Acta
Technica
Jaurinensis

Vol. 8, No. 2, pp. 96-112, 2015
DOI: 10.14513/actatechjaur.v8.n2.328
Available online at acta.sze.hu

The Initiation of Safety-Enhancing Actions in Railway Crossings Using Modelling Procedures

L. Szabó[1], M. Somogyi[2], G. Horváth[1]

[1]Széchenyi István University, Department of Transport
Egyetem tér 1, 9026 Győr, Hungary
E-mail: szala@sze.hu

[2]Széchenyi István University, Department of Automation
Egyetem tér 1, 9026 Győr, Hungary

Abstract: This study is examining the application opportunities of models, and it is revealing a possible solution which can be the basis of creating a catalogue which can establish the reconstruction of railway crossings. This reconstruction catalogue is suggesting infrastructure changes and new traffic management procedures with taking account of the controlled model outputs. Based on the reconstruction catalogue, feasibility studies and construction project documentations can be made.

Keywords: hazard ranking, degree of traffic safety risk, system of revision aspects, Evaluation Matrix, reconstruction catalogue

1. Introduction

The dangerous characteristics of railway crossings come from the fact that a crossing is a contact point of different traffic systems, where structurally different traffic paths cross each other. These different kinds of paths are used by vehicles with significantly different technical parameters. Therefore the railway crossings have prominent importance in terms of traffic safety. Because of the facts above, and in addition due to a ministerial instruction released in 2003, the most dangerous railway crossings must be determined. The basis of this determination is an establishment of a hazard ranking. This hazard ranking can be created using a total score per crossings indicator generated by a data processing algorithm. Within the project called 'Safety Inspection of Railway Crossings in the Area of Railway Track Sections Extended to the Border of Burgenland and West Hungary' (Sicherheitsinspektionen von Eisenbahnkreuzungen entlang der grenzüberschreitenden Bahnlinien in Burgenland und Westungarn - SIEBaBWe) by the examination of several possible modelling techniques we created a model which is suitable not only for the determination of the railway crossings' hazard ranking in our project, but also it is applicable for every railway crossings. The point creation was the common element in the potentially attractive models; the method of this point creation meant the difference. Regarding to certain models a mentionable common feature was the demand of counting with the environmental characteristics of railway crossings, the

accident and traffic data and the way of insurance The railway crossing, where the implementation of safety-enhancing actions are the most reasoned, can be designated by model outputs which are checked and filled with proper data.

2. The Accidental Status of Railway Crossings on the GySEV's Hungarian Network

The Government Regulation Number 168 of 2010 (V.11.) is about the listing of national mainline network, regional and other railway tracks. Within the project SIEBaBWe mentioned in the introduction, 23 railway crossings were mulitcriterially examined on the Hungarian Network of GySEV. The Table 1 contains the line group distribution of railway crossings.

Table 1. Distribution of the Examined Railway Crossings

Line type	***Number of Railway Crossings (pc)***
1	14
2	7
3	2
1	14

1: Railway track operated as a part of the Trans-European rail freight traffic network

2: National mainline network track which is not a part of the Trans-European rail freight network

3: Regional railway track

The selected railway crossings are showing a pretty various image in the aspect of technological features and other safety-enhancing factors, however these crossings are totally applicable for every existing provisions and regulations according to the special authorities' qualification. Table 2 contains the accidental data from 2001 to 2012 within the framework of the project.

Table 2. The Change of Accidents in Railway Crossings on the GySEV's Network

Year	***Accident (pc)***	***Died (pc)***	***Wounded (pc)***	***Property damage (HUF)***
2001	2	0	0	2 623 018
2002	8	4	6	2 865 554
2003	3	0	3	272 102
2004	4	0	5	1 857 166
2005	5	1	3	1 863 862
2006	4	0	2	1 547 281
2007	8	1	4	136 368 412

2008	7	7	2	137 344 914
2009	1	1	2	61 437 611
2010	0	0	4	565 330
2011	2	2	3	1 456 929
2012+	9	8	11	12 150 303

+ : Jan. 2012 – Nov. 2012

The average distance of railway crossings is 1430 meters on the 434.7 kilometres long network, and 89 217 trains has travelled on the examined network per year. There are 87 station and 217 line crossings of the railway crossings.

Analyzing the accidents we made the following major statements:

The following conventions should be taken into account.

- the accidents happened mainly in secured railway crossings.
- the most serious accidents happened in crossings which have small or average traffic
- the essential reason of accidents (99%) comes from the violation of Road Traffic Regulations
- more and more frequently the accidents are caused by improper behaviour
- the accidents also happened despite the locomotive's sound signal
- the property damages comes from accidents have significantly increased
- the technological developments couldn't significantly reduce the number of accidents
- peaks and troughs can be observed in the changes of the accident numbers
- property damages have increased in more modern railway crossings.

Analyzing the accidents had confirmed more the necessity of raising awareness in every age groups of the society.

3. The Examination of the Practicable Model Versions

In this chapter we introduce two modelling techniques which had been considered suitable by the working committee established within the project's framework for determining the hazard raking of the railway crossings. Speaking about the content and the actuality of the reachable data is necessary, regardless of the suitable modelling technique's type. The source of these data is the National Transport Authority, the Track Facilities Department of the MÁV Zrt., the Magyar Közút Nonprofit Zrt. (Hungarian Public Road Non-profit Ltd.), and the recordings of the railway companies. The railway companies are registering every event which happens in a railway crossing, whether it is an accident comes with personal injury or an accident which causes

property damage. The ranking practice of the past years met incomplete, wrong, old or non-recorded data. According to the experts' concordant opinion, a central database would be needed which should be based onto the road crossings' GPS coordinates assigned to the sum of digitally recorded data of the given railway crossing. Controlling the updates of databases would lead to further improvement.

3.1. „The GySEV model" [1]

The model made available by GySEV generates a ranking which shows the railway crossings' degree of traffic safety risk compared to each other. In this model the maximum given score of the certain railway crossings is determined. This maximum score will be distributed among the crossings' specific factors. These factors are:

- the numbers and results of accidents in the crossing
- the road and rail traffic of the railway crossing
- the safety equipment installed in the crossing
- other characteristics (Fig. 1).

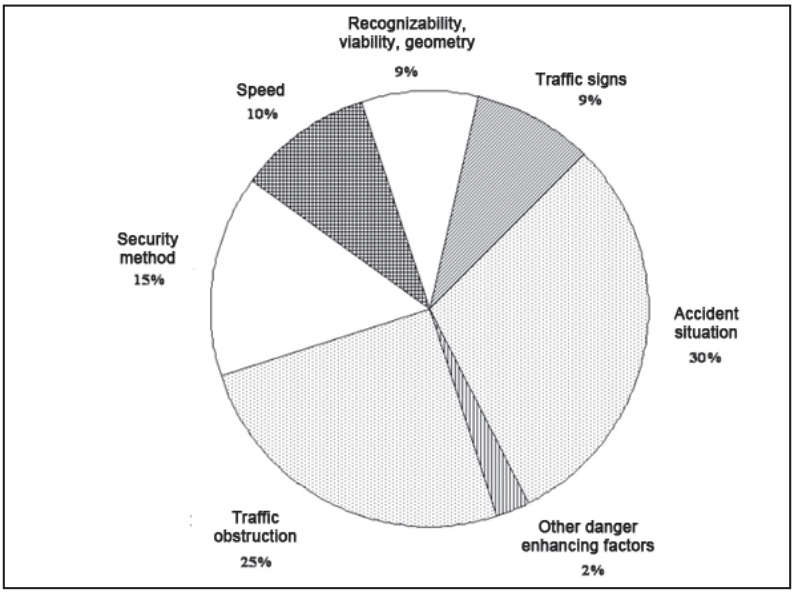

Figure 1.
Percentage Distribution of Maximum Assignable Scores by Indicator Groups [1]

The ranking system uses correlations in the case of every factor.

3.1.1. Accidental status

The assignable score contains events in the railway crossing from the past ten years. The events (accidents) which happened more than five years ago are counted with 50% relevance.

$$S_a = 300 * \frac{\ln(N_e+0{,}05)}{V_{max}} \quad (1)$$

where:

S_a: accidental status

N_e : corrected number of events

V_{max}: maximum value

Value: a) if there aren't any established crosswalks:

N_e = number of hits by car + 0.3*(number of pedestrians and cyclists hit)

b) if there are any established crosswalks:

N_e = number of hits by car

3.1.2. Traffic status

In the railway companies' traffic records the number and the division by type and time period of the trains were travelled in a certain railway line within 24 hours are recorded. Knowing the examined road crossing's railway gauge number the daily traffic can be extrapolated to the crossing.

$$S_t = 50 * \frac{ADT*t_i*N_t}{V_{max}} + \frac{ADT}{150*ADT_{max}+50*B} \quad (2)$$

where:

S_t: traffic status

ADT: Average Daily Traffic

It's value: in E/day unit of measurement, the maximum value means the maximum value of the numerator.

t_i: average interference time

N_t: number of trains

B: bus traffic

It's value:

a) 1, if there is any bus traffic
b) 0, if there is no bus traffic

3.1.3. Traffic Management Characteristics and Other Hazard-enhancing Factors

The 'railway speed limit' in context (3) is taking account of four speed rates. Among the 'other hazard-enhancing factors' the score calculation is based on:

- the type of the road where the railway crossing is situated (territorial nature)
- public lightning, the existence of separated sidewalk footpath

the existence of crossroads within 30 meters where the vehicles come from the railway crossing has no priority, and the security track is missing

$$C_{tm} = SM + SL_{rail} + SL_{road} + R_v + S_{road} + Sb + F_h \quad (3)$$

where:

C_{tm}: traffic management characteristic

SM: Security Method

SL_{rail}: Railway Speed Limit

SL_{road}: Roadway Speed Limit

R_v: Recognizability, viability, geometry, angle of crossing

S_{road}: Road Signs

Sb: Signboards

F_h: Other hazard-enhancing factors

3.2. Conceptual model

If we examine the railway crossing as a given physical object, then we mean the concept of system as a model which can be described using physical variables. The contents of these physical variables also can be heterogeneous such as a combination of physical, chemical or economical characteristics. A part of the variables are given, these are the inputs or excitations. A common attribution of the other group of variables is the aim that we want to determine their behaviour. These are the outputs or outlets. We can speak about a third group of variables which is responsible for describing the connection or connections between inputs and outputs. In that case it is true that we describe an object by a system. In other words the system is the model of a given physical object.

Basically two main types of models can be distributed in conjunction with the type of processes we want to map. The geometrical, physical and mathematical models are belong to material models. The so-called mental models are functioning by the logical connections established by people, regarding to their methods and forms they are subjective, regarding to their contents they are objective. The so-called conceptual model [2] also can be classified as a mental model. This modelling method is applicable when a given problem's possible solution is requiring the creation of different scenarios. In our case each scenarios are equivalent to the hazard raking of the railway crossings. This procedure can be particularly suitable in the case of railway crossings, because some parts of the inputs are very hard – or impossible – to express in monetary terms. In the case of railway crossings the possible inputs are involving a quite wide range.

Narrowing the input's range is necessary, and only those inputs can be taken into account which can basically influence the expected output. During the modelling process the railway crossing can be created as a replica which contains the essential features of the examined system.

3.2.1. The Model Creation Process

The first step of model creation is the determination of the inputs which are suitable for describing the railway crossing as a system and the necessary narrowing of the inputs' range is also must be performed – in such a way we described it in the previous chapter. The exact definition of the certain inputs are also belong to the first steps. The main groups of inputs can be the following:

- the environment of the railway crossing
- the security method of the railway crossing
- the accidental status of the railway crossing
- the traffic data of the railway crossing
- the way of traffic management and traffic arrangement in the railway crossing.

After that we can determine the most favourable (Z_{maxi}) and the worst (Z_{mini}) values of the specific inputs. Using the most favourable, the worst and the effective input values we can generate the input score (4) which can be given for certain inputs.

$$l_x = \left\{1 - \left[\frac{Z_{ix} - Z_{mini}}{Z_{maxi} - Z_{mini}} \times \left(1 - \frac{Z_{mini}}{Z_{maxi}}\right)\right]\right\} \times 100 \qquad (4)$$

where:

l_x: input scoring number

Z_{ix}: the effective value of the input (the i-th input within the x-th input group)

For setting up the hazard ranking the determination of the total score per crossing is necessary, according to the aggregation of the input indicators (5).

$$Al_{xe} = \sum_{i=1}^{n} I_{ix} \times W_{ei} \qquad (5)$$

where:

AI_{xe}: the aggregated value of the inputs

I_{ix}: the score of the i-th input

W_{ei}: the preference weight of the i-th input from the perspective of the given expert group

In the conceptual model the preference weight of the inputs is determined by an expert group, the aggregated value can be understood as the total score of the road crossing. The higher value means a higher hazard rate.

4. The Establishment of the Hazard Raking with the Usage of Evaluation Matrix

After the examination of the potentially attractive models we described in the previous chapter, a so-called Evaluation Matrix was created. Considering practical and theoretical aspects it can be stated that the Evaluation Matrix is a flexible method which provides in every case of railway crossings a useful solution and a basis for reconstruction catalogues for railway crossings.

4.1. The Structure of the Evaluation Matrix

The Evaluation Matrix is based on the simplest data sorting procedure; on the sorting by magnitude. The elements of the sample are corresponding to the railway crossings involved to the examination. The scores given for crossing features are summarised by crossings, and the received amount is being sorted by their size. The resulting sequence is the ranking and the scores mean the rank numbers.

The Evaluation Matrix separates three main input groups (Fig. 2):

- accidents
- traffic status
- traffic management characteristics

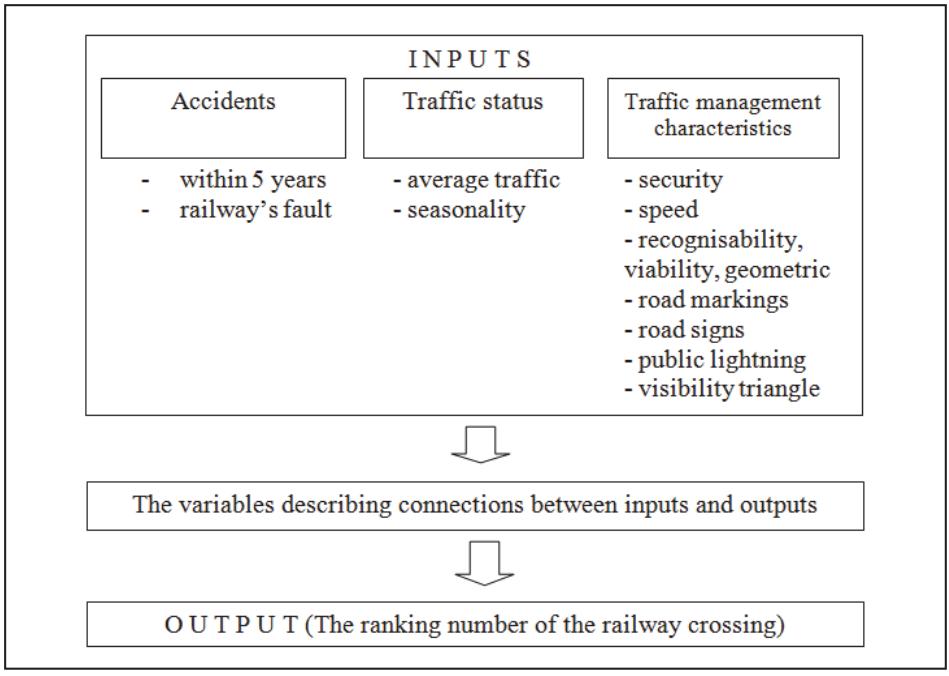

Figure 2. The Input System of the Evaluation Matrix

The main contents of the input groups are essentially equal to the groups discussed in the case of the 'GySEV model', but also there are some significant differences. In the score values' calculation extra scores are given for the conditions which are the railways' fault. The past five years are the basis of the calculation when we calculate the accidental status' score values. The seasonal traffic growth in the crossing is taken into account in the traffic status input group. The Evaluation Matrix calculates with the daily average roadway and railway traffic. This element allows the easier practical application of the matrix. The Government Regulation No. 20 of 1984 about 'the traffic control of roads and the placement of road sings'- which is still in force – is mentioning the 'standard railway crossing's traffic'. The feature of traffic standard had been built into the model with the existence or the non-existence of the seasonal traffic growth. Contrary to the government regulation, the average daily railway traffic isn't counting with that trains only which are 'passing through the crossing by schedule'. In relation to speeds, the basis of the calculations is the highest allowed speed (80 km/h) of regional railway lines; therefore only two speed groups had been determined. The highest road speed allowed in a railway crossing is divided into four classes. New elements were picked up into traffic management characteristics such as:

- the contrast effect of sunshine persists for several hours due to the crossing's location in the case of road traffic
- the 'Start of level (railway) crossing' board is faded or damaged (even in a single board's case)

4.2. The Modelling Process

When the Evaluation Matrix was established, the input range was demarcated and the certain inputs were defined on the first stage, and after that they were merged to input groups. The Evaluation Matrix's processing algorithms are taking into account the requirement of the simple, fast and reliable practical application. The existence or non-existence of any input elements can be evaluated by given scores. Scores can be given to the railway crossing's traffic status by reference (6).

$$\mathrm{ADT} = \mathrm{ADT}_{\mathrm{road}} \times \mathrm{ADT}_{\mathrm{rail}} \qquad (6)$$

where:

$\mathrm{ADT}_{\mathrm{road}}$: Daily Average Roadway Traffic

$\mathrm{ADT}_{\mathrm{rail}}$: Daily Average Railway Traffic

If the value of NÁF is below 800.00 then the score can be given for traffic status will be allocated proportionally, but if NÁF is more than 800.00 then the maximum score number must be given. The outputs and the score values which can be given to each input elements (input groups) can be generated by the help of the processing algorithms. After the model is designed, it will be filled with values, and the feasibility of the outputs will be checked (Figure 3).

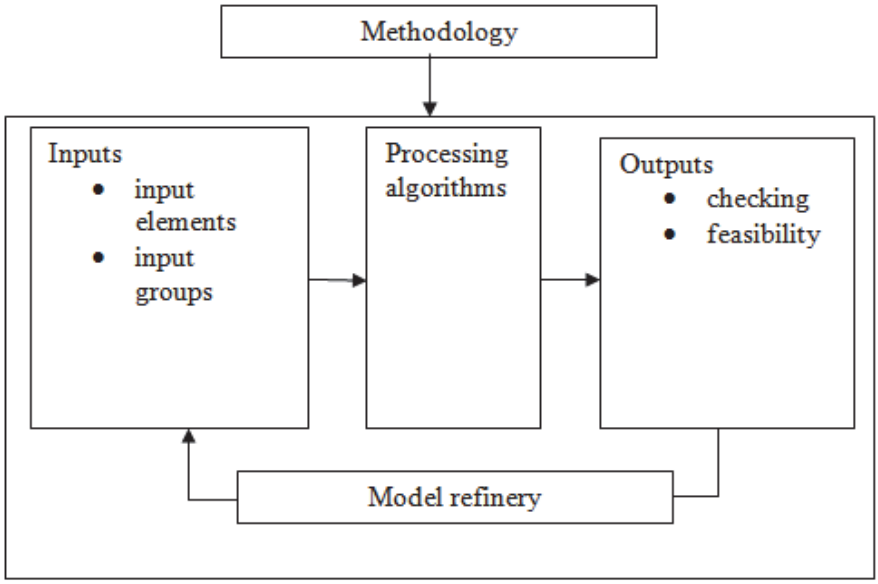

Figure 3: The Modelling Process

5. Describing the Crossing Designated for Reconstruction

The reason of designing the Evaluation Matrix we discussed in the previous chapter was the fact that in many cases the railway crossings' classification by conventional examination elements was unable to give a realistic hazard picture for the given crossing (for example according to some studies and analyzes the serious accidents are usually happening in crossings with smaller traffic [3]). There was a single score number or value limit determined for the inputs (evaluation criteria) of the Evaluation Matrix. The weighting of the given score value system is based on the concerning standards, existing regulations and legislations [4] [5] [6] [7] [8].

In the determination of some input factors' value system some results was also used which had been calculated in previous work packages (as connection points) such as AP3, traffic counts etc. Beside the above-mentioned, more important criteria elements, which were used in previous evaluations, other input factors are parts of the matrix, which are also must be followed (for example consumer questionnaires, observations and expectations raised by opinions –as input factors).

According to the aggregation of each input elements' score value, every crossings got a kind of so-called 'hazard total score'. Based on this score in terms of the examined crossings' safety the crossing which has the most score number is the 'most dangerous', and the crossing which has the lowest score is the 'most secure'. However it is very important to highlight that every of the selected 28 (23+5) crossings in the project are suitable for the currently existing provisions and regulations by the special authorities' classification.

5.1. The 'Hazard Ranking' of the Examined Crossings and the Most Dangerous Crossing

From the examined crossings (GySEV railway lines) - according to the hazard ranking was established by the results of the crossings' evaluation matrix - the most score value (1574 points) and 'the most dangerous crossing' title was given to the AS765 railway crossing on the Line No. 8. (the 'most secure' crossing is the AS759 on the Line No. 15, which had got 306 points.) The life-likeness of the theoretic criteria system developed by the working group was justified by the unfortunate fact that there were two accidents on that crossing in the past half years, and one of them had a fatal outcome. At the same time it is important to remember that in its current state the crossing is suitable for every regulatory requirements and regulations, just as we mentioned it above (the last special authority traversal was on 24th July, 2013) [9]

5.2. The Security Features of the Crossing

In relation to the safety features of the crossing - based on each output's scores - the following statements can be determined:

- 'Accidents' input: from the inputs – as one of the most significant examined aspects - the examined crossing (AS765) had got the maximum score (800 points; two or more accidents within 2 years) in terms of accidental scores.
- 'Traffic status' input: the traffic status in terms of input factors (daily average railway/roadway traffic, seasonal traffic growth etc.) can be considered average or slightly below it compared with the other examined crossings (total score: 184 points).
- 'Traffic Management Characteristics': In terms of traffic management characteristics the crossing is the fifth most dangerous from the examined crossings based on the amount of each input factors' score (total score: 590 points).

From the matrix's columns which are containing the total scores ('Ranking by total scores') it can be obviously determined that with its 1574 points (which is the amount of each inputs' scores) this crossing is the most dangerous not only from the selected 23 GySEV crossings, but from all of the examined ones (GySEV + Raaberbahn).

5.3. Establishment of the Crossing No. AS765, Environmental Characteristics, Accidental Statistics

The crossing selected for reconstruction is situated in the level junction of the No. 8526 public road's 1 + 534 km gauge (the link road of Kópháza – Balf – Fertőrákos) and the No 8. GySEV railway line's 765 + 94 gauge (GPS coordinates: N47, 38' 25,1"; E16, 39' 40,9").

5.3.1. Traffic Management and Environmental Characteristics

The railway and roadway characteristics of the No. AS765 crossing according to the condition-measuring before the reconstruction proposal:

- Railway:
    - The crossing is equipped with four masts and a light barrier, and it is operating automatically by the train (omnidirectional)
    - there is no pedestrian passage or built traffic safety island
    - One track; the path's speed is max. 100 km/h
    - The pavement of the railway crossing is STRAIL

- Roadway:
    - number of lanes: 2x1
    - 'Start of level (railway) crossing' signboard in the crossing ('St. Andrew's cross')
    - 'Level (railway) crossing' predictor and additional boards from both directions (repeated on the left and right travel directions)
    - the allowed speed of the public road is 90 km/h (public domain), speed limit before the crossing (40 km/h) positioned on the triband predictor board from both directions
    - 'Dangerous bend' predictor board from both directions
    - Road bike passage to the public road before the crossing (from Kópháza)
    - connecting dirt roads (within 22 meters) from both directions
    - the crossing angle is 90°
    - the visibility triangle is reduced

5.3.2. Accidental Statistics

The accidental statistic of the crossing in terms of the past 12 years can be summarized in the following:

- 30/08/2002 - fatal accident, 2 people
- 19/09/2009 - slight injuries, 1 person
- 29/06/2013 - slight injuries, 1 person
- 16/07/2013 - fatal accident, 2 people

6. Proposals for the Reconstruction

From the actions aimed to improve the safety of railway level crossings the barrier-mounting program and the LED-program had got big boost in Hungary after 2003. According to experts' statements the LED reduces the possibility of accidents with

50%, and the barrier + LED construction reduces it with 85%. Beyond all of that it is advisable to examine the possibility of deploying further safety-enhancing actions.

6.1. Safety-enhancing Actions from the Recent Past

Because of the accidents from the past few years in the crossing designated for reconstruction (which were too frequent compared to the traffic conditions – two accidents, one of them is fatal within a month in the summer of 2013), there were more authority traversals in the near past. [9] [10]. Each time on these site inspections it was determined that the crossing is suitable for all of the concerning regulations and legislations.

As a result of these traversals different safety-enhancing actions had been implemented (replacement of the worn predictors, wood felling, thinning of vegetation on account of better visibility, placement of predictor boards to the connecting municipal roads, reducing the distance between 'striped' predictor boards and placing them to the same axis line and the same distance from the roadway, the replacement of the railway light signalling device to LED optics etc.).

Despite all of these actions the accidental statistics confirms that the crossing is remained extremely dangerous. The rating of the Evaluation Matrix also refers to this fact.

Considering these facts we can determine that further changes and modifications are required in the crossing in order to increase the safety and to reduce the emergency phase – also from the railway and the roadway part.

6.2. Further Security-Enhancing Suggestions for Modifications and Improvements

The preparation of a reconstruction concept is also a part of the SiEBaBWe project's working package, which is intended to taking action recommendations in order to increase the safety of the object classified as the most hazardous from the 23 (+5) examined crossings. For this purpose the partners involved to the established professional working group measured the potentially attractive opportunities on a common traversal, and they worked out proposals together. The accepted draft proposal has affect on the railway infrastructure and also on the roadway infrastructure.

6.2.1. Proposals for Railways

Based on the current railway infrastructure the following modifications and additions had been recommended:

- the installation of a half-barrier in the crossing and the necessary replacement of its signalling masts within the prescribed distance [11] as its consequence (comment: this installation is a part of the medium-term development program of GySEV) /Figure 1, points marked with (a)/
- the moving of the additional light signalling mast to the prescribed [11] maximum allowable 8 meters distance from the path axes (Kópháza side, mast marked with 'D'), and its rotation to the appropriate direction (viewing angle) in order to the better visibility and the earliest perceptibility for the road traffic (a

hazardous road bend with small radius curving to the left, with the crossing immediately on the end of the curve) /Figure 1, point marked with (b)/

- the moving (increasing) to the appropriate distance of the barrier-actuator equipment's so-called interlocking system's exposure points to ensure the prescribed pre-flashing and pre-closing time. This action's necessity is justified by the requirement of ensuring the prescribed value which had been changed due to the supplement of the half-barrier [11]. The current distance of the exposure points (light assisted security only) and the required values calculated with the supplement of the half-barrier are shown on Table 3.

Table 3. Determining the Installation Distance of the Exposure Points [11]

AS 765	Currently		Calculated		Core data			
	Middle	**765+48**	**A2 transmitter**	**A2 receiver**	Angle (°)	J90	$\mathbf{B_t}$	2
-2080,0	**A2**	**777+11**	**756+31,0**	**756+13,0**	Number of Masts	4	$\mathbf{L_v}$	6,01
	A4				Barrier	Yes	$\mathbf{t_{min}}$ (c.)	29,4
-1993,0	**A3**		**A1 transmitter**	**A1 receiver**	Road Width (m)	6	$\mathbf{t_{min}}$	33
	A1	**754+72**	**774+65,0**	**774+83,0**	Track Speed (km/h)	100	$\mathbf{l_{beh.}}$	917

$\mathbf{b_t}$: Safety distance (m)
$\mathbf{l_v}$: the length of the endangered road section (m)
$\mathbf{t_{min}}$ (c.): minimal pre-closing time / calculated/ (sec)
$\mathbf{t_{min}}$: minimal pre-closing time (sec)
$\mathbf{l_{beh.}}$: exposure distance (m)

6.2.2. Proposals for Roadways

The following recommendations had been placed into the roadway's modifications package as the part of the reconstruction concept:

- The curve widening of the public road No. 8526 in the road bend before the crossing on the Kópháza side. This widening is necessary because long vehicles (for example hinged buses) are also travelling on this section of the public road, and a possible bidirectional encounter on the given place in the same time could cause a dangerous situation (particularly regarding to the relative short distance of the unloading section before the bend). It is also confirmed by the fragmented and dimpled roadside we experienced during the site visit. /Figure 4, point marked with (1)/

- Putting the roadway's Ferro-concrete crash barriers and the concerned mast to the same line in the crossing while the prescribed distance is maintained [6]. These crash barriers can be exchanged with guardrails where necessary. In case of exchange the incorporation of reflective prism elements onto the barrier would be advisable for better visibility. /Figure 4, point marked with (3)/

- Establishing ribbed transversal shaker lanes (single, double and triband) on the pavement (in vivid yellow colour) in the travel direction's band from both directions ('hazardous area predicting' lines). For the better visibility it is advisable to place reflective prisms on the shaker lanes' edge as well as the fixing of modern and cost effective solar powered flashing prisms. The application of these elements are highly increase the visibility and it has a powerful awareness raising role - mainly in darkness or among bad or restricted visibility conditions (Nowadays these latter accessories are more and more often used in those areas which are particularly dangerous in terms of public road traffic). /Figure 4, points marked with (2)/

- Moving the striped predictor boards which are indicating the railway crossings (single, double or triband boards) to the possible prescribed [4] minimum distance still fits to the given circumstances. This modification is necessary because according to previous experiences the settlement structure and the traffic conditions and habits have similar features than populated areas have, in spite of the rural classification of this public road section. /Figure 4, points marked with (6)/

- Moving (from both directions) the Speed Limit board (40 km) to the single band predictor board (because of the reasons mentioned in the previous section) /Figure 4, points marked with (6)/

- Placing a 'board indicates the bend's curve' (red herringbone-like herding boards with yellow framework) directly to the start of the road bend /Figure 4, points marked with (7)/

- Placing 'no overtaking', 'no turning' and 'mandatory driving direction' boards on the public road and on the connecting dirt roads because of the barrage line /Figure 4 points marked with (4)(5)(9)/

- Moving the bicycle path's current exit lane (aimed to the public road) from Kópháza preferably to a destination which is farther from the crossing in order to reduce the concentrated emergency situation. More variants are emerging in relation to the exit lane (adjusted to the traffic and usage habits): on one hand the placement of the exit lane to the existing opposite-directional connection point from Balf (passed through the public road), and on the other hand the restructuring of the current vectoring point in the bend and its 'shifting' towards the connecting dirt road. /Figure 4, points marked with (8)/

- The placement of an informative, awareness raising board or poster, which has appropriate hazard warning and awareness raising affect with its concentrated message-bearing graphical and image-based elements on every traveller who uses the railway crossing. /Figure 4 points marked with (11)/

- The equalization of the road surfaces' level difference, its minimization between the road surface and the crossing's surface (STRAIL) with the proper modification of the connecting public road's surface ('trimming') /Figure 4, point marked with (12)/
- The replacement of the 'level crossing with half barrier and light barrier ahead' board on the triband predictor board /Figure 4, points marked with (10)/.

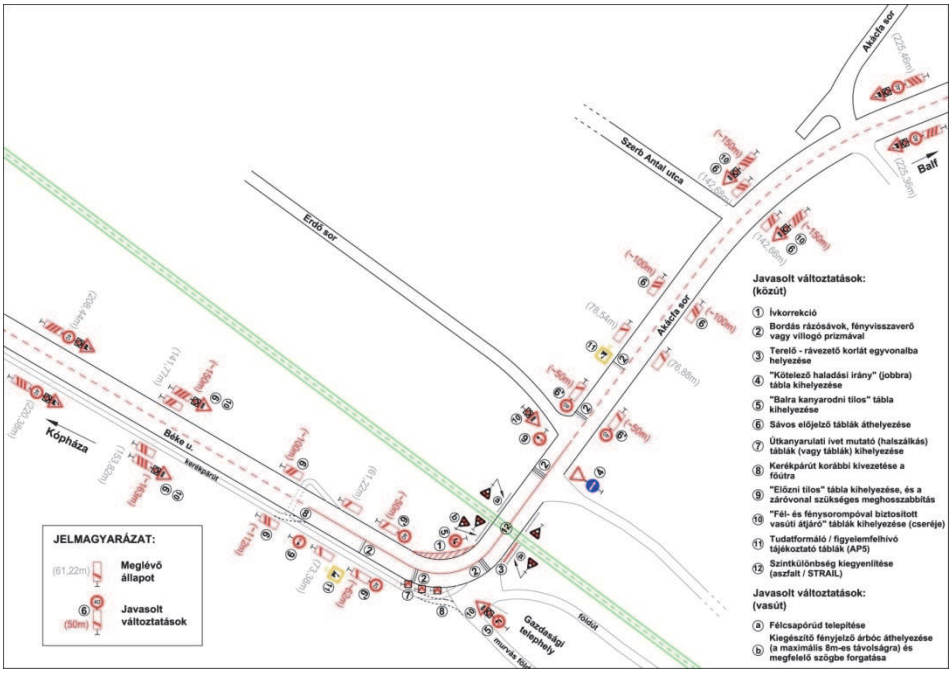

Figure 4:
The Site Plan of the Crossing No. AS765 and the Railway-Roadway Connection

For the implementation of the listed proposals the preparation of a preliminary plan is required both from the railway and the roadway side.

6.3. Preliminary Cost Estimation and Time Schedule Plan

From the alternatives of the reconstruction concept the preliminary cost estimation - which contains the planning and the implementation costs in relation to the finally accepted modification proposals - will be performed by the project partner (GySEV) within their respective powers.

The expected realization time of the accepted reconstruction concept's complex implementation will be the end of the year 2015, depending on the availability of the sources.

Conclusion

Considering the expected effects in relation to the reconstruction at first we have to take into account the fact that based on statistical studies the crossing's supplement with a half-barrier is significantly increases the safety in the given crossing [3] [13]. In addition we can speak about a financially beneficial investment, which can be justified by economical calculations [13]. However the other modifications and additions proposed in the project are providing more information to the user of the crossing: in one hand the traveller will become more aware of the hazardous zone's approaching, on the other hand these information are helping in conscious formation of the necessary and desired traffic behaviour, therefore the chance of occurring accidents will be further reduced. This latter conclusion indicated the renewal of the program series called ILCAD (International Level Crossing Awareness Day), which is organised annually from 2009 with the coordination of the International Union of Railways (UIC) and with the support of the European Commission (EC) and the United Nations Economic Commission for Europe (UNECE). The slogan of the ILCAD is: 'Act safely at level crossings'. The main goal of this international conference series is the same as the project SiEBaBWE's main objective; to effectively increase the safety of the level roadway-railway crossings, to reduce the number of accidents and in long term to reach the totally accident-free traffic.

References

[1] Tigyi Sz, Gábor M: The Safety Status of the Railways in Hungary and the Development of Their Hazard Ranking. XI. RODOSZ Conference, Kolozsvár 12–14 November 2010.
[2] Pokorádi L: Modelling Systems and Processes. Campus Publisher, Debrecen, 2008.
[3] Konrád Gy: Accidents and Railway Crossings Operated on the lines of GYSEV Zrt. GYSEV Zrt., Study on 15/05/2013
[4] Government Regulation 20/1984 (Xll.21)
[5] Standard of KPMSZ. Kk No. 108-73 and It's Modification (E 07MSZ.-07 Kk.108-73 M/1983)
[6] Road Standards No. ÚT2-1.201
[7] Road Standards No. ÚT2-1.225
[8] Technical Guidance of MÁV No. D54. (Technical Data and Regulations of Construction and Track Maintenance I-II.)
[9] Memorandum of Police Headquarters, Győr-Moson-Sopron Country, Law Enforcement Department, Case No.: 0800/4088-7/2013.), Győr, 30/07/2013.
[10] Report, No. NKH KÜI KU/VF/1853/0/2009 Report and the Associated Report, Sopron, 16/11/2009.
[11] 'Light Barrier Requirements' E07 MSZ.-07 Kk.108-73 M/1983 Standard and It's Additional Modification
[12] Traffic Safety 02/2011, NKH, 2011.
[13] Tremmer T: A sorompó megvéd a haláltól. Napló Online, veol.hu [Last visited: 19/05/2014]

Acta
Technica
Jaurinensis

Vol. 8, No. 2, pp. 113-130, 2015
DOI: 10.14513/actatechjaur.v8.n2.332
Available online at acta.sze.hu

Historical Development and Special Building Structures of In-earth Embedded Houses

D. Bozsaky

Széchenyi István University
Department of Architecture and Building Construction
Egyetem tér 1, 9026, Győr
E-mail: bozsaky@sze.hu

Abstract: Earth houses have been traditional residential buildings for a long time and their three different types (in-hill house, cave dwelling and atrium house) were separated during the early history. Modern earth houses appeared in the 1960s in the USA but spread only after the first oil crisis. Earth-sheltered houses became popular nowadays because of their several benefits (environmentally friendly, economical, energy savings, landscape protection). Unfortunately they have some disadvantages (negative prejudices, high building costs, special building constructions, complicated design and construction). Several design aspects (orientation, ventilation) are extremely important for in-earth embedded houses. Because of the non-conventional loads and landscape situations they have special building and supporting structures. They had also extraordinary damp-proofing and thermal insulation problems, this is why special solutions are needed to have pleasant and comfortable indoor atmosphere.

Keywords: in-earth embedding, earth-sheltering, special building structures

1. Introduction

In-earth embedded houses are not recent forms of the residential buildings. This tradition can be traced back thousands of years. The earth architecture today has been developed so far that it is now regarded as an architectural guideline, which can become a key element of the sustainable development of built environment. This approach may be able to solve the aesthetic and environmentally harmonization with the landscape; the smart energy consumption and the healthy living environment can be taken into consideration in the same time.

This paper shows the historical archetypes of the kinds of in-earth embedded houses from ancient times to modern earth architecture. It summarizes the most important specific design aspects, advantages and disadvantages of these extraordinary buildings. It introduces the special supporting structures and the main damp-proofing and thermal insulating problems of them.

2. The Historical Archetypes of In-earth Embedded Houses

Residential buildings that were embedded into earth always served as human shelters. In the beginning, people used naturally shaped natural formations (caves) used as protection against elements and wild animals. Later people were able to create habitation for themselves with their own creative activities. Since then, people built residency with their own hands that we call construction activity.

At first, natural materials for construction were used. Then earth became important building material such as wood and stone. From hillsides, where the hardness of rock allowed artificial caves were able to slot with the use of simple mining tools [6]. Where soil conditions allowed in-ground pitted houses could be made (Fig. 1). Their walls were entirely or partially formed by the side walls of the pit dug into the ground [11][12]. The roofs of pit houses were probably made of stone or wood that were covered with earth.

Figure 1: Draft of a half-pitted house [6]

This ancient type of residential house was applied for a long time, because the construction was easy and cheap; in winter it was easy to be heated; it ensured pleasantly cool indoor climate in the summer and protected against wild animals, and, in addition, it was hard to burn down during a potential robbery attacks [4].

3. The Types of In-earth Embedded Houses

Three types of in-earth embedded houses were developed throughout ages that are called cave house, in-hill house and atrium house (Fig. 2).

Figure 2: Types of in-earth embedded houses: cave house, in-hill house and atrium house [4]

3.1. The Cave House

The basic concept of man-made cave house comes from the image of natural cave. The cave dwellings can be built easily in hilly, sloping terrain. Typically it has one front facade. The other facades, side walls and roof are covered with earth. It warrants the

freest architectural forming, because cutting into the hillside any form of architectural plan is available and it is suitable for multi-level layout.

In mining areas every citizens lived in such houses. Best known example is the city of Matera (Fig. 3) in Southern Italy. The ”sassi” dwellings – that were declared as a part of the World Heritage in 1993 by UNESCO – were inhabited until 1950s [19]. In some cities of (e.g. Troo) the Loire Valley (France) and in Granada (Southern Spain) part of the population still live in cave houses hollowed into the rock.

Figure 3: Cave dwellings called ”Sassi” in Matera (Southern Italy) [19]

The underground cities in the area of Cappadocia (Turkey) are also well-known worldwide. In the relatively tractable volcanic tuff cut cave dwellings were shelters of the early Christians. The most famous in-rock villages in this area are Kaymakli, Derinkuyu.and the fortress of Uchisar. Similar homes we can find near Kandovan (Iran), Gondrani (Pakistan), in several provinces of India (Aurangabad, Rajagriha) and in the Caucasus (Georgia, Armenia).

Along the Yangtze River (China) we can meet numerous cave dwellings cut into the loess wall. In Chinese called "Yaodong” homes have been used for centuries, especially by the poorer classes (Fig. 4). Nearly 20 million Chinese still live in cave houses such along the river [14].

Fig. 4: „Yaodong" cave dwellings cut into loess wall (Yangtze river, China) [14]Lots of rock-cut cave dwellings are located in the rocky and arid regions of North America. The best known region among these is in the Mesa Verde National Park (Colorado) in Santa Clara Canyon (New Mexico) in the United States. The finest examples of cave dwellings are the nearly 1000 year-old Quarenta Casas in Mexico, Chihuahua province that were built by the Tarahumara Indians.

At present, many people live in cave dwellings in North African countries. A well-known example is located near the city of Bandiagara (Mali), where the members of Tellem tribe built a complete settlement of rock-cut cave dwellings.

Many cave houses erected since the 13th century are located in Hungarian mountain areas (such as Gerecse Region, Northern Mountains, Upper Balaton Lake Region, Sokoró Hills, Tolna Hills). Two main types of these dwellings are: natural caves used for housing (hermit caves) and artificially formed cave dwellings cut into limestone, loess, or volcanic tuff.

Best known hermit caves (Fig. 5) are the Saint Ivan Cave, the Báthory Cave (Buda Mountains), the hermitages at St. Michael's mountain (Börzsöny Mountains), the hermitages at Mátraverebély-Szentkút (Fig. 6), and the monk apartments in Tihany (Lake Balaton) [9].

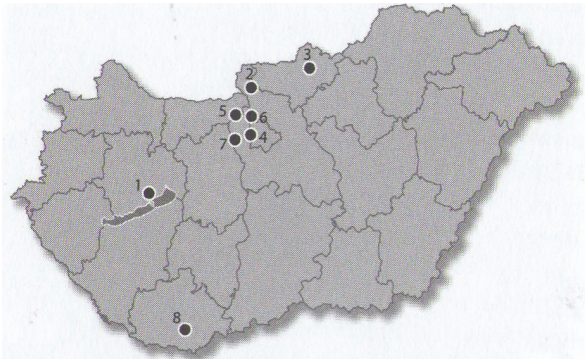

Figure 5: Hermitages in Hungary. 1: Tihany; 2: Saint Michael's mountain; 3: Mártaverebély-Szentkút; 4: Saint Ivan Cave; 5: Máriaremete; 6: Báthory Cave; 7: Budaörs; 8: Jakab Hill [9]

Figure 6: Hermitage in Mátraverebély-Szentkút (self-made photo)

Man-made cave dwellings (Fig. 7) were applied at territories where population dealt with the earth (mining, quarrying and viniculture). Several cave dwellings are located in Nógrád and Borsod-Abaúj-Zemplén County cut into volcanic tuff. These residents were mainly created for the poors in Cserépváralja and Cserépfalu. 10-15 of the cave dwellings in Szomolya (near Mezőkövesd) are still in use [9].

In Heves County (particularly at Eger area) several cave dwellings are cut into volcanic tuff. From the 19th century the majority of the population in the village Noszvaj lived in cave dwellings and even in the 1970s 40 existing in-earth apartments were registered. In Egerszalók in the 1980s were many one-room cave dwellings that were built in the turn of the 19th and 20th century. Many of them are now under protection [9].

The cave dwellings and wine cellars were cut into rhyolitic tuff in Andornaktálya. Numerous wine cellars were used as dwelling-house in the first half of the 20th century in Sirok and Ostoros [9].

Several notifications from the 18th century mention the (mainly limestone) rock-cut cave houses in Buda Mountains. 24% of the population of Budafok (2700 people) lived in such houses [2]. In the 1950s in Budafok and Budatétény 374 cave houses were registered; many of them were used as residential buildings until the 1970s [2].

In the area of Transdanubia we can find man-made cave dwellings cut into loess walls in Upper Balaton Lake Region (e.g. Balatonkenese, Balatonendréd, Ádánd), Sokoró Hills (e.g. Pannonhalma, Nyúl), Danube Region (e.g. Dunaföldvár, Paks, Dunaszekcső) and in Tolna Hills (e.g. Szekszárd, Iregszemcse, Miszla, Ozora).

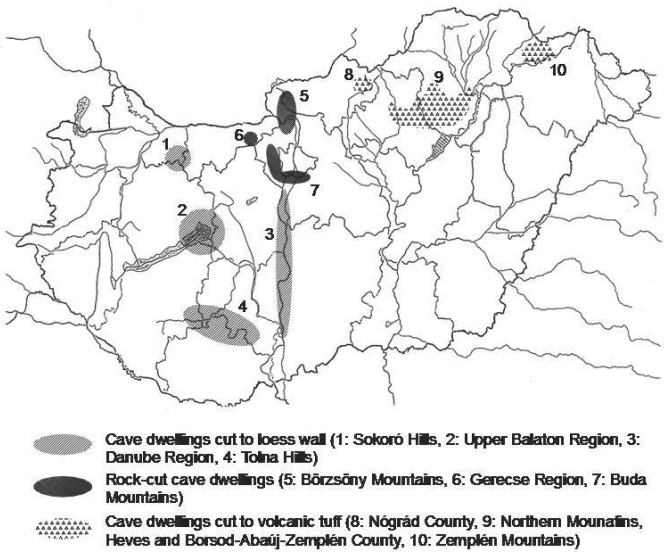

Figure 7: Cave dwellings in Hungary (self-made figure)

3.2. The In-hill House

The pit houses are considered as the archetype of in-hill houses. The floor of this ancient type of earth houses was sunk underneath the earth, the side walls were made of stone or wood and the roof was covered with earth.

In-hill houses can be formed typically on flat or on gently sloping areas. The architectural design, orientation and surfaces of the facade is more freely formed than in the case of the cave houses. It may happen that the building has only a front surface, but if required, additional openings may be implemented. Light and ventilation conditions may be more beneficial than the cave house conditions. The disadvantage is that mainly single-story floor plan is available and it has potential for large-area needs. The most ancient (more than 3000 year-old) in-hill houses called Skara Brae are located in Orkney Island (Scotland) (Fig. 8) [22].

Figure 8: The ancients earth houses of Skara Brae (Orkney Island, Scotland) [22]

This type of housing was also used and is still used today in North America by the Navajo Indians (Utah, USA). This traditional Navajo house called "Hogan" generally shaped like spherical dome with a transom light on the top (Fig. 9) [18] [20]. Shape of the ground plan is ordinary round or polygon, the frame structure is timber with earth covering. In California State (USA) a similar type of housing called "Mongollon" was used by the native Indians.

Figure 9: A Navajo "Hogan" [18]

While in-hill houses located in California and Utah were primarily covered with earth to protect interior against summer heat, in Europe and North America in-earth embedded houses and earth covering used as protection against hard winter frosts. In Alaska, Scandinavia, Iceland (Keldur, Sænautasel) and Russia (called "foxhole") were built in-earth embedded and soil-vegetation-covered houses.

3.3. The Atrium House

An atrium house can be built underground in flat landscapes. Side walls and roof covering are made of earth. The reason of the birth of atrium house - unlike the other two housing types - especially not winter cold, but protection against summer heat. The building has a central atrium space (patio) which is usually open at the top, and the residential rooms around it are grouped around the courtyard. Solid side wall surfaces completely lock the outside world out. The room lighting is possible only with roof skylight windows and openings to the inner courtyard.

We can meet with the ancient type of atrium house sin Mediterranean region, embedding into earth is more typical in North Africa and in the Middle East. The best known example of atrium houses we can find near the village of Matmata (Tunisia) (Fig. 10) [7]. Although primarily intended to protect against the summer heat, but the old days as an excellent hiding place they were able to mislead the enemy. During construction, at first approx. 5-15m in width and the same depth, round-shaped hole is dug. Then starting from the side wall of the pit the rooms were cut into earth like tunnel building. Similarly buildings can be found in the village of Rocheménier (France).

Figure 8: The courtyard of an atrium house is Matmata (Tunisia) [10]

4. The Modernist Earth Architecture

The roots of modern earth architecture can be found in the works of modern architect in the early 20th century. In 1926 Le Corbusier (1887-1965) formulated a principle in his work entitled Five Points of a New Architecture that says:

"Buildings should stand on pillars accordingly not to take place in nature, and allow vegetation below. Roofs should be flat and they need to be installed with vegetation."

These rules he applied for the design of Ville Savoye in 1929. The American architect, F. L. Wright (1867-1959) can be considered as a precursor of modern earth architecture. One of his most famous works is the Fallingwater house (Mill Run, Pennsylvania, USA), that was designed for E. Kaufmann (1885-1955) in the 1930s.

Designing this building the purpose of the architect was to bring up man and nature as close as possible. It came true with the mastery utilization of terrain and the conduction of stream through the house [4].

However the works of these outstanding figures of architecture were only harbingers of the campaign that we call modern earth architecture today. The main principles of this new architectural trend were born in the USA in the middle of the 20th century.

Malcolm Wells (1926-2009) is sometimes regarded as the „father of modern earth-sheltered architecture". His first modern in-earth embedded residential building was built in Cherry Hill (New Jersey, USA) in 1964 (Fig. 11) [5].

Figure 11: The earth house of Malcom Wells in Cherry Hill (New Jersey, USA) [5]

He found many followers after the oil crisis of 1973. This time people realised that the amount of fossil fuels are running out, and excessive energy consumption leads to a level of greenhouse gas emissions, that results global warming and climate change [5].

Since then, lots of designers proclaimed themselves as an earth architect and started their activities. In the United Kingdom the most important deputy of earth architecture was A. Quarmby, who designed the first modern British earth house in 1975 in Huddersfield, near Peak District National Park. Another important earth architects are the American M. Muenning (1935-), the Argentinian E. Ambasz (1943-), the Serbian V. Milković (1949-), and the Swiss P. Vetsch (1943-), who planned 47 earth houses since 1978. Vetsch's unique architectural design is a successful alloy of the style of A. Gaudí and Art Nouveau. Its famous Hungarian followers are Zs. Hegedűs, Zs. Kassai and I. Kovács (Fig. 12) [8] [17].

Figure 12: In-hill house planned by Zs. Hegedűs (Bőny, Hungary, 2007) [8] [17]

5. The Design Aspects of In-earth Embedded Houses

One of the most important aspects is the configuration of terrain. On a steeply sloping terrain a cave dwelling is advisable. In-hill house and atrium house is more conceivable on a flat terrain [4].

The direction of sloping is also determining, because an in-earth embedded house has limited ways of facade openings. The southern side of the mountain is ideal for cave dwellings (on the northern hemisphere). From south the house more sunshine from south and it is easier to ensure enough natural lighting for the living spaces [15] [21].

Important influencing factor of designing is the specific climatic conditions dominating in the country. In Scandinavia the protection against cold and the maximization of winter sunshine utilization, in the Mediterranean region protection against summer sunshine and the shielding is primary.

The average rainfall and air humidity is also an important point of view. Planning the orientation of an earth house the main direction of wind is also a serious design aspect. On the northern hemisphere cold winter winds blow from north and northwest, so facades should not be opened in this direction [13].

It is essential to have regard to the soil conditions. Most essential soil properties to pay respect to are density, water permeability and thermal insulation quality. An ideal soil is relatively light (less gross weight to take into account), has low potential to erosion, leads easily the unnecessary rainwater away (no lasting moisture effects and risk of frost), has good thermal insulation quality and retains the vegetation well [3].

Unfortunately, these needs often lead to contradictions. Hard ground (e.g. clay) has advantage that it has less potential for erosion and good thermal insulation quality. At the same time water permeability of clay is low, this is why thermal motions and effects of frost are more dangerous.

The level of groundwater is also an important design aspect. The designer needs to find an area, where moisture can leek over easily otherwise serious groundwater pressure must be accounted during the structural design. Up-dammed water needs increased attention. It is recommended to ensure the excess moisture in the shortest possible way to escape away from the structure. It is available with plantation, landscape planning, a building out an appropriate drainage system.

6. The Advantages of In-earth Embedded Houses

In-earth embedded houses represent a special architectural trend, because they have not only aesthetic appearance, but they can be key elements of sustainable architecture because of their long lifespan and low energy consumption [4] [21].

Compared to conventional buildings, earth houses fit perfectly into their surroundings. A nicely designed earth house blends into the environment like it has always been there because of its low built-up area requirement. Large green areas can be developed on the building site. Proximity to nature provides a healthy home for residents. Unlike traditional buildings they can be more easier formed and built on sloping terrain [13].

Because of the thick earth-covering on the roof an in-earth embedded house is much less exposed to environmental effects like a traditional building. Because of the earth-covering an in-earth embedded house keep the heat in much longer, and it cools down slower. This thermal lag provides low temperature fluctuations inside. Air can filtrate through earth-covered walls and roofs uneasily this is why heated inside air cannot wind

out so quickly. Consequently energy consumption of in-earth embedded houses is much lower than traditional buildings.

Earth-covering provides protection against extreme weather conditions like storms, hurricanes, tornadoes. Owing to its fire resistance, earth-covering features efficient fire protection. This is why earth embedded houses is preferred for fire safety reasons.

Not to forget either the excellent sound insulating properties of earth. A 60-300 cm thick earth-covering is able to absorb the noises of a busy highway. Inmates are completely isolated from the surroundings, so pleasant, relaxed living space guaranteed.

Illumination and ventilation problems can be solved with correct orientation and smart utilization of facades. With the help of careful planning an in-earth embedded house can be lighter and more airy like a traditional building.

7. The Disadvantages of In-earth Embedded Houses

It is a false belief that underground living areas are dark, narrow cold and damp. Due to this misconception most people are deals critical with the problems of in-earth housing that blocks them to spread widely [4].

They are designed with great care, skill and high level of knowledge in building construction, building materials and building physics. If the designer does not have these competences the interior climate of a poorly planned in-earth embedded building can be really unpleasant.

Inadequate ventilation can abuse the indoor air quality. It can cause humidity problems that can be solved only by installing expensive mechanical ventilation equipment. Wrong orientation and improper design of lantern-lights or windows lead to solar access troubles.

Maintenance costs of an in-earth embedded building are usually much lower than a traditional building an exception of one point of view.

Special attention should be paid for the attendance of roof gardens and to choose the quality of soil used for roof-covering (special mixture of soil is necessary) if we want to see a well-kept vegetation.

Another disadvantage might be that potential reparations (e.g. water leakage due to problems with the damp-course) can be performed only circuitously at the sacrifice of high costs. Unfortunately, the construction of an in-earth embedded residential building can be much more expensive than a conventional building because of the more careful planning, the better building material qualities and the more accurate construction standards.

It can frighten a lot of investor away especially if he is not aware to the relatively fast payback of construction costs during the operation.

Disadvantage is that most of the materials used in the construction of earth houses are artificial (concrete, plastic sheet waterproofing), so their production involves environmental pollution, and recycling is difficult. There are also ongoing experiments

to these less environmentally friendly materials should be replaced (e.g. Eco-Flex waterproofing, earth and adobe bricks).

Although it can be expected architecturally unique that the ground plan of the most earth houses – mainly due to structural considerations – curved or irregularly shaped, but therefore these houses are more difficult to be furnished.

Among disadvantages should be mentioned the incompleteness in special literature and design specifications. Currently effective building regulations contain no specific passage for earth houses. Certain parts of paragraphs (e.g. floor level below terrain of rooms intended for permanent residence) can be interpreted in various ways.

Urban architecturally the spawn of earth houses may raise several doubts. The building instructions of resettlement plans (e.g. development in unbroken row, scattered development method) can hardly be observed by earth houses.

Also raise questions whether or not can be achieved, and if so, in what manner in densely built in and exposed to heavy traffic and air pollution metropolitan areas, where eco-building design is most needed.

8. The Peculiarities of the Building Structures of In-Earth Embedded Houses

Building in-earth embedded house needs much more earthwork, excavation and landscaping than a conventional building. Their supporting structure is most often dome or shell structure because this structure is the most appropriate for resisting stress (earth pressure for the walls and roof), keeping the unnecessary moisture away and elimination of illumination problems (e.g. skylights placed on the top of the dome structure).

It is important to emphasize to pay careful attention to the planning and implementing of insulation (thermal insulation and damp-proofing). Ideal insulation materials (because of the unusually and irregular surfaces) are the coatings].

8.1. The Supporting Structures of In-earth Embedded Houses

The choice of structure materials is mainly affected by the building design, the in-site conditions and the extent of bedding into earth.

Thicker layer of earth necessitates stronger loam-bearing structure. Reinforced concrete is the most common material for creating supporting structures of in-earth embedded houses. Although it is not the most environmentally friendly building material its strength, durability and fire resistance is a great benefit compared to other materials.

Prefabricated and monolithic reinforced concrete is also applicable. A structure made of reinforced concrete can be combined with other materials (e.g. reinforced concrete walls and wood roof structure).

Steel structure is also conceivable as supporting structure of an earth house. It has relatively small structural self-weight and high tensile and compressive strength. Its main disadvantage is poor fire resistance and susceptibility to corrosion.

Rarely masonry structure (stone, brick, adobe, earth brick) can be used, but at higher load and stresses it should be strengthened with steel insets or additional reinforced concrete structures. As the reinforced concrete the most common form of masonry structures is also spherical dome, however it is more environmentally friendly and cheaper. Small bricks and (with cement or lime) stabilized earth brick masonries (e.g. BIOECO bricks) are the most popular (Fig. 13) [17].

Figure 13: Masonry structure of the earth house in Tököl (Hungary) [17]

Wood is also an ordinary structural building material. Construction of a timber frame structure can be attractive for environmentally conscious architecture because of its natural origins (Fig. 14). Supporting power of it is not as high as a steel or ferro-concrete construction so it has applicability limitations. In addition it is less resistant to moisture, heat (e.g. fire) and pests (insects, rodents). Wood structures can be guarded against them with wood preservatives additives and coatings.

Figure 14: Wood structure of the Navajo Hogan (Utah, USA) [20]

Steel and reinforced concrete structure can be used together as a plastically formed shell structure. At first with spacers reinforced double steel frame structure is built. Then a fine woven steel mesh is attached. The external and internal side gets shot-concrete layer (Fig. 15).

The space between the two shot-concrete layers can be filled by thermal insulation. Less thermal bridge, quick construction, possibility of large spans and unnecessary of formworks are its main benefits (Fig. 16).

Figure 15: Supporting structure of an in-hill house (Egerszalók, Hungary) [21]

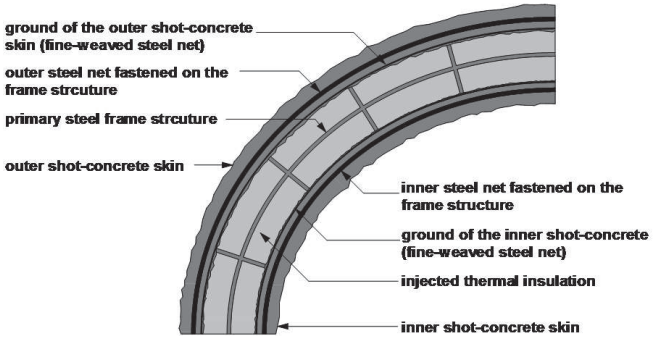

Figure 16: The stratification of the plastically formed shell structure (self-made figure)

8.2. The Problems with Damp-Proofing of In-earth Embedded Houses

As any structure contacting with soil, the interior of an in-earth embedded house also needs protection against harmful effects of moisture. Similarly to green-roofed constructions the protection of damp-proofing membrane against roots is substantial.

At lower damp-proofing requirements, when only particular impermeability required, the supporting structure itself can be watertight (e.g. watertight concrete). Moreover we can use watertight plaster on the inner or outer side of the wall structures [3].

In all residential buildings total impermeability should be provided. Most prevalent damp-proofing technologies are membrane and coating insulation. One of the most important damp-proof membrane insulations is the modified bituminous membrane During their preparation synthetic rubber additives is mixed with bitumen and then this mixture is carried up to a polyethylene mesh. They have long life expectancy and they are stickable to outer walls. Unfortunately unusual, irregular surfaces and shell structures are uneasy to insulate with it [15][16].

Useable plastic damp-proofing materials are polyethylene (PE), chlorinated polyethylene and polyvinyl chloride (PVC) membranes. They are cheap and durable but UV radiation can damage them, so almost immediately after their placement protection from degradation due to solar radiation is necessary. The most commonly used elastic materials are polyisobuthylene (PIB), polychloroprene and ethyl-propylene-diene (EPDM) plastic damp-proofing membranes. EPDM membranes have integrated protection against roots, but they are heavily stickable and some pests (e.g. fire ant) can bite it through. Plastic waterproof materials are not so environmentally friendly because they are petroleum products.

The other important group of damp-proof materials are coating insulations. This technology means that the liquid waterproofing material is applied (spraying, brushing) on the surface to be insulated. After drying it forms damp-proofing membrane.

Coatings are easily practicable moreover they have no splices. Nevertheless they can be easily damaged and their implementation is proposed under relatively warm and dry weather (so attention to weather conditions should be paid). Generally they are used in places where implementation of membrane insulations is impossible. Most often used coating insulations are bituminous emulsion and liquid polyurethane.

Sometimes bentonite damp-proofing occur insulating in-earth embedded houses. Bentonite is clay mineral from which insulating plates are prepared which are fixed on the outer side of the wall. They are rarely used because of their high self-weight and difficult installation. There is also bentonite coating insulation.

8.3. The Thermal Insulation of In-earth Embedded Houses

Thermal insulation quality of earth is a special problem. Depending on moisture content, density and compound thermal conductivity (λ) of earth is ranged between 0,15-2,00 W/mK, which is much lower than the thermal conductivity of the most common building insulation materials. But – unlike thermal insulation materials – the thickness of earth-covering is much larger (it can be more than 1,0 m). Because of its large mass (and thermal lag) it is able to reduce the effects of outer temperature changes, so it can prove adequate thermal protection for residential building. Due to this beneficial effect thermal insulation of in-earth embedded houses much less energy needs to be paid a sin a conventional building [1][4].

The earth-sheltering acts positively upon the interior climate of the building. It gives protection from the solar radiation. By reducing thermal fluctuation on the outer surface of the roof and by increasing their thermal capacity, earth-sheltering contributes to the cooling of the spaces below the roof in the summer and to the increase of their heat during the winter. Due to the decrease of the thermal losses earth-covering saves the energy consumption [10].

On basis of researches the surface temperature of earth-covering varies according to the different kind of vegetation. Lower temperature measured in the spaces, which are covered by thick, dark green vegetation and higher in spaces, which are covered by sparse red vegetation or only soil. During the summer period the external surfaces with the earth-covering are heated less than traditional roofs. Besides, in the winter period the planted roofs reduce the heat loss. Greek scientists showed that the estimated

heating and cooling loads are lower in the building with earth-sheltering [10]. The greatest energy savings during the whole year period were calculated in the case of non-insulated buildings 37%, which increased to 48% when advanced cooling was applied. The total energy saving consumption in case of moderate insulated buildings varied from 4% till 7% (with or without advanced cooling). The impact of earth-covering on the energy savings of well-insulated buildings was almost 2%.

Several buildings have been built on the basis of assumptions that earth-covering is itself sufficient and no added thermal insulation layer was built in. However, this is only allowed between specific climatic conditions. In cold and temperate climates appropriate indoor air temperature cannot be ensured in all seasons without extra thermal insulation.

Basically, two types of thermal insulation technology used. One option is to place the insulation outside of the waterproofing membrane. In this case, the insulating layer is considered as soil-contacting structure. Thus, only a material is allowed to use for this purpose, which does not lose its insulating ability under permanent moisture effects. For this purpose there are some plastic foam insulation materials (such as extruded polystyrene foam) [16].

Another way of insulation is the on-site foamed (sprayed or brushed) thermal insulation products (such are polyurethane foam) with which practically any surface can easily be insulated. In order to keep water away the protective layer must be overlaid on it.

9. Conclusions

Residential buildings embedded into earth always served as human shelters. Natural caves, in-hillside cut cave dwellings and in-ground pitted houses were the archetypes of earth houses that modern earth architecture calls cave, in-hill and atrium houses today.

Basic principles of modern earth architecture were expressed by the first modernist architects in the 1930s (such as Le Corbusier and Wright) but the really idea of in-earth embedded houses came to forward only after the first oil crisis in the 1970s. In this time the environmental and energy conscious architecture became a matter of common knowledge which greatly contributed to revalue several advantages of in-earth embedded houses: e.g. the excellent fit to the environment, low energy consumption and protection of indoor living space against external environmental effects (e.g. noises, wind and the fluctuation of temperature). In order to utilize the benefits of in-earth building more thoroughly examined design is required comparing to the design of any conventional residential buildings.

As an example particular attention should be paid to the topographical, climatic and soil characteristics. In addition to this, the architect must have excellent knowledge of building construction and building physics to have capacities in solving the structural, lightning, solar, ventilation and humidity questions derived from the peculiar relation with the earth.

The correct orientation, the floor plan design, the landscaping and the selection of the materials are the essential questions, because an incorrectly designed building may be easily narrow, damp, dark and cold, therefore unsuitable for human habitation.

For supporting the structure of in-earth embedded houses, dome or shell structures are the most suitable. Structural material is commonly reinforced concrete, but other materials (steel, wood, stone, adobe, and clay bricks) are also adequate. The most important issues are damp-proofing and thermal insulation.

Earth-covering has positive effect on indoor climate. It protects the outdoor surface from solar radiation and reduces the indoor thermal fluctuation. Earth-covering decreases thermal losses in winter and warming up in summer. This is why it saves energy consumption. Without extra thermal insulation heating and cooling loads are 37-48% lower, than in case of a non-insulated building and 2-7% lower than in case of insulated buildings. Sometimes in cold and temperate climate extra thermal insulation is needed. It can be solved with extruded polystyrene foam or monolithic foamed, sprayed or brushed insulation materials, such as polyurethane foam.

Because of their disadvantages (relatively high building costs, lack of standards and regulations, strict design aspects) in-earth embedded houses are built relatively rarely nowadays. But as they spread widely the negative stereotypes (damp, dark and cold) wear away. Slowly they are becoming increasingly popular and expectedly succeeding architects have to design more and more in-earth embedded houses in the future.

References

[1] Al-Temeemi AA, Harris DJ: A guideline for assessing the suitability of earth-sheltered mass-housing in hot-arid climates. Energy and Buildings Vol. 36, No. 3, pp. 251-260, 2004.

[2] Batár A: The Invisible Cave Dwellings (in Hungarian). Ab Ovo Publisher, 2005.

[3] Bica S, Rosiu L, Radoslav R: What Characteristics Define Ecological Building Materials. 7th IASME/WSEAS International Conference on Heat Transfer, Thermal Engineering and Environment, Moscow (Russia), 20.08.2009-22.08.2009, pp. 159-164, 2009.

[4] Bozsaky D: In-earth Embedded Residential Buildings (in Hungarian). Magyar Építőipar, Vol. 60, No. 3, pp. 109-115, 2010.

[5] Braham W: Malcom Wells, 1926-2009, Williambraham.net – Ecology, Technology and Design, 06-12-2009 http://williambraham.net/?p=192 (downloaded: 19.10.2009)

[6] Dám L: In-earth Embedded Residential Buildings in the Native Architecture of the Alföld (in Hungarian). A nyíregyházi Jósa András Múzeum évkönyve, 33-35 (1990-1992), Jósa András Múzeum, Nyíregyháza (H), pp. 133-151, 1993.

[7] Fajszi B: Discovering Tunisia Independently (in Hungarian). Útikalauz.hu/Tunézia/Fajszi Bence: Önállóan felfedezni Tunéziát, 2004. szeptember. http://www.utikalauz.hu/index.php?p=folap&id=868 (downloaded: 02.04.2010)

[8] Hegedűs Zs: In-hill Houses in Bőny (in Hungarian). Magyar Építőművészet 4/2007.
http://magyarepitomuveszet.mm-art.hu/hu/paholy_design.php?lapszam=2007-4&id=859
(downloaded: 19.10.2009)
[9] Mednyánszky M: Cave houses in Hungary (in Hungarian). Terc Kiadó, Budapest, 2009.
[10] Niachou A, Papakonstantinou K, Santamouris M, Tsangrassoulis A, Mihalakakou G: Analysis of the earth-covering thermal properties and investigation of its energy performance. Energy and Buildings Vol. 33, No. 7, pp. 719-729, 2001.
[11] Ortutay Gy (ed.): Cyclopaedia of Hungarian Ethnography Vol. 2 (in Hungarian). Akadémiai Kiadó, Budapest, 1979.
[12] Ortutay Gy (ed.): Cyclopaedia of Hungarian Ethnography (in Hungarian). Akadémiai Kiadó, Vol. 4, Budapest, 1981.
[13] Roy R: Earth-Sheltered Houses: How to Build an Affordable Underground Home. New Society Publishers, Gabriola Island (Canada), 2006.
[14] Yong X: Inside the yellow earth. Beifan.com/China Postcards/Xu Yong: Inside of the Yellow earth, July 1996.
http://www.beifan.com/044postcards/page01.html (downloaded: 04.06.2011)
[15] Anon: An introduction to earth sheltering. The British Earth Sheltering Association/An introduction to earth sheltering
http://www.besa-uk.org/intro.html (downloaded: 04.06.2011)
[16] Anon: Building Earth Sheltered Homes. Earth Sheltered Homes – Residential, Commercial & Multi-Family Construction
http://www.earthshelteredhome.com/index.htm (downloaded: 02.04.2010)
[17] Anon: Zsolt Hegedűs (in Hungarian).
Élőépítészet.hu/Épületek/Építészek/Egyesülésen kívüli építészek
http://eloepiteszet.hu/hu/epuletek/epiteszek/egyesulesen-kivuli-epiteszek/hegedus-zsolt
(downloaded: 19.10.2009)
[18] Anon: Hogan – The Navajo Dwelling, Max Bertola's Southern Utah/Feature stories about southern Utah/Hogan – The Navajo Dwelling
http://www.so-utah.com/feature/hogan/homepage.html (downloaded: 04.06.2011)
[19] Anon: Matera, Wikipedia – the free encyclopedia /Matera
http://en.wikipedia.org/wiki/Matera (downloaded: 19.10.2009)
[20] Anon: Native American Shelters – Southwest, Minnesota State University/Emuseum/Prehistory/Settlements/Southwest
http://www.mnsu.edu/emuseum/prehistory/settlements/regions/southwest.html
(downloaded: 02.04.2010)
[21] Anon: Free-formed Houses Under The Earth (in. Hungarian). Koos.hu/zöld/2007
http://koos.hu/2007/11/28/szabadformaju-szendvicshej/ (downloaded: 25.08.2014)
[22] Anon: Orkney's Underground Earth-houses, Orkneyjar.com/History/Earth Houses
http://www.orkneyjar.com/history/earth-houses/ (downloaded: 19.10.2009)

Acta
Technica
Jaurinensis

Vol. 8, No. 2, pp. 131-144, 2015
DOI: 10.14513/actatechjaur.v8.n2.350
Available online at acta.sze.hu

State Space Estimation for an Electromechanical Actuator Valve using Extended Kalman Filter

Zs. Horváth

Production Planning V-Engines,
AUDI Hungaria Motor Kft., 9027 Győr, Kardán utca 1.
zsolt2.horvath@audi.hu

Abstract: This paper discusses a state space estimation for an electromechanical actuator valve using extended Kalman filter (EKF). A proposed actuator model includes a Tustin's friction model with strong nonlinearities, hence it represents an accurate model for describing the friction phenomenon in an electromechanical actuator valve. At first the state equations of the model were converted into a time-discrete. Then the EKF is used for estimation of three states, namely the motor current, valve angular velocity - and position. Both the actuator model and the EKF algorithm were implemented in MATLAB®. The measurements for input data in the EKF are performed in National Instruments CompactRIO system. Outline of our examination should be to take a conclusion, how much EKF is capable of accurate estimation of the desired states. Furthermore we will use this method for further developing the fault detection in the electromechanical actuator valve.

Keywords: extended Kalman filter, electromechanical actuator valve, state space estimation, Tustin's friction model

1. Introduction

Processing integrated electromechanical systems such as actuators and sensors are fundamental part of automotive vehicles. With increasing complexity of them, the methodology of state space estimation becomes an essential part of the field of fault diagnosis and condition monitoring. Generally the electromechanical actuators product cost such as the electromechanical throttle valve must be low. On the other hand their structure is complex and they have a relatively large number of components. The measurement of desired state signals such as current, pressure, velocity, position are often physically not possible or application of additional sensors in them is too expensive.

A typical case in most dynamical systems with complex structure is that the state vector can not be measured. In these situations a suitable approximation to the state

vector is needed. The extended Kalman filter is a very frequently used tool for performing state estimation on nonlinear dynamical systems.

This paper discusses a state space estimation for an electromechanical actuator valve using extended Kalman filter (EKF). A Tustin's friction model can be found in our proposed actuator valve model which includes nonlinearities too, that is why it represents an accurate model in an electromechanical actuator valve for describing the friction phenomenon. It is necessary because the EKF requires an accurate model of the plant. In this work I have used the EKF for the estimation of the motor current, valve angular velocity - and position.

For modelling and control of electromechanical actuators nowadays are used many different observer concepts. A nonlinear Variable Structure System (VSS) observer for throttle systems is presented in [1]. Paper [2] uses a sliding-mode observer for a robust position control of these actuators. An additional study is [3], it describes applying Luenberger observer for sensor monitoring in active front steering systems. Another study [4] is about using the Kalman filter for the observation of the DC-motor. The most common approaches are using extended Kalman filter (EKF) or unscented Kalman filter (UKF) [5]. These two methods are preferable solutions in our case now too, when the process nonlinearities are strong and noise is associated with the real system. Research [6] used the EKF for parameter estimation of a hydraulic proportional valve. Another research [7] extended it to fault detection.

The throttle valve is a type of an electromechanical actuator, it is advanced in several applications of combustion engine control (intake manifold, exthaust gas recirculation, variable turbine geometry, ect.).

The presented paper consists of : the electromechanical throttle valve model is introduced in Section I, Section II is about the design of the extended Kalman filter, Section III shows the experimental results and finally Section IV presents the main obtained results.

2. Basic Model of an Electromechanical Throttle Valve

The electromechanical throttle valve (Fig. 1.) regulates air flow in the engine of the car. It consists of following parts: DC-motor, gearbox, return spring, throttle plate and sensor for valve position. A bipolar chopper supplies the DC-motor, a gearbox to the throttle plate transmitts the motor shaft rotation. The position of the throttle plate is given to the ECU (Engine Control Unit). A reset spring places the de-energized throttle plate in its default position („limp-home"). The valve is slightly opened here, and the engine can operate at minimum power.

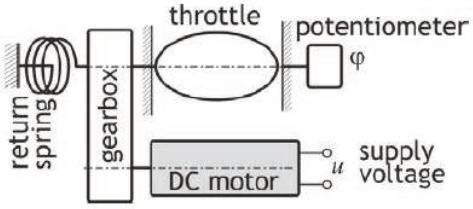

Figure 1. The scheme of the electromechanical actuator [8]

2.1 Model Equations of the Electromechanical Throttle Valve

The relation between input voltage u_a and current i in the armature circuit can be described as

$$u_a = Ri + L\frac{di}{dt} + K_e n\omega. \tag{1}$$

where L is the inductance, R is the resistance in the armature circuit, K_e is the inductive voltage constant, n is the gear ratio and ω *is* the angular velocity of the throttle plate.

The throttle valves are usually equipped with two springs. One spring is acting on the valve when the angular positon is greather than the „limp-home" position, whereas the other is acting when the angular position is smaller than the „limp-home" position. In our approach we have considered that the two springs may have some spring stiffness. The signal of the spring pretension torque will be changed depending on the signal of the angular position of the throttle. Further in our model we assume the „limp-home" postion at angular position by zero. The realisation of the spring torque is depicted on the Fig. 2.

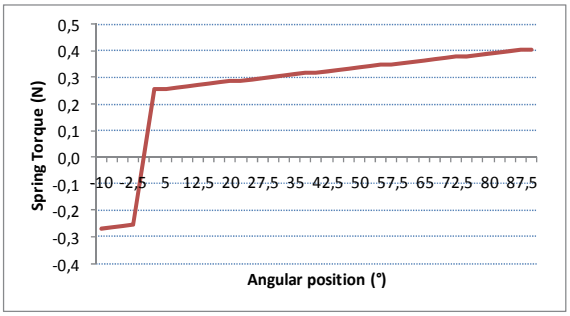

Figure 2. Nonlinear spring torque

The motion Equations of the throttle plate (related to the plate axel) can be desribed as

$$J\frac{d\omega}{dt} = K_t i - M_o sign(\varphi) - c\varphi - M_f. \tag{2}$$

where φ is the angular position of the throttle plate, J is the rotary inertia at the throttle shaft, which is composed of inertia of the DC-motor. M_o denotes a torque due to the prestressed springs. Finally K_t denotes the motor torque constant.

When the Stribeck effect is taken in consideration, the friction model is called Tustin's model. The Tustin's model is defined as

$$M_f = \left[M_c + (\mathrm{M}_s - M_c)e^{-|v|/v_s} \right] sign(\dot{x}) + k_v\dot{\varphi}. \tag{3}$$

where
M_f denotes the static friction torque, M_c is a torque of the coulomb friction of the system, M_s is the stiction force, v represents the velocity, v_s is the Stribeck velocity, c denotes a linear spring constant. K_v is a positive constant for viscousus friction.

Table 1. EMA parameters

Parameter	***Values***	***Units***
R	4.3	Ohm
L	0.002	H
K_e	0.017	Vs/rad
K_t	0.017	Nm/A
J	0.0016	kgm^2
n	40	-
M_c	0.2	Nm
M_0	0.25	Nm
c	0.059	Nm/rad
K_v	0.01	Nms/rad
M_s	0.3	Nm
v_s	1	rad/sec

Table 1 shows the values of the model parameters, some of them (*R, L, Ke, Kt, J, M_0, c*) are provided by the manufacturer (Continental AG). The remaining parameters are identified with the well known identifications procedures, with the static load test and viscous frictional coeffecient test, see e.g. [9].

2.2 Discrete State Space Model of the Electromechanical Throttle Valve

A discrete state space model of the system is required for applying the extended Kalman filter. Hence at first the state equations of the model must be converted into a time-discrete.

With having system and measurement noise matrices as *w(t)* and *v(t)* respectively, the continuous state space model for a linear system can be written as :

$$\dot{X}(t) = AX(t) + BU(t) + w(t), \tag{4}$$

$$Z(t) = CX(t) + v(t). \tag{5}$$

where *X(t)* is the state vector, (n x 1) matrix, *A* is the system, (n x n) matrix, *B* is the input matrix, (n x p), *C* is the output matrix, (m x n), *U(t)* is the input vector, (p x 1), *Z(t)* is the output vector, (m x 1), *w(t)* is the system noise, (p x 1), and *v(t)* represents the measurement noise, (m x 1).

To get a discrete state space model, I have used the "forward difference" approximation to calculate the derivative, which given by

$$\dot{X}(t) \approx \frac{X(k+1)-X(k)}{T_s}. \tag{6}$$

where T_s is the sampling time, k is a step, $(k+1)$ is a next step of the iteration.

Using the Equation (6) and applying it to the Equations (4), (5), the discrete state space model is [7]:

$$X(k+1) = \Phi X(k) + GU(k) + W(\mathrm{k}), \tag{7}$$

$$Z(k) = HX(k) + V(k), \tag{8}$$

$$\begin{aligned} \text{with } \Phi(k) &= T_s A + 1, \\ G(k) &= T_s B, \\ W(k) &= T_s w(k), \\ V(k) &= v(k). \end{aligned} \tag{9}$$

where Φ is the discrete system matrix, G represents the discrete input matrix, H stands for the discrete output matrix, $W(k)$ represents the process noise and $V(k)$ is the measurement noise. It is assumed, that process noise and measurement noise (elements of $W(k)$ and $V(k)$) are respectively uncorrelated and white .

Thus,

$$W \approx N\big(0, Q(k)\big), \text{ and } V \approx N\big(0, R(k)\big). \tag{10}$$

where $Q(k)$ and $R(k)$ are known covariance matrices of $W(k)$ and $V(k)$ respectively.

Finally we define the noise covariance matrices as

$$\begin{aligned} E\left[W_i(k)W_j(k)^T\right] &= Q(k), \quad i = j, \\ &= 0, \qquad i \neq j, \end{aligned}$$

$$\begin{aligned} E\left[V_i(k)V_j(k)^T\right] &= R(k), \quad i = j, \\ &= 0, \qquad i \neq j. \end{aligned} \tag{11}$$

Using the forward difference approximation Equation (6) and applying it to the continuous state space model Equations (1) to (3), the discrete state space model of the electromechanical throttle valve is given by

$$x_1(k+1) = x_1(k) - \frac{R}{L}x_1(k)T_s - n\frac{K_e}{L}x_2(k)T_s + \frac{1}{L}U(k)T_s + w_1(k)$$

$$\begin{aligned} x_2(k+1) &= \frac{nK_t}{J}x_1(k)T_s + x_2(k) - \frac{K_v}{J}x_2(k)T_s - \frac{c}{J}x_3(k)T_s \\ &- \frac{M_0}{J}sign\left(x_3(k)\right)T_s - \frac{1}{J}(M_c + (M_s - M_c)e^{-\left|\frac{\omega}{\omega_s}\right|})sign\left(x_2(k)\right)T_s \\ &+ M_{Stop}(\varphi,\omega) + w_2(k) \end{aligned} \tag{12}$$

$$x_3(k+1) = x_2(k)T_s + x_3(k) + w_3(k).$$

where $x_1(k)$ is the state variable for motor current, $x_2(k)$ is a state variable for angular speed of the throttle, $x_3(k)$ is the state variable for trottle angle, w_1, w_2, w_3, are the system noise elements, T_s is the sampling time.

3. Applying Extended Kalman Filter to State Estimation of the Electromechanical Throttle Valve

3.1 Extended Kalman Filter

To perform state estimation on nonlinear dynamical systems, the extended Kalman filter is a very commonly used tool. The extended Kalman filter is used for parameter and state estimation, it is a modified form of the Kalman filter. The Kalman filter is an optimal state estimation process applied to dynamic systems with random perturbations. If the process model is nonlinear, the Kalman filter is applied by continually updating the linearization around the most recent reference $X(k)$ for the system function and $X^{-}(k)$ for the measurement function as shown on the Figure 3.

A Taylor approximation of the system function is made at the refined state estimate $\hat{X}(k)$ and an approximation of the measurement function is made at the unrefined state estimate $\hat{X}^{-}(k)$.

The nonlinear model can be represented with:

$$X(k+1) = f\left(X(k), U(k)\right) + W(k), \tag{13}$$

$$Z(k) = h\left(X(k)\right) + V(k). \tag{14}$$

where $f\left(X(k), U(k)\right)$ and $h\left(X(k)\right)$ are the nonlinear system and measurement functions respectively.

The linearization is made using the most recent obtained state estimate, which is using the unrefined estimate for the system function and the refined estimate for the measurement function.

Therefore, doing a linear Taylor series approximation of $f\left(X(k),U(k)\right)$ at an operating point $\hat{X}(k)$ and of $h\left(X(k)\right)$ at the operating point $\hat{X}^{-}(k)$, and ignoring the higher-level term yields, the linearised form is:

$$X\left(k+1\right)=\Phi\left(k\right)X\left(k\right)+U_1\left(k\right)+W\left(k\right), \tag{15}$$

$$Y\left(k\right)=H\left(k\right)X\left(k\right)+V\left(k\right), \tag{16}$$

where:

$$\begin{aligned}\Phi\left(k\right)&=\frac{\partial f\left(X\left(k\right),U\left(k\right)\right)}{\partial X\left(k\right)}\Big|_{\hat{X}(k)},\\ H\left(k\right)&=\frac{\partial h\left(X\left(k\right)\right)}{\partial X\left(k\right)}\Big|_{\hat{X}^{-}(k)},\\ U_1\left(k\right)&=f\left(\hat{X}\left(k\right),U\left(k\right)\right)-\Phi\left(k\right)\hat{X}\left(k\right),\\ Y\left(k\right)&=Z\left(k\right)-h\left(\hat{X}^{-}\left(k\right)\right)+H\left(k\right)\hat{X}^{-}\left(k\right).\end{aligned} \tag{17}$$

At step k a refined estimate of $\hat{X}(k)$ is now used to predict what $\hat{X}^{-}(k)$ is at step *k + 1,* based on the plants modeled dynamics.

The five Kalman filtering equations can be summarized as follows:

Estimation equations:

$$K(k)=P^{-}(k)H^{T}\left[HP^{-}(k)H^{T}+R(k)\right]^{-1}, \tag{18}$$

$$\hat{X}(k)=\hat{X}^{-}(k)+\mathrm{K(k)}\left[Z(k)-H\hat{X}^{-}(k)\right], \tag{19}$$

$$P(k)=\left[I-K(k)H\right]P^{-}(k). \tag{20}$$

Prediction equations:

$$\hat{X}^{-}(k+1)=\Phi\,\hat{\mathrm{X}}(k)+GU(k), \tag{21}$$

$$P^{-}(k+1)=\Phi\,\mathrm{P(k)}\Phi^{T}+Q(k). \tag{22}$$

The EKF is a recursive algorithm with a set of prediction (time update) and correction (measurement update) is shown in the Figure 3.

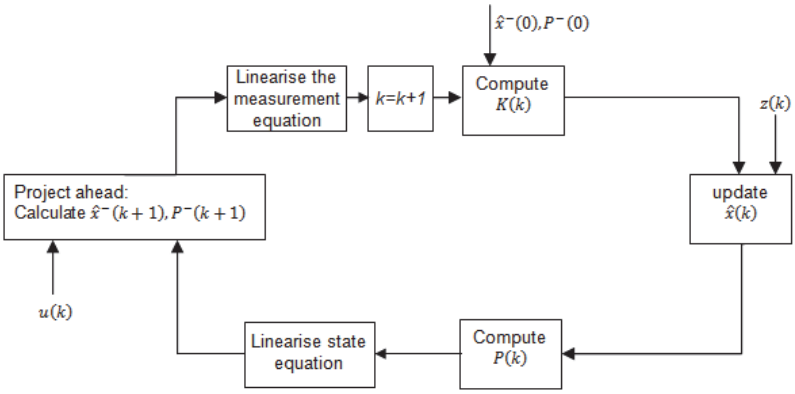

Figure 3. Flowchart of the recursive extended Kalman filter [6]

3.2 Observability Test

In state estimation of dynamic systems, observability is one of the most important requirements. In the case when the system is linear, it can be characterized with the rank of the observability matrices, a methology for this can be found in [10]. If the system is nonlinear observability is more complicated. The nonlinear observability is highly dependent on the Lie derivative, which is a derivative of a scalar function along a vector field. We can define the Lie derivative of scalar function *h* with respect to a vector field *f*, denoted $L_f(h)$.

It is given by

$$L_f(h) = \frac{\partial f}{\partial x} f(x) \tag{23}$$

The examination of the observability is not discussed in this paper, a method for it is presented in [11]. The theory of the observability of nonlinear systems is discussed in [12].

3.3 State Estimation for the Electromechanical Throttle Valve by the EKF

As above mentioned, the extended Kalman filter is used to estimate three states of the electromechanical throttle valve, namely motor current, valve angular velocity and valve angular position. In order to apply the EKF, firstly the Equations (15), (16) must be calculated. The discrete output matrix is:

$$H = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{24}$$

The output Equation is:

$$Z(k) = HX(k) + V(k)\,,$$

$$\mathrm{Z}(k) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} v_1(k) \\ v_2(k) \\ v_3(k) \end{bmatrix}. \tag{25}$$

where $V_1(k)$, $V_2(k)$, $V_3(k)$ are the measurement noise elements.

Just as next step should be the Jacobi matrix calculated, with using Eq. 12. On them the function *sign(x)* is not able to take the derivative, because it is not continued at zero.

Therefore we use the approximation $\mathrm{sign(x)} \approx \frac{2}{\pi} arctg(10^3 x)$,

The linearized (Jacobi) system matrix is:

$$\Phi(k) = \frac{\partial f\left(X(k)\right)}{\partial X(k)} = \begin{bmatrix} \Phi_{11}(k) & \Phi_{12}(k) & \Phi_{13}(k) \\ \Phi_{21}(k) & \Phi_{22}(k) & \Phi_{23}(k) \\ \Phi_{31}(k) & \Phi_{32}(k) & \Phi_{33}(k) \end{bmatrix}, \tag{26}$$

where the elements of the matrix are

$$\begin{aligned} &\Phi_{11} = 1 - \frac{R}{L} T_s\,,\ \Phi_{12} = n\frac{K_e}{L} T_s\,,\ \Phi_{13} = 0\,, \\ &\Phi_{21} = n\frac{K_t}{J} T_s\,, \\ &\Phi_{22} = 1 - \frac{K_v}{J} T_s - 10^3 T_s \frac{M_c + (M_s - M_c)}{J\pi\left(1 + 10^6 Xest(2)^2\right)} e^{-\left|\frac{Xest(2)}{\omega_s}\right|}\,, \\ &\Phi_{23} = -\frac{c}{J} T_s - \frac{10^3 T_s M_0}{J\pi\left(1 + 10^6 Xest(3)^2\right)} T_s\,, \\ &\Phi_{31} = 0\,,\ \Phi_{32} = T_s\,,\ \Phi_{33} = 1. \end{aligned} \tag{27}$$

where *Xest* denotes a vector, that includes the estimated state space variables.

It is assumed that the initial values of the states has a known mean value and variance. Hence the initial state vector and the initializing matrices were set as follows:

$$x_0(k) = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}^T$$

$$P(0) = \begin{bmatrix} 10^2 & 0 & 0 \\ 0 & 10^2 & 0 \\ 0 & 0 & 10^2 \end{bmatrix}. \tag{28}$$

The system noise covariance matrix and the measurement noise covariance matrix where set by trial and error.

$$Q(k) = \begin{bmatrix} 10^{-2} & 0 & 0 \\ 0 & 10^{-2} & 0 \\ 0 & 0 & 10^{-2} \end{bmatrix} \tag{29}$$

$$R(k) = \begin{bmatrix} 1x10^{-4} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1x10^{-8} \end{bmatrix}. \tag{28}$$

Measured signals have different accuracy, which are represented on the measurement covariance matrix. As above mentioned, the angular velocity was obtained by derivation of the angular position. On account of the derivation procedure, the signal of the angular velocity became noisy, hence it has got a big value in the matrix.

Under normal conditions the estimated state vector $\hat{X}$ from the EKF converges to the system state vector *X(k)* such as all elements of the residual vector *e(k)* will be relatively small.
The residual vector can be described

$$\mathrm{e}(k) = Z(k) - H\hat{X}^{-}(k). \tag{30}$$

It causes a fault, hence one or more system parameters change, and the estimated state vector diverges from the actual state trajectory. The residual vector will increase.

3. Experimental Results

In the following the discrete state space model of the electromechanical actuator valve, was used by the Kalman filter to estimate the three states, the motor current, valve angular velocity and angular position. We will show how the simulation results collaborate with the experimental data. Both the throttle valve model and the extended Kalman filter algorithm are computed in MATLAB® SIMULINK®. The experimental measurements were made using National Instruments CompactRIO® 9014 system. The

actuator was excited through PWM-Signal with square wave of 1kHz and an Amplitude of 12V.

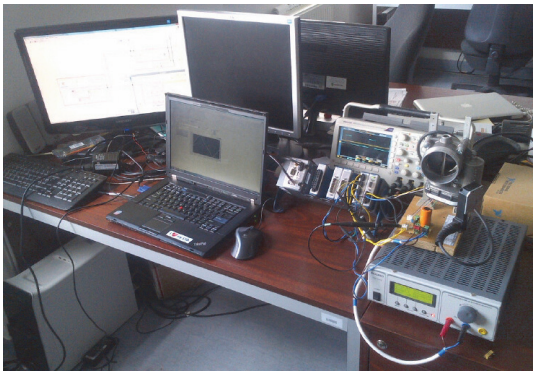

Figure 4. Experimental setup

Input signal for the parameter estimation was a sine wave with frequence of 5Hz and amplitude of 12V. The measured data (motor current, valve angular position) where recorded with sampling time of 1ms. The block diagram of the experimental setup is shown on the Fig. 5. Since the throttle valve was not equipped with any velocity sensor, the angular velocity of the throttle plate has been calculated by differentiating of the measured values of the angular position. Henceforth the angular velocity is treated as that where measured.

The measured data was used by the EKF to estimate the three states, the approach works iterative. Finally the estimated results were compared with data obtained from simulation of the Actuator model.

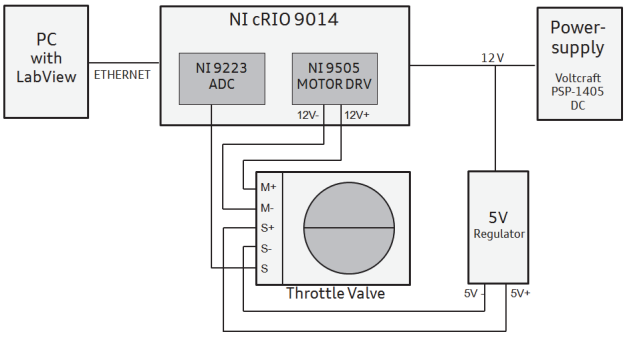

Figure 5. Block diagram of the experimental setup

The initial state vector and the initializing matrices for the EKF were set as mentioned in chapter 3.3.

The graphs in the Figures 6, 7, 8 show the estimated state curves with simulated ones. It can be seen that the results collaborate well with the results of the simulated data.

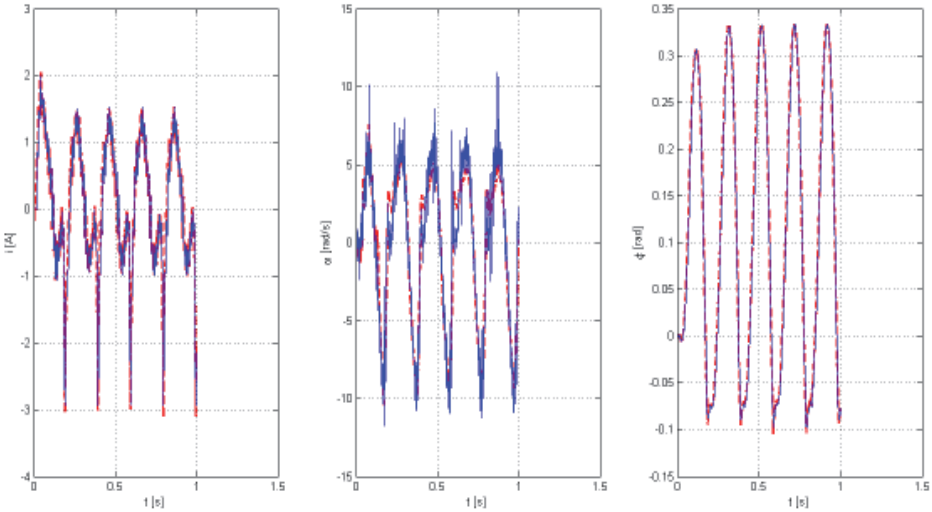

Figure 6. Experimental results measured vs. estimated states of the throttle valve blue line: measured state, red dashed line : estimated state, left: motor current, middle : angular velocity, right: angular position of the valve

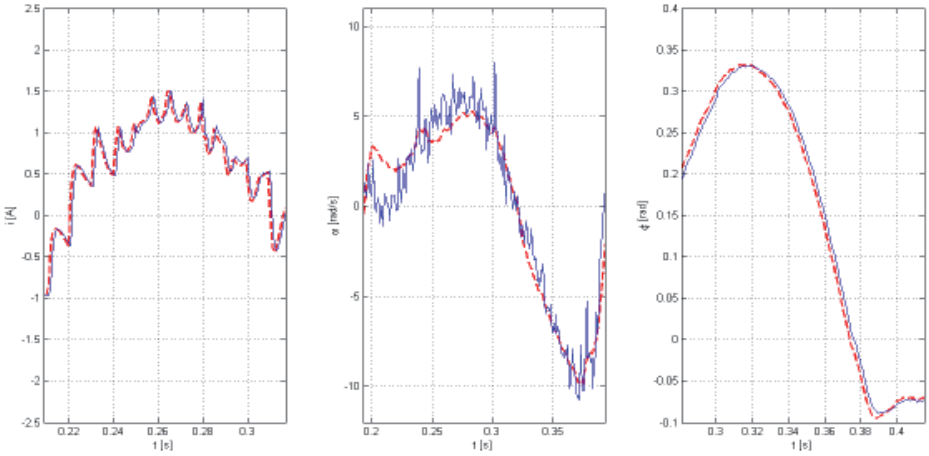

Figure 7. Comparison measured and estimated states of the throttle valve blue line: measured state, red dashed line : estimated state, left: motor current, middle : angular velocity, right: angular position of the valve

As we can see it in the Figures the estimated values of the states, motor current and valve angular position matches the measured states very closely as it is proved by the curve of the residual too. The estimated angular velocity matches not so good with the mesaured data curves. The reason for this is, that the angular velocity was not measured, just obtained by derivation of the angular position. The measured data of the velocity was not filtered. The initial matrices used for initializing the EKF were changed

before and it was discorvered that the estimation of the angular velocity was significantly affected. A best fit between the estimated and simulated curves was obtained by taking initial matrices as presented in chapter 3.3. Fig. 8 shows the residue signal of the motor current - and angular position. It can be seen, that the curve of motor current is different from zero. That can be caused by uncertainties of the model parameters. In spite of them, the EKF estimates both motor current and valve angular position successful.

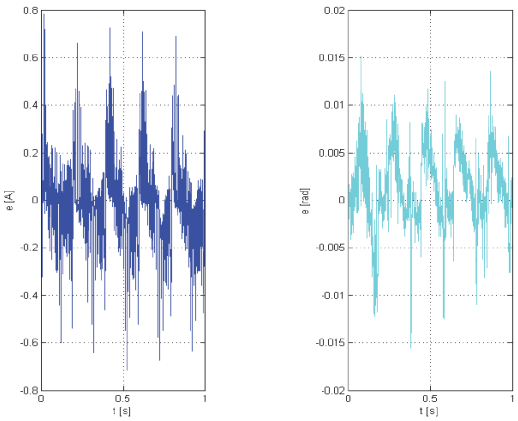

Figure 8. Residue computed during EKF iterations
blue line: motor current, cian dashed line : angular position of the valve

4. Conclusion

The state space estimation for an electromechanical actuator valve using extended Kalman filter (EKF) is discussed in this paper. Our actuator model captures the Tustin's friction model which includes nonlinearities. It is necessary because the EKF requires an accurate model of the plant. The first step before applying the EKF was converting the state equations of the actuator model into a time-discrete. Than the EKF was constructed and used to estimate three states of the system, namely the motor current, valve angular velocity - and position. After this the described actuator model and the EKF algorithm were implemented in MATLAB®. The measurements for the input data for the EKF were performed in National Instruments CompactRIO® 9014 system.

The curve of the residual proves, that the estimated values of the states, motor current and valve angular position matches the measured states very closely.

With our examination we have proved, that the proposed iterative EKF approach is able to estimate the states of the electromechanical actuator valve with strong nonlinearities successfully.

References

[1] Nakano K, Sawut U, Higuci K, Okajima Y: Modelling an Observer-based Sliding-Mode Control of Electronic Throttle Systems. ECTI Transactions on Electrical Eng., Electronics, and Communications, Vol. 4, No. 1, 2006.

[2] Reichhartinger M, Horn M: Robust Position Control of an Electromechanical Actuator for Automotive Applications. World Academy of Science, Engineering and Technology, Vol. 5, 2011.

[3] Reinelt W, Lundquist C: Observer Based Sensor Monitoring in Active Front Steering System using Explicit Sensor Failure Modelling. Proceedings of the 16th IFAC World Congress, Vol. 16, No. 1, 2005.

[4] Padmakumar S, Vivek A, Kallol R: A Comparative Study into Observer based Fault Detection and Diagnosis in DC Motors: Part-I. World Academy of Science, Engineering and Technology, Vol. 3, No. 3, 2009.

[5] Vašak M, Petrović I, Perić N: State Estimation of an Electronic Throttle Body. Industrial Technology IEEE International Conference, Vol.1, pp. 472-477, 2003.

[6] Wright G: Parameter Estimation of a Hydraulic Proportional Valve Using Extended Kalman Filtering. M.Sc. Thesis, University of Saskatchewan, Canada, 2001.

[7] Yuvin AC: Fault Detection In the Electrohydraulic Actuator Using Extended Kalman Filter. Doctoral Thesis, University of Saskatchewan, Canada, 2004.

[8] Grepl R, Lee B: Modelling, Identification and Control of Electronics Throttle Body Using Dspace Tools. Technical Computing Prague, 2008.

[9] Loh RNK, Thanom W, Pyko JS, Lee A: Electronic Throttle Control System: Modeling, Identification and Model-Based Control Designs. Engineering, Vol. 5, No. 7, pp. 587-600, 2013.
DOI: 10.4236/eng.2013.57071

[10] Horváth Zs, Molnárka Gy: Design Luenberger Observer for an Electromechanical Actuator. Acta Technica Jaurinensis, Vol. 7, No. 4, 2014.
DOI: 10.14513/actatechjaur.v7.n4.313

[11] Nor Hazadura H, Sazali Y, Hariharan M, Norhizam H: Nonlinear observers for attitude estimation in gyroless spacecraft via Extended Kalman filter algorithm. International Journal of Scientific and Research Publications, Vol. 4, 2014.

[12] Slotine J-JE, Li W: Applied Nonlinear Control. Prentice Hall, Englewood Cliffs, New Jersey, pp. 229-234, 1991.

Acta
Technica
Jaurinensis

Vol. 8, No. 2, pp. 145-152, 2015
DOI: 10.14513/actatechjaur.v8.n2.365
Available online at acta.sze.hu

Iterative Calibration of VISSIM Simulator Based on Genetic Algorithm

T. Tettamanti[1], A. Csikós[2], I. Varga[1], A. Eleőd[3]

[1]Budapest University of Technology and Economics, Department of Control for Transportation and Vehicle Systems,
Stoczek utca 2., 1111, Budapest, Hungary
Phone: +36 1 463 2255
e-mail: tettamanti@mail.bme.hu

[2] Systems and Control Laboratory, Research Institute for Computer Sciences and Automation, Hungarian Academy of Sciences
Kende utca 13-17, 1111, Budapest, Hungary

[3] Budapest University of Technology and Economics, Department of Vehicle Elements and Vehicle-Structure Analysis
Stoczek utca 2., 1111, Budapest, Hungary

Abstract: In the recent decade, computer simulation has been developed enormously in all engineering fields and has become a basic step of design. Accordingly, simulation plays a fundamental role in road traffic engineering offering a basis for the effective analysis of intelligent traffic control systems. A reliable simulator is able to provide effective analysis of a given traffic network if the applied simulation scenarios properly converge to the real-word situation. This requirement can be achieved based on the mixed use of prior real-world traffic measurements and proper simulation settings. The latter one, however, is not straightforward. Accordingly, the paper proposes a potential calibration method to create true-to-life VISSIM simulations. VISSIM is one of the most popular microscopic traffic simulator, a well-known software applied in engineering and scientific practice as well. Basically, a calibration technique with genetic algorithm optimization is proposed to reproduce realistic traffic on a network based on floating car speed data.

Keywords: Vissim simulator, calibration, genetic algorithm

1. Introduction

The continuous maintenance and development of road traffic control systems is indispensable due to the increasing traffic demands. Appropriate simulation represents the preceding stage of such engineering tasks. Moreover, the concept of intelligent transportation systems (ITS) of our days also requires the advanced use of informatics.

As a consequence, when traffic engineers design intelligent solutions, the use of traffic simulators and mathematical optimization software is more and more expected during the development process and the validation phase as well. However, the proper use of simulators is important in order to avoid false simulation results. Namely, beside the several advantages of simulation, simulators also contain the danger of providing erroneous results in case of inappropriate simulation settings in the scenarios. Accordingly, calibration of simulation parameters is expected. Numerous researches have been conducted in this field providing efficient methods to optimize the most important settings, such as travel times, driving behavior parameters, saturation flow rates, etc. These parameters are mostly microscopic variables which can be successfully tuned based on the research results, e.g. [1][2][3].

In this paper, however, we focus only on the calibration of a macroscopic variable, the traffic demand which is one of the most important network parameter in the microscopic simulation. Moreover, traffic input is calibrated based on real-world floating car data (FCD) participating as sensors in the road network. FCD is typically collected from fleet cars equipped with a GPS receiver providing accurate speed data and GPS position logs. Beside fleet cars, cellular phone based methods (server and client side as well) are also available to collect individual speed data of travelers [4]. Another emerging technology is represented by the Bluetooth-based vehicle detection [5], which is already applied in few cities.

Apparently, the plethora of the novel measurement technologies offers the possibility to gather link average traffic speed for large networks. A well-known example is the Traffic functionality of Google Map. The sources of Google data are FCD information of fleet management companies and mobile data of private smartphone users. This service determines speed categories by using color codes displayed on links (see Figure 1).

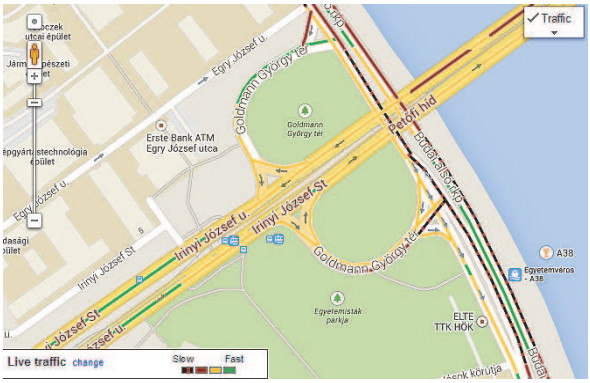

Figure 1. Google Map screenshot in Budapest

As demonstrated above, the traffic speed on road links is already available. Therefore, the problem is simply given: how to create a reliable traffic model based on FCD speed information and without any traffic flow measurement? The solution is not straightforward as the traffic demand (traffic input flows) must be determined only from

speed data. Note, that a given speed can be induced by different traffic inputs. Moreover, in case of a bigger traffic network, a combination of several demand flows has to be taken into account. Practically, this is a reverse engineering problem which analytical solution is not possible as it is highly underdetermined. As a result of these reasons, an online and iterative optimization is proposed to be applied for the problem.

The paper is organized as follows. First, the applied microscopic traffic simulation environment is introduced. Then, the proposed iterative calibration method is described. Finally summary and evaluation is provided.

2. Microscopic road traffic simulation with VISSIM

Road traffic simulations can be realized on macroscopic and microscopic level. On the former level, traffic is described as a continuum by macroscopic parameters (e.g. traffic flow, traffic density, average travel times, etc). The latter modeling level provides a high-resolution description of traffic via dynamic modeling of each vehicle. Macroscopic modeling level clearly has the advantage in terms of computational demand, therefore it is mainly used on a large spatial and temporal domain while microscopic modeling comes into prominence when analyzing the traffic of small networks, covering a few intersections. On both levels, the proper knowledge of basic parameters the process is essential in the design of a valid traffic simulation, i.e. traffic demands and origin-destination data of the network. However, these data are not fully available in most cases. Therefore, intelligent calibration of such parameters is suggested (see next section).

Users may choose among many different road traffic simulators. The greater part of them is commercial software. Additionally, open source simulators are also available, developed by universities or research institutes. Each of them has advantages or drawbacks depending on the individual demands of the user. In our paper, the applied simulation environment is designated to be designed by using VISSIM [6]. Furthermore, the integrated VISSIM-MATLAB environment is applied as described in [7] as a state-of-the-art technology for advanced simulation.

VISSIM is a microscopic simulator based on the individual behavior of vehicles. The goal of the microscopic modeling approach is the accurate description of traffic dynamics. Thus, the simulated traffic network may be analyzed in detail. The simulator uses the so-called psycho-physical driver behavior model developed originally by Wiedemann [8]. VISSIM is widely used for diverse problems by traffic engineers in practice as well as by researchers for developments related to road traffic. VISSIM offers a user friendly graphical interface (GUI) through of which one can design the geometry of any type of road networks and set up simulations in a simple way. However, for several problems the GUI is not satisfying. This is the case, for example, when the user aims to access and manipulate VISSIM objects during the simulation dynamically. For this end, an additional interface [9] is offered based on the COM (Component Object Model) which is a technology to enable interprocess communication between software [10]. Via the VISSIM COM the user is able to manipulate the attributes of most of the internal objects dynamically. The first step of a COM based simulation is to create the COM client. Then, one can realize the parts or even the whole process of the simulation.

3. Genetic algorithm based calibration

3.1. The basic problem and assumptions

The basic task is to reveal traffic demand, i.e. vehicle input flows entering the network. To achieve, this FCD based average speeds of road links can be applied where a link is determined as a road stretch between two signalized intersections. Moreover, measurement data of average turning rates of junctions need to be available. The microscopic parameters such as driving behavior, lane change are assumed to be well calibrated based on any existing methods cited in the introduction part for example. Real-world signal settings (green time split, offset, cycle time) are known and used in the simulation. The whole problem is depicted in Figure 2.

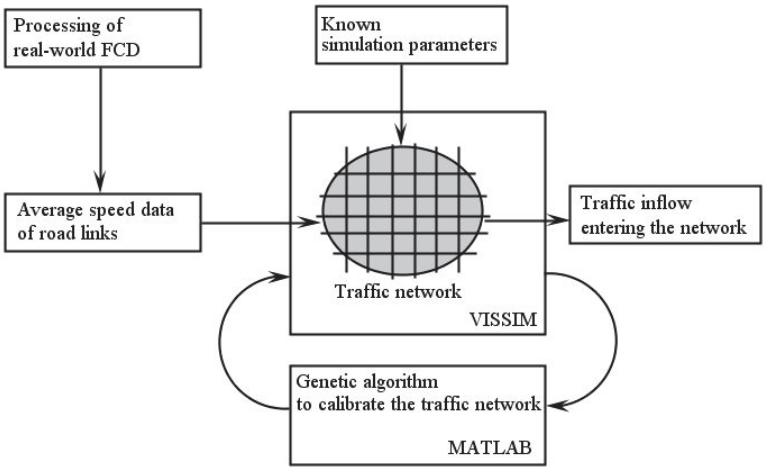

Figure 2. The basic calibration problem

3.2. Iterative method for calibration

The traditional method to optimization applies the derivative-based approach. However, if the objective function does not have a derivative, this approach cannot be used. In our case, the objective function contains a non-derivative term, i.e. the results of the VISSIM simulation run. A potential solution is to formalize and solve the optimization problem as a genetic algorithm (GA) problem.

The idea of using soft computing [11] for simulator calibration has already been recognized. However, the approach has only been used for the tuning of model parameters. In [3] a GA is used to optimize vehicle performance models in traffic simulators. In [12] the parameters of the car following and lane-changing dynamics of VISSIM are tuned using the particle swarm optimization method. The traffic simulator PARAMICS is calibrated in [13] by using a GA approach. [14] investigates a method for parameter optimization using a GA approach as well, tested in CORSIM and VISSIM simulators.

In our contribution, an iterative method is proposed. By applying the *Snapshot* option of the simulator, the algorithm tries to fit the average link speeds to the traffic inflows (*Vehicle Inputs* in VISSIM). Basically, during the calibration process a discrete time window is iteratively simulated. The time window only moves on if the optimization criteria is fulfilled, i.e. the average speeds measured in the simulator do not differ from the real-world FCD speeds. The applied fitness function of the optimization is as follows:

$$J(k) = \sum_{i=1}^{n} \left\| \frac{\bar{V}_i^{FCD}(k) - \bar{V}_i^{VISSIM}(Q(k))}{\bar{V}_i^{FCD}(k)} \right\|_{\infty} \rightarrow min$$

$\bar{V}_i^{FCD}(k)$ is the average speed of link i originated from FCD information at time step k. $\bar{V}_i^{VISSIM}(Q(k))$ denotes the average speed of link i produced by VISSIM in the previous simulation time window k. $Q(k)$ represents the applied vehicle input parameter.

Practically, the fitness function penalizes the relative deviation between simulated and real-world speed of each road link. The function is expressed as the infinity norm of the vector of traffic speeds. The GA based optimization runs within each time window until the termination criterion is achieved, i.e. $J(k) < \varepsilon$. For example, if termination parameter is set to $\varepsilon = 0.2$, the maximal admissible relative deviation of the average speed is 20%.

The complete calibration method is given by Figure 3. The calibration is realized in VISSIM-MATLAB environment by using COM programming, see [6].

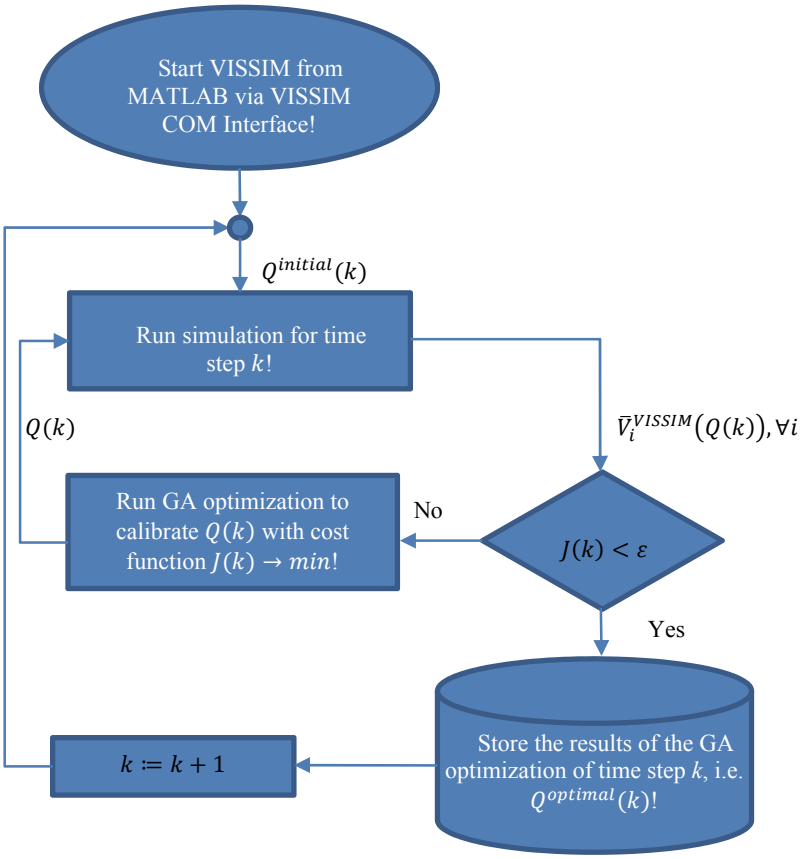

Figure 3. Workflow of the iterative calibration of VISSIM simulator

3.3. Example for the calibration

To demonstrate the proposed calibration technique, a simple example was investigated. Based on average speed functions deduced from real-world FCD data (iData Ltd.), a test link (Budapest, at Oktogon square) was calibrated for a peak period between 7:00 and 09:30. The calibration method has been applied with 1 minute sample time and termination parameter ε=0.2.

The variation of the fitness function $J(k)$ can be observed in Fig. 4. This reflects the maximum relative error obtained during the calibration process, i.e. the difference between the link speed based achieved from the calibrated parameters and the real-world mean speed. Practically, $J(k)$ always remains under 22%, which can be considered a satisfying accuracy for calibration.

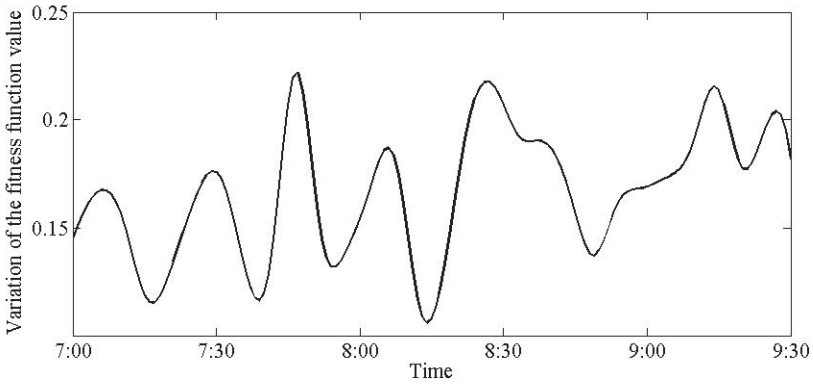

Figure 4. Fitness function value during the calibration

4. Summary

In our work, a genetic algorithm based calibration method is suggested for traffic simulators, reproducing realistic traffic conditions on a network represented by FCD data. The tuned variables of the simulation are the traffic flow inputs of the network. The obtained network link speed values should fit to the FCD speed data.

By using real-world floating car traffic data and a microscopic traffic simulator, the proposed calibration algorithm is efficiently applied and tested under different traffic conditions.

The proposed method was demonstrated by using the well-known VISSIM simulator. Nevertheless, it is emphasized that the technique can be applied for any type of microscopic traffic simulators.

Acknowledgement

This work is connected to the scientific program of EITKIC-12-1-2012-0001 project (supported by the Hungarian Government, managed by the National Development Agency, financed by the Research and Technology Innovation Fund), and TÁMOP-4.2.2.C-11/1/KONV-2012-0012: Smarter Transport project (supported by the Hungarian Government, co-financed by the European Social Fund). The FCD measurement data set, used for the simulation test, has been provided by iData Ltd., which is gratefully acknowledged.

References

[1] Columbia River Crossing, VISSIM calibration and validation. Technical report, 2006.

[2] Park B, Schneeberger JD: Microscopic Simulation Model Calibration and Validation, Case Study of VISSIM Simulation Model for a Coordinated Actuated Signal System. Transportation Research Record: Journal of the Transportation Research Board, Vol. 1856, pp. 185-192, 2003. DOI: 10.3141/1856-20

[3] Cunha AL, Bessa JE Jr, Setti JR: Genetic Algorithm for the Calibration of Vehicle Performance Models of Microscopic Traffic Simulators. Progress in Artificial Intelligence, Lecture Notes in Computer Science, Vol. 5816, pp. 3-14, 2009.
DOI: 10.1007/978-3-642-04686-5_1
[4] T. Tettamanti, I. Varga, Mobile phone location area based traffic flow estimation in urban road traffic. Columbia International Publishing, Advances in Civil and Environmental Engineering, 1(1):1-15, 2014.
[5] Qing O: Fusing Heterogeneous Traffic Data: Parsimonious Approaches using Data-Data Consistency. PhD Thesis, Delft University of Technology, 2011.
[6] Tettamanti T, Varga I: Development of road traffic control by using integrated VISSIM-MATLAB simulation environment. Periodica Polytechnica - Civil Engineering. Vol. 56(1), pp. 43-49, 2012.
DOI: 10.3311/pp.ci.2012-1.05
[7] PTV, VISSIM 5.4 User Manual. PTV Planung Transport Verkehr AG, Germany, 2012.
[8] Wiedemann R, Simulation des Straßenverkehrsflusses. Schriftenreihe des Instituts für Verkehrswesen der Universität Karlsruhe, 8, 1974.
[9] PTV, VISSIM - COM Interface Manual 5.4, PTV Planung Transport Verkehr AG, Germany, 2012.
[10] Box D: Essential COM. Addison-Wesley, ISBN 0-201-63446-5, 1998.
[11] Weise T: Global Optimization Algorithms – Theory and Application, 2008
[12] Aghabayk K, Sarvi M, Young W, Kautzsch L: A novel methodology for evolutionary calibration of Vissim by multi-threading. Australasian Transport Research Forum 2013 Proceedings, Australia, 2013
[13] Chu L, Liu HX, Oh JS, Recker W: A calibration procedure for microscopic traffic simulation. Proceedings of the 2003 Intelligent Transportation Systems. Vol.2, pp. 1574 – 1579, 2003.
DOI: 10.1109/ITSC.2003.1252749
[14] Park B, Won J, Yun I: Application of Microscopic Simulation Model Calibration and Validation Procedure: A Case Study of Coordinated Actuated Signal System. Transportation Research Record, Vol. 1978, pp. 113-133, 2006.
DOI: 10.3141/1978-16

Acta
Technica
Jaurinensis

Vol. 8, No. 2, pp. 153-164, 2015
DOI: 10.14513/actatechjaur.v8.n2.368
Available online at acta.sze.hu

Safety Considerations Regarding to the Shielding of Electric Fields during High Voltage Live-line Maintenance

G. Göcsei, I. Berta, B. Németh

Budapest University of Technology and Economics,
High Voltage Laboratory
Hungary, 1111 Budapest, Egry József u. 18.
Phone: +36 1 463 3236
e-mail: gocsei.gabor@vet.bme.hu

Abstract: Live-line maintenance (LLM) has several technical and economic benefits. Planned works can be executed without any consumer disturbance, while the network is energized. Strict regulations regarding to the working method, well-educated workers and well-considered safety factors ensures the high level of safety of this technology. Although statistics show that the number of injuries is lower than in case of de-energized works, the unseen short- and long term effects of electric and magnetic fields also have to be taken into consideration.

Keywords: live-line, maintenance, electric, field

1. The future of maintenance

Live-line maintenance can be referred as the "future of maintenance". It is more and more popular, because of its technical and economic benefits and is preferred from both the side of the Distribution/Transmission System Operator (DSO/TSO) and the consumers. This technology ensures planned reparation and maintenance works to be executed while the network is energized. No need of switching-off increase consumer satisfaction and is advantageous from the side of the operator as well [1].

1.1. Technical benefits

However most of the high voltage networks are looped to ensure proper safety factors of electrical energy distribution, there might be special conditions when switching-off a given line leads to violate "rule *n-1-m*", where *n* is the sum of the elements of the network and *m* is the number of elements which are out of operation because of a planned maintenance. In case of the failure of a specific element, network shall still be operating reliably.

In case of switching-off a given line for the purpose of de-energized work, current paths increase; increased length and current also increases losses. Operation of circuit breakers, disconnectors, etc. shortens the service period of the given equipment while degradation increases.

1.2. Economic benefits

Economic losses caused by de-energized works and not occurring during any kind of live-line work are the sum of the cost and fines of energy not served. Fines usually consists predictable and non-predictable parts. Non-predictable part (e.g. as a result of rainstorms, icing, snow, etc.) cannot been foreseen. Cost of predictable part can be minimized by keeping the length of de-energized periods as low as possible. Consumer satisfaction as an indicator of service quality can also be expressed as a financial quantity.

2. Live-line work related risks

Speaking about safety of any kind of live-line work, the most important question is the protection against electric shock. Depending on different technologies applied at different voltage levels it can be executed by different ways: rated insulating materials (e.g. blankets, gloves, rods, etc.) or air – as an insulator – can also be used. In case of high voltage live-line maintenance barehand method is widely applied. The main principle of this method is that keeping the proper distances prescribed can guarantee to avoid electric shock caused by touching or approaching specific network elements with different potentials. Safety distances include a flashover/breakdown component and an ergonomic component (as a function of voltage level). Switching and lightning overvoltages are also taken into consideration and worst-case scenarios are assumed. Because the most critical cases regarding to transmitted power, voltage level, current load, etc. occur during high voltage live-line maintenance the main topic of this paper is to focus on this field of LLM. Regarding to power arcs – might be caused as a result of multiple failures – thermal effects are the most critical. Electric shock also may lead to fatal injuries as well. The examination of these dangers and the analysis of the related risks is out of the scope of this paper. The unseen – especially long-term – effects of extra-low frequency (ELF) electric and magnetic fields is inspected instead.

3. Electric and magnetic field during high voltage LLM

Electric and magnetic fields together are often referred as “electromagnetic fields” but from the aspect of the topic it is essential to clarify the basic differences between electric, magnetic and electromagnetic fields. Static electric/magnetic fields have two criteria to meet: drift currents have to be negligible compared to the total current density and the length of the line cannot been in the same scale than the wavelength. In case of typical power lines with a frequency of 50 Hz, typical wavelength is 6000 km, which is significantly longer than the length of any conventional AC power line regardless of voltage level. Drift currents are determined by the material and frequency; in case of aluminium-steel conductors with a typical conductivity of 10^7 Ω/m, critical frequency can be determined and is about 10^{17} Hz.

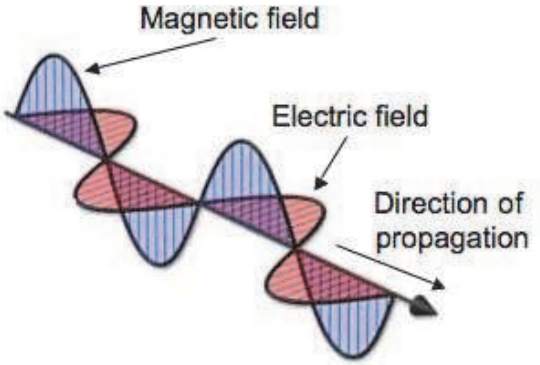

Figure 1. Electromagnetic wave [2]

Both length- and material-related conditions shows that in case of industrial frequency and conventional material selection, none of the fields around power lines can be handled as an electromagnetic field; electric and magnetic fields – and the effects of them – have to be taken into consideration separately from each other. These properties are often summarized by referring them as extra low frequency (ELF) fields – usually below a frequency of 100 kHz. Both electric and magnetic fields have several short- and long term health effects above a given limit. These limits are defined by the International Commission on Non-Ionizing Radiation Protection (ICNIRP). Because of the non-ionizing nature of these kinds of exposures, exposure doses cannot been defined. Current values together with the previous limits (before 2010) are summarized in Table 1.

Table 1. Current and previous limits of ICNIRP regarding to ELF electric and magnetic fields [3], [4], [5]

	Electric field strength limit [kV/m]		Magnetic field strength limit [μT]	
	Before 2010	Currently	Before 2010	Currently
Public (24 h/day)	5	**5**	100	**200**
Occupational (8h/day)	10	**10**	500	**1000**

4. ELF electric fields

4.1. Health effects of ELF electric fields

Short-term effects of ELF electric fields above the valid exposure limits are described as a discharge on the skin of the human body which may make any activity requiring high accuracy hard. This is the main reason of the application of conductive clothing in case of barehand method. However electric fields themselves do not have any direct long-term effects on the human body, drift currents occurring as a result of the variation of electric field has similar effects as the above-the-limit magnetic fields categorized as possibly carcinogenic to humans (2B) by the International Agency for Research on Cancer (IARC) of World Health Organization (WHO) [6].

4.2. Protection against ELF electric fields

A Faraday-cage is an enclosed metal surface. Theoretically inside an ideal Faraday-cage without any openings electric field strength is zero from outer source. Conductive clothing worn by the workers during high voltage live-line maintenance as an essential accessory of barehand method is acting as a Faraday-cage against ELF electric fields in the vicinity of the phase conductors of high voltage power lines. Typically the material of these clothing is a special mixture of a flame-retardant textile and conductive metal threads which guarantee proper screening efficiency.

Figure 2. Conductive clothing during an inspection in the High Voltage Laboratory of Budapest University of Technology and Economics

Effectiveness of a conductive clothing is currently determined by the ratio of two currents. Capacitive currents flow through both in the conductive clothing and the mannequin with a conductive surface placed in high electric field. There are several issues [7] regarding to the current arrangement of the valid standard [8]. There might be practical cases with above-the-limit electric fields inside the clothing, while the clothing pass the current ratio-based inspection of the standard. Laboratory inspections with direct electric field strength measurement in the High Voltage Laboratory of Budapest University of Technology and Economics (BUTE) prove that inside a conductive clothing without any face mesh ELF electric field strength may be above the current limits defined by ICNIRP. Results for electric field strength values at different voltage levels are shown in Fig. 3. The distance between the conductor and the mannequin was 30 cm, while the conductive clothing was energized (practical case after the connection of the potential clamp).

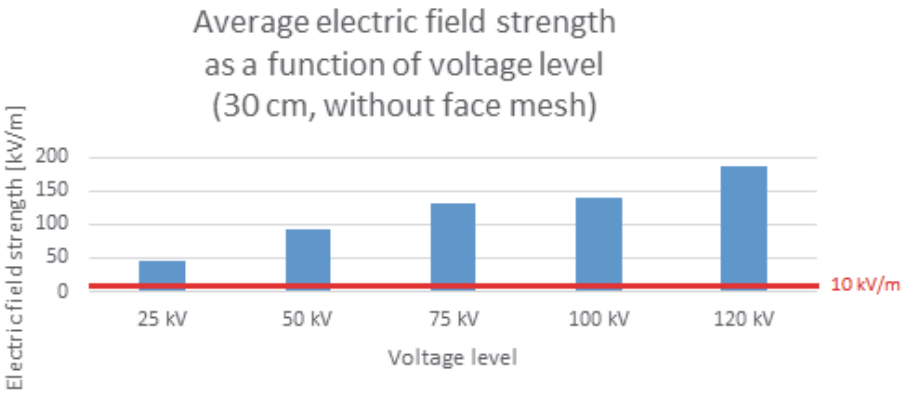

Figure 3. Electric field strength inside the conductive clothing in front of worker's face without any face mesh; laboratory measurement results (conductor-face distance: 30 cm)

As it can be seen in the Fig. 3, electric field was always higher than the current limit in the front of the face of the worker – even at the lowest inspected voltage level (25 kV). The same measurement have been repeated in the same arrangement, the only difference was the application of a face mesh with an opening of 1.75 cm (commonly applied in Hungary as a part of Dr. Béla Csikós' technology of high voltage live-line maintenance). Electric field strength results are shown in Fig. 4. Results show that face mesh has reduced the electric field strength below the valid limits. Fig. 5 shows that this shielding effect is highly effective even in an extreme case of a distance of 1 cm between the face and the conductor. It can be determined that face mesh is an essential part of any kind of conductive clothing [9-13]. Using a properly designed face mesh is always necessary to ensure the proper level of protection of LLM personnel at any time of any kind of high voltage live-line work. Finite element simulations have been executed to investigate critical face mesh opening size as a function of voltage level.

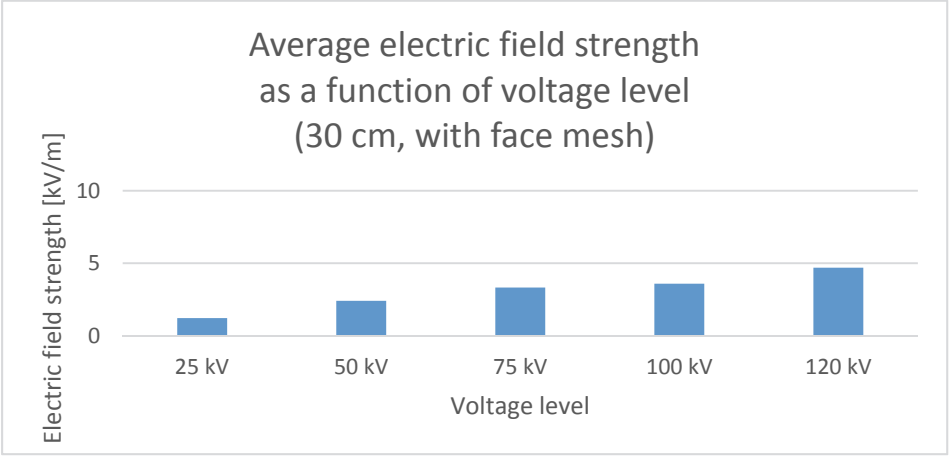

Figure 4. Electric field strength inside the conductive clothing in front of worker's face with a commonly used type of face mesh (Hungarian "Csikós", average opening size: 1.75 cm; laboratory measurement results, conductor-face distance: 30 cm)

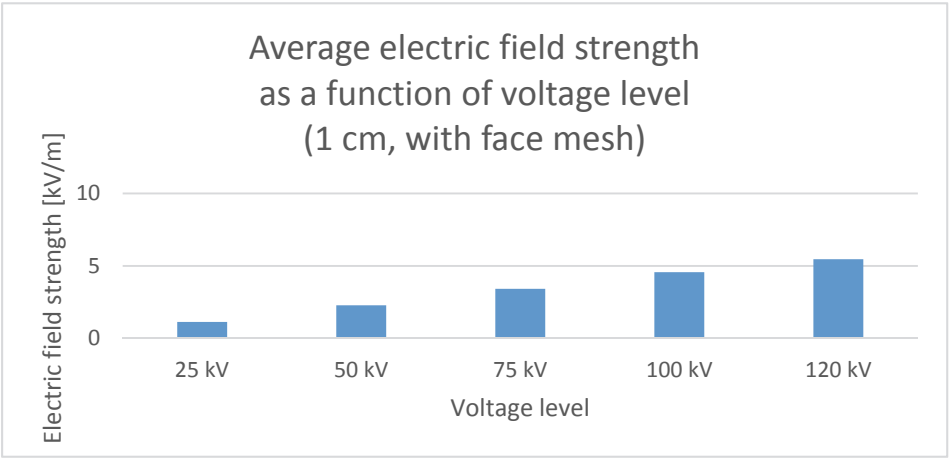

Figure 5. Electric field strength inside the conductive clothing in front of worker's face with a commonly used type of face mesh (Hungarian "Csikós", average opening size: 1.75 cm; laboratory measurement results, conductor-face distance: 1 cm)

3D CAD model used for the calculations is shown in Fig. 6 (grounded structures e.g. towers in the vicinity of the working site have also been modelled).

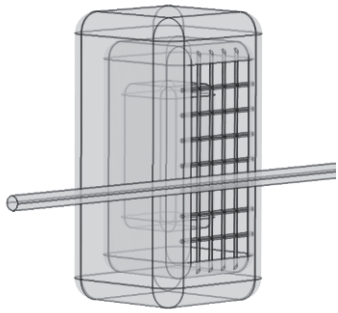

Figure 6. 3D CAD model for finite element simulations

Typical arrangements have been inspected in two cases. The first case is the approach of the phase conductor without the potential clamp being connected to the line (the potential of the conductive clothing floats). Three different mesh designs have been examined:

- A regular face mesh with an opening of 1.94 cm (horizontal) and 2.97 cm (vertical)
- A coarse face mesh with an opening of 6.82 cm (horizontal) and 6.44 cm (vertical)
- Conductive clothing without any face mesh.

During the examination of the approach of phase conductor, the distance between the line and the face of the worker was 50 cm. The voltage level of the conductor was the phase voltage of a 400 kV power line (about 231 kV). Electric field distribution is shown in Fig. 7.

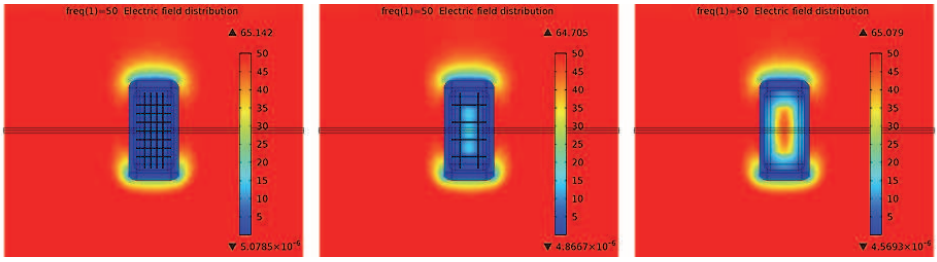

Figure 1. Electric field distribution during the approach of the conductor (from left to right: with normal, coarse and without any face mesh)

Minimal, average and maximal electric field strength values are summarized in Table 2 and in Fig. 8.

Table 2. Minimal, average and maximal electric field strength values for each cases (conductor approach)

Electic field [kV/m]	Normal face mesh	Coarse face mesh	No face mesh
Minimum	0,62	1,49	6,38
Maximum	2,47	**16,73**	**45,68**
Average	1,59	6,75	**24,89**

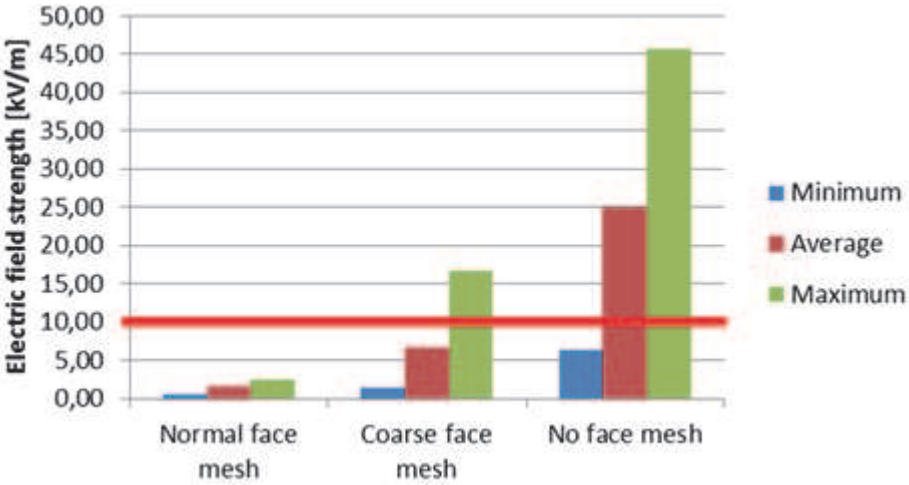

Figure 2. Values of electric field strength during the approach of the conductor

Fig. 9 shows the same cases during working at the potential of the phase conductor (with the potential clamp being connected to the line). Distance between the conductor and the face of the worker was 10 cm in this case. Electric field distribution is shown in Figure 9.

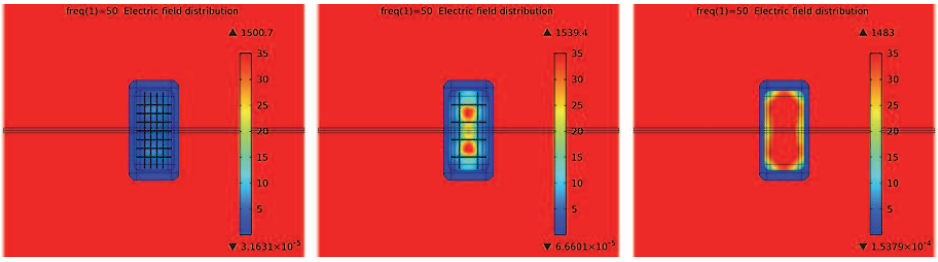

Figure 3. Electric field distribution while the conductive clothing is energized (from left to right: with normal, coarse and without any face mesh)

Minimal, average and maximal values of electric field strength is shown in Table 3 and in Fig. 10.

Table 3. Minimal, average and maximal electric field strength values for each cases (energized clothing)

Electic field [kV/m]	Normal face mesh	Coarse face mesh	No face mesh
Minimum	1,70	4,53	**21,05**
Maximum	8,06	**41,03**	**112,93**
Average	4,17	**16,23**	**60,00**

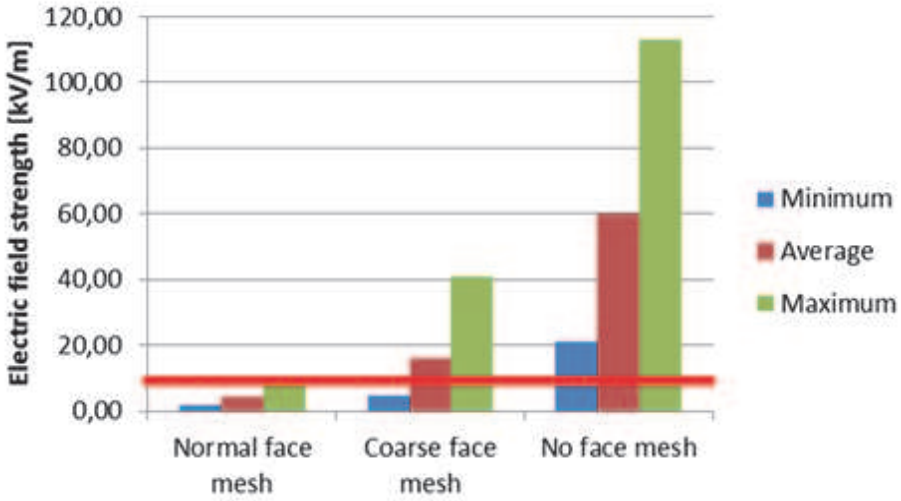

Figure 4. Values of electric field strength while the conductive clothing is energized

As it can be seen from the figures above face mesh has an essential shielding effect, which is the key of the safe work from the aspect of the protection against electric fields. Another important moral of the simulations is that mesh distribution has a significant effect on the minimal, average and maximal electric field strength values on the face of the worker during high voltage LLM.

Other simulations were executed to determine the optimal size of openings on the face mesh. Electric field strength as a function of mesh opening radius can be seen in Fig. 11 (overview) and in Fig. 12 (in the range of 0-10 kV/m) for different nominal voltage levels (the distance between the conductor and the face was 10 cm).

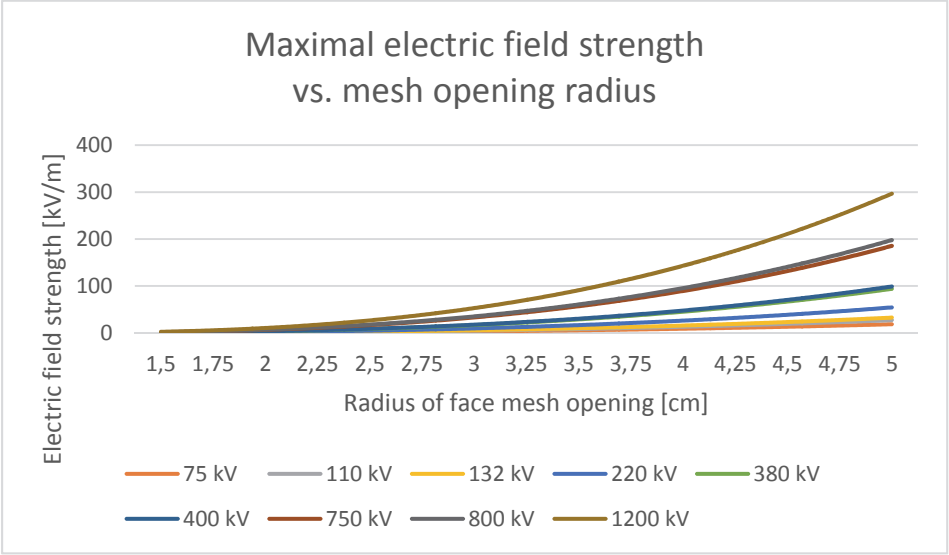

Figure 51. Electric field strength as a function of the size of mesh openings (overview)

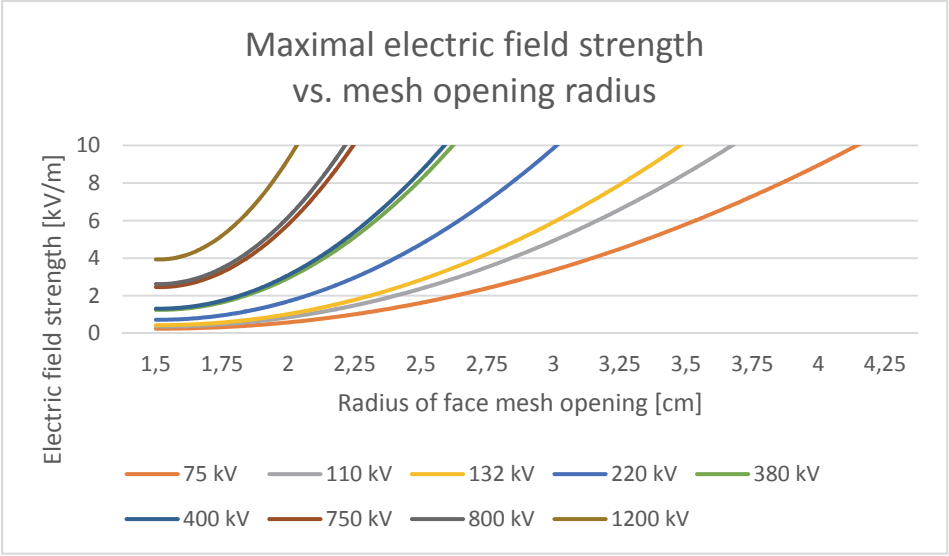

Figure 62. Electric field strength as a function of the radius of mesh openings (range of 0-10 kV/m)

Besides the proper electric shielding properties, other aspects have also to be taken into consideration to design a face mesh, such as visibility, ergonomics, or ventilation. In case of decrease of personal comfort, fault rate of LLM personnel might increase. This is the main reason why it is especially important to design a face mesh which keeps the electric field below the limits – even in critical cases – and guarantee proper ergonomic aspects at the same time. As it can be seen from Fig. 11 and Fig. 12, face meshes with an opening radius of 1.75 cm can guarantee these two criteria to meet simultaneously: electric field strength remains below 10 kV/m even in case of a voltage level of 1200 kV and based on the practical experiences it does not have any disadvantageous effect on the accuracy of the work – even in case of long periods of activity in the vicinity of high voltage power lines.

5. Summary

Electric fields can be shielded effectively in practice during live-line maintenance with conductive clothing acting as a Faraday-cage [14 - 16]. To guarantee the safety of the workers strength of electric field has to be reduced below the current exposure limits defined by ICNIRP at any time of the work. Measurements and simulations executed in the High Voltage Laboratory of BUTE have proved that during normal working conditions ELF electric fields might exceed their limits, even when a given clothing pass the screening efficiency test of the valid standard. Suggestions have been made and are under consideration by the committee of IEC for the revision of the current arrangement to simulate worst-case scenarios in practice.

Only conductive clothing with a properly designed face mesh is effective enough to reduce electric field strength. Protection of the face is essential and has to be ensured to guarantee the safety of the work. Proper design reduces the electric field effectively, but ensures proper ventilation and visibility as well. A practical recommendation for face mesh opening radius has been defined in this paper.

Live-line maintenance is the future of maintenance: numerous planned activities can be executed safely, economically and without any consumer disturbance. Safety of LLM personnel can be guaranteed by properly designed and frequently inspected live-line and personal protective equipment. Safety of workers has always to be handled as a first priority during any kind of work.

References

[1] International Council on Large Electric Systems (CIGRÉ): Report 561, Live Work – A Management Perspective, Joint Working Group, B2/B3.27, December 2013.

[2] http://fab.cba.mit.edu/classes/MIT/862.06/students/alki/GA.html, date of visit: 02 February, 2015.

[3] International Commission On Non-ionizing Radiation Protection (ICNIRP): Statement, General Approach to Protection Against Non-Ionizing Radiation Protection. Health Physics 82(4):540-548, 2002.

[4] International Commission On Non-ionizing Radiation Protection (ICNIRP): Guidelines for Limiting Exposure to Time-Varying Electric and Magnetic Fields (1 Hz – 100 KHz). Health Physics 99(6):818-836, 2010.

[5] Directive 2013/35/EU of the European Parliament and of the Council of 26 June 2013 on the minimum health and safety requirements regarding the exposure of workers to the risks arising from physical agents (electromagnetic fields) (20th individual Directive within the meaning of Article 16(1) of Directive 89/391/EEC) and repealing Directive 2004/40/EC.

[6] World Health Organization (WHO) International Agency For Research On Cancer (IARC): Monographs On The Evaluation Of Carcinogenic Risks To Humans. Volume 80, Non-Ionizing Radiation, Part 1: Static And Extremely Low-Frequency (ELF) Electric And Magnetic Fields, IARCPress, Lyon, France, 2002.

[7] Göcsei G, Németh B: New challenges in live-line maintenance. IEEE Electrical Insulation Conference, Seattle, WA, USA, 2015 (publication in progress).

[8] International Electrotechnical Commission (IEC): Live working. Conductive clothing for use at nominal voltage up to 800 kV a.c. and ±600 kV d.c. (IEC 60895:2002 + corrigendum 2003, modified).

[9] Göcsei G, Németh B, Kiss I, Berta I: Shielding efficiency of conductive clothing in magnetic field. Journal of Electrostatics, 71:(3) pp. 392-395, 2013. DOI: 10.1016/j.elstat.2013.01.001

[10] Göcsei G, Németh B, Tarcsa D: Extra low frequency electric and magnetic fields during live-line maintenance. IEEE Electrical Insulation Conference, Ottawa, Ontario, Canada, 2013. DOI: 10.1109/EIC.2013.6554212

[11] Göcsei G, Németh B, Kiss I, Berta I: Health effects of magnetic fields during live-line maintenance. International Conference on Live Maintenance (ICOLIM) Budapest, Hungary, Paper 235, pp. 1-6, 2014. DOI: 10.1109/ICOLIM.2014.6934329

[12] Göcsei G, Németh B, Tamus Á, Kiss I, Meixner J: Shielding efficiency of conductive clothing during live-line maintenance. International Conference on Live Maintenance (ICOLIM), Budapest, Hungary, Paper 306, pp. 1-5, 2014. DOI: 10.1109/ICOLIM.2014.6934339

[13] Göcsei G, Németh B, Tamus Z, Kiss I: Face Protection Investigation Against Electric Field On Live Line Workers (keynote). IEEE International Symposium of Electrical Insulation, San Juan, USA, pp. 535-539, 2012. DOI: 10.1109/ELINSL.2012.6251527

[14] Hotte P, Gela G, Mitchell J. D. Jr., Lyons P.F.: Electrical performance of conductive suits. IEEE Transactions on Power Delivery, Vol. 12, No. 3, pp. 1193-1201, 1997.

[15] Malgesini R, Valagussa R, Villa A, Carraro R,De Donà G, Milanello C, Parizia A: Conductive clothing for live line working. International Conference on Live Maintenance (ICOLIM), Budapest, Hungary, pp. 72-75, 2014.

[16] Neira L, Pascual H, Portillo Belinchón M, Pérez F, Albanese A, Fata O, Franchini R, Burna A, Stivanello I: A research on conductive clothing for live working. International Conference on Live Maintenance (ICOLIM) Budapest, Hungary, Paper pp. 1-7, 2014

Acta
Technica
Jaurinensis

Vol. 8, No. 2, pp. 165-178, 2015
DOI: 10.14513/actatechjaur.v8.n2.373
Available online at acta.sze.hu

Study of 2D and 3D Methods for Worn Surface Analysis of Tool Materials

I. Hatos[1], H. Hargitai[1], L. Solecki[2]

[1]Széchenyi István University, Department of Materials Science and Technology
Egyetem square 1, H-9026 Győr, Hungary
Phone: +36 96 613 572
e-mail: hargitai@sze.hu

[2]Széchenyi István University, Department of Vehicle Manufacturing
Egyetem square 1, H-9026 Győr, Hungary

Abstract: This paper is dealing with the comparison of the results given by different worn surface analysing methods. Maraging steel tool materials were used for the experiments. Specimens were produced by direct metal laser sintering from metal powder and by conventional way from rod having equivalent chemical content. The samples were age-hardened and surface treated by nitrocarburising and oxynitriding and they were tested by pin-on-disc type tribometer. The tribological behaviour was compared by using different methods for worn surface analysis. Worn area of the surface was determined by the surface profiles of 2D measurements and the wear volume was calculated by using 3D images of focus variation microscopy measurements.

Keywords: worn surface analysis, focus variation microscope, maraging steel, heat treatment

1. Introduction

Direct metal laser sintering (DMLS) belongs to the rapid prototyping technologies. By producing the part layer by layer from metal powder extremely complex metal parts can be made in a relatively short time. A promising field of the applications of DMLS is tooling by producing mold insert for injection molding with special cooling systems, which offers so called conformal cooling, acurved cooling channels with various cross sections that can follow the surface of the part [1-3].

The increasing demand for tool applications in polymer industry often required to improve the surface properties, such as higher hardness and enhanced resistance against wear and corrosion. Effective ways of surface hardening are thermo-mechanical surface treatments or PVD coatings or combining two of them in duplex treatments [4-6]. The most important treatments are nitriding, nitrocarburising or carburising [7-8]. Nitriding and nitrocarburising are also widely used in case of molds for injection molding.

Increase in hardness, having lower resistance of erosion, corrosion, abrasion at low and high temperature and by preventing adhesion and reducing friction increase of the life of working components can be achieved [9].

Cajner et al. presented an overview about the maraging steels used in mold manufacturing and focuses on the wear resistance. They concluded that the thermo-chemical heat treatments such as nitriding, nitrocarburizing, boriding and carburizing improve the wear resistance of maraing steel. Among of these processes the best results was given by plasma nitriding [6].

Wear properties are very important and staying in the middle of the interest in case of tool applications. Wear resistance of the surfaces can be characterized different parameters, e.g. mass loss (as wear loss) [10-11], wear coefficient [8, 12-14], and by analyzing the worn surface. For qualitative characterization to study the worn morphology and thus the wear mechanism, microscopic techniques are generally used, like conventional optical microscope (stereo microscope) or scanning electron microscope [10-11, 15-17] and there are different kinds of confocal microscopes [13].

Psyllaki et al. used SEM not only for quantitative analysis of the worn surface of PE-CVD diamond-like carbon coatings on tool steel substrates. They calculated wear rates, which is the volume loss per unit of applied load and unit of sliding distance [$mm^3 \cdot N^{-1} \cdot m^{-1}$] based on the volume losses from SEM observations of sections perpendicular to the wear tracks [18].

AL-Bukhaiti et al. determined the wear rate by using a 3D-surface profilometer (InfiniteFocus, Alicona) and measure cross-section area (A) (mm^2) and diameter (d) (m) of the wear track by using the equation (1):

$$W_c = \frac{A \cdot \pi \cdot d}{F_n \cdot l}, \tag{1}$$

where F_n (N) is the applied normal load, and l (m) is the total sliding distance. The volume loss was calculated by measuring the average of four cross-section areas of the wear track at four different points (90° intervals) [13].

Karamboiki et al. calculated first also the wear volume by using a profilometer. They measured cross-sectional area ten different locations along the wear track and then by multi-plying the average track area by the circumference of the slide cycle [$mm^3 \cdot N^{-1} \cdot m^{-1}$] [17].

Wang et al. determined the specific wear rate (W) by measuring mass loss (Δm (g)) according to the following equation (2) [10]:

$$W = \frac{\Delta m}{F_n \cdot l} \tag{2}$$

In some cases the wear behavior was characterized by volumetric wear loss [mm^3] [15]. Bressan et al. calculated cumulative lost volume by the division of the measured lost mass by the sample density [11]. Fontalvo et al. used white-light interferometry by taking measurements in three locations to determine the volume of transferred material [μm^3] using the software of the profiler [16].

As it can be clearly seen after the literature survey there are many methods and parameters for characterization the wear properties. Present paper is dealing with the tribological characterizing of maraging steel materials and focusing on the comparison of different methods which can be used for analysing the worn surface and thus qualifying the wear resistance of the materials. In our experiments DMLS MS1 and Böhler W722 samples with different surface hardness were tested after ageing, nitrocarburising and nitorcarburising with post oxidation (oxynitrided).

2. Experimental

2.1. Materials

For the experiments MS1maraging steel (1.2709) powder from EOS widely used in DMLS systems and W722 VMR (~1.2709) maraging tool steel from Böhler-Uddeholm having nearly the same chemical compositions (see Table 1.) were used. Both of them are ideal for injection molding applications.

Table 1. Chemical compositions of maraging steel tool materials

Material	***C***	***Cr***	***Ni***	***Mn***	***Si***	***Al***	***Co***	***Mo***	***Ti***
MS1	<0.03	<0.5	17-19	<0.1	<0.1	0.05-0.15	8.5-9.5	4.5-5.2	0.6-0.8
W722	<0.005		18				9.25	4.85	1.00

2.2. Sample preparation

Machining

For the experiments disk specimens having diameter of 30 mm and thickness of 5.5 mm were produced by DMLS technology from MS1 metal powder and fabricated from rod in case of W722 material. All samples were ground and polished with the same conditions after machining (Figure 1.). Samples were polished using 1 μm diamond paste before heat treatments, and no subsequent polish was carried out after nitrocarburising and oxynitriding. After age-hardening the samples were re-polished.

Figure 1. The laser sintered samples (left), ground specimens (middle) and polished sample (right)

Heat treatment

Bulk and surface heat treating methods were used to achieve higher strength and hardness of the steels and enhance the wear resistance on the surface.

Age hardening, which is the common heat treating technology of maraging steels and thermochemical treatments, such as nitrocarburising and oxynitriding were applied. The main parameters of the heat treatment methods are summarized in Table 2.

Table 2. Heat treatment conditions

Heat treatment	***Abbrev.***	***Temperature[°C]/ Duration [h]***	***Atmosphere***
age-hardening	H	500°C/4h	air
nitrocarburising	N	550°C/8h(Σ12h)	50% N_2; 45% NH_3; 5% CO_2
oxynitriding	ON	550°C/6.5h (Σ13h) 450C/1h	50% N2; 45% NH3; 5% CO_2 water

2.3. Test and evaluation methods, equipment

Wear test

Abrasion resistance was determined by ball-on-disc type tribology test. For the experiments an UNMT-1 Universal nano & micro tester was applied (see Figure 2) and zirconium oxide ceramic balls were used as test tools. For each test a new ball was applied.

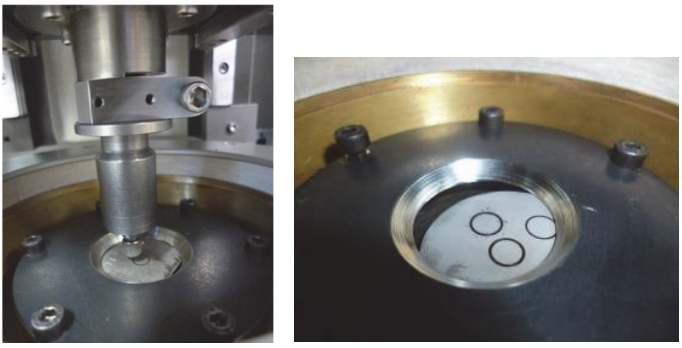

Figure 2. The pin (ball)-on disc type tribometer (UNMT-1 Universal nano & micro tester)(left) and the test specimen after the wear test(right)

The test parameters applied at the measurements were the followings:

- sliding speed:100 mm/s (318.47 1/min),
- set force: 20 N,
- duration: 360 m (time: 60 min)
- room temperature and
- atmosphere: laboratory air.

The typical characteristics of the tribological behaviour measured and calculated were friction coefficient and the area/volume of worn surface.

Characterizing of worn surface by 3D techniques

In present study wear resistance of the heat treated materials is characterised by worn area and wear volume. The area of the free-form surfaces can be determined directly by non-contact 3D optical surface metrology. The results of this technique have to be processed are data files having "stl" format, which describes the surface by triangles. This file contains the direction of the normal vectors and x, y and z coordinate value of vertices of the triangles forming the surface.

Basically white light laser (WLL) interferometry and confocal microscopy are used to visualize the surface topography. For our experiments two kind of device were used; an optical 3D measurement system, ALICONA InfiniteFocus confocal microscope and LEICA DCM 3D system, which unites the advantages of high definition confocal microscopy with interferometry by using WLL source.

To determine the wear volume Geomagic Studio software was applied including the following steps:

1. A 3D image was created by scanning the worn surface of the sample.
2. A plane was fitted to the measured surface near the wear track.
3. The data of the measurement range are intersected with the fitted plane, then the volume above and below the plane is subtracted from each other.

Characterizing of worn surface by 2D profiles

The worn area was determined by series of 2D profile measurements of the wear tracks. As in our case, the wear track torus (see Figure 2), thus the centre of the circle generated by the edge of the wear trace should be first determined. Then the profile measurements are performed along the diameter.

For the experiments a Taylor Hobson – Talysurf CLI2000 scanning surface topography instrument was used, by which both contact and non-contact (optical method) measurements can be carried out. In case of our samples the contact method was used, where a diamond stylus (angle: 90°, radius of curvature: 5 μm) is traversed across the test part to detect variations in the texture of the surface. The test results x, z coordinates: the displacement in the measuring direction and the corresponding height values, respectively.

Because of the long and multi-stage evaluation process of general engineering software an own software was developed to evaluate the profile measurement data and determine the worn area.

The evaluation includes the following steps:

1. Data (x-z) are plotted, and then connected with straight line or polynomial curve.
2. A straight line is fitted to the data points near the wear track.
3. The data of the measurement range connected with a curve are intersected with the fitted straight line, then the area above and below the line is subtracted from each other.

The evaluation process is illustrated by a screenshot on Figure 3. After input and show the coordinates stored in a text file, arbitrarily graphic fields (blue boxes) are placing onto the intact surface to designate the points for which the line is inserted. The position of the line is significant in terms of determining the size of the worn surface area. After placing the measurement field (brown box), the software calculates the worn area.

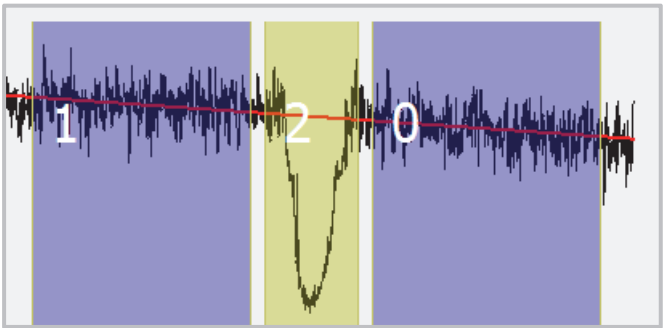

Figure 3. Screenshot of the own developed software for determination the worn surface area from the 2D profile measurement data

3. Results

3D Analysis

Based on the resulted topography images of the different samples by using different equipment the following conclusions can be stated.

- Optical confocal microscope (Alicona) is less suitable for mirror-like or optically variable surface digitization (see Figure 4 and 5).
- Setting parameters of optical confocal microscope are significantly affecting the numerical results of the measurements.

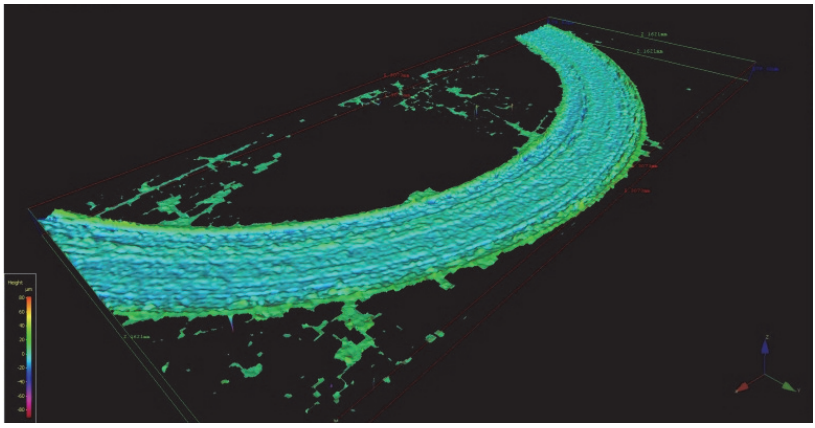

Figure 4. Worn surface topography of W722 materials by using Alicona microscope: after age hardening

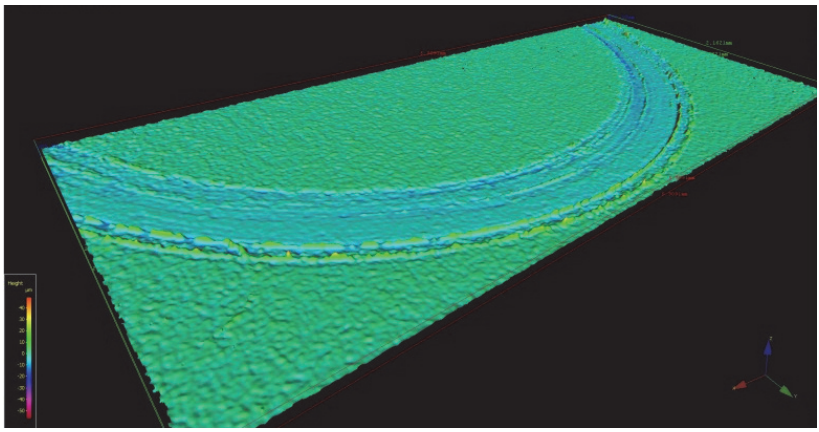

Figure 5. Worn surface topography of W722 materials by using Alicona microscope: after nitrocarburising

The resulted images of the surface analysis made by Leica microscope using combined mode of WLL interferometry with confocal microscopy can be seen in Figure 6.

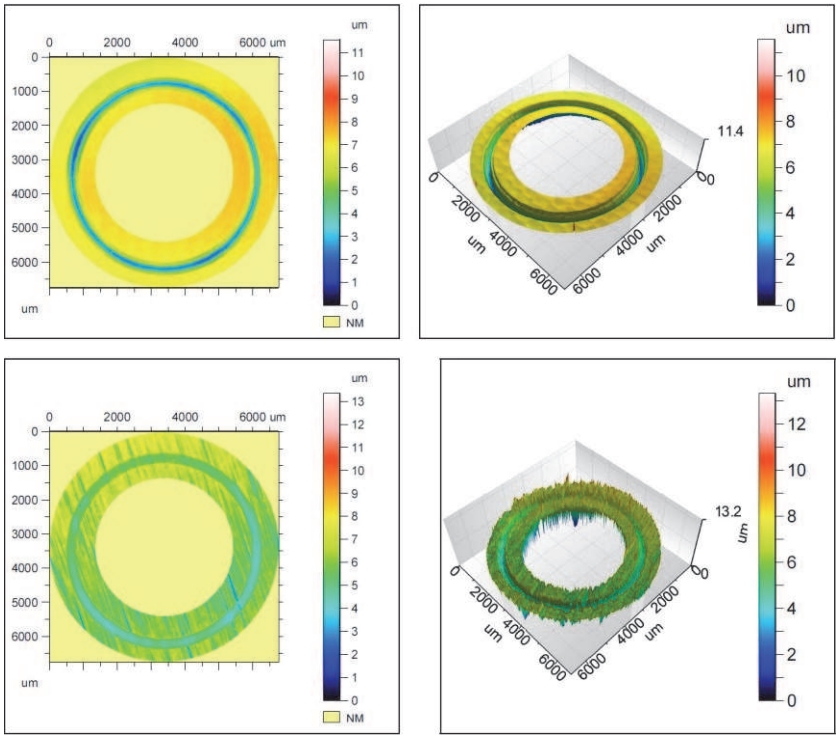

Figure 6. Worn surface topography of W722 materials by using Leica microscope: (top) after age hardening, (bottom) after nitrocarburising

The wear volume of the different samples was calculated from the results of Leica measurements and the comparison can be seen on Figure 7.

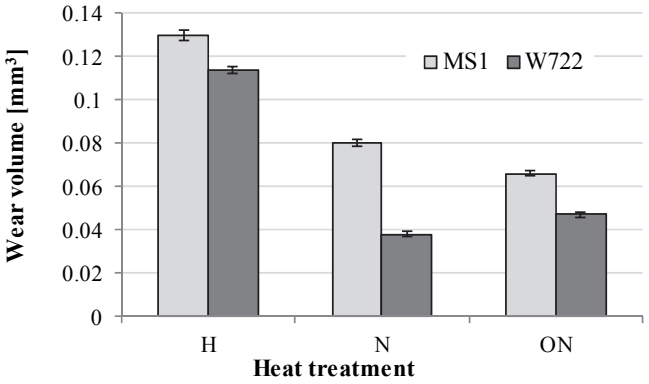

Figure 7. Worn surface volume calculated from 3D measurements data of MS1 and W722 materials having different heat treatments (H-age-hardening, N-nitrocarburising, ON-oxynitriding) by using Leica microscope

Results of 3D measurements show the same tendency, higher wear resistance of the samples made from rods of W722 material and having surface treatment.

3.1. 2D Analysis of the worn surface

The surface of the test specimens was scanned and profile curves were evaluated to calculate worn area as it was described in Chapter 2.3.

In generally the worn area is not the same value in each part of the surface because of the imperfect specimen fixation and the irregular wear process. These imperfections are often not visible to the naked eye but they can influence the result. In our experiments the robustness of the process were also studied by analysis of the measurement system uncertainties.

To make correct evaluation the appropriate number of the profile measurements have to be determined and influence of the location of the measurements need to be examined.

Influence of the number of measurements on the accuracy of the result

Based on the 3D measurements a specimen (the “worst”) was selected having the highest differences between the profiles in the different cross section areas to determine the optimal (appropriate) number of 2D measurements provide exact result. Figure 8. shows 3D images of the tested (“worst”) wear track and a “good” one.

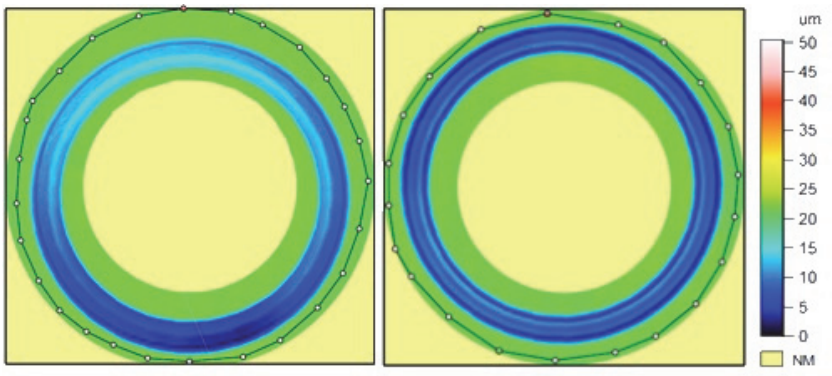

Figure 8. 3D topography of the "worst" wear track used for the test series to determine the optimal cross section profile number for correct evaluation (left: MS1 after age hardening) and a "good" sample having only slight differences between the profiles measured in different cross sections (right: W722 after age hardening)

On the wear track showed big differences in the different cross sections and obviously visible to the naked eye (see Figure 10) profile measurements were carried out along 4 diameter of the track (45° between the two neighbouring diameters), thus altogether 8 cross sections were scanned (No.1.-No.8.) (see Figure 9).

Figure 9. The locations of the profile measurements along the worn trace

Four type of measuring process was defined including different number of single results to determine the worn area and the standard deviation (accuracy the process) to characterize test specimen, as the followings:

- 8 data to compare: measurements in eight independent cross sections (No.1,2..8), characterizing the fault of choosing only one location to measure,
- 4 data to compare: average of measurements 2 cross sections along a diameter (No.1-2, No.3-4, No.5-6, No.7-8), characterizing the fault if taking measurement along one diameter.
- 2 data to compare: average of measurements 4 cross sections along 2 perpendicular diameters (No.1-4, No.5-6),
- 1 data: average of measurements 8 cross sections along 4 diameters 45° between the two neighbouring diameters (No.1-8)

The results of the measurements are presented in Figure 10.

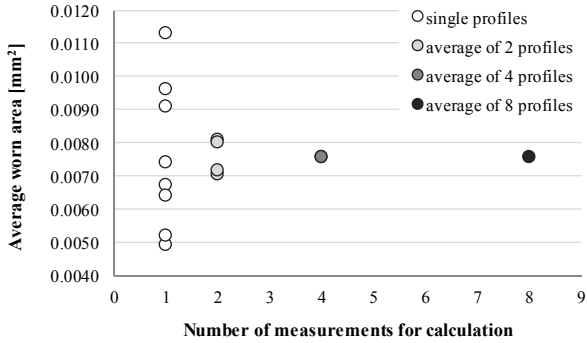

Figure 10. Influence of the number of profile measurements on the accuracy of the resulted worn area (material: MS1 after age hardening). Worn area values calculated as 8 single measurements or as an average of 2, 4 or 8 data.

In the first case having 8 independent data (white points) big differences can be seen between the resulted values. As expected one measurement is not enough for the precise characterization. If the measurement is carried out along one diameter and the worn area is calculated as an average of two data, more precise evaluation can be made (light grey). The four results (dark grey) are very close and having lower standard deviation. Doing profile measurements along two perpendicular diameters of the worn trace seems to be optimal for the precise characterization and more measurements (black point) cannot enhance the accuracy of the evaluation.

Results of worn surface analysis of DMLS and W722 materials

Worn traces of samples made by DMLS and from W722 were measured, and average worn areas were determined as average of measurements in four cross sections along two perpendicular diameters. The results are summarised in Figure 11.

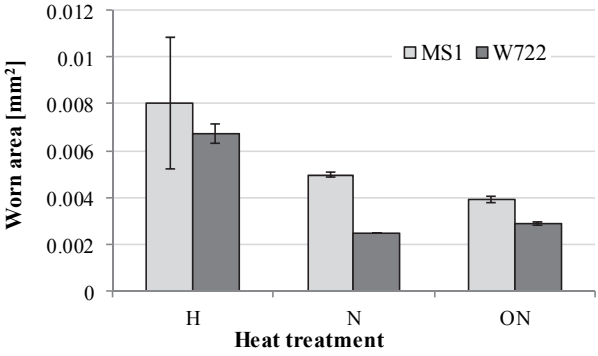

Figure 11.Worn surface area calculated from 2D measurements data of MS1 and W722 materials having different heat treatments

As it is clearly seen the results are in good agreement with data of 3D measurements: W722 material has better wear resistance in each case and surface treatments have positive effect, by reducing significantly the worn area.

Comparison the results of 2D and 3D analysis

When 2D measurements are carried out, the result is a distinct area of the worn trace or can be calculated as an average of some measurements, while in case of 3D measurements the whole volume of the worn trace can be measured. The surface and volume results can be compared if the results are related to a reference value. In our case the age hardened MS1 specimen was selected as reference value of both 2D and 3D measurements and each value was related to these values (average value), thus age hardened MS1 material has the maximal relative value of 1. Relative values of worn characteristic for both 2D and 3D measurements are shown in a common diagram (Figure 12).

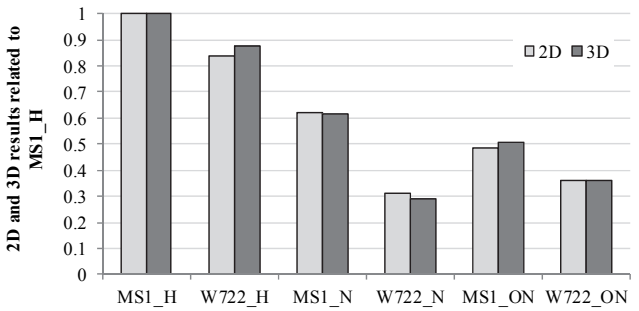

Figure 12. Comparison of related values of 2D and 3D measurements (the data are related to the age hardened MS1 material in case of both test method)

It is clearly seen by the results that both 2D and 3D methods are suitable for the correct characterization of the wear resistance of the materials. Only a slight difference can be seen between the results calculated from the different test results.

The wear volume was calculated by using the 2D profile results (as average of measurements 4 cross sections along 2 perpendicular diameters) and a good correlation was found between the measured and calculated data (Figure 13).

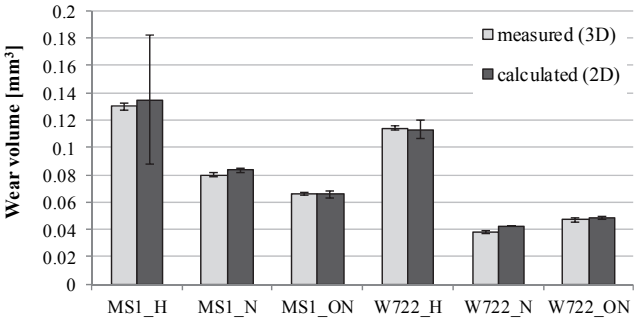

Figure 13. Comparison of wear volume data of 3D measurements and calculated from 2D measurement results

4. Conclusions

In our experiments one the other hand we focused on tribological characterization of maraging steels to compare the effectiveness of different heat treatments. Samples for the wear tests were produced by DMLS method from EOS MS1 material and from rod made of W722 tool steel with equivalent chemical composition. The samples were tested by ball-on-disc equipment after age-hardening, nitrocarburising and nitrocarburising with post oxidation.

The other aim of our experiments was to compare the reliability of the results derives from data of 2D and 3D measurements, and to determine the optimal number of measurements in case of the used 2D method.

Based on the results the followings can be concluded:

- Profile measurements along two perpendicular diameters of the wear track ensure good accuracy in determination worn area to characterise the wear resistance of the surface.
- In comparison the relative wear characteristics of 2D and 3D results of Leica measurements the same tendencies and very slight differences were found between the related values of the surface and volume measurements. Both of the evaluation methods can be used for precise characterisation and comparison of the wear resistance tested by ball-on-disc tribometer.
- The wear volume determined by 3D measurements show good correlation with the results calculated from 2D measurement data.

Acknowledgement

This paper was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. The research work presented in this paper was carried out as part of the TÁMOP-4.2.2.A-11/1/KONV-2012-0029 project in the framework of the New Széchenyi Plan. The realization of this project is supported by the European Union, and co-financed by the European Social Fund.

References

[1] Au KM, Yu KM, Chiu WK: Visibility-based conformal cooling channel generation for rapid tooling. Computer-Aided Design, Vol. 43, No. 4, pp. 356-373, 2011.
DOI: 10.1016/j.cad.2011.01.001

[2] Nickels L: Channelling quality for moulded parts using fast manufacturing. Metal Powder Report, Vol. 64, No. 8, pp. 8-12, 2009.
DOI: 10.1016/S0026-0657(09)70201-2

[3] Kumar S: Selective Laser Sintering/Melting. in Comprehensive Materials Processing, Editors: Hashmi S., Batalha G. F, Van Tyne C. J, Yilbas B, Elsevier, Oxford, pp. 93-134, 2014.
DOI: 10.1016/B978-0-08-096532-1.01003-7

[4] Zeghni AE, Hashmi MSJ: The effect of coating and nitriding on the wear behaviour of tool steels. Journal of Materials Processing Technology, Vol. 155–

156, pp. 1918-1922, 2004.
DOI: 10.1016/j.jmatprotec.2004.04.281

[5] Haftlang F, Habibollahzadeh A, Sohi MH: Comparative tribological studies of duplex surface treated AISI 1045 steels fabricated by combinations of plasma nitriding and aluminizing. Materials & Design, Vol. 60, pp. 580-586, 2014.
DOI: 10.1016/j.matdes.2014.04.026

[6] Cajner F, Landek D, Leskovsek V: Surface modifications of maraging steels used in the manufacture of moulds and dies. Materials and technology, Vol. 44, No. 2, pp. 85–91, 2010.

[7] Hussain K, Tauqir A, ul Haq, Khan AQ: Influence of gas nitriding on fatigue resistance of maraging steel. International Journal of Fatigue, Vol. 21, No. 2, pp. 163- 168, 1999.
DOI: 10.1016/S0142-1123(98)00063-2

[8] Karamboiki C-M, Mourlas A, Psyllaki P, Sideris J: Influence of microstructure on the sliding wear behavior of nitrocarburized tool steels. Wear, Vol. 303, No. 1–2, pp. 560-568, 2013.
DOI: 10.1016/j.wear.2013.04.002

[9] Jacquet P, Gołąbczak M, Lourdin P: Experimental characterizations of non-seizing solutions for plastic moulding tools. Applied Surface Science, Vol. 261, pp 824-829, 2012.
DOI: 10.1016/j.apsusc.2012.08.110

[10] Wang YX, Yan MF, Li B, Guo LX, Zhang CS, Zhang YX, Bai B, Chen L, Long Z, Li RW: Surface properties of low alloy steel treated by plasma nitrocarburizing prior to laser quenching process. Optics & Laser Technology, Vol. 67, pp. 57-64, 2015.
DOI: 10.1016/j.optlastec.2014.09.012

[11] Bressan JD, Battiston GA, Gerbasi R, Daros DP, Gilapa LM: Wear on tool steel AISI M2, D6 and 52100 coated with Al2O3 by the MOCVD process. Journal of Materials Processing Technology, Vol. 179, No. 1–3, pp. 81-86, 2006.
DOI: 10.1016/j.jmatprotec.2006.03.070

[12] Podgornik B, Majdic F, Leskovsek V, Vizintin J: Improving tribological properties of tool steels through combination of deep-cryogenic treatment and plasma nitriding. Wear, Vol. 288, pp. 88-93, 2012.
DOI: 10.1016/j.wear.2011.04.001

[13] AL-Bukhaiti MA, Al-hatab KA, Tillmann W, Hoffmann F, Sprute T: Tribological and mechanical properties of Ti/TiAlN/TiAlCN nanoscale multilayer PVD coatings deposited on AISI H11 hot work tool steel. Applied Surface Science, Vol. 318, pp. 180-190, 2014.
DOI: 10.1016/j.apsusc.2014.03.026

[14] Siow PC, Ghani JA, Ghazali MJ, Jaafar TR, Selamat MA, Haron CHC, Characterization of TiCN and TiCN/ZrN coatings for cutting tool application. Ceramics International, Vol. 39, No. 2, pp. 1293-1298, 2013.
DOI: 10.1016/j.ceramint.2012.07.061

[15] Navas C, Conde A, Fernández BJ, Zubiri F, de Damborenea J: Laser coatings to improve wear resistance of mould steel. Surface and Coatings Technology, Vol. 194, No. 1, pp. 136-142, 2005.
DOI: 10.1016/j.surfcoat.2004.05.002

[16] Fontalvo GA, Humer R, Mitterer C, Sammt K, Schemmel I: Microstructural aspects determining the adhesive wear of tool steels, Wear, Vol. 260, No. 9–10, pp. 1028-1034, 2006.
DOI: 10.1016/j.wear.2005.07.001

[17] Karamboiki C-M, Mourlas A, Psyllaki P, Sideris J: Influence of microstructure on the sliding wear behavior of nitrocarburized tool steels. Wear, Vol. 303, No 1–2, pp. 560-56815, 2013.
DOI: 10.1016/j.wear.2013.04.002

[18] Psyllaki PP, Jeandin M, Pantelis DI, Allouard M: Pin-on-disc testing of PE-CVD diamond-like carbon coatings on tool steel substrates. Surface and Coatings Technology, Vol. 130, No. 2–3, pp. 297-303, 2000.
DOI: 10.1016/S0257-8972(00)00711-8

Acta
Technica
Jaurinensis

Vol. 8, No. 2, pp. 179-187, 2015
DOI: 10.14513/actatechjaur.v8.n2.367
Available online at acta.sze.hu

Longitudinal Behaviour of Embedded Rails

Z. Major

Széchenyi István University, Department of Transport Infrastructure
Egyetem tér 1, 9026 Győr, Hungary
Phone: +36 96 503 038
e-mail: majorz@sze.hu

Abstract: The longitudinal behaviour of embedded rails significantly differs from that of ballasted and slab track with direct rail fastening. As the ambient temperature rises and falls, rails expand and contract, imposing stresses on the channel and the embedding compound. In this article the author presented the creation of a method that makes the calculation and the checking of embedded rails simpler and becomes closer to everyday engineering mentality. The author's design process includes an analysis of displacement and normal forces along the rail axis.

Keywords: embedded rail, longitudinal behaviour, breathing length of ERS

1. Introduction

The dilatational behaviour of embedded rails significantly differs from that of ballasted and slab track with direct rail fastening [1]. While the firstly flexible resistance is typical of the order of ~10 mm displacement, in the second and the third cases after a relatively small displacement (≤ 2 mm) the linear flexible period is followed by an ideally plastic period [2].

This latter behaviour can be modelled with software that enable non-linear calculation (providing spring stiffness and limit force), and in practical calculation the flexible period can be ignored and the plastic ballast resistance can be applied in well-known formulas [3,4].

In Fig. 1 there are force-displacement diagrams valid for ballasted tracks and slab track with direct rail fastening according to Hungarian standard MSZ EN 1991-2:2006 [5].

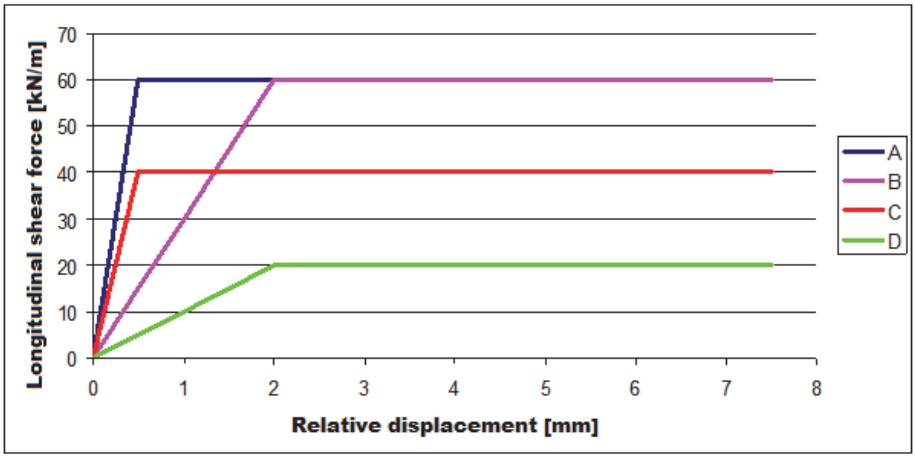

Figure 1. Value of shear resistance in case of ballasted tracks and slab track with direct rail fastening [5]

Note: Fig. 1 is applied to 2 rails!

Table 1. Features of bilinear spring characteristics [6,7]

Sign	***Name***	***Spring constant***	***Limit force***
		[kN/m/m]	***[kN/m]***
A	Rail fastening on loaded track	120,000	60
B	Ballast resistance on loaded track	30,000	60
C	Rail fastening on unloaded track	80,000	40
D	Ballast resistance on unloaded track	10,000	20

The two different kinds of dilatational behaviour (flexible and plastic) show a significant difference when the Figures of dilatational forces and rail end movements are compared. The figures show the behaviour of a 60E1 rail in case of a 45 °C-temperature change. 10 [kN/m/rail] value as plastic shear resistance was applied, and the embedded rail with the value of 5,000 [kN/m/m/rail] was described. When the length of the breathing period was established, 0.01 mm displacement as a limit condition was set and not complete stillness [8].

The modulus of elasticity of the rail is 206,000 [N/mm^2], the value of the linear heat expansion factor is 1.2*10^{-5} [1/°C].

Figures 2-5 describe the changes of dilatational forces and the evolving displacement along the length of the rail.

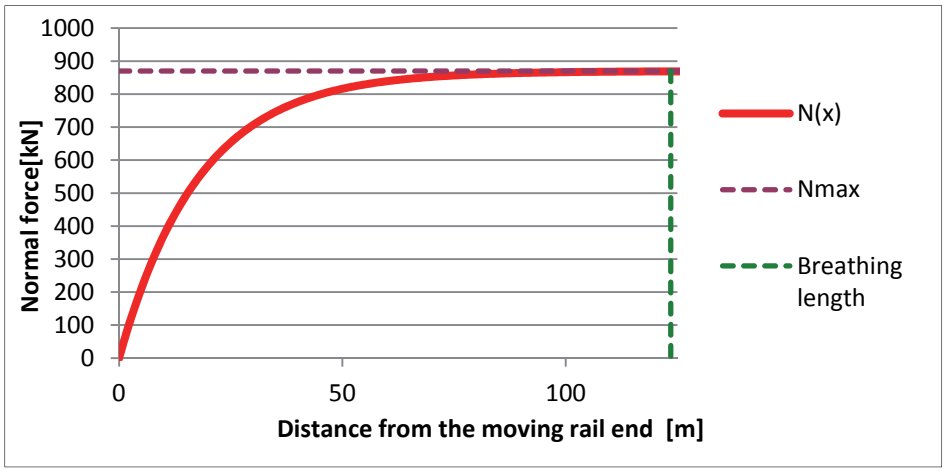

Figure 2. The change of dilatational force along the length of the rail in case of flexible shear resistance

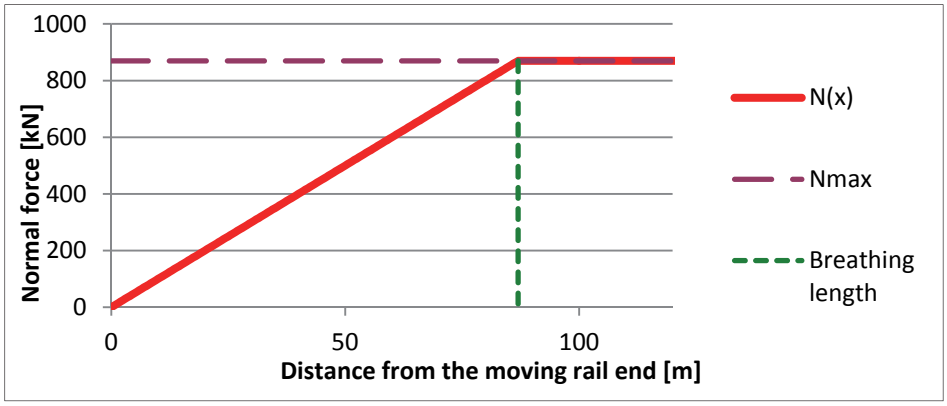

Figure 3. The change of dilatational force along the length of the rail in case of plastic shear resistance

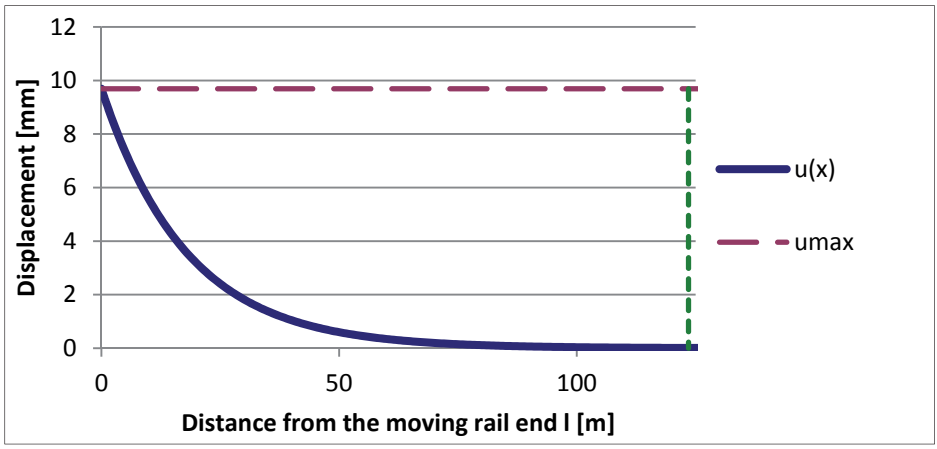

Figure 4. The displacement of the rail along the length of the rail in case of flexible shear resistance

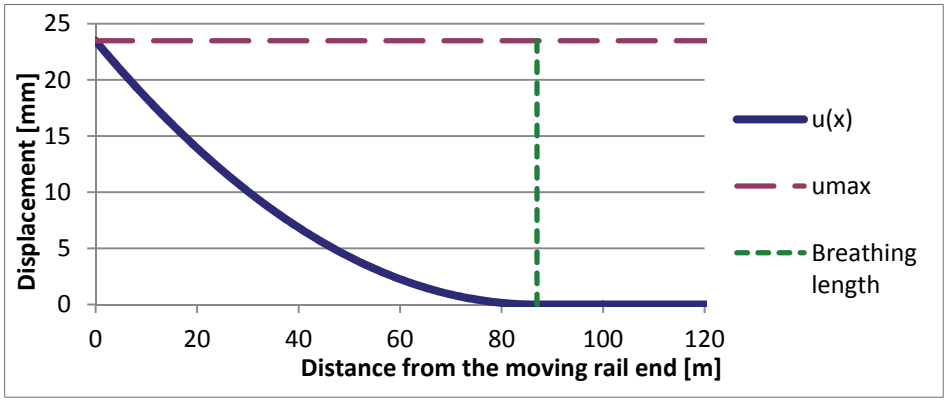

Figure 5. The displacement of the rail along the length of the rail in case of plastic shear resistance

Based on the Figures 2-5 it can be stated that the hypothesis of the linear change of the dilatational force in case of embedded rails can not be applied as in the cases of ballasted tracks and slab tracks with direct rail fastening, and the evolving rail end movements are significantly smaller than the ones expected in ballasted track and slab track with direct rail fastening.

2. The analytical description of longitudinal behaviour

A method for the analytical description of longitudinal behaviour can be found in Coenraad Esveld's book titled Modern Railway Track [9].

The occuring linear shear resistance is described with the following formula (Equation 1):

$$\tau = ku \ , \tag{1}$$

where:

τ: flexible shear/sticking resistance [kN/m],

k: longitudinal spring constant [kN/m/m],

u: displacement [m].

The differential equation describing the analysed problem can be given with the following formula (Equation 2):

$$\frac{d^2u}{dx^2} - \frac{k}{EA}u = 0 \ , \tag{2}$$

where:

x: distance from the moving rail end [m];

E: modulus of elasticity of the rail material [kN/m^2];

A: cross sectional area of the rail [m^2].

After solving the differential equation the following equation (Equation 3) can be received for the displacement along the length of the rail (leaving out the deduction):

$$u(x) = \pm\frac{\alpha\Delta T}{\mu}e^{-\mu x} \ , \tag{3}$$

$$\mu = \sqrt{\frac{k}{EA}} \ , \tag{4}$$

where:

α: linear expansion coefficient of the rail material [1/°C],

ΔT: change of temperature [°C].

Formula is valid for normal force change:

$$N(x) = \pm EA\alpha\Delta T\left(1 - e^{-\mu x}\right) . \tag{5}$$

In engineering practice the analysis of two factors is essential. The first one is the size of the complete moving length of the rail (if there is a rail fracture, the breathing length is examined), the other one is the extent of the displacement of the moving rail end (u_{max}).

Determining the breathing length in a closed form is impossible on the basis of Equation 5, the value of the maximum displacement can be counted from Equation 6:

$$u_{\max} = u(0) = \pm\frac{\alpha\Delta T}{\mu} = \pm\frac{\alpha\Delta T}{\sqrt{\frac{k}{EA}}} = \pm\alpha\Delta T\sqrt{\frac{EA}{k}} \quad . \tag{6}$$

An Excel program was prepared on the basis of the above calculation, which can calculate and present the changes of the normal force and displacements along the rail line graphically, and gives the complete moving length of the rail with u_{lim}=0.01 mm limit displacement.

With the purpose of validation the results of this program was compared to the results of a former FEM calculation [10] done at Széchenyi István University. The used material characteristics, spring constant and cross sectional data were in Table 2.

Table 2. Input parameters of the model examined

Parameters	***Values***	***Units***
E	206,000	[N/mm^2]
A	7,247	[mm^2]
α	1.2*10^{-5}	[1/°C]
ΔT	45	[°C]
k	10	[kN/mm/m]

The results are summarized in Table 3.

Table 3. Results I.

Parameters	***Value in [10] literature***	***Value after Z. Major***
Nmax [kN]	806.16	806.16
umax [mm]	6.33	6.59
z [m]	~60.00	79.50

There is only one significant difference in the breathing length among the results in Table 3. The reason of this is that the technical literature estimated the limit of the breathing length from a figure, it put it to where there is no substantial change in the normal force along the length of the rail, while in the author's Excel program the u_{lim}=0.01 mm limit displacement sets the limit. Therefore the extent of the normal force at 60 m in both cases was set. These figures are summarized in Table 4.

Table 4. Results II.

Parameter	***Value in [10] literature***	***Value after Z. Major***
N(60 m) [kN]	800.22	800.21

It can be seen that the running of the normal force is the same in both cases so the author's Excel program can be used for practical calculations, too.

In connection with the methods above it can stated that planning needs the purchase / creation of a FEM program or the calculation of relatively complicated interrelations has to be done.

Instead of the former interrelations the author is going to make a suggestion to apply a calculation method which is simpler from an engineering point of view but provides correct results.

3. Longitudinal behaviour from an engineering viewpoint

In the previous paragraph the analytical aspect of the dilatational behaviour of the rail in embedded rail structures was decribed, on the basis of which the author deducted the following interrelation (Equation 7):

$$u_{\max} = u(0) = \pm\frac{\alpha\Delta T}{\mu} = \pm\frac{\alpha\Delta T}{\sqrt{\frac{k}{EA}}} = \pm\,\alpha\Delta T\sqrt{\frac{EA}{k}} \quad . \tag{7}$$

Instead of applying the Equation 7 a „c" system factor for each rail system was defined, and it is used to rewrite the correlation into the following formula (Equation 8):

$$u_{\max} = u(0) = \pm\, c\Delta T\sqrt{\frac{1}{k}} \quad , \tag{8}$$

where:

$$c = \sqrt{\alpha^2 \cdot EA} \quad . \tag{9}$$

In the „c" system factor the material features of the rail was contracted which can be considered constant (α, E), and the cross sectional area in each rail profile. (A). In this form the relationship is simpler and the need for calculation is much lower. The introduced correlation makes the calculation of the rail end movement clear and free from the difficulties of transformation.

To determine the factors the written Excel program was used. Table 5 shows an example for „c" system factors belonging to applied rail profiles. Due to content limit all the calculated values will not be given. 12 rail profiles were examined while the method was being created. (Besides flat bottom rails grooved rails were also examined that are applied on light rail systems.)

Table 5. Values of „c" system factors in case of 54E1 and 60E1 rails

Rail profile	***c [kN^{0,5}/°C]***
54E1	14.386
60E1	15.084

Note: During the calculation the modulus of elasticity of the rail material was E=206,000 N/mm^2. The linear expansion coefficient was $\alpha=1.2*10^{-5}$ [1/°C].

Having done the validation exercise introduced earlier the result below was received (Equation 10):

$$u_{\max} = u(0) = c\Delta T\sqrt{\frac{1}{k}} = 14.662 \times 45 \times \sqrt{\frac{1}{10000}} = \frac{14.662 \times 45}{100} = 6.5978 \quad [mm] \quad (10)$$

which matches the author's result.

In order to be able to determine the value of the complete moving length of the rail in the relation of the longitudinal spring constant the power functions is determined describing the change considering the 45 °C temperature change in the following formula (Equation 11):

$$z(k) = ak^b \quad [m] \ . \quad (11)$$

The given general correlation by examining the results from the author's Excel program was received. A power function to the breathing length values determined at a spring constant changing in each rail system was adapted, whose „a" and „b" parameters are summarized in Table 6.

Fig. 6 shows the relation between the occuring breathing length values in the case of 54E1 rail and the longitudinal spring constant.

The applied formula provides the calculation of the breathing length in a closed form, for which there has not been a possibility before.

Due to content limit all the calculated values will not be given. 12 rail profiles were examined while the method was being created. (Besides flat bottom rails grooved rails are also examined that are applied on light rail systems.)

Table 6. Values of a and b factors of the power functions in case of 54E1 and 60E1 rails

Rail profile	***a***	***b***
54E1	15,716	–0.5763
60E1	16,598	–0.5763

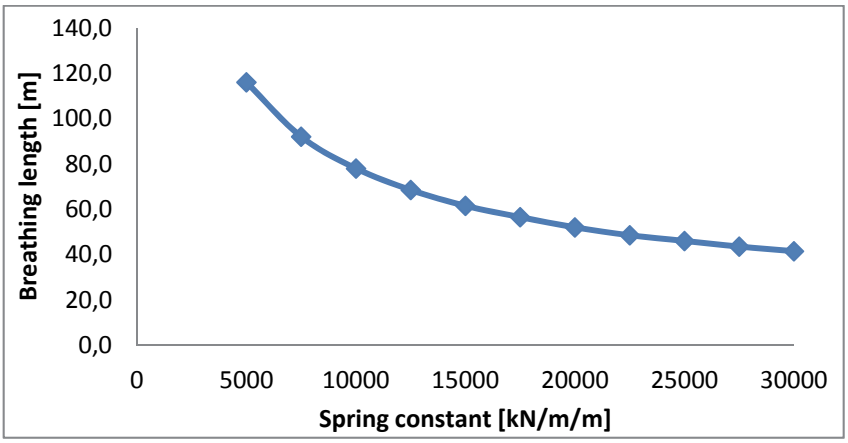

Figure 6. Breathing length occuring in case of 54E1 rail in relation of the spring constant (temperature change is 45 °C)

Doing the calculation the following result was received (Equation 12):

$$z(10000,35GPB) = 16{,}053 \times 10{,}000^{-0.5762} = 79.57 \quad [m], \qquad (12)$$

which result matches the former result. (Table 2)

4. Summary

In this article the author presented the creation of a method that makes the calculation and the checking of embedded rails simpler and becomes closer to everyday engineering mentality. The other great advantage of the created method is that it does not require a FEM program, it is time-saving compared to it, because it needs no model-building.

References

[1] Ludvigh E: Elastic behaviour of continuously embedded rail system. Periodica Polytechnica, 2001.
[2] Nemesdy E: Railway Superstructure. Tankönyvkiadó, Budapest, 1966.
[3] Lichtberger B: Track Compedium. Eurailpress, Hamburg, 2005.
[4] Führer G: Oberbauberechnung, Transpress, Berlin, 1978.
[5] MSZ EN 1991-2:2006: Actions on structures – Traffic loads on bridges.
[6] Major Z: A vasúti híd és vágány kölcsönhatása, Sínek Világa, Budapest, 2012.
[7] Major Z: Special problems of interaction between railway track and bridge, Pollack Periodica, Akadémiai Kiadó, Budapest 2013. DOI: 10.1556/Pollack.8.2013.2.11
[8] Kormos Gy: Longitudinal behaviour of rail embedded in elastic material. Periodica Polytechnica, 2001.
[9] Esveld C: Modern Railway Track – Second Edition. MRT-Productions, 2001, Zaltbommel, p. 186.
[10] Horvát F, Németh Gy: Technical Report – Edilon type ERS with 35GPB (35LPG) type rail, Győr, 2004.

Acta
Technica
Jaurinensis

Vol. 8, No. 2, pp. 188-200, 2015
DOI: 10.14513/actatechjaur.v8.n2.369
Available online at acta.sze.hu

Decision Support Model to Select Cushioning Material for Dynamics Hazards During Transportation

Á. Mojzes, P. Böröcz

Széchenyi István University, Department of Logistics and Forwarding
Egyetem tér 1., 9026 Győr, Hungary
Phone: +36 96 503 400
e-mail: mojzesa@sze.hu

Abstract: Designing a product-packaging system is a complex challenge for engineers. Building a suitable and sustainable system the packaging engineers have to take into consideration many aspects, demands and requirements. To protect the product by packaging-system from physical events during transportation, handling and warehousing the cushioning has got a significant role, because this means the primary protection. This paper aims to give a model help to decide the applicability of different cushioning materials based on their protective, sustainable and economical attributions. This model ranks the different types of packaging materials choosing the final optimal solution. Our study also introduces a determinant, which is in connection with the mechanical characteristic (cushion curve) of the material on different environmental factors.

Keywords: packaging, cushioning, decision support model, product-package system,

1. Introduction

The aim of our paper is to present a model, which can be a support mechanism to the packaging engineers. During a packaging designing process, a lot of parameter and requirements have to be taking into consideration, like product protection or logistic environmental factors. It's well known that the primary protection of product can be ensured by any kind of known or innovative cushioning solution. By the described method in this paper, the decision can be performed in the early phase of designing process, between different cushion materials using primary. Comparing the possible versions by different aspects, engineers can decide which should be the best or optimal material. This paper introduces a parameter to modify and enlarge the equitation used in practice during testing and comparing cushion materials. This parameter is in connection with mechanical characteristic of material on different environmental parameters, so the ranking and comparison going to be more realistic for those cases where the material specially sensitive for relative humidity changes or

temperature changes, and moisture can be appeared in very characteristics ways (like transcontinental transports). The model also can compare the continuously appearing new, innovative and environmentally friendly materials.

In this paper a solution is given, which can support the development processes of a product–packaging system. Groups were defined giving one of the most significant influences during a designing process. Also a basic function is performed, which can be a right solution, if the primary protection of product- cushioning systems should be compared and ranked. By this model all the parameters of influence group can take into account, except the behaviour of the mechanical characteristic on different environmental conditions. The basic model can be enough, when the materials non-sensitive for humidity or the cushioning characteristic is not a major issue.

2. Model set-up and the its components

When a product's designing processes is investigated, many of methods can be configured by former knowledge to make easier, more effective, make cheaper the product itself [1]. These designing processes mainly investigate the product design. Just some of the processes mention and work with the logistic and packaging aspects. [2] [3] Beside that, these models are become more and more complex; none of them deal with the packaging system as a complex system to protect product. These models are not able to give a help hand when the engineer have to choose the right, cost-effective, optimal and mechanical suitable packaging protection system, like cushioning. Nowadays, this decision is getting more and more important, as the environmental regulations, waste requirements get higher accent, and the material science develops new materials, which are possible applicable for the packaging industry.

The described method in this paper helps and supports the decision policy among the possible suitable move-damping systems. By this, the possible cushion materials can be re-parameterized or modified if the consumer or the logistic environment requires subsequently this. To build up the model, at first, categories (parameter groups) have to be grouped which influence the product package system. Fig. 1 shows these groups.

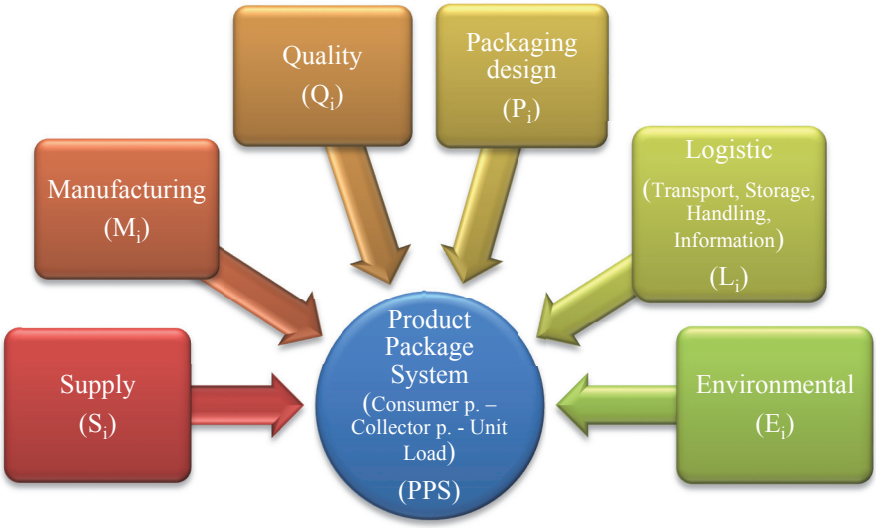

Figure 1. The Product - Package System

In the system, the groups mean the followings:

- PPS: (Product Package System): The complex system, which includes each package components and product protect devices. The result and sum of sub categories (described below) have to be the maximum, as the result of development.
- S_i: (Supply): Category includes those information and parameters, which are important during the supply and purchasing process of PPS components. In case of this sub-category, the minimum (for example costs in EUR/m^3, etc.) of amounts is the most important ($S_i \rightarrow min$).
- Q_i: (Quality): The quality parameters mean the most important factors in this sub-category. The maximizing of the criteria's ($Q_i \rightarrow max$) is the most important, like the percentage of fault free income packaging devices (%/ 1000 pcs), etc.
- M_i: (Manufacturing): Sum of parameters can be connected with the producibleness. In this category the minimization of the values ($M_i \rightarrow min$) is the aim, like manufacturing time (pcs/min), switch over time (min/pcs), etc.
- P_i: (Packaging): It contains the entire packaging system, built up around the product. Maximizing the contains ($P_i \rightarrow max$) of this sub category is the goal, like resistance of package devices, passing the functional requirements, percentage of intact products, etc.
- L_i: (Logistics): Those parameters are in this sub category, which are able to affect the effectiveness of transportation, handling and storage, like stackability (kg/package), possibility of mechanical handling, etc. To maximize the parameters is the goal ($L_i \rightarrow max$).

- E_i: (Environment): This sub-category contains the parameters, which are important when the whole package is at the End-of-Life and becoming waste. To minimize the factors ($E_i \rightarrow min$), like package waste per product (kg/pc) or the degree of non-recyclable materials in the system, are the goals.

The contents of each sub-category can include many factors, so in the followings they are not detailed as a complete list. They are important when the model going to be improved in an exact task. The function (1) of the PPS is going to be the following, based on the Fig. 1:

$$PPS_i = f(S_i, Q_i, E_i, L_i, M_i, P_i) \quad (1)$$

Building up a product package system can be modified by each sub categories. The weight of the sub categories can be changed in each designing task, so a weighting coefficient has to be adopted.

For each of sub-categories the following coefficients (2) can be applied:

$$a_S = \frac{S_{\min}}{S_i};\ a_Q = \frac{Q_i}{Q_{\max}};\ a_M = \frac{M_{\min}}{M_i};\ a_P = \frac{P_i}{P_{\max}};\ a_L = \frac{L_i}{L_{\max}};\ a_E = \frac{E_{\min}}{E_i};\ \quad (2)$$

Based on the above written, the function (3) can be expressed for the value analysis:

$$PPS_i = (a_S S_i + a_Q Q_i + a_M M_i + + a_L L_i + a_E E_i) \cdot a_P P_i \quad (3)$$

Where:

S_i: the Supply, etc. sub category markings

a_S: the weight coefficient concerning to the sub-category, like Supply (S_i), etc.

As from the function can be seen, the $a_P P_i$ block is factored out, because the protection of product is the major and primary task of a product package system. The parameters, which are applied to the weighting coefficients, can be modified or extended in order to fine the analysis. It is important that the coefficients have to be defined individually in each design development task. To verify the model test calculations was performed. The bases of calculations were the values of three types of different cushioning materials (Fig. 2).

Sample 1. Polyethylene | Sample 2. Expanded Polystyrene | Sample 3. Thermo-Plastic Strach

Figure 2. Tested and compared cushions

Two of them are common plastics foams (Polyethylene: PE and Expanded Polystyrene: EPS), and the third one is a new innovative and environmentally friendly material (Thermo-Plastic Starch: TPS). Comparing these different material versions the weighting values (a_i) are same for each cushion, and the sub-category values are pre-defined and based on previous investigations and researches. With the calculation the types can be ranked and compared. Based on the results, it can be decided whether the new environmentally friendly material capable to be a final cushioning material or not. The comparison of the materials can be seen in Table 1.

Table 1. Calculating and comparison the materials by categories

	S_i	a_S	Q_i	a_Q	E_i	a_E	L_i	a_L	M_i	a_M	P_i	a_P	**Cum. Value**	**Δ values (%)**
PE	4	35	8	20	5	30	9	30	5	25	10	60	507000	-9,1
EPS	2	35	7	20	7	30	10	30	7	25	10	60	537000	-3,8
TPS	9	35	8	20	1	30	10	30	5	25	10	60	558000	0,0
	$S_i \rightarrow min$		$Q_i \rightarrow max$		$E_i \rightarrow min$		$L_i \rightarrow max$		$M_i \rightarrow min$		$P_i \rightarrow max$			

The last column of the Table 1 is able to show the rank and the differences between the investigated materials. Actually, between the results, there are not so huge differences.

3. Discussion of the modified model

When a new material appears, many suitability tests, verification processes and investigations are required in order to be applied the given material in practice. It means a long time, huge amount of measurements, and of course it is a very expensive process. To make a decision about a new material, whether it is suitable or not, information are required from two aspects. The first aspect is the information and parameters written in the previous chapter, by a basic comparison. The second one is a knowledge about the given supply chain, where the possible material going to be applied as a product protection system.

Many papers investigated those influential logistical and environmentally parameters, which can have major influence on both product and packaging [4][5]. From two groups of logistic stresses, which are the mechanical and environmental affects, I had to choose the affects, can make the highest risk for the packaged product. Investigating these, the shock/drop and the temperature/humidity stayed as major effects.

There are papers, which investigated the cushion characteristics of different materials [6][7]. Several testing methods show relevant information, but one of the most important is the "cushion curve" test method [8]. A classical cushion curve can be seen on Fig. 4, when the material quality is tested in the function of thickness, drop height and static stress [9] [10]. If the engineers want to know the all-possible variation, it means a lot of tests and it requires a very long time. To solve this and to

shorten the duration, the so-called Stress – Energy method can be a suitable solution [11] [12].

The dynamic stress can be defined in $G \cdot s$ (G times static loading), and the dynamic energy can be defined as $s \cdot h/t$ (static loading times drop height divided by cushion thickness). Both of them have units of Pascal [kPa].

This method says that for any calculated energy, G can be predicted. Predicted G levels have to be compared (from $G \cdot s$) to actual G levels from the cushion curve in different combinations of s, h and t. So, G levels can be predicted very accurately.

1st step: The maximum and minimum limits on the energy absorbed have to be set. As Energy = $s \cdot h/t$, the minimum energy corresponds to the smallest s, the smallest h, and the largest t that data to be wanted for. The maximum energy corresponds to the largest s, the largest h, and the smallest t that data to be wanted for.

2nd step: Divide the energy range in step first into about 5-10 approximately evenly spaced points. If the range 5 to 100 *kPa* is used, then test for energies in steps of about 20 *kPa*. So, 9 different energies could be chosen for equal to 20, 40, 60 . . . . . and 100 *kPa*.

3rd step: For each of the energy step have to be chosen in the second step, select five-six different combinations of (s), (h) and (t) values that give this energy. Next, minimum 6 drops have to be performed with cushion tester (or drop tester). For the first drop, the cushion tester has to be set up for an equivalent free fall drop height of 400 millimeters, and a cushion sample can be selected with an actual thickness of 10 *mm*. Enough weight has to be added to the platen in order to achieve a static stress of 1 *kPa*, and drop on the platen. The shock pulse can be captured the peak acceleration (G) by recording machine (in this way we apply HBM-Spider 8). These have to be completed for the six drops corresponding to an energy of 20 *kPa*. Now we can summarize the experimental data in a table.

4th Step: Repeat step third for each of the energies (doing on the six levels of the total range) in the range chosen in second step and construct the stress vs. energy relationship shown below. The stress values listed are the means for the 5 replicates tested for each energy class.

5th step (optional: Fit an equation to the stress (G) vs. energy data. The relationship between stress and energy can usually be described to a high degree of correlation by the exponential relationship:

$$stress = a \cdot e^{b(energy)}$$

where (a,b) = constants specific to foam type and density and ($e = 2,71$ constant)

This regression can be used to best fit this equation to the data. The next step is to plot dynamic stress versus dynamic energy, and apply a simple exponential curve fit to the data points.

Fig. 3 shows the applied instrumentations during the tests, which were are the following:

- COMETECH QC-113B1 drop tester (1)
- PCB piezoelectric tri-axial accelerometer (range: 0 – 300g) (2)
- Hottinger-Baldwin Messtechnik SPIDER 8 Analyser and Software Package (3)
- Loads (4)
- Cushioning sample material (5)

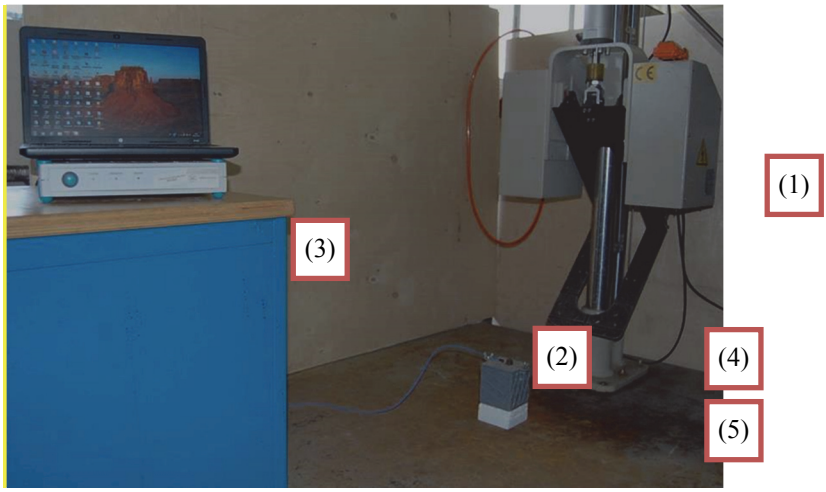

Figure 3. Measurement system for step 3 by drop test

By this 5 steps method, the points of the curve can be estimated considering the different variables. This estimation method can be seen in the Fig. 4 and 5.

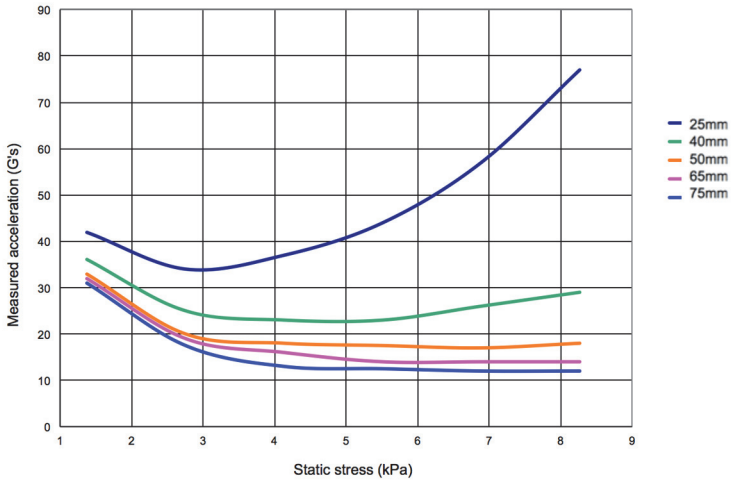

Figure 4. The cushion curves of TPS material (with different thickness')

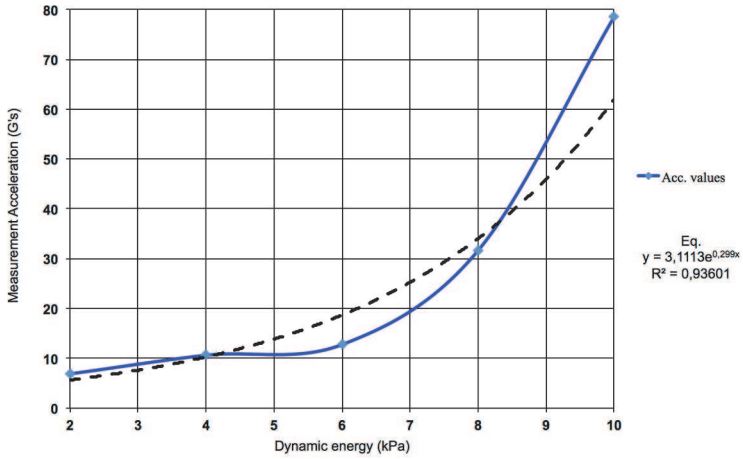

Figure 5. The Stress energy model of TPS foam on 23°C and 50%RH

It can be seen at first sight that an exponential function fits best to the measured values, and this is proved by both industrial and research projects. In this calculation the R^2 also had showed that function assimilating to the measurements.

Nowadays, the continuous development and the environmental regulations more and more prefer to use natural materials, like corrugated fiberboard, honeycomb or innovative biodegradable materials as TPS foam is. The widely known problem with these materials is the sensitivity for changes of relative changes means some kind of mechanical property decrease.

To be able to apply these materials as a real packaging material on transportations with different circumstances where the humidity and temperature values changes so

fast, the equation (3) has to be modified. The modification is necessary to be able to show the changes on those development projects, where a sensitive material has to be compared with others. It can be seen from the basic equation (3) that the final value affected majorly by the values of Packaging category (P_i). So, if the mechanical behaviours of material on different environmental versions want to be taken into consideration, the new modifying factor has to be grouped into this category. This factor will be the modification factor α. The $\alpha = f(R^2)$, where R^2 is the determinant, and the Stress – Energy method will calculate the given function. A modification factor should be found, which is able to modify the full model and be able to show the suitability and applicability too. The function of modification factor (eq. 4) is going to be as follows:

$$\alpha = R^2 - \frac{1}{(R^2)^m} \tag{4}$$

, where $0 < R^2 \leq 1$ and $R^2 \rightarrow 0$, so the $\alpha \rightarrow -\infty$. The connection can be seen on Fig. 5.

The power of *m* ($1 \leq m$) counted with *m*=1. Other tasks, the m can be varied to any other value in connection with the given sensitiveness of product, or any other special characteristic. This can be seen on Fig. 6.

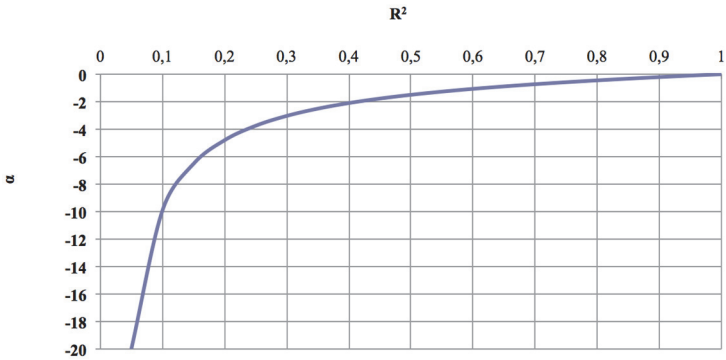

Figure 6. Connection of R^2 when m=1

According to the written above, the α is going to be implemented as a power of Packaging category. So, the basic equation extended with α, going to be expressed as follow (5):

$$PPS_i = (a_S S_i + a_Q Q_i + a_M M_i + a_L L_i + a_E E_i) \cdot a_P \cdot P_i^{\alpha} \tag{5}$$

In the practice, it means the followings. Most of cushion curve and Stress-energy tests are performed on standard climate conditions (23ºC/50%RH). To be able to implement and investigate the humidity sensitiveness on materials, a 6th step has to be add to the Stress – Energy method. This step is the investigation on variable

environmental parameters, like different temperature and humidity values. Theoretically, it means small change in R^2 value in case of common plastics foams, but for the natural and sensible materials can result significant decrease in R^2 value.

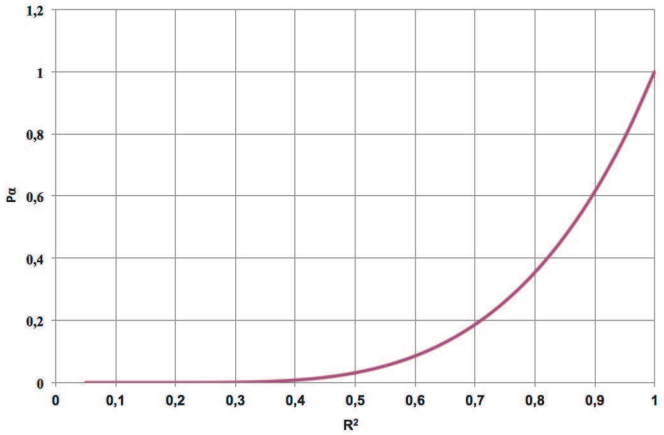

Figure 7. The connection of R^2 and α

It seems well on the Fig. 7 using the exponent, that how the decrease of R^2 affects the P_i. Consequently the modified and abated $P_i\alpha$ has a great role for the final result and can be decided its applicability.

4. Verification of the model

To be able to verify the model three types of materials were chosen. These materials have already been defined in the previous chapter. The tests of Stress energy model had been done on same density (25 kg/m^3) foams. To investigate the behaviour of R^2 the 6 steps method was performed on each foam type with the following environmental conditions in described Table 2. Table 3 contains the results of R^2 after performing all tests and calculations.

Table 2. Combinations of climate conditions

Versions	Temperature [°C]	Relative Humidity [%]
1	23	50
2	30	65
3	30	90
4	40	65
5	40	90

Table 3. Function of R^2 values on varied temperature and relative humidity combinations

Version	Temperature [°C] / Relative Humidity [%]	R^2 values in connection with °C and % RH		
		EPS	PE	TPS
1	23/50	0,9536	0,9670	0,9360
2	30/65	0,9737	0,9771	0,8968
3	30/90	0,9345	0,9652	0,8882
4	40/65	0,9370	0,9663	0,7994
5	40/90	0,9397	0,9890	0,7812

More researchers have also written about that the common plastic cushions haven't showed major change for the temperature and humidity combinations [12] [13]. In case of TPS, which is a humidity sensitive material, the R^2 values showed critical decrease. It means that the function fits less and the environmental factors affect significant the cushion characteristic. Based on the previously written, the calculations were accomplished, which is summarized in the Table 4.

Table 4. The cumulated values regarding α and environmental conditions

23°C/50%r.H

	S_i	a_S	Q_i	a_Q	E_i	a_E	L_i	a_L	M_i	a_M	P	a_P	R^2	α (m=1)	P^α	Cumulated values	Δ values (%)
PE	4	35	8	20	5	30	9	30	5	25	10	60	0,967	-0,071	0,849	43067,86	-0,18
EPS	2	35	7	20	7	30	10	30	7	25	10	60	0,9536	-0,095	0,803	43143,62	0,00
TPS	9	35	8	20	1	30	10	30	5	25	10	60	0,936	-0,132	0,737	41139,42	-4,65

40°C/65%r.H

	S_i	a_S	Q_i	a_Q	E_i	a_E	L_i	a_L	M_i	a_M	P	a_P	R^2	α (m=1)	P^α	cumulated value	Δ values (%)
PE	4	35	8	20	5	30	9	30	5	25	10	60	0,9652	-0,022	0,95	48182,09	0,00
EPS	2	35	7	20	7	30	10	30	7	25	10	60	0,9345	-0,136	0,732	39299,17	-18,44
TPS	9	35	8	20	1	30	10	30	5	25	10	60	0,8882	-0,238	0,579	32282,09	-33,00

40 °C/90%r.H

	S_i	a_S	Q_i	a_Q	E_i	a_E	L_i	a_L	M_i	a_M	P	a_P	R^2	α (m=1)	P^α	cumulated value	Δ values (%)
PE	4	35	8	20	5	30	9	30	5	25	10	60	0,989	-0,022	0,95	48182,09	0,00
EPS	2	35	7	20	7	30	10	30	7	25	10	60	0,9397	-0,124	0,751	40318,55	-16,32
TPS	9	35	8	20	1	30	10	30	5	25	10	60	0,7812	-0,499	0,317	17691,00	-63,28

It is clearly seen from the last column of the Table 4 that the difference between humidity sensitive material and the non-sensitive materials is getting more and more. From this verification process, the following consequences can be drawn:

- The environmentally friendly TPS foam can be a good alternative for replacing the common plastic foams, if the logistic flow happened on controlled normal climate (cc: 23°C/50% RH)

- If the climate parameters show high relative humidity during the transportation, the sensitive materials are even less suitable, namely when the mechanical characteristic of a cushion gets high priority in the PPS.
- When the final aggregated value of a sensitive material is very low comparing the others, the required packaging protection should be solved with the modification of cushioning characteristic (thickness, etc.) or other packaging technology solutions, like moisture resist outer packaging, etc. Naturally, these modifications degrade the other category values like Logistics (L_i), Supply (S_i), etc.

5. Conclusion

For that cases, when the environmental conditions (temperature / humidity combination) show extremely varies or the possible applicable material is humidity sensitive, like paper, paperboard, TPS, etc., or the cushioning characteristic of the packaging system is important, our extended and modified model can be applied. This extended function takes into account the cushioning behaviour of material as well, which is based on cushion curve tests and Stress-Energy calculations on different environmental conditions.

The verifications of both models showed that these functions can be applicable in practice and can give helping hand for the packaging engineer, to accelerate the packaging development and designing processes.

References

[1] Dowlatshahi S: A modeling approach to logistics in concurrent engineering. European Journal of Operational Research, Vol. 115, No. 1, 59-76., 1999. DOI: 10.1080/10170660609508994

[2] Bucci DZ; Forcellini FA: Sustainable packaging design model, Complex Systems Concurrent Engineering. Springer London, p. 363-370., 2007.

[3] Barmklev C, Bjärnemo R, Jönson J, Johnsson M: Towards an Integrated Design of Product and Packaging. International Conference On Engineering Design, ICED 05, August 15-18, Melborne, 2005.

[4] Pánczél Z: The significance of logistic package system design. Acta Technica Jaurinensis, Vol. 1, No. 2, pp. 247-257, 2008

[5] Rouillard V, Sek MA, Perry T: Analysis and simulation of road profiles. Journal of transportation engineering, Vol. 122, No. 3, pp. 241-245, 1996. DOI: 10.1061/(ASCE)0733-947X(1996)122:3(241)

[6] Singh SP, Chonhenchob V, Burgess G: Comparison of Various Loose Fill Cushioning Materials Based on Protective and Environmental Performance, Packaging Technology and Science, Vol. 7, pp. 229–241, 1994. DOI: 10.1002/pts.2770070504

[7] Sek M, Kirkpatrick J: Characteristics of corrugated fibreboard as a cushioning material in protective packaging, 10th IAPRI World Conference on Packaging, pp. 257-266, Melbourne, 1997.

[8] Sek MA, Minett M, Rouillard V, Bruscella B: A new method for the determination of cushion curves. Packaging Technology and Science, Vol. 13, No. 6, pp. 249-255, 2000.
DOI: 10.1002/pts.517

[9] Mojzes Á, Földesi P, Böröcz P: Define cushion curves for environmental friendly packaging foam. International Journal of Engineering Vol. 10, No. 1, pp. 113-118, Hunedoara, Romania, 2012.

[10] Mojzes Á, Böröcz P: Predicting Cushion Characteristic on New Type of Environmental Friendly Foam. Acta Technica Jaurinensis Vol. 3, No. 3, pp. 395-404, 2010.

[11] Burgess G: Generation of Cushion Curves from One Shock Pulse. Packaging Technology and Science, Vol. 7, No. 4, pp. 169-174, 1994.
DOI: 10.1002/pts.2770070403

[12] Mills NJ, Gilcrist A: Creep and recovery of polyolefin foams – deformation mechanisms. Journal of cellular plastics, Vol. 33, No. 3, pp. 264-292, 1997.