R. D. Brennan, M. Y. Silberberg: Two Continuous System Modeling Programs, IBM Sysems Journal, 6(4), pp.242-266, 1967

[14] R. Nilsen, W. Karplus: Continuous-System Simulation Languages: A State-of-the-Art Survey, Mathematics and Computers in Simulation, 16 (1), pp.17-25, 1974

[15] E. L. Edward, Mitchell: Using a Continuous Simulation Language (ACSL) to Model and Control a Hybrid Simulation, Mathematics and Computers in Simulation, 19 (2), pp.133-140, 1977

[16] J. Strauss, et al: The SCI Continuous System Simulation Language (CSSL), Simulation, 9, pp. 281-303, 1967

[17] A. Demiroren, H. L. Zeynelgil: Modelling and Simulation of Synchronous Machine Transient Analysis using SIMULINK, International Journal of Electrical Engineering Education, 39 (4), pp. 337-346, 2002

[18] The Mathworks Inc: Using SIMULINK, Dynamic System Simulation for MATL AB, 5h Ed., Mathworks Inc, USA, 1997

[19] Z. Ghassemlooy, R. Saatchi: Software Simulation Techniques for Teaching Communication Systems, International Journal o Electrical Engineering Education, 36 (4), pp. 287-297, 1999

[20] S. E. Mattsson, H. Elmqvist, M. Otter: Physical System Modeling with Modelica, Control Engineering Practice, 6, pp. 501-510, 1998

[21] D. Matko, R. Karba, B. Zupancic: Simulation and Modelling of Continuous Systems - A Case Study Approach, Automatica, 30 (11), pp. 1808-1810, 1994

[22] G. Lipovszki, P. Aradi: Simulating Complex Systems and Processes in LabVIEW, Journal of Mathematical Sciences, University of Tokyo, 132 (5), pp. 629-636, 2006

[23] H. Klee: Some Novel Applications of a Continuous System Simulation Language, Computers and Industrial Engineering, 11 (1-4), pp. 385-389, 1986

[24] M. Alfonseca, J. Lara, E. Pulido: An Object-oriented Continuous Simulation Language and Its Use for Training Purposes, Proceedings of 5th International Workshop on Simulation for European Space Programs, Noordwijk, Netherlands, pp. 49-54, 1998

[25] R. Thomas, A. Close, P. McAndrew: Tailoring Simulations for Teaching and Learning: the Potential of the MultiVerse Environment, International Journal o Electrical Engineering Education, 37 (1), pp.1-12, 2000

[26] A. Canizares, Z. Faur: Advantages and Disadvantages of Using Various Computer Tools in Electrical Engineering Courses, IEEE Transactions on Education, 40 (3), pp. 1-7, 1997

[27] M. Despotovic, A. Markovic, Z. Bogdanovic, D. Barac and S. Krco, Providing Adaptivity in Moodle LMS Courses, Educatoional Technology and Society, 15 (1), pp. 326-338, 2012

[28] M. Trenas, J. Ramos, E. Gutierrez, S. Romero, F. Corbera: Use of a New Moodle Module for Improving the Teaching of a Basic Course on Computer Architecture', IEEE Transactions on Education, 54 (2), pp. 222-228, 2011

[29] P. Jacobs, N. Lang, A. Verbraeck: D-SOL; a Distributed Java-based Discrete Event Simulation Architecture, Simulation Conference, 2005 Proceedings of the Winter, Orlando, FL, USA 1, pp. 793-800, 2002

[30] P. Sauter, G. Vögler, G. Specht, T. Flor: A Model–View–Controller Extension for Pervasive Multi-Client User Interfaces, Personal and Ubiquitous Computing, 9 (2), pp 100-107, 2005

Forecasting Rail Transport Petroleum Consumption Using an Integrated Model of Autocorrelation Functions-Artificial Neural Network

Hanialhossein Abolfazli[1], Seyed Masoud Asadzadeh[2], Salman Nazari-Shirkouhi[3,*] Seyed Mohammad Asadzadeh[4], Kamran Rezaie[5]

[1]PMP®, RMP®, Business Development Division, Enermonde AB. Signe Tillischgatan 15, Stockholm, Sweden, 16973, Solna, Email: hani.abolfazli@enermonde.com

[2]Department of Software Engineering, University of Science & Culture, Tehran, Iran. P.O. Box: 13145-871; Email : masoud.asadzadeh@hamfekrangroup.org

[3,*]Young Researchers and Elite Club, Roudbar Branch, Islamic Azad University, Roudbar, Iran, P.O. Box 14579-44616, E-mail: snnazari@ut.ac.ir

[4] School of Industrial and Systems Engineering, College of Engineering, University of Tehran, Tehran, Iran. P.O. Box 11155-4563. Email: smasadzadeh@ut.ac.ir

[5]School of Industrial and Systems Engineering, College of Engineering, University of Tehran, Tehran, Iran. Email: krezaie@ut.ac.ir , P.O. Box 11155-4563

Abstract: This paper presents the application of time-series and artificial neural network for improvement of energy forecasting in rail transport section. An integrated artificial neural network (ANN) model is presented that uses autocorrelation and partial autocorrelation functions to determine the best input variables for ANN. The proposed ANN uses autocorrelation function (ACF) and partial autocorrelation function (PACF) extracted from time-series data to select appropriate inputs for ANN. For validation of the ANN results, they are compared with two auto regressive models using analysis of variance (ANOVA) technique. Weekly total gasoil consumption in Iran railway transportation from January 2009 to October 2011 is used for constructing and comparison time-series and ANN models. Result shows that ANN with inputs extracted from ACF and PACF analysis, performs well for estimation of railway gasoil consumption. The interaction between input variables as well as non-linearity in consumption data is properly handled by the proposed integrated ANN.

Keywords: Gasoil Consumption Forecasting; Railway; Artificial Neural Network; Time-series; Autocorrelation

1 Introduction

Economic development boosts different activities amon them transport activity. One of the pressing problems related to this increasingly important economy sector, is petroleum dependence. A great portion of energy (around 95%) used in the transportation sector is supplied by petroleum. Therefore, managing the use of petroleum in this sector is crucial for secure supply of energy.

During the last decade several new techniques have been used for energy demand management and specifically for accurate prediction of the future energy needs. Suganthi and Samuel [15] reviewed the existing models of energy demand forecasting. Among them, they discussed traditional methods as well as soft computing techniques such as neuro-fuzzy, GA, and ANNs. As new techniques of energy demand forecasting they discussed Support vector regression, ant colony and particle swarm [15]. These models can be categorized into three main approaches: time-series approach [3, 9], econometric approach [7, 12,] and artificial intelligence (AI) approach [14].

This variety of techniques also can be found in the literature related to the modeling and forecasting energy demand in transportation sector. Haldenbilen and Ceylan [10] used genetic algorithm (GA) and developed three forms of the energy demand equations (linear, exponential and quadratic) to improve transport energy estimation. Their model uses population, gross domestic product and vehicle-km as inputs. Forouzanfar et al. [4] proposed a multi-level genetic programming approach for demand forecasting of transport energy in Iran. It was shown that multi-level genetic programming approach outperforms neural network and fuzzy linear regression approaches for transport energy forecasting. Al-Ghandoor et al. [1] used ANFIS-double exponential smoothing to model and forecast the demand of transport energy in Jordan. Petrol and diesel consumption in the transpor sector of China is analyzed by Chai et al. [2]. They used the Bayesian linear regression integrated with Markov Chain Monte-Carlo method to establish a demand-forecast model.

AI techniques are increasingly diversifying today and among them Artificial Neural Networks have gained special attention in energy forecasting. Geem and Roper [6] developed an ANN model with four independent variables of *i*) gross domestic product (GDP), *ii*) population, *iii*) import, and *iv*) export to forecast transport energy demand. They showed that ANN performs better than linear regression model or an exponential model. Geem [5] developed ANN models to forecast South Korea's transport energy consumption and demonstrated that ANN produced more robust results. Murat and Ceylan [11] illustrates an ANN approach for the transport energy demand forecasting using socio-economic and transport related indicators.

Generally, econometric models which take into account the socio-economical variables are more powerful than time-series models. However, it should be noted that econometric models need accurate data regarding its explanatory variables (e.g. production data, stock exchange market data, etc). In some countries, especially in developing countries accessing to accurate and reliable data is limited. Moreover, for forecasting purposes, the problem of forecasting the explanatory variables (e.g. production data, stock exchange market data, etc) is central and in fact an unsolved sub-problem for the main problem of forecasting the dependent variable (here gasoil consumption). Therefore, time-series models that directly model and forecast the dependent variable are of great interest because they don't need data of other variables. Time-series models may not capture the non-linearity and complex relationships in the modeling environment. Because of that, time-series modeling concept is integrated with ANN to improve forecasting in more complex cases.

To the best knowledge of the authors, the application of time-series integrated with ANN modeling is limited for railway transport energy forecasting and there is a research gap to study the advantages of combining time-series modeling and ANN modeling to reach improved models for railway energy forecasting.

The main objective of this paper is to use ACF and PACF of time-series data to construct ANN model to be used for gasoil consumption forecasting in rail transport sector. To this end, this paper presents a working algorithm to analyse autoregressive time series data and decide on input variables for ANN.

This paper is organized as follows. In section 2, the integrated ANN algorithm is described. Section 3 presents a case study which illustrates an experiment with the proposed algorithm. Section 4 presents the comparison, validation, and analysis of the results. Conclusions are presented in the last section.

2 Working Algorithm for ANN Modeling

In this section the working algorithm for ANN modeling is presented. At its early stages, this algorithm proposes to collect data and preprocess them using ACF and PACF. This preprocessing involves the calculation of confidence interval for ACF and PACF. With refer to Gujarati (2004), the algorithm decides on the most appropriate autoregressive moving average (ARMA) model. This decision is made based on the pattern in ACF and PACF. Table 1 provides the guidelines to choose ARMA models [8].

Moreover, significant ACFs and PACFs can be specified with reference to their confidence intervals. In the integrated algorithm of this paper, the inputs of ANN are selected as the lagged variables of the output. For modeling gasoil consumption, ANN output is the amount of gasoil consumed in time period t (Y_t). It is noted that the number of lagged variables that should be selected as ANN inputs is of question.

Table 1
How to choose ARMA model and its parameters

ACF pattern \ PACF pattern	- Exponentially decreasing	- Very big for *p* lags and cut after *p* lags
- Exponentially decreasing	ARMA (p,q)	AR (p)
- Oscillating and damping sin curve	-	AR (p)
- Very big for *q* lags and cut after *q* lags	MA (q)	-

p is the number of autoreggresive lags and *q* is the number of moving averages

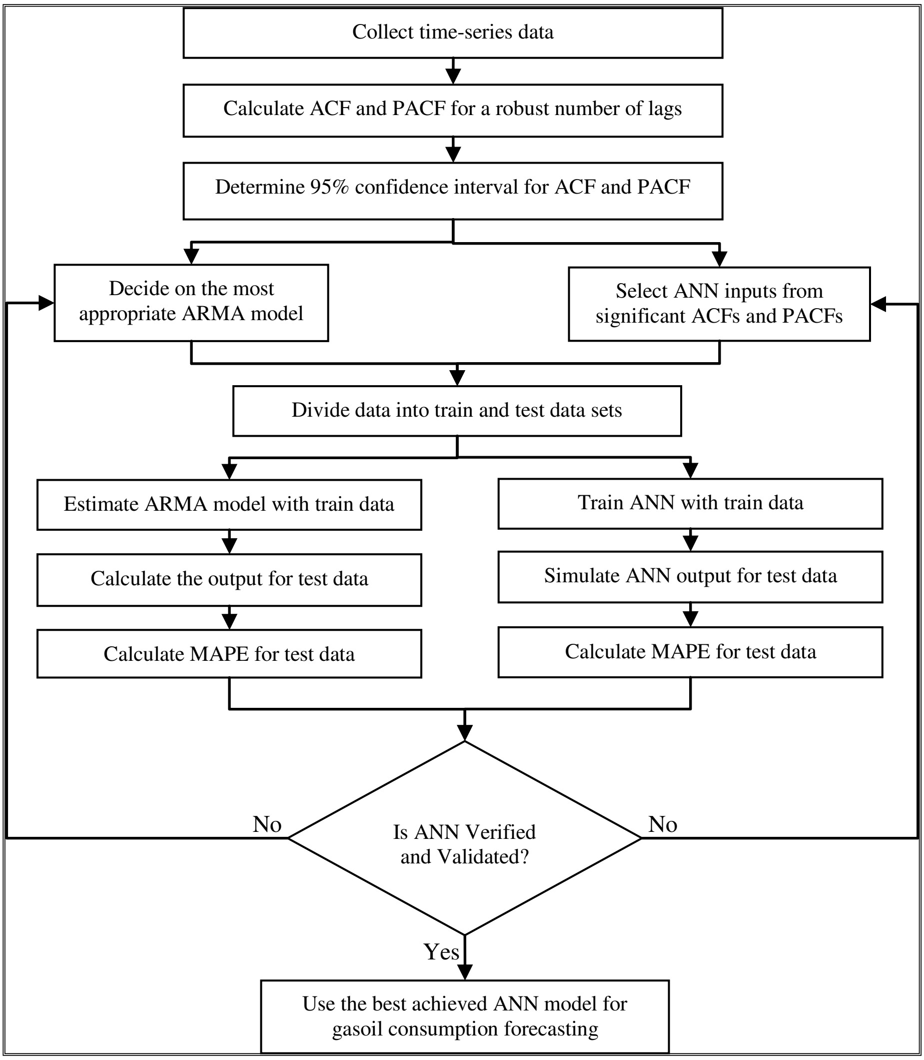


Figure 1
Working Algorithm for ANN Modeling

The inputs of ANN are specified as the lag variables which their ACF or PACF is significant. In some cases, the number of significant ACF or PACFs may be very high. In these cases, a robust number of input variables can be preset by the modeler. This procedure will determine the best inputs that could be used in ANN modeling. After that the data set is divided into two sets, train and test. With the train set, both ARMA and ANN models are trained and estimated and the performance of the models is examined with test data according to their Mean Absolute Percentage Error (MAPE). Finally, the result of ANN is verified and validated by comparison with the result of ARMA model. If the results be satisfactory, ANN could be used for gasoil consumption forecasting. Otherwise, more input variables should be used for both ANN and ARMA models. This integrated algorithm is presented in Figure 1.

2.1 Artificial Neural Network

Neural Networks can be configured in various arrangements to perform a range of analysis including function estimation and forecasting. ANNs are well suited for applications where we need to estimate a nonlinear function. ANNs consists of an inter-connection of a number of neurons. In Multi Layer Perceptron (MLP) network the data flows forward to the output continuously without any feedback. Figure 2 shows a two-layer feed forward model used for gasoil forecasting. The input nodes are the previous *m* time lagged consumptions while the output is the gasoil consumption for the current time period. Hidden nodes with linear or nonlinear transfer functions provide a basis for estimate the non-linear relation between inputs and output.

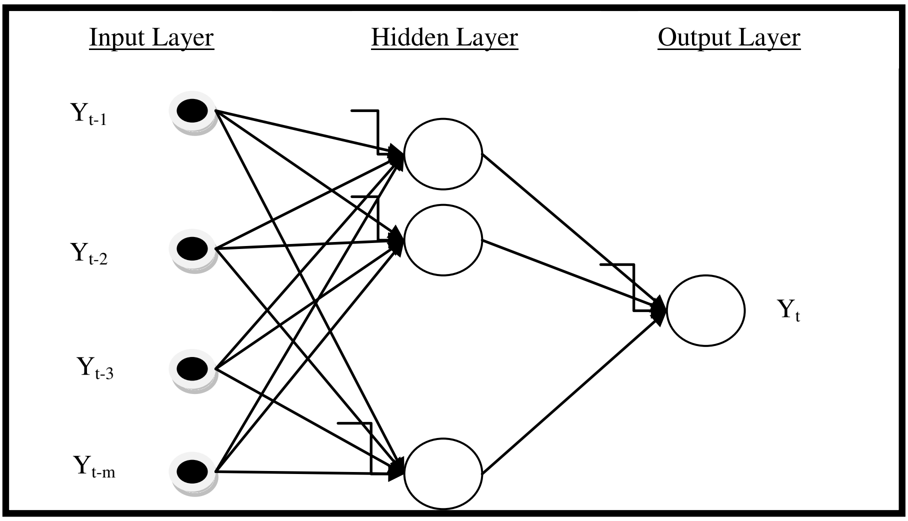


Figure 2
The structure of the ANN

The model can be written as:

$$y_t = \alpha_0 + \sum_{j=1}^{n} \alpha_j f\left(\sum_{i=1}^{m} \beta_{ij} y_{t-i} + \beta_{0j} \right) + \varepsilon_t \qquad (1)$$

where m is the number of input nodes, n is the number of hidden nodes. f is a transfer function that can be tansig with $f(x) = 2/(1+\exp(-2x)) - 1$ or logsig with $f(x) = 1/(1+\exp(-x))$. Moreover, $\{\alpha_j,\ j = 1, \ldots, n\}$ is a vector of weights from the hidden to output nodes and $\{\beta_{ij},\ i = 1, 2, \ldots, m;\ j = 0, 1, \ldots, n\}$ are weights from the input i to hidden node j. α_0 and β_{oj} are bias weights. Note that in Equation (1) a linear transfer function is employed for the output layer and this is a good choice for forecasting problems. The MLP's most popular learning rule is the error back propagation algorithm. Back Propagation learning is a kind of supervised learning introduced by Werbos [16] and later developed by Rumelhart and McClelland [13].

2.2 Dividing Data into Train and Test Sets

There are two aspects in the procedure of data division that should be noted. First, the percentage of test data (hereafter call it Test Ratio) and second, which of data rows should be selected for test (hereafter call it Test Selection). Here, Test Ratios is selected randomly between 10%, 20%, 30%, 40%, and 50%. Block selection is an appropriate selection rule for forecasting purposes. In block selection, test data are selected from the most recent data rows.

2.3 Error Estimation Method

Error Estimation Method used in this study is Mean Absolute Percentage Error (MAPE). It can be calculated by the following equation:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{x_t - x'}{x_t} \right| \qquad (2)$$

In (2), x and x' are actual and estimated data, respectively. Scaling the output, MAPE method is the most suitable method to estimate the relative error because input data may have different scales.

3 Case Study

3.1 Gasoil Consumption in Iranian Railway Transport

To show the applicability and usefulness of the proposed ANN, an experiment of rail gasoil consumption estimation is conducted. Weekly total gasoil consumption in railway transportation of Iran from January 2009 to October 2011 is collected. The time-series is shown in Figure 3.

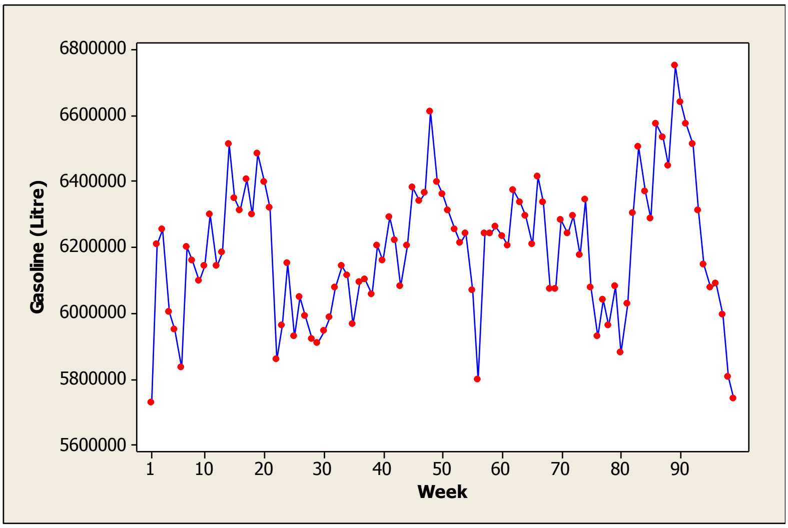


Figure 3

Weekly total gasoil consumption in railway transportation of Iran

3.2 ACF and PACF Analysis

Figure 4 shows the ACF and PACF of the gasoil consumption data. This figure also presents the lower and upper bound of 95% confidence intervals for the significance of ACF and PACF. In other words, those ACFs or PACFs which lie between these two bounds are not significant. For example the first four ACFs are significant but not the fifth ACF.

Analysis of the ACF graph shows that ACF has undergone a sin curve which is damped for lags after 15 (Y_{t-15}). For all lags, except for the first one, PACF has damped and the majority of partial autocorrelations are not significant or if significant, they are very close to the boundaries. Therefore, with respect to the guidelines in Table 1, this suggests that an autoregressive model AR(4) is the best time series for the data series under study.

3.3 Specifying ANN Input Variables

According to the significant ACFs, we can see that lagged variables 1, 2, 3, 4, 8, 9, 10, 11, 12, 13, and 14 are significant and can be used as ANN inputs. However, to have a robust number of input variables we select the first four variables as ANN inputs. Therefore, four lagged variables Y_{t-1}, Y_{t-2}, Y_{t-3}, and Y_{t-4} are selected as ANN inputs. In other words, gasoil consumption in 1 week before (Y_{t-1}), 2 weeks before (Y_{t-2}), 3 weeks before (Y_{t-3}) and 4 weeks before (Y_{t-4}) have been selected as ANN inputs where the gasoil consumption in the current week (Y_t) is the output.

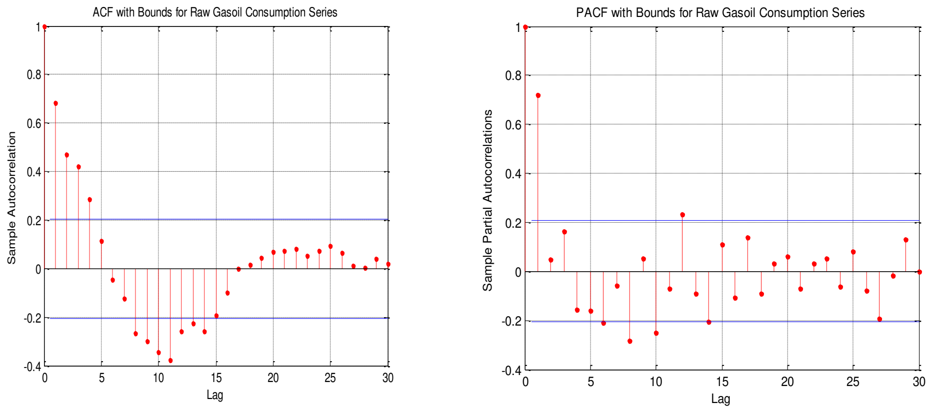


Figure 4
ACF and PACF for Gasoil consumption time-series

3.4 Auto-Regressive (AR) Modeling

Two autoregressive models are constructed for modeling gasoil data. The first model is a autoregressive model with 4 lags, AR(4), because the first four ACFs are the greatest of all. AR Model 1 is as equation (4)

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \varepsilon_t \quad (3)$$

The second autoregressive is an extension of AR Model 1 in which the interactions between every pair of lags are considered as explanatory variables. This formulation allows us to model the non-linear effects of interaction between variables.

$$\begin{aligned} Y_t = {} & \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \beta_4 Y_{t-4} + \\ & \gamma_{12} Y_{t-1} Y_{t-2} + \gamma_{13} Y_{t-1} Y_{t-3} + \gamma_{14} Y_{t-1} Y_{t-4} + \gamma_{23} Y_{t-2} Y_{t-3} + \\ & \gamma_{24} Y_{t-2} Y_{t-4} + \gamma_{34} Y_{t-3} Y_{t-4} + \varepsilon_t \end{aligned} \quad (4)$$

4 Results and Analysis

4.1 ANN Results

The ANN model is used for estimation of weekly gasoil consumption and shows satisfactory results in terms of MAPE for test data. ANN model is trained with two different hidden transfer functions tansig and logsig (see section 2.1). Another parameter that is suspected to affect ANN performance is the number of neurons in the hidden layer. Different number of hidden neurons between 10 and 50 are considered to find the best ANN. Tables 2 and 3 show the results of MAPE for these ANNs and the best number of neurons in the hidden layer.

Table 2
Test MAPE of ANN with tansig hidden transfer function

Percent of test data	Hidden neurons					Best MAPE
	10	20	30	40	50	
10	2.2%	2.2%	2.5%	2.2%	1.8%	1.8%
20	2.3%	2.3%	2.3%	2.4%	2.5%	2.3%
30	2.0%	2.6%	2.1%	2.5%	2.5%	2.0%
40	2.1%	2.1%	2.2%	2.2%	2.2%	2.1%
50	2.3%	2.4%	2.4%	2.3%	2.3%	2.3%

Table 3
Test MAPE of ANN with logsig hidden transfer function

Percent of test data	Hidden neurons					Best MAPE
	10	20	30	40	50	
10	2.1%	1.4%	2.0%	1.6%	1.7%	1.4%
20	2.4%	2.3%	2.4%	2.6%	2.4%	2.3%
30	2.5%	2.6%	2.5%	2.2%	2.4%	2.2%
40	2.3%	2.2%	2.2%	2.3%	2.2%	2.2%
50	2.1%	2.3%	2.2%	2.3%	2.3%	2.1%

Figure 5 shows the actual and estimated gasoil consumption by ANN-tansig for train and test data when 30% of all data are used as test data. The MAPE error is calculated as 2.5%.

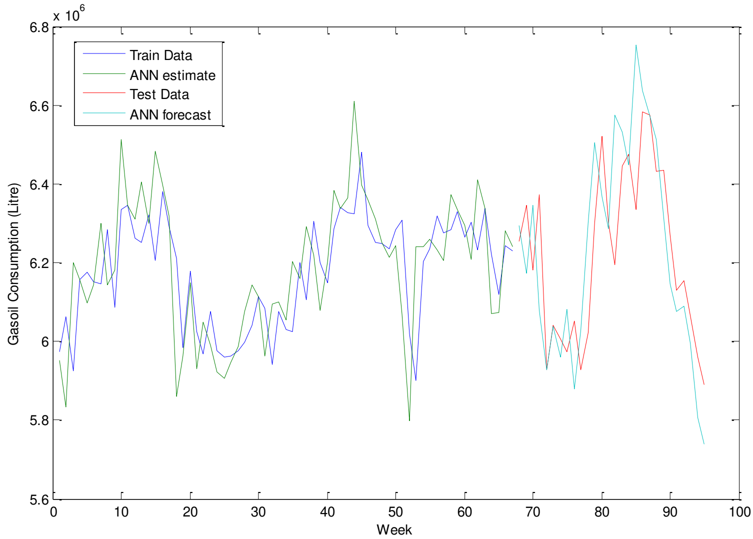


Figure 5

Actual and estimated gasoil consumption by ANN

4.2 AR Model Results

Two autoregressive models 1 and 2 are estimated with use of train data and then the estimated models are used to predict gasoil for test period. For these AR models the percentage of test changes from 10% to 50%. The results of MAPE are reported in Table 4.

Table 4

Test MAPE of ANN with logsig hidden transfer function

Percent of test data	Model 1 (MAPE)	Model 2 (MAPE)
10	2.3%	2.5%
20	2.3%	2.4%
30	2.3%	2.3%
40	2.0%	2.0%
50	1.9%	2.0%

4.3 ANN Verification and Validation

ANN results are verified with respect to the low value for average MAPE. The average MAPE of different ANN models is 2.1% that is a satisfactory result for rail gasoil consumption forecasting. To have a robust validation, we used a analysis of variance (ANOVA) technique to test the difference between ANN and AR results. Since the percentage of test data could affect MAPE, we design a complete block design with four techniques: ANN-tansig, ANN-logsig, AR model 1, and AR model 2.

The results of ANOVA for this experiment are presented in Table 5. The null hypothesis for this ANOVA is:

H0: the MAPE for all techniques is the same

H1: at least two techniques have different MAPEs

The P-value is the risk of rejecting the null hypothesis *H0*. Because this risk is very high, 71%, we cannot reject *H0* and conclude that the MAPE for all techniques is the same. This conclusion serves as a validation for the ANN results.

Table 5
Two-way ANOVA: MAPE versus Technique, test ratio

Source	DF	SS	MS	F_0	P-Value
Technique	3	0.0000066	0.0000022	0.47	0.710
testratio	4	0.0000200	0.0000050		
Error	12	0.0000564	0.0000047		
Total	19	0.0000830			

Conclusion

This paper presented an integrated autoregressive-ANN algorithm to improve gasoil consumption estimation and forecasting in rail transportation sector. The proposed ANN uses autocorrelation function (ACF) and partial autocorrelation function (PACF) extracted from time-series data to select appropriate inputs for ANN. ANN results are assessed according to an error index namely mean absolute percentage error (MAPE). ANN results are validated by comparison with the results of two linear and non-linear auto regressive models. It was concluded that ANN provides satisfactory results hence it can be used for forecasting purposes. This is a unique study that introduces the application of ACF and PACF analysis for defining ANN inputs in time series modeling. Due to its mechanism, the integrated ANN of this study is capable of dealing data correlation, autocorrelation, complexity, and non-linearity.

References

[1] Al-Ghandoor, A., Samhouri, M., Al-Hinti, I., Jaber, J., Al-Rawashdeh, M., Projection of future transport energy demand of Jordan using adaptive neuro-fuzzy technique, Energy, 2012, 38 (1), pp. 128-135

[2] Chai, J., Wang, S., Wang, S., Guo, J., Demand forecast of petroleum product consumption in the Chinese transportation industry, Energies, 2012, 5 (3), pp. 577-598

[3] Forouzanfar M., Doustmohammadi A., Bagher Menhaj M., Hasanzadeh S., Modeling and estimation of the natural gas consumption for residential and commercial sectors in Iran, Applied Energy, 2010, 87, pp. 268–274

[4] Forouzanfar, M., Doustmohammadi, A., Hasanzadeh, S., Shakouri G, H., Transport energy demand forecast using multi-level genetic programming, Applied Energy, 2012, 91 (1) , pp. 496-503

[5] Geem, Z.W., Transport energy demand modeling of South Korea using artificial neural network. Energy Policy, 2011, 39, 4644–4650

[6] Geem, Z.W., Roper, W.E., Energy demand estimation of South Korea using artificial neural network. Energy Policy, 2009, 37, 4049–4054

[7] Gorucu, F. B. , Gumrah, F. 'Evaluation and Forecasting of Gas Consumption by Statistical Analysis', Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 2004, 26(3), pp. 267-276

[8] Gujarati D.N., Basic Econometrics, Fourth Edition, McGraw−Hill Companies, 2004

[9] Gutierrez, R. A. Nafidi, R. Gutierrez Sanchez, Forecasting total natural-gas consumption in Spain by using the stochastic Gompertz innovation diffusion model, Applied Energy, 2005, 80, pp. 115–124

[10] Haldenbilen, S., Ceylan, H., Genetic algorithm approach to estimate transport energy demand in Turkey. Energy Policy, 2005, 33, pp. 89–98

[11] Murat, Y.S., Ceylan, H., Use of artificial neural networks for transport energy demand modeling. Energy Policy, 2006, 34, pp. 3165–3172

[12] Primoz Potocnika, Marko Thaler, Edvard Govekar, Igor Grabeca, Alojz Poredos, Forecasting risks of natural gas consumption in Slovenia, Energy Policy, 2007, 35, pp. 4271–4282

[13] Rumelhart D.E., McClelland J.L., Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Foundations, MIT Press, Cambridge, MA. 1986

[14] Shakouri H., G., R. Nadimia and F. Ghaderi, A hybrid TSK-FR model to study short-term variations of the electricity demand versus the temperature changes, Expert Systems with Applications, 2008, 36, pp. 1765-1772

[15] Suganthi, L., Samuel, A.A., Energy models for demand forecasting - A review, Renewable and Sustainable Energy Reviews, 2012, 16 (2) , pp. 1223-1240

[16] Werbos P.I., Beyond Regression: new tools for prediction and analysis in the behavior sciences. Ph.D. Thesis, Harvard University, Cambridge, MA. 1974

New Test Results in Cycloid-Forming Trochoidal Milling

István Szalóki, Sándor Csuka, Sándor Sipos

Óbuda University, Donát Bánki Faculty of Mechanical Engineering
Népszínház u. 8, H-1081 Budapest, Hungary
szaloki.istvan@bgk.uni-obuda.hu, csuka.sandor@bgk.uni-obuda.hu,
sipos.sandor@bgk.uni-obuda.hu

Abstract: The different types of trochoidal milling are widely used in order to produce grooves and slots. These procedures can be purchased as a built-in option of modern controls or they can be generated by modelling the movements of the tool path. In addition to the advantages, arising during machining (lower load on the milling tool and machine, the possibility to apply increased cutting parameters, raised productivity and tool durability) there is a disadvantage, to be mentioned: it is the forming of laces on the machined surface. In the present paper our trials will be introduced where a modelled cycloid-forming milling method is used to determine the effect of the cutting data, performed on the result parameters (force effects, accuracy, surface microgeometry). The tests were executed systematically, with the help of Experiment Design. Based on the results, gained during the tests, the force, accuracy and microgeometrical limits can be determined, being characteristic to this machining type.

Keywords: slot milling; semicircular milling; zigzag milling; trochoid milling; productivity; force effects; microgeometry

1 Introduction

The slot (or groove) milling is used in the production of axis symmetric parts to form torque transmitting surfaces. Based on the accuracy of the machining, there are operations of general and increased accuracy. The slots, acting as necessary for the cirlclip of bearings, match the firstly mentioned group, while the slots of fixed and movable locking connections, having a tight tolerance (more precise than tolerance class IT9), belong to the second type [1]. The necessary depth of groove of some components, machined with groove milling, may achieve or even exceed the value of the tool diameter (a_p>d). In this case the machining may be a roughing operation of a groove, used as a part of a large-sized cog-wheel or a surface for fixing blades in a turbine. The so-called one-step general milling is inaccurate due to the fact that the tool contacts the workpiece with a very long edge section, as a result of it, an increase in the cutting force and temperature can be noticed. The momentaneous chip thickness causes an unsteady load on the edge; the disadvantageous chip transport often leads to chip re-cutting. Although general milling is usually a productive operation, it is accompanied with unfavourable tool edge durability and increase of the machining costs. The earlier described unfavourable phenomena may lead to even more disadvantageous results in case of hard-to-machine materials.

We use the expression of trochoidal milling if the tool path contains the combination of circle and straight line, or it is cycloid-like; the width of the groove (bw) is made with a milling tool, having a diameter of at least 1.15 x d, the value of the radial step-over is between 0.02…0.25 x d [2]. The trochoidal movement cycle can be considered as a type of cylindrical milling, the realizable accuracy and the surface roughness (waviness and roughness) is subject to the applied strategy. The radial depth of cut (a_e) is continuously changing during the machining, the radial step over, indicated during the design process of the tool path, and influences the momentary value of the maximal radial contact (a_{emax}) to a great extent. From the aspect of the process reliability it is essential to have exact information of this value.

According to the literatures [2-4], the trochoidal milling methods have the following benefits:

- it is effective in case of high strength pieces, or even in case of hardened workpieces with small wall thickness,
- the axial depth of cut is relative great (its typical value is in the range of 10 and 20 mm), while the radial depth of cut is smaller as usual,
- the productivity can be raised considerably by applying increased cutting speed and feed values,
- the tool life and the surface roughness are generally favourable,
- the programmed milling strategies prevent the machining system from damage,
- the utilization and the productivity of the devices will increase.

Without doubt, an intensive acceleration and deceleration is demanded from the machine tool, while a look-ahead function is required from the control in case of application of this milling type. The rigid connection of the tool clamping system and the possible smallest overhang of milling cutter are also important facts, to be considered.

This research topic was developed with the ***purpose*** to determine the dependence of the characteristics, to be examined via instrumental measurements, on the cutting parameters with analysing the precisily modelled cycloid-forming trochoidal milling for the determination of the force effects and the productivity; moreover, the development of the quality of the grooves and of the surface roughness were also analysed.

1 Modelling the Cycloid-Forming Trochoidal Milling

The cycloid-forming milling is a built-in option in several modern CNC controls, but this application is control specific and the activation of the movement cycles means huge costs. Also for this reason it is reasonable to create a uniform program, enabling the possibility to serve most controls. Microsoft Office Excel application was used to program the tool paths. The description was prepared in a parametric way, so the re-calculation of the tool path points is automatically made if there is a change in any of the initial data (d, b_w, w etc.); after that the NC program, written in ISO programming language, is created in a form, being easily adjustable to the applied control [3].

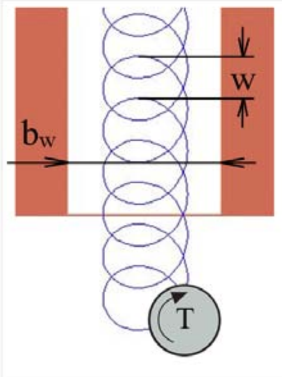


Figure 1
Cycloid-forming milling

The tool movement paths of the cycloid-forming milling strategy are shown in Figure 1. The tool is controlled on a path, having a looped cycloid form; it is really advantageous from the aspect of the tool load.

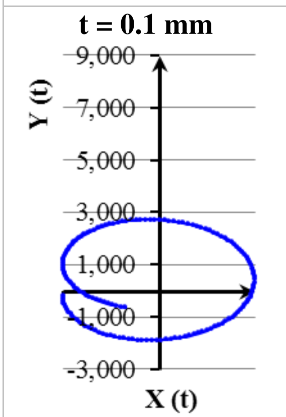
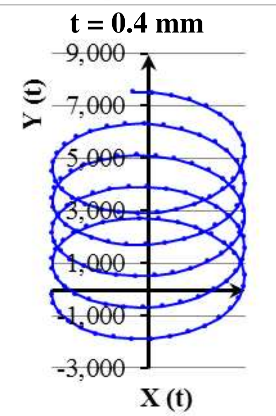
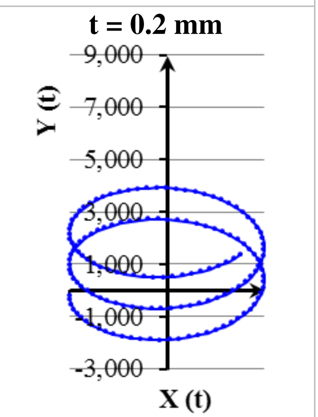
<table>
<tr><td>Too much points</td><td>Insufficient number of points</td><td>Sufficient number of points</td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
</table>

Figure 2
Possible versions of the path accuracy factor (t)

The most important characteristic is the length of the chord (t) to create a cycloid path; it is called as the path accuracy factor. This feature means the length of all lines, forming the generated path, and it determines the accuracy, „angularity" of the path. Its optimal value is between t = 0.1…0.3 mm (Figure 2). If this value is greater than 0.3 mm, then it may lead to unfavourable wear and to a decreasing tool life because of the uneven load. If the chord length value is smaller than 0.1 mm, then the model contains too many generated points and it may lead to problems when reading, storing and processing data. Thus the tool path generation was made with a value of t = 0.2 mm in our tests [5].

2 The Research of the Cycloid-Forming Trochoidal Milling (CTM)

The most important ***aim*** of the tests was to analyse the effect of the cutting parameters and to answer the question how the test conditions affect the features (for example, force effects), developing during the trochoidal milling, and how they may influence the result parameters (accuracy, microgeometry), being in connection with product quality. Considering the fact that the cutting speed value was kept on a constant level, so the number of data, to be set, was reduced to three; in this way the Design of Experiment (DoE) contains the analysis of three variables (a_p, a_{em}, f_z), varying them on three (minimal, medial and maximal) levels. From the possible 27 settings only 6 were applied, in this way material, tool, time and, of course, money could spared without losing important information. With the consistent execution of DoE (shown in Table 1) appropriate mathematical models were established, being able to describe the probable results of cutting operations even in case of not measured settings. This fact was verified by executing two check tests (7E and 8E).

The test were executed in the workshop of Donát Bánki Faculty of Mechanical and Safety Engineering of Óbuda University, on a vertical machining center (type: MAZAK Nexus 410A, control: Mazatrol™), which is installed to carry out tool test experiments, equipped with a device for blowing cold, compressed air. The force examinations were simple by using a dynamometer (type: Kistler 9257) and by *DynoWare*™ software to evaluate the results. The microgeometrical results were gained by using *Mahr-Perthen Perthometer PRK* Concept-2D, 3D surface analysing device.

Workpieces with dimensions of 160×50×50 mm were used to carry out the tests, they had the following parameters: grade 40CrMnMo7 (DIN standard W. Nr. 1.2391), pre-tempered to approx. 1000 MPa and its hardness was always verified (HB285±5).

The monolith cemented carbide HPC shank milling cutter (type: R215.36-12060-AC26L, Figure 3), having a diameter of ø12 mm, was provided by SANDVIK Magyarország Kft.

- fine-grained cemented carbide;
- modern TiAlN coating;
- λ_s = 60° helix angle

Figure 3 HPC shank milling tool (CoroMill® Plura)

During the tests, the tool was clamped in a cold shrink adapter, with EMUGE-FRANKEN clamping system (PGR „powRgrip®”). The tool overhang was 40 mm; the run-out of the milling cutter was 0.01 mm.

The applied settings and test conditions are shown in Table 1. The feed (F_f, N) and axial force (F_z, N) components were measured with the force measuring cell. The tests were completed with the following result parameters: time and productivity analysis, accuracy examination and roughness analysis.

Table 1 Machining conditions

Constant values	Cutting speed, v_c, m/min	Rpm, n, min^{-1}	l_w, mm	b_w, mm	Cooling
	~ 120	3180	45	16	ColdAirGun™

Changing conditions	COMMON VALUES			1ST BLOCK		2ND BLOCK	
	O. Nr.	**a_p, mm**	**a_{em}, mm**	**f_z, mm**	**v_f, mm/min**	**f_z, mm**	**v_f, mm/min**
	1.	9	3.2	0.1	477	0.14	668
	2.	9	3.9	0.08	382	0.12	573
	3.	12	3.2	0.12	573	0.16	764
	4.	12	3.9	0.1	477	0.14	668
	5.	12	4.5	0.08	382	0.12	573
	6.	16	4.5	0.12	573	0.16	764
	7E.	9	4.5	0.12	573	0.16	764
	8E.	16	3.9	0.1	477	0.14	668

where: l_w – groove length; b_w – groove width; a_p – axial depth of cut, a_{em} – maximal contact width; f_z – feed per tooth; v_f – feed rate

2.1 The Force Analysis and Modelling of the Cycloid-Forming Trochoidal Milling (CTM)

Based on the three components, registered by KISTLER force measuring cell and processed by Dynoware software, the minimum/average/maximum values can be determined. Analysing the high resolution (ZOOM) registratums of force sample values (Figure 4), the following conclusions can be drawn:

- According to the axial force (Fz) sample, the tool is continuously trying to lift up the workpiece, it is confirmed by the negative value range as well (it is shown on the upper part of Figure 4). This can be traced back to the very high helix angle of milling cutter. At the moment before the tool tooth leaves the workpiece, the force value becomes positive, it means that the clamping down effect becomes dominant (see „Max”). Considering the fact that the greatest part of the forces has a lifting character, therefore the positive values may cause just a small distortion when calculating the average values („Mean”). The values of six settings describe the effect of the cutting parameters, performed on the axial force components with quite certainty when processing the results of the trials, executed based on the Design of Experiment.

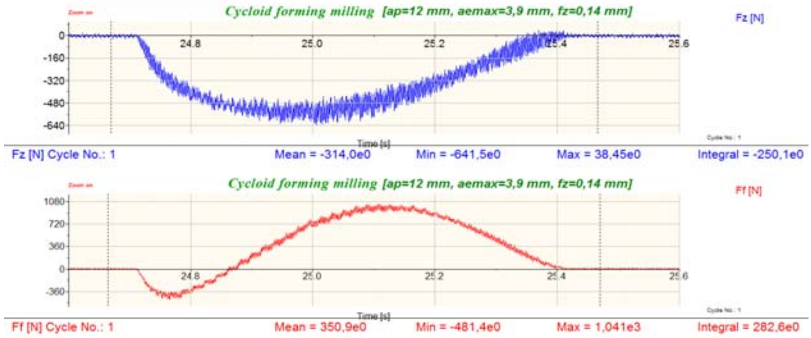


Figure 4
Development of force demand, being characteristic to the milling process

- Analysing the feed force samples (the lower part of Figure 4) it can be noticed that the tool firstly is trying to pull (-), after that trying to push (+) the workpiece. Considering the fact that the negative values play a dominant role in the force development therefore they have a significant distortion effect when calculating the average values („Mean”). Processing the results of the trials, executed based on DoE, the values of six settings *are not able to describe* the effect of cutting parameters, performed on the feed force components, with enough certainty therefore it may be necessary to carry out the test for the seventh setting, planned originally as a check test.

The force effect results, measured in case of the planned settings, are processed with a three-factor regression program (HAFARE), and the adequacy of the proposed power function model was verified, with the help of two check tests.

It is known that the productivity can be raised by increasing the feed rate per tooth (f_z) in case of constant cutting speed (v_c) and allowance values ($a_p \times a_{em}$). Therefore it is necessary to set values of the second block (with increased productivity) beside the first one (with lower efficiency). Having executed the trials in „accidental” sequence we had the results of 2x6=12 settings. The program flow is not presented due to space limitations. The describing functions of the average force (Eq. (1) and (2)) components are shown in Figure 5.

Ff_{ave}	$\overline{F}_f = 15.86 \cdot a_p^{0.84} \cdot a_e^{1.28} \cdot f_z^{0.56} \quad [N]$ Dispersion= ±11 N; Determining index (R^2) = 0.979; Correlation = 0.9894	(1)
Fax_{ave}	$\overline{F}_{ax} = -9.28 \cdot a_p^{1.13} \cdot a_e^{1.5} \cdot f_z^{0.88} \quad [N]$ Dispersion = ±7 N; Determining index (R^2) = 0.994; Correlation = 0.9973	(2)

Figure 5

Established models of average values of F_f and F_{ax}

The settings Nr. 7E and 8E (i.e. the check tests see in Table 1) confirmed a quite good correspondence and the adequacy of the chosen mathematical model, especially in case of axial force effects. Illustrating the modelled and measured (marked with star) values of the average force components in function of the feed per tooth (Figure 6), it can be seen that the measured values show a very good correlation with the calculated values of Figure 5. Analysing the diagram it can be stated that the double increase of the milling process productivity (from f_z = 0.08 to f_z = 0.16 mm) increased the average values of feed force only by 30 percent, while the growth is 40-50 percent in case of axial force. Based on the steady chip detachment, observed during the milling tests, and on the decreasing tendency of the noises, we can declare that the increase of the feed value – despite of the increasing force components – had a positive effect on the dynamic process and on the phenomena in connection with the vibrations.

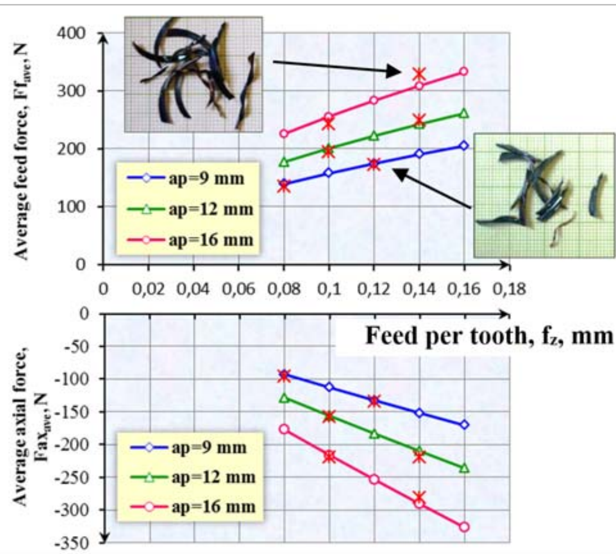


Figure 6

Calculated and measured force values

(Cutting conditions: w=1.6 mm, a_{emax}=3.9 mm)

During the general milling tests, executed parallel with these tests (cutting conditions are: a_p=16 mm, f_z=0.1 mm), even fivefold greater average feed and axial force values were measured; furthermore, much more intense vibrations and ear-splitting noises were observed. If only the discolouration and deformation of the detached chips were compared, then the conclusion that the heat deformation of the trochoidal milling process is more favourable, compared to the traditional (general) groove machining, could be drawn.

The approach to the limit load occurred in case of the sixth setting (see Table 1).

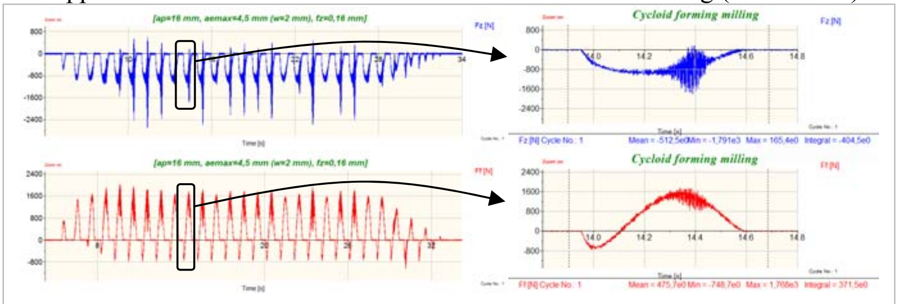


Figure 7

Development of the force demand in case of max-max-max setting

In case of setting the maximal values (Get the maximum from everything!) intense vibrations developed during the tests. It is confirmed not only by the registered force development curve (Figure 7), but by the vibration marks on the side walls and on the bottom of the machined groove as well. Keeping the further usability of the tool in mind, it is recommended not to apply these cutting parameters.

2.2 The Productivity and Accuracy of the Cycloid-Forming Trochoidal Milling (CTM)

The material removal rate (MRR, V′ cm^3/min) is used mainly to characterise the productivity: it establishes a clear order among the different types of trochoidal and similar milling [4]. From the aspect of the productivity it is determinant in case of CTM movement cycle that approx. 50 percent of the tool path can be considered as dead time (it means the tool does not work at this time), therefore it is very strongly recommended to minimise down time. Maybe a good solution is to use the so-called parametrical programming, dispersing more and more: in this case the dead-time lengths can be covered by the tool, applying an increased feed.

The function connection of the material removal rate can be determined by the formula (3) with a low dispersion (±0.17 cm^3/min) and a very good correlation (0.9995):

$$V' = 1.11 \cdot a_p^{\,0.98} \cdot a_e^{\,1.53} \cdot f^{\,1.03} \quad \left[\mathrm{cm^3/min}\right] \tag{3}$$

Analysing the curve of function it can be seen that there is a (nearly) linear connection between the axial depths of cut and feed per tooth from the cutting parameters, while the greatest effect is performed by the length of radial contact. It is logical as the radial step over (w) is the base when programming the trochoidal tool path, the radial depth of cut (a_e) is just a consequence of the step over. The step over has a tight correlation with the tool diameter and groove width, from all of these parameters the maximal value of the radial depth of cut (a_{emax}) can be determined.

Table 2

The maximal values of the radial depth of cut (a_{emax})

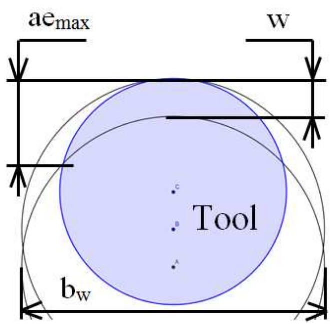
The radial step overs and the calculated maximal values of radial depth of cut			
w = 1.2 ($a_{emax} \approx 3.2$)	w = 1.6 ($a_{emax} \approx 3.9$)	w = 2 ($a_{emax} \approx 4.5$)	
The maximal values of the radial tool load, expressed in percent			
27%	33%	38%	

Table 2 shows the maximal values of radial tool load in function of the applied step overs, i.e. it is shown how great is the percentage of the tool diameter, being in contact with the workpiece in the moment when the tool has the greatest contact with the material.

During the tests not only the productivity, but the accuracy variantions of the machined grooves were measured as well. The following statement is valid for all trochoidal milling versions: the tool path is curved (it is well demonstrated in Figure 2), as its result, on the side wall of the groove „laces" develop and it may lead to dimensional fault. As its consequence, these milling methods may be used only as roughing machining. The dimensional fault always means a size, being smaller than the groove width; it depends mainly on the step over and on the tool diameter. In order to study the degree of the differences, the accuracy of the machined grooves, prepared according to the settings of design of experiment, was measured by a calibrated sliding caliper with a resolution of 0.01 mm, produced by Mitutoyo and equipped with a dial gauge. Figure 8, drawn based on the gained results, shows the deviation from the nominal groove width of 16 mm in function of the radial step over. (The calculated maximal values of radial contact are indicated in round brackets, next to the values of step over.)

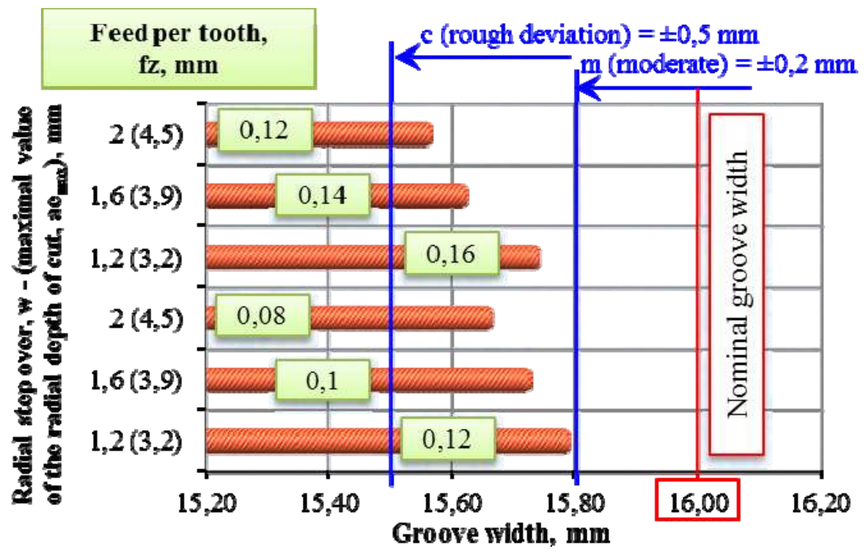


Figure 8
The width values of the machined grooves (a_p=12 mm)

The measured values, shown in Figure 8, can be classified into tolerance classes; they belong to the class „c" (rough deviation) and „m" (moderate). Analysing the diagram, it can be stated that during the trials grooves, having „c" tolerance class, were produced. From the tendency it became clear that by optimising data settings an accuracy of ±0.2 mm may be achieved, belonging to tolerance class „m".

Processing the inaccuracies of the machined grooves (it is the double value of the degree of the lace development) by a three-factor regression model, a very low dispersion (±0.02 mm) and an outstanding good correlation (0.98) were received. The mathematical model is the following:

$$\Delta b_w = 0.3 \cdot a_p^{\ 0.31} \cdot w^{1.3} \cdot f_z^{\ 0.65} \quad [\mathrm{mm}] \tag{4}$$

It means that the fault degree is influenced by the radial step over to the greatest extent, while the feed rate per tooth and the axial contact have only a degressive effect. Based on the model a clear picture may be drawn which from the applied cutting data and to what degree is „responsible" for the calculated and measured inaccuracy. Starting from this information and applying the equation (4), describing inaccuracy, the grooves may be milled also with CTM technology within the prescribed fault limitation [6]. A report on the data generation, executed with multipurpose optimisation, and its practical tests will be presented in our next paper.

Based on the experiences, gained from the tests, there is an absolutely different situation in case of general milling. This finishing technology, used to produce more accurate grooves, generates groove width that is greater than prescribed. For example, by applying machining parameters of $d=a_p=12$ mm and $f_z=0.1$ mm, grooves were made – one after another – with a width of 16.3 mm. It can be clearly explained by the fact that even a solid carbide tool runs „tottering" due to the high radial load.

2.3 The Microgeometrical Analysis of the Cycloid-Forming Trochoidal Milling (CTM)

In the previous chapters of the present paper the reason of lace formation and the effect, performed by the adjusted data, were introduced. The high-resolution presentation of waviness and roughness of the laces and their characterisation will follow in this section. During the measurements, executed on a Perthen-Mahr large-scale device, with tracing length (Lt) of 17.5 mm, measuring cut-off wavelength (cut-off, Lc) is 2.5 mm and measuring length (Lm) is 12.5 mm. The research included the analysis of the filtered (so-called R-system) and unfiltered, full profiles as well (P-system).

On the enlarged images of Figure 9, the deviations are shown well. The surface profile was registered – due to the degree and size of the lace formation – in a length of ±50 µm on the side wall and in a length of ±30 µm on the bottom of the groove (the R-system is shown on the upper part, the P-system with full profile on the lower part of figure).

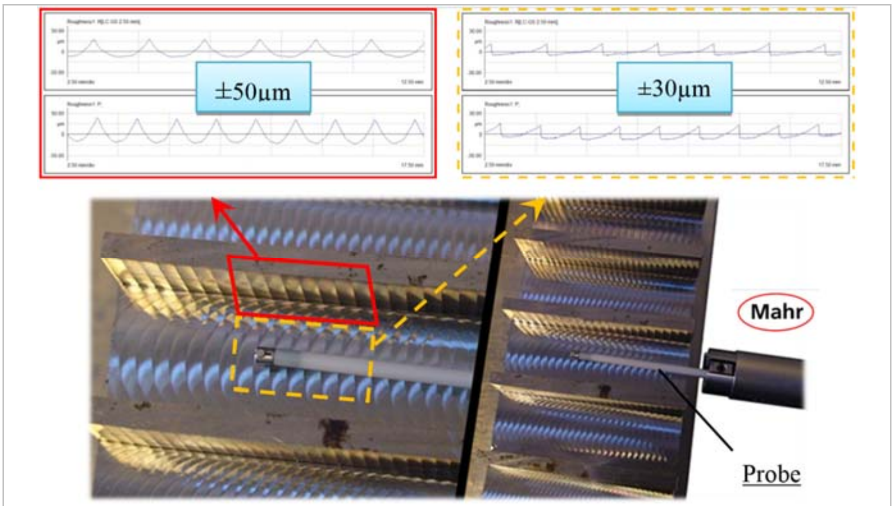


Figure 9

Macro- and microgeometrical deviations, developed on the side wall and bottom of the groove (Cutting data: $a_p=12$ mm, $w=2$ mm, $f_z=0.12$ mm)

Considering the fact that the roughness, formed on the side wall of groove, cannot be classified as microgeometrical deviations, therefore its modelling as roughness may lead to a wrong result. The lace development on the side wall is considered as a first-class deviation, i.e. as a form error, and according to DIN 4760 it is to be classified as macrogeometrical deviations. This form error – resulting from the characteristics of this machining process – may cause even a size error and it has a tight connection with the groove accuracy, mentioned in the previous chapter.

The analysis of the roughness values, measured on the bottom of the machined grooves (in the further part of the present article we are going to deal only with them), was set back only by one circumstance, it was the limit load (the 6th setting in the first block), already mentioned in the present paper. Its negative effect on the process could have been measured even in case of the surface roughness on the bottom of the groove. The average (-410 N) and the maximal values (-2100 N) of the axial tool load were so significant that they caused vibrations and the marks of the tool edges were simple „copied" on the surface. Compared to the registered roughness parameters, resulting from other settings, three-five times worse roughness values were measured in this case.

Excluding the limiting load values, the connection between the cutting parameters and the measured unevenness height (Rz) was analysed. The surface of the machined grooves was described with a low dispersion and with an unusual high correlation with the help of the following models:

Roughness model of profile, measured in R-system

$$R_z = 7.6 \cdot a_p^{0.64} \cdot w^{1.04} \cdot f_z^{0.68} \quad [\mu m] \tag{5}$$

Dispersion = ±0.4 μm; Determining index R^2 = 0.988;
Correlation = 0.994

Roughness model of profile, measured in the full profile P-system

$$R_z(\mathrm{JIS1982}) = 10.33 \cdot a_p^{0.59} \cdot w^{0.95} \cdot f_z^{0.66} \quad [\mu m] \tag{6}$$

Dispersion = ±0.3 μm; Determining index R^2 = 0.995;
Correlation = 0.9977

Both models (Eq. (5) and (6)) certify that the roughness, formed on the bottom of the groove, is linearly influenced by the step over and the feed rate per tooth but the axial contact has a degressive character (with the exception of the limit load).

The Figure 10 shows the Rz roughness characteristics, evaluated by both surface profile measuring systems. The calculated and measured (marked with x) values show that by increasing the feed parameters, the roughness values degrade, showing a slightly degressive character. The best roughness can be achieved in case of machining grooves with a depth of 9 mm, while the worst is in case of grooves with 16 mm depth. This statement confirms the fact that the max-max-max setting is to be avoided from the aspect of the roughness as well.

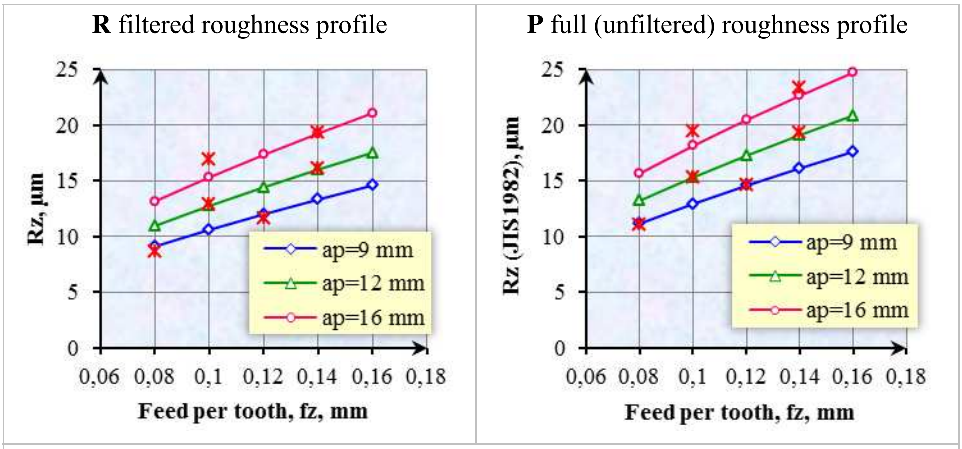


Figure 10

Development of the filtered and unfiltered profile on the groove bottom
(Cutting conditions: w=1.6 mm , a_{emax}=3.9 mm)

In connection with Figure 10 it must be emphasised that the Rz values of unfiltered profiles are by 2 – 5 µm greater, compared to the numbers, resulting from the filtered profile. Further research results, concerning this topic will be published in the next releases within the frame of TÁMOP-project.

Conclusions, further tasks

The main goals of the test series, executed based on Design of Experiment, were realised. The connection between the applied data (a_p, w, a_e, f_z) and certain technical-economical features (force demand, productivity, accuracy, roughness) has been modelled with the help of power functions as it is used in the technological practice. By applying the proposed mathematical models, it can be described which parameters and to what degree influence the process characteristics. Having processed the results of two blocks, based on the applied DoE, not only functions with low dispersion and high correlation were received, but they were able to determine the limiting values of CTM, when the productivity is increased.

As further task it is definitely reasonable to carry out researches, determining the data settings of the most productive cycloid-forming milling, taking the prescribed accuracy into consideration. The programming and preliminary examinations of other trochoidal milling methods have already been executed, in this way we have the possibility to carry out detailed, or even long-lasting tests of any of these research tests in the future. A good possibility would be to test extreme material grades (for example, heat-resistant super-alloys) as well.

Acknowledgements

The project was realised through the assistance of the European Union, with the co-financing of the European Social Fund, namely: TÁMOP-4.2.1.B-11/2/KMR - 2011 - 0001: Researches on Critical Infrastructure Protection.

References

[1] Bakondi, K. – Kardos, Á.: A gépgyártás technológiája III. Tankönyvkiadó, Budapest, 1974

[2] Präzisionswerkzeuge Klaus-D. Dung GmbH & Co KG: High - Performance – Cutting, HPC Fräser, Ausgabe HPC-010-1, Wolfsburg (Germany), 2012

[3] Szalóki I.: Trochoidális pályák programozása Microsoft Office Excel alkalmazásban, ÓE / BGK / Forgácsolástechnológia számítógépes tervezése feladat, Budapest, 2012. p. 30

[4] Szalóki, I. – Csesznok, S. – Csuka, S – Sipos, S.: Can trochoidal milling be ideal? Manufacturing 2012 The XXI. Conference of GTE on Manufacturing and related technologies, Budapest, November 15-16. 2012. Session #5, CD version, S6_08., pp. 10. ISBN 978-963-9058-35-4

[5] Seco Tools AB: Trochoidal Milling Excel application, Fagersta (Sweden), 2005

[6] Mózes, A. – Tállai, P. – dr. Sipos, S.: Új fejlesztésű száras maró-szerszámok alkalmazási sajátosságai, Óbuda University, e-Bulletin Vol. 2, No. 1, 2011, pp. 385-390

Both Sides of CSR Practice: A Case from Oil and Gas Industry in Kazakhstan

Gulzhan Buldybayeva

Abai Kazakh National Pedagogical University
Bukhar-Zhyrau 23, 9, Almaty, Kazakhstan

Abstract: Given article covers some special aspects of Corporate Social Responsibility in the oil and gas industry of Kazakhstan. The research deals with the analisys of activities of Kazakhstani multinational companies similarly to other transnational corporations of exploration and production industry in developing countries, which belongs to good or the darker side of the responsible acts. Recommendation demonstrating with a case study how to develop a better strategy for corporate social responsibility which helps leading Kazakh oil and gas joint venture company to avoid the possible occurrence of the CSR Dark side is presented in this article.

Keywords: corporate social responsibility; developing countries; oil and gas industry, exploration and production; case study; Dark side

1 Introduction

Since the 1990s, the concept of corporate social responsibility (CSR) has gained strong popularity within the business industry of developing countries, saying and doing both theoretically and practically. In the beginning, the concept had an Anglo-American feature based on „stakeholder economy" by British Prime Minister Tony Blair in 1996 [10] and followed by first, the US view of market driven operation of business, and later became, the general use of social and environmental aspects of business operations. Porter and Kramer [36] were leading proponents of the view that different types of social issues linked to core business of the company: generic social issues, value chain social impacts, and social dimensions of competitive context [5]. These thoughts were spread with dizzy quickness towards developing countries of the world generating an endless debate on the concept and meaning without any chance for final consensus on it. There are at least two explanations of this lack of consensus. Ghosal in 2005 [14]. First of all there is an argument that CSR concept is only a social marketing idea, a general speech from corporations which take care of their public image but avoid making any important changes, or even worse, it is managed by business people who have no moral responsibility for the business and the community. The publication of Berthoil and Sobczak [6] raises this issue, the term „corporate social responsibility" doesn't necessarily have the same meaning in culturally different places (countries, regions, continents) of the Globe. The Asian and European traditions are different from those of the Anglo-American.

„Is corporate social responsibility the privilege of developed market economies?"- question has been proposed by Koleva and her co-authors [23]. The answer is evidently: it is not a privilege of dedicated developed countries and there is less evidence of its existence in other geographical areas. Due to economic system change, the horrible amount of foreign direct investments to these regions, and the dominant presence of multi- and transnational companies (MNCs and TNCs) have resulted in the manifestation of CSR concepts in the new market economies including the (new) countries of Central and Eastern Europe and Kazakhstan (a state of former Soviet Union).

CSR practice has a great developing potential in Kazakhstan. This issue is important not only for the government but for the multinational companies which run business in Kazakhstan. What's important to distinguish today is that the companies see CSR as an obligation before the government of the country they are bringing their business into. Of course it is understandable that the giant multinationals enter the third world countries for reasons other than charity, however they must change their attitude towards CSR. The emergence of CSR concept in domestic and joint venture and multinational companies of the developing countries offer different implications of post-socialist compromises related to CSR. The convergent features of CSR in these countries have partly been explainable by post-socialist traditions. Different levels of interaction of actors produces very different CSR models, like hierarchical, participative, minimalist, endogenous, exogenous or hybrid [23]. The old general rule of management studies, „no one best way", is succeeding in this respect, too. The result is that CSR models and protocols have a big selection: the company managers, researchers can have a very broad variety of practices for application and field for research, for preparing case studies.

Traditional approaches to corporate social responsibility describe it as a contribution to society of businesses, apart from economic sector, industry, products or services the company operates and delivers. In every case the CSR wants to say something to the community about the operation of the company. In this regard such industries where the „content of saying" [33] has a long–term dimension, on the one hand, and the number of addressed people is extremely large, on the other hand, they are placed in distinguished position. The exploration and production industries have to be classed among this group of industries in different parts of the world. This business field is a very common target of CSR researches and publications.

The Oil and Gas Industry nowadays is the most robust actor of the exploration and production industry. The responsibility of owners, managers and the employees of these firms, MNCs, TNCs for the present and for the future, is huge. The deal with the topic of CSR in the developing countries is such that, corporations in the oil and gas industry propounds some delicate questions. CSR is a positive concept, a very nice result of positive thinking of mankind.

However, the sources and international practice has demonstrated many losses and mistakes, misconceptions, conspiracy of silence and scandals (e. g. Exxon Valdez, Mexico Gulf accident, Bonga Onga exploration area disaster). Behind these events, very often the bad CSR solutions are hidden. It is called the other side or the Dark Side of CSR, summarized by Baghi at al. [4]. Since, the technical development of this industry is not able to preclude the uncertainties, the CSR programs and its communication bear almost all the responsibility to avoid the occurrence of the dark side of CSR.

2 Overview of the Oil and Gas Industry in Kazakhstan

Oil and gas sector of Kazakhstan is the most dynamic sector of its economy. It provides a significant portion of national GDP. Since independence from Soviet rule in 1991, Kazakhstan has received more than 18 bln. US$ in 2010 of foreign direct investment - the highest per capita indicator in the former Eastern Bloc. The exploration of the largest oil field in the world, Kashagan, will make Kazakhstan one of the major producers of hydrocarbons not only on the regional, but also on the international level. Production of crude oil and natural gas condensate from the oil and gas basins of Kazakhstan amounted to 79.5 million tons in 2010, up 4% from the production in 2009. Kazakhstan raised oil and gas condensate exports to 72 million tons in 2010, which makes almost 91% of the produced oil in 2010. Gas production in Kazakhstan in 2010 amounted to 37.4 billion cubic meters. At the moment Kazakhstani gas is supplied to Russia only, however it is said that another independent pipeline to be built in the direction of China in 2013. Expected amount of Kazakh gas to be supplied to China is 30 billion cubic meters per year [50].

On the territory of the Republic of Kazakhstan 202 oil and gas fields are located. They are placed between the 6 out of 14 regions of the country. They are Aktobe, Atyrau, Western-Kazakhstan, Kyzylorda, Karaganda and Mangistau regions. The production is conducted from the 55 fields, the main ones are Tengiz and Karachaganak [43].

Today the oil and gas sector is represented by numerous companies. The largest national petroleum company is “Kazmunaigas” and it is engaged in a full cycle from exploration and production to the petroleum stations around Kazakhstan. There are many of the worlds majors involved in the development of three giant fields such as Tengiz, Karachaganak and Kashagan. Chevron holds important stakes in Tengiz and Karachaganak and a number of different international Oil Companies (primarily Chinese) also have significant stakes in the large Kazakh onshore fields. Other companies invested in the area include ENI, BG, Total, LUKoil, Shell, ExxonMobil, ConocoPhilips and Inpex [13].

The oil and gas industry is usually divided into three major sectors: upstream, midstream and downstream. The upstream segment of the industry, often called exploration & production, is traditionally what comes to mind when people think of the oil and gas industry. Companies will search prospective areas for potential reserves of oil and gas and perform geological tests called seismic tests, to determine the size and composition of the resource. Initial wells are often drilled to "explore" the basin, and if satisfied with results, a company will enter the production phase to extract the hydrocarbons.

Often coupled with the downstream segment, the midstream sector involves the transportation, storage and marketing of various oil and gas products. Depending on the commodity and distance covered, transportation options can vary from small connector pipelines to massive cargo ships making trans-ocean crossings. While most oil can be transported in its current state, natural gas must be either compressed or liquefied for transport. The midstream sector also includes the storage of oil and natural gas, which balances the fluctuations between supply and demand and helps ensure a secure supply of energy products.

The downstream sector involves the refining and processing of hydrocarbons into usable products such as gasoline, jet fuel and diesel. Refining is required since "raw" hydrocarbons extracted from the ground are rarely useful in their natural form. The refining process is a complex chemical process that helps separate the hundreds of hydrocarbon molecules into useful forms. Petrochemical plants also break down hydrocarbons into chemical compounds that are used to create a myriad of products ranging from plastics to pharmaceuticals [1]. Through the period of independence of Kazakhstan the oil and gas sector has made a significant growth. In 1999, Kazakhstan has celebrated 100 year anniversary of the oil and gas sector. The first oil has been produced in 1899 in the field called Karashagul. Passing this uneasy path, Kazakh oil and gas sector makes a significant contribution to the economic, social and cultural development of the country. Only in the last decade Kazakhstan has doubled its oil production [54]. By the level of oil production Kazakhstan is amongst the 20 producers of the world's oil and gas industry (With future development of the Kashagan field Kazakhstan will enter the top five oil-producing companies in the world).

During this short period the oil and gas sector along with the national economy has experienced the difficulties of transition and market reforms. The Kazakh economy lacked sufficient financial and technical capabilities to develop the gigantic oil fields such as Tengiz and Kashagan. That's when it was decided to attract foreign investors, companies that have the experience and the finances necessary to develop these fields. However, along with the positive aspects of attracting foreign investments we must mention the drawbacks of the given policy. Often, foreign companies forget about their obligations mentioned in the contracts, they become unsympathetic to the problems of the region they enter with their business. Companies do not use the local labour force; pay no attention to the social sphere of the region.

In order to come out of the dependency on foreign investments, Kazakhstan is coming out with its own investment opportunities; domestic investors are willing to invest their money into promising projects. The share of foreign capital has been reduced to half of the total investment. This means that this country has emerged strongly and is capable to finance major projects on its own. All of this suggests the need to change the strategy of the oil and gas sector of Kazakhstan. It is important, without compromising the interests of foreign investors, to line up our national interests, the situation in the oil and gas sector.

The oil and gas industry is capable to support all the monetary fund into Kazakh treasury, it contributes to the development of different fields of economy, and of course to the standards of living, because the development of oil and gas industry affects the economic rise not only the spheres directly related to the oil and gas industry. Production, transportation of large volumes of oil and gas, during the certain economic policy, develop grounds for developing domestic energy, chemistry, petroleum chemistry, telecommunication, transportation, construction of roads and infrastructure sectors. All these economic sectors contribute to the development of science and innovative technology. That is why the Republic of Kazakhstan connects all its future plans and hopes to the oil and gas industry. In order to increase the effective use of oil and gas resources Kazakhstan has to be determined to extract maximum income from the export of oil and gas without affecting the stable development of the oil and gas industry.

The world has not yet come to the point of stop using the oil and gas. The petroleum Era has not come to its finish. The market is still going strong stated a world wide well known expert at this industry Tippee [44].

3 Methodology of the Study

Case study methodology has been applied for the research. This method is a well known effective tool to represent and analyze real business activities and projects as well as the social movements from early at entering social science researches to business life [42], [15]. This methodology is frequently used in the oil and gas industry with different objectives [26], [32]. The firstly quoted case about the

Norwegian industry is partly applicable for the Kazakh oil and gas companies. The study mainly focused on the analysis of CSR activities and performances at the Tengizchevroil. To meet the need of the objectives of the study, data were collected and analysed from the secondary sources of the sample company in Kazakhstan. Previously, an extensive desk and library work was conducted to get a clear idea about the concept of corporate social responsibility in general and especially within the exploration and production industry with an outlook into the developing countries. The semi-structured interviews with competent managers of the company and different leaders of community assisted us to understand the inner main goals and ideas of the companies on the one hand. The daily practice of CSR, at the Tengizchevroil joint venture company on the other hand. The investigation of CSR reports of Chevron Corporation in the previous years and in addition TCO gave useful information regarding current joint community development programs within the Caspian region of Kazakhstan. The proposed framework can support practitioners in the oil and gas industry (and other industries) to develop their CSR system.

4 TCO Case Study

4.1 Good (Positive) Side of CSR

Tengizchevroil (TCO) – joint Kazakh-American Corporation a leading exploration, development, production and marketing of petroleum and related products. TCO is the largest oil producer of Kazakhstan's oil and gas sector. The Tengiz oil field was first discovered in 1979 and is one of the deepest and largest oil fields in the world located in the Western part of Kazakhstan. In April of 1993 with the signing of an agreement between the Government of the Republic of Kazakhstan and the "Chevron" Corporation the Tengizchevroil has been created. Today, the shareholders of the Tengizchevroil are: Chevron – 50%, KazMunaiGas (Kazakh National oil company) – 20%, ExxonMobil – 25%, and LukArco – 5%. The duration of an agreement is 40 years [35].

From the day Chevron and the Government of the Republic of Kazakhstan signed an agreement; they have undertaken a responsibility before the community they where they are acting. "We strive to work so that the result of our work was a real improvement in people's life" [35]. Throughout its stay in Kazakhstan, TCO has shown a great respect towards the agreement and has developed several social investment programs which were included in its Annual Corporate Social Responsibility Report. The terms of corporate social responsibility is well known within the interviewed managers of TCO. (The sample was relatively small, 11 persons who are working at the company in "CSR-related" positions.) Among the tested definitions had a first priority that the CSR is „the CSR covers the relationship between corporation (or other large organization) and the societies with which they interact. CSR also includes the responsibilities that are inherent on both side of these relationships. CSR defines society in its widest sense and on many levels, to include all stakeholders and constituent groups that maintain an ongoing interest in the organization's operation" [19,18]. The definition from the early history of CSR concept had been accepted as basic point of understanding the corporate social responsibility at TCO. The CSR Reporting is a daily practice of existence of the TCO model and the majority of these managers are personally involved in report writing process. From the major CSR programs at company was mostly mentioned two of the TCO applications. They are the Egilik and the SIP. Next paragraphs of the paper will highlight the results of these prominent programs. Both of these programs were mentioned as an output of positive side of the CSR operation at TCO.

4.1.1 Egilik Voluntary Program in Atyrau Oblast

At the start of Tengizchevroil joint venture in 1993, the company created a five-year, $50 million dollar program called Atyrau Bonus Fund. At that time, TCO developed a tradition of working collaboratively with the Atyrau Oblast Akimat leadership to choose social infrastructure projects in Atyrau Oblast to be funded with Tengizchevroil funds.

In 1999, TCO developed a first voluntary program Egilik (Kazakh for benefit) program to succeed the Atyrau Bonus program. It began in the year 2000 with a $4 million budget, growing steadily until 2003 when the budget was increased to $8 million for three years and then again increased to $12 million for another three year period.

Today, the budget is $20 million per year. The majority of funds are currently dedicated to replacing the water system in Kulsary, the nearest population center to the Tengiz field and home to many TCO employees and contractors. This project will take three years and compliments one of the TCO Community Investment Program key focus areas; health. TCO health and social impact assessment studies sites that Kulsary water quality and water service is undependable and low. In 2008, Tengizchevroil was bestowed the Silver Paryz Award for another Egilik project in Kulsary, the vocational school built to educate the future of the oil industry.

Over the life of the program, TCO has funded various social infrastructure projects for water systems. By the end of 2011, TCO have invested $185 million in the Egilik and Atyrau Bonus programs since 1993.

4.1.2 The SIP is to Improve the Quality of life of Employees

The Social Investments Program (SIP) has been an initiation of TCO in the development of the social sphere of the region with an annual budget of $ 1 million dollars. As the result of recent organizational decisions, TCO improved social investments program, which meets the public and industry needs and goals. The company has created a Council on Social Investments, which consists of representatives from different departments of the company, whose competence includes the distribution of the SIP budget. The Council also developed a philosophy and strategy of social investment, changing the vector from simple charity to target actions by which the urgent problems are resolved, and important social goals are achieved with priority to Atyrau region and TCO.

The aim of the company is to find direct connection of the projects necessary to the public without missing its own interests. Thus, the priorities have been identified, priorities in education, supply of drinking water, environmental protection, as well as the provision of quality health care.

When it comes to education, the company aims to educate its own staff for the further development in order to invoke them in the strategy of nationalizing the staff. To accomplish this TCO has a strong training and development plan for its employees. TCO has spent over 28 million USD $ for its training and development plan over the 9 year period. Every year the company trains the employees at training courses abroad.

Nationalization of the staff is a practice that all the TNCs use when they enter developing countries. In TCO is already taking place, today it is almost 85% of the staff are local employees compared to 1993 when it was only 50%. TCO believes in taking care of its own, so they offer great housing projects for its employees, support for employees' families and healthcare. The most popular benefit among employees is the housing loan program. TCO has invested over 32 million USD $. This is done in order to improve the quality of life of employees.

Continue with educational program, TCO has invested on scholarship program. It grants 70 scholarships a year to high school graduates, four of them are granted to gifted children from the Atyrau region and rest is split amongst the children of the employees. The company also provides internships; about 3000 students per year are able to enter the program.

TCO has invested over 450 million USD $ in social projects, which include building hospitals, schools also funded improvements of roads, water and etc.

Once again, TCO has become a silver winner of 2008 Paryz honour award for corporate social responsibility. This award is a creation of the President of the Republic of Kazakhstan Nursultan Nazarbayev, to recognize all the good the companies do for the community where they operate [39].

4.2 The Dark Side of CSR and How to Avoid It

The history of Corporate Social Responsibility concept proves, CSR has the potential both positive and negative impacts [4], [7], [28]. The negative impacts have been said to be the dark side of this management concept. The appearance of the dark side is especially dangerous within environment – in the industries such as oil and gas mainly in developing countries of the world emphasises Belal [5], and the 2012's report of the WTO [51]. The social and environmental impact (health and safety risks, large-scale air pollution) of the oil and gas industry affects the image or reputation of the company [46]. Therefore, building a sustainable corporate image is important to the industry. When one company suffers from image loss, the industry in total can be affected. In certain respect the situation seems to be similar to the one that has been outlined by the Hungarian edition of Kotler and Lee book about the CSR practice of Hungarian Oil Company (MOL) [24]. Oil activities very often have caused heavy damages to the environment more than in one occasion. The oil production process consists of upstream and downstream operations that include activities such as drilling, well completion, well testing, oil treatment and transportation through pipelines to consumers. Lands, farms, animals, forests and seas are affected from extensive growth of oil and gas production, transportation and distribution. The impacts of these processes may be partly natural, partly social. The natural consequences are perceptible and tangible (like air pollution, low farm production, polluted water). However, social impacts seem to be indirect. Negative impacts of CSR are illustrated by these social effects on local communities. The oil and gas industry show very broad examples from the developing world. The firm's relationship with their community via operation and community development programs hide many possibilities of conflicts. Local individuals and groups are fighting for donations of international oil companies, for services and infrastructure provided by MNC'S, community leaders try to exclude the other actors of community.

Originating from aspirations of these countries new concepts for tailoring CSR to the demand of developing countries are continually emerging. For example Ite suggests that the oil TNCs are making considerable contribution to community development in the Niger Delta in Africa that reflects the changes to new CSR model [21]. Idemudia says: „A South-centered CSR perspective has both been instrumental to the maturation of contemporary CSR agenda as well as generating fresh critical insights into the limits and strengths of CSR initiatives in developing countries" [20, 11]. The black side of CSR should successfully be counterbalanced with other new models that had been published, like Global Memorandum of Understanding (GMoU). The implementation of GMoU is Chevron, one of the main partner of this initiative, shows a good example of community development as a part of modern CSR concept [21]. Chevron model of the GMoU represents a radical departure from the past, from the earlier models in terms of benefits, both to the communities and company. These experiences are applicable for other Chevron investments like Tengizchevroil in Kazakhstan.

Similar outcome of new type of corporate social responsibility programs has developed for the oil and gas industry in India under the scheme of Oil and Natural Gas Commission. Oil companies in India have manifested CSR in a variety of ways, most of which have been quite innovative in the field of community development, promoting excellence, environmental management, preserving national heritage and other CSR-related issues [37].

As it was presented in earlier part of this paper the CSR practice is a new challenge for Kazakh Big Firms. It raises a question: how could be interpret both sides of this business phenomenon? Are there symptoms of black side? Is there any chance for corporations to avoid them? First of all, it must be stated that the Kazakh Presidency and Government do everything to avoid other (dark) side of CSR practice:

- Kazakhstan joined the EITI (The Extractive Industries Transparency Initiative) in October 2005 [51]. Consequently, Kazakhstani Government has accepted the multi-stakeholder initiative, involving business, government, international organizations, promoting transparency and accountability of the revenues from exploration and production industries;
- the national legislation has guaranteed protection for all business actors (domestic and international investors, local producers, labour force), at the same time the guarantee has been expanded to direction of national security, defence, environmental safety, healthcare as has pointed out Yerkebulanov and Saginova [52].
- the other special project within the oil and gas industry (e.g. Caspian Revenue Watch) that started for the whole geographical region, is to reduce the dark side of industrialization [18]. However, the initiative for the community development conducting by the Open Society Institute finally was not totally successful [45], the main idea of transparency and social responsibility has been converted into the daily business in Kazakhstan;
- due to many reasons, in Kazakhstan the Transnational Oil Companies have not been the target of negative or anti-corporate campaigns, as in the case of other developing countries (mainly in Africa and Latin-America). Many times in these countries the poor CSR performance of transnational oil corporations and the inadequate CSR package has overshadowed the benefits of good initiatives [5], [31], [38], [46]. One of the most famous researchers of this topic Michael, J. Watts pointed out that one and a half decades ago, that " … the current political economy of the oil industry, new bodies of research on the relations between oil and human rights violations, which include case studies of the human rights records of transnational and joint-venture oil operations" [49,375].

Taking into consideration many reasons, naturally to have a declared legal system, to join international movement and to be part of high-sounding projects – it is not enough to avoid the impending dangers and mistakes either for the society or for the companies. It is necessary to take permanent, significant efforts for the sake of averting CSR aspects that want to be effective. The focus for business community investment tends to be on some of the further issues around health, the environment and economic development, and much more focus on building or supporting civil society in Kazakhstan [16], [40], [46].

4.3 New Challenges for Tengizchevroil: Recommendations

Learning from company history and experiences of CSR practice conducting interviews on the field with combination of external (domestic and international/foreign) knowledge, the Tengizchevroil faces new challenges. The creation of a sustainable corporate image as an essential part of a proactive strategy and core value creation is addressing series of potential challenges in the near future. Eventually, these challenges can use this list of recommendations: how the corporate social responsibility system at TCO should be developed in five areas.

Development of relation with the Government:

- because there is the absence of agreed definition and clear common understanding of CSR among the state, business and society, Tengizchevroil as a transnational company has to be pioneer for this clarification process which is badly needed.
- effective implementation of the Government CSR Strategy desires a powerful support from leading corporations such as TCO so that co-operation between the State and the Company should have been successful for long term. In light of the existing “Public Responsibility Deficit” [47,53] and the growing power and dominance of big corporations they are wearing more and more responsibility for supporting the Government that this deficit should be reduced. In creating attention focusing on the role of companies in public service provision – one of the best strategy for TCO.
- the requirement for “social investments” of the Company can pose a further challenge to a company’s development of a business strategy that takes CSR as significant part of TCO’s core business in Kazakhstan.

Communication channels and priorities:

- the Tengizchevroil is a massive business actor of Atyrau Oblast. The activity and voice of this company determines the economic and social life of many other actors (both organizations and private people).

- the main issue of CSR communication is the reporting. CSR reporting has developed in quality and quantity since the first reports were published. To further explore this issue, the literature specifically focusing on The CSR reporting of the exploration and production industries such as oil and gas. For this industry, a special publication on reporting was published in 1999 by Sustainability [17]. At that time 50 companies were surveyed, 34 were publishing environmental reports. The oil and gas industry is active in this area, for the year 2005 80% of all companies from exploration and production industries produce reports on CSR activities – state the above mentioned German researchers. As a suggestion, another good methodology that is well known within the exploration and production industry is the GRI (Global Reporting Initiative) guidelines. This method was developed by one of the big consulting firms, KPMG [25].

Above listed initiatives are applicable to demonstrate CSR achievements within the sections of: vision and strategy of the company, profile of the company performance indicators and the economic, environmental and social perspective." They are the basic elements of a well-structured CSR reporting system. Although, Tengizchevroil constantly, year by year fulfills its liability with reporting, but it is mainly not more than a forced obligation. The challenge is for the TCO top management to work out a new mode of CSR reporting. The applicable model seems to be a „CSR information package" of Morsing [34], which ideally should address following key words:

- promise: show CSR as a shared concern – Invisible contribution but visible success in various fields
- proposition: link CSR to core business – direct link between business and safety/environment
- evidence: demonstrate organizational support – "we and our" stakeholders (partners and customers).
- results: demonstrate objective claims – High level of detail to increase credibility, sifting on various technology used by TCO.

Improvement of relation to business environment and community:

- the primary interest of TCO CSR practice is to keep solid relationship with its stakeholders from business environment. The stakeholder analysis could demonstrate elementary importance of local population. Concerning to this very evident fact, to improve the public relations in the given region of TCO operation represents high priority for corporate social responsibility process.
- as it was mentioned by the interviewed managers and the secondary sources, TCO supports many civil organizations in the Caspian, Atyrau Oblast [27]. Despite of flowering character of nonprofit and non-governmental sector in Kazakhstan, it is permanently facing new challenges [2], [19]. In newly published empirical study by Makhmutova and Akhmetova the lack of

professional and skilled managers and qualified staff and volunteers within the sector has been pointed out [30]. According to the results of these studies the NGO sector and the powerful actors of the competitive sphere (e.g. from oil and gas industry) must be strong partners. (Good example is the success of Atyrau Bonus Fund.) The for- profits can have direct benefits if they utilize the social power of not- for-profits if they strengthen their efforts for leading community development, environment protection-oriented projects. For example (Green Line Buzz, EcoCenter, and especially the earlier demonstrated Egilik in Kazakh Caspian region) and big oil corporations (as the TCO is) have to be key players on this stage.

Contribution to the community wellbeing:

- this theme of community wellbeing represents a very attractive and challenging goal structure for TCO focusing on its involvement in the community. Following the findings of Cowper-Smith and Grosbois [11] by this concern at the Tengizchevroil the most popular initiative would be: supporting training and education possibilities for youth, donating arts and culture projects and events, supporting environmental education in local schools. All this should be offered on higher quality and quantity level.
- the internal context of community wellbeing relates to employees of the company. The annual average number of TCO's employees is about 15,000. They are aspiring for good working and living conditions and for healthy environment as citizens of Kazakhstan. The physical circumstances and even the mental conditions of workforce should have an important indicator of CSR construction of TCO. The corporate social responsibility agenda cannot ignore the personal feeling of company's people: they are the primary stakeholders of the organization.

Communication and communication:

There is a common feature of every CSR concept. Growing public demand for information regarding corporate social responsibility plans and actions using the most appropriate communication channel. Positive messages from the TCO to the society should beconveyed through such channels like:

- logos
- short banners
- informational materials, presentations
- commercials and media advertisements
- independent reports on companies
- records of journalists
- publication of corporate reports
- CSR-award
- CSR-event
- CSR-webpage
- forums and blogs
- CSR-applications in social media
- publication of the code of ethics [3].

Understanding the local relevance of these possibilities, the communication is extremely needed, because the communication is "the soul of the CSR". The TCO has a very nice technical presentation of CSR on the company website, notwithstanding, shows lack of a well-constructed plan for communication of CSR actions. The recommendation regarding this statement has first priority: a correct plan for two different types of cases to be worked out. Firstly, the regular daily operation of CSR actions and processes. Following action such as distribution must take place, every stakeholder and group of community should be informed on how Tengizchevroil thinks about the corporate social responsibility and what it does for its success? Secondly, it is important to emphasize that communication of conflicts and handling of different conflict situations is also a part of CSR communication system [28], [48]. There are lot of possibilities of emerging conflicts at business operation of TCO. The predictable issues are:

- diverse expectations create tensions between ethical and economic interests and values. These two views are grounded historically contradictory, but the short history of Kazakh oil and gas industry was not able to offer effective mechanisms to cope with this paradox, yet. The time has arrived for working out a system of knowledge intensive professional business service model which is able to meet the needs of international CSR concepts argues for this concept the professors of University of Pécs Faculty of Economics and Business Katalin Dobrai and Ferenc Farkas who are privileged researchers of KIBS topic [12].
- the extractive industry is a dangerous one. The natural crises and disasters seem to be real dangers with non-predictable probability. The Kazakh society has to be sure that Tengizchevroil Co. would be able to handle this unforeseen situations and to save the people and the nature of Kazakhstan. That is the Complex Concept of Corporate Social Responsibility.

Conclusions

This paper has verified the relevance of corporate social responsibility application at one of the flagship company of Kazakh Oil and Gas Industry the Tengizchevroil. By the statistical data this industry offers a proper field for the concept in an international framework: In the 2013 ranking Global 100 Most Sustainable Companies have listed dozens of big corporations from energy sector; the trend is positive from the viewpoint of the energy sector in comparison with previous years (the representation of the sector was less); TCO, Chevron, over the six years invested approximately 1 billion US dollars in areas essential to sustainable communities, including health, education, and economic development-CEO message published on the company's website [35].

Connecting to the strategy and activities of TCO, who launched 1 million $ USD on Social Investment Program, and recently, Egilik the main subject of the delivered case study. These initiatives to be considerable starting-point to direction of advanced CSR system at Tengizchevroil.

This kind of business such as Tengizchevroil plays a major role in creating wealth and innovations, supplying markets, and generating employment, while at the same time contributing to the community in which it operates. Society on the other hand is expected to provide environment where enterprises could develop and prosper, allowing investors to earn profits without fear of unfair actions from the government [53]. This general business hypothesis for this matter [29] was also proved with demonstrated case study.

It is important to note that the responsibility of the corporations to behave socially shall be spread on all companies, both local and foreign companies to avoid any discrimination.

Unsocial activities are penalized, and are grounds for very harsh sanctions starting from administrative fines the amount of which could be estimated in millions, temporary suspension of the petroleum activity in the country, and what worse is the permanent license and contract termination or even criminal liability of the managers of the company. Therefore, compliance with the law is the minimum what companies shall provide to legitimately stay with successful business in especially in a rural environment when the companies of the oil industry are generally operating [8].

In the oil and gas sector, CSR activities represent an attempt to fill that void. It is no secret that many oil-rich nations have been poorly governed, and the international oil companies have extracted valuable resources from such countries in the past, while paying inadequate attention to the attendant environmental and social costs states Spence, in his paper [41]. His study shows that there are no boundaries of CSR, such as cultural, racial and etc., the danger everywhere is the same, namely, the giant multinational companies producing annual CSR reports showing their active involvement in the regions, but how they operate practically in the local business can be said different. The TCO case also represents many special characteristics from both sides of CSR operations. The recommendation of the paper is serving the objective: the dark side of CSR should remain only as a threat that should be avoided. The author from Nigeria of a very new research paper enforces the joint activities of academic people and business men to design a new model which will be applicable within any circumstances and to different developing countries [22].

Organizations around the world and their stakeholders, are becoming increasingly aware of the need for and benefits of socially responsible behaviour. The objective of social responsibility is to contribute to sustainable development. It is important that local authorities create non-governmental organizations that work for the interests of the community. It is important that they know exactly the problematic areas and act accordingly. These agencies will be responsible and also produce different kinds of CSR reports which will consist of the study of the required necessities and the proper distribution of the resources, donations received from companies for the benefits of common goals of community development.

Limitations and Further Research

The result of this paper was based on the overview of relevant CSR literature and on a field study at a joint venture transnational petroleum company in Atyrau Oblast in Kazakhstan, the TCO. Any efforts to draw general conclusions on this very delicate management and social topic are premature. The further research as a subject of a PhD dissertation will be put into a broader social perspective. The investagation of social influence of these multinationals on community and its development means more than an empathic discussion on social responsibilities of corporations. Without the understanding of policy makers' interests (both local and on government level) further steps of research can not offer generalized findings for the Kazakh Oil and Gas Industry. A benchmarking profile will follow as the next phase of the research. Benchmaring is a constant and systematic process for gaining results of the total industry. This first study - as a pilot case – offers a good direction to clarify the real objective of an industry level research.

Special thanks

Special thanks are due to managers of TCO who were very helpful in providing information regarding CSR practice and reports, to Professor Ferenc Farkas at University of Pécs Faculty of Business and Economics for his help as External Consultant, and to the Director of the Graduate Programme at Abai Kazakh Natonal Pedagogical University Gulmira B. Nurlihina for supporting and sponsoring this study.

References

[1] A Dictionary for Petroleum Industry. Second edition. The University of Texas at Austin (2007)

[2] Amagoh, F.; Kabdiyeva,A.: Management Issues for Improving NCO Sustainability in Kazakhstan. World Journal of Social Sciences (2012), Vol. 3, May 2012, pp. 31-40

[3] Ásványi, K.: CSR-marketingmix a gyakorlatban. Marketing & Menedzsment. The Hungarian Journal of Marketing and Management (2012), XLVI. évf. 3. sz. 32-41. o.

[4] Baghi, I; Rubaltelli, E.; Tedeschi, M.: A Strategy to Communicate Corporate Social Responsibility: Cause Related Marketing and its Dark Side, Corporate Social Responsibility and Environmental Management (2008) 16, 15-26

[5] Belal, A. R.: Corporate Social Responsibility reporting in developing countries: the case of Bangladesh, Aldershot England Burlington Vt. Ashgate (2008)

[6] Bertoil, A., A.; Sobczak, A.: "Endogenous and Exogenous Factors for Change in Concept and Practice of Corporate Social Responsibility, Corporate Citizenship and the Politics of Stakeholder Influence: Learning

from the Case of France, EGOS Annual Conference (2004) Ljubljana, Slovenia, July

[7] Bird, F.; Smucker, J.: The Social Responsibility of International Business Firms in Developing Areas. Journal of Business Ethics (2007) 73:1-9

[8] Bowler, I. R. ; Bryent, C. R.; Cocklin, Ch.: The Sustainability of Rural Systems: Geographical Interpretations (2002) Dordrecht; Boston: Kluwer Academic Publishers 280 p.

[9] Brudnay, J ; Nezhina, T.: "What is Old is New Again. Achieving Effectiveness with Volunteer Programs in Kazakhstan.", Voluntas: International Journal of Voluntary and Non-profit Organizations (2005) Vol. 16, No. 3, pp. 293-308

[10] Burkitt, B.; Ashton, F.: The Birth of Stakeholder Society. Critical Social Policy (1996) Vol. 16, No. 49, 3-16

[11] Cowper-Smith, A.; de Grostois, D.: The Adaptation of Corporate Social Responsibility Practices in the Airline Industry. Journal of Sustainable Tourism (2011) Vol. 19, No. 1, 59-77

[12] Dobrai, K.; Farkas, F.: „Professional Business Organizationals Meeting the Challenges of a Changing Business Environment", International Journal of Knowledge, Culture and Change Management (2007) Vol. 7, No. 3, pp. 73-84

[13] Egemberdiyeva, S. M.: Tendency of Development of the World Petroleum Market in Terms of Globalization (2010) Eurasian National University Journal of L.N. Gumilev – No. (73) 247-252

[14] Ghosal, S.: "Bad Management Theories are Destroying Good Management Practice", Academy of Management Learning & Education (2005), 4, 1, 75-91

[15] Gomm, R.; Hammersley, M.; Foster, P. (ed.): Case Study Method. (2000), Sage Publications Ltd.

[16] Grajzi, P.; Murrell, P.: Fostering civil society to build institutions. Why and when?, Economics of Transition (2009), Vol. 17 Issue. 1, 1-14 pp.

[17] Guenther, E.; Hoppe, H.; Poser, C.: Environmental Corporate Social Responsibility of Firms in the Mining and Oil and Gas Industries: Current Status Quo of Reporting Following GRI Guidelines, Greener Management International (2007), Issue. 53, 7-25. pp.

[18] Haufler, V.: "Corporate Transparency: International Diffusion of a Policy Idea?", University in Maryland Workshop Paper (2006)

[19] Hutchins, M., J.; Walck, C.,L.; Sterk, P., D.; Campbell, G., A.: Corporate Social Responsibility. A Unifying Discourse for the Mining Industry. Grener Publishing Ltd. (2007), pp. 17-30

[20] Idemudia, U.: Corporate social responsibility and developing countries: moving the critical CSR research agenda in Africa forward, Progress in Development Studies (2011), Vol. 11 Issue. 1, 1-18. pp.

[21] Ite, U.E.: Partnering with the state for sustainable development: Shell's experience in the Niger Delta, Nigeria Nigeria Sustainable Development (2007), Vol.15 Issue. 4, 216-228. pp.

[22] Kiikpoye, K. A.: New corporate social responsibility models for Oil Company in Nigeria's delta region: What challenges for sustainability?, Progress in Development Studies (2012), Vol. 12 Issue. 4, 259-273. pp.

[23] Koleva, P.; Rodet-Krichvili, N.; David, P.; Marasova, J.: Is corporate social responsibility the privilege of developed market economies? Some evidence from Central and Eastern Europe, The International Journal of Human Resource Management (2010), Vol 21., No.2, 274-293

[24] Kotler, P.; Lee, N.: Corporate Social Responsibility. John Wiley and Sons Inc. Hungarian Edition: A vállalatok társadalmi felelősségvállalása. HVG Kiadó Budapest (2007)

[25] KPMG: International Survey of Corporate responsibility Reporting (2005), Amsterdam: KPMG

[26] Kumar, R.; Markeset,T.: Development of performance-based service for the oil and gas industry: a case study, Journal of Business & Industrial Marketing (2007), Volume 22 issue 4. 272-280

[27] Lee, J.: NGOs in Kazakhstan & Caspian See Region: an Overview. Centre for Applied Studies in International Negotiations. Geneve, Switzerland (2007) 31.pp.

[28] Lehtimaki, H.; Kujala, J.; Heikkinen, A.: Corporate Responsibility in Communication: Empirical Analysis of Press Releases in a Conflict, Business Communication Quarterly (2011), Volume 24, Number 4, December 432-449

[29] Lindberg, Y.: "The Global CSR agenda: implications for Kazakhstan" How can CSR performance be improved in Kazakhstan?, Central Asian Studies Centre Panel debate co-organized with SIGLA, 1. March (2012)

[30] Makhmutova, M.; Akhmetova, A.: Civil Society in Kazakhstan (2011), Civicus, Almaty

[31] Martinez, E. O.; Crowther, D.: Is disclosure the right way to comply with stakeholders? The Shell Case, Business Ethics (2008), Vol. 17 Issue. 1, 13-22. pp.

[32] Matos, S.; Hall, J.: Integrating sustainable development in the supply chain: The case of life cycle assessment in oil and gas and agriculture biotechnology, Journal of Operations Management (2007), Volume 25, Issue 6, 1053-1102

[33] McWilliams, A.: Siegel, D.: Corporate social responsibility. A theory of the firm perspective, Academy of Management Review (2001), 26, 117-127

[34] Morsing, M.: Strategic CSR communication: The challenge of telling others how good you are, Berlin. Germany: Springer Verlag M. De Witte & Jonker, J. (Eds.) Management models for CSR: A comprehensive overview (2006), 237-245. pp.

[35] Official "Tengizchevroil" site www.tengizchevroil.com (03.04.2013)

[36] Porter, M.E.: Kramer, M.R.: Strategy and society: The link between competitive advantage and corporate social responsibility, Harvard Business Review (2006), 84 (12), 78-92

[37] Prasad, A.: CSR as nash Equlibrium, Journal of Management Research (2005), Vol. 5 Issue. 2, 59-71. pp.

[38] Prepared by: Ettenbotough, M.; Shyne, J.: World Bank Report. Corporate Social Responsibility Public Policy and the Oil Industry in Angola, Boston USA (2003), 19. pp.

[39] Report on TCO's CSR activities. www.csrkz.org (02.03.2013)

[40] Saltmarshe, D.: Civil society and sustainable development in Central Asia, Central Asian Survey (1996), Vol. 15 Issue. 3-4, 387-398. pp.

[41] Spence, D., B.: "Corporate social responsibility in the oil and gas industry: the importance of reputational risk"(2011), Chicago-Kent Law Review Volume 86:1

[42] Stake, R. E.: The Case Study Method in Social Inquiry. Educational Researcher (1978), Vol. 7. No. 2. pp. 5-12

[43] Strategic development plan of Ministry of oil and gas of Kazakhstan for 2011-2015. Astana, 2009

[44] Tippee, B.: „Where is the shortage? A nontechnical guide to petroleum economics". PennWell Publishing Company 1993

[45] Tsalik, S.; Ebel, R.: Caspian Oil Windfalls: Who Will Benefit?, Caspian Revenue Watch Open Society Institute (2003), 215. pp.

[46] Tuodolo, F.: Corporate Social Responsibility Between Civil Society and the Oil Industry in the Developing World, ACME: In International E-Journal for Critical Geographies (2009), Vol. 8 Issue. 3, 530-541. pp.

[47] Valenta, M.;Crane, A.: Public responsibility and Private Enterprise in Developing Countries, California Management Review (2010) Vol. 52 Issue. 3, 52-78. pp.

[48] Verboven, H.: Communicating CSR and Business Identity in the Chemical Industry Through Mission Slogens, Business Communication Quarterly (2011), Vol. 74 Issue 4, 415-431. pp.

[49] Watts, M. J.: Righteous Oil?, Human Rights the Oil Complex and Corporate Social Responsibility Annual Reviews Environment Resources (2005), Vol. 30 Issue. 9, 373-407. pp.

[50] Wikipedia materials on Kazakhstan www.wikipedia.org (05.04.2013)

[51] World Trade Organization (WTO), White Paper Corporate Social Responsibility (CSR) in Kazakhstan, (2012), 11. pp.

[52] Yerkebulanov, Y.; Saginova, D.: "Kazakhstani Content in Procurement of Subsoil Users", Kazakhstan International Business Magazine (2011), Issue. 5. 3-6. pp

[53] Zarsky, L.: "Human Rights and the Environment. Conflicts and norms in a globalizing world." (2002) London: Earthscan Publishing LTD.

[54] Zhumaly R., "Running a spiral: the history of Kazakh oil". Exclusive 2009 No. 90