

VERIFICATION OF THE RISK ASSESSMENT MODEL THROUGH AN EXPERT JUDGMENT

RENÁTA TURISOVÁ, JOZEF MIHOK, JAROSLAVA KÁDÁROVÁ

1 INTRODUCTION

Model uncertainty is sometimes described as uncertainty about the truth of the model. However, since all models are false, this definition does not seem very useful. Still, some false models are more useful than other false models. A model with a poor scientific basis can still give reasonable predictions. Indeed, for consequence analysis, the quality of a model (which means in this paper the predictive quality of the model) is the only thing that is important. One way to give model uncertainty a meaning is to view it as a special case of parameter uncertainty, by introducing a new discrete parameter indicating which model is being used. It should be stressed that the interpretation of the model probability is not as the probability that the model is correct. Since the probabilities must sum to 1, this would mean we are assuming exactly one model to be actually correct. However, no model is exact, and if we allow models to be approximately correct then more than one model may satisfy this criterion.

There are plenty of methods for an assessment of expert judgments. Older methods like the Delphi method or the Nominal group techniques work with the point expert estimates of unknown quantities. Cooke (1991) described a method based on the assessment of expert efficiency (ability to make a successful estimation) based on the variability of their assessment from the actual value obtained post-hoc, i.e. after the occurrence of the assessed phenomenon.

These methods, which are based on efficiency weights, are increasingly applied in practice. The experience has shown their better accuracy with respect to classical methods of expert assessment (Goossens, 1998).

The main goal of these methods is to make a foundation for reaching a rational consensus. In the presented article, we will show an example of actual usage of the given method for the verification of a probabilistic model for the assessment of adequacy of a fire prevention assistance service in a large metallurgical complex.

The underlying principle of Cook's method of weighing based on efficiency consists in the fact that the weights used in the combination of distributions of expert judgments are selected by the so-called *expert efficiency*. It is a numerical assessment of their ability to answer the so-called calibration questions, i.e. the answers to the questions that are known only to the assessors, not to the experts.

The inputs for determination of efficiency weights are quantile estimates of experts on requested variability, whereas both the variability of unknown variables and the calibration variability are assessed. Calibration variability is variability of deviation of estimates from the actual values of the variable, which are known to the assessor (post hoc). The expert estimates are weighted based on their calibration ability and the informativeness of their estimation. Consequently, these values meet the given conditions with an asymptotic strictness. That means that an expert reaches a maximal expected weight in a longer period of assessment, if the estimates long-lastingly correspond to the actual values. The result of evaluation by such system of weighing is subsequently processed by the examiner. The acquired estimation is weighted with respect to calibration and informativeness of the estimation. The examiner determines the so-called inherent range, i.e. the lower and upper bound that is usable for a good approximation of the distribution of an analyzed quantity (Tkáč, 2000).

2 CALIBRATION AND INFORMATIVENESS

The quality of an expert's calibration can be measured based on the differences between the empirical distribution of calibration variable and the distribution determined by the expert; thus the calibration is a probabilistic characteristic of statistical hypotheses tests that are defined for each expert. Realizations can be understood as independent samples from a distribution corresponding to the quantiles estimated by an expert. The assessor prioritizes those experts, whose statistical hypotheses correspond to the data acquired from an empirical estimation of the distribution of calibration variable.

Let's assume that we observe a set of N calibration variables, such as s_1N realizations are from the interval 0-5%, s_2N realizations are from the interval 5-50% etc. Then the empirical density has a form (s_1, \dots, s_4) , and we want to measure its proximity to the hypothetical density $(p_1, \dots, p_4) = (0.05, 0.45, 0.45, 0.05)$. The way how to measure this proximity is offered by the so-called **relative information** with respect to p given by the formula:

$$I(s; p) = \sum_1^4 s_i \log \left(\frac{s_i}{p_i} \right).$$

It is a non-negative value that reaches its minimum, i.e. 0 if $s = p$. A good expert should have his empirical density (s_1, \dots, s_4) close to (p_1, \dots, p_4) and his relative information should be close to 0. It is a well known fact that, for large N , the distribution of relative information (with the size of $2N$) is well approximated by a χ -square distribution with three degrees of freedom

$$P(2N I(s, p) \leq x) \approx \chi_3^2(x),$$

where χ_3^2 is a distribution function of a χ -square distribution with three degrees of freedom.

The calibration of an expert e is defined as the probability of giving (acquiring) worse information (greater or equal) than the actually acquired information providing that the expert distribution is (p_1, \dots, p_4) . Thus,

$$C(e) = 1 - \chi_3^2(2N I(s, p)),$$

the empirical density s equal to the hypothetical density p gives us the best possible calibration, which is equal to 1. Informativeness is assessed considering each variable and each expert by the calculation of relative information of an expert's density for this variable with respect to the primary measurement.

Inherent range is acquired by adding $k\%$, i.e. by increasing the smallest interval containing all quantiles and realizations. k is generally determined by the assessor (the most common value is $k = 10\%$) Densities of distribution are connected with the assessments of each expert for every requested variable as follows:

- densities correspond to the expert quantile estimates,
- densities are minimally informative with respect to the basis of measurement given by the quantile boundaries.

If the primary measurement is uniform, it means that an expert interpolating distribution with respect to the inquired question is uniform between 0-5% and 5-50%, etc. Relative information of an expert e for a given requested variable is

$$I(e) = \sum_{i=1}^4 p_i \log \left(\frac{p_i}{r_i} \right),$$

where $p = (0.05; 0.45; 0.45; 0.05)$ is the expert probability and values r_i are the primary measurements of corresponding intervals. The general informativeness of each expert is a mean of all the information over all the variables. This mean is proportional with respect to the relative information with expert continuous distribution over all the variables considering the fact that these variables are independent.

3 DETERMINATION OF WEIGHTS

For determination of a weight that is based on the efficiency of each individual expert, the information about his informativeness and calibration will be used. When enumerating the above-mentioned weights, the examiner will set a definite basic success level α . Each expert, whose calibration will be lower than the α level, will automatically be assigned the weight of 0. Weighing rule R for determining an unknown variable that reaches values 1, ..., n is a function in a form of $R(p, i)$ for a probabilistic prediction p during the realization of i . The expected value for the subjective probability p , when an expert believes that the

actual value has a distribution of q , is $E_q R(p/i) = \sum_1^n q_i R(p,i)$. We say that the evaluation rule is suitable if, for every p and q , there is one maximized $E_q R(p/i)$ and $q = p$. That means that if there is used a suitable evaluation rule, an expert minimizes his weight by determining a probability that he believes is right.

An example of such a suitable evaluation rule is $R(q,i) = \log q_i$. Then the expected value assigned to the subjective probability p is $\sum_i p_i \log(q_i)$, which is

known as the relative information. In the model, we will use more than one calibrating quantity. Thus, the generalization of an idea of a suitable evaluation rule is used in a way that gives us an assessment based on a group of estimations and realizations. Supposing that an expert believes that a set M of unknown values X_1, \dots, X_m reaches values $1, \dots, n$ and has a Q distribution. Expected relative frequency of the result i is

$$q_i = E \frac{\#\{X_j = i\}}{m} = \frac{1}{m} E \left(\sum_j 1_{X_j=i} \right) = \frac{1}{m} \sum_j Q(X_j = i).$$

Supposing that we have the evaluation rule $R(p, M, s)$. If the expert determines the expected relative results with frequency p in the set of M variables, whereas the observational relative output frequency is s , then the result expected by the expert is:

$$E_Q(R(p, M, s)).$$

4 THE PROBABILISTIC MODEL ASSISTANCE ADEQUACY ASSESSMENT

One of the projects of Safety Improvement – SI using PRA (Probabilistic Risk Analysis), which was realized in practice, was the project focused on the adequacy of fire assistance in a large metallurgical company. In the company, there is a need to execute miscellaneous activities in various places that create an increased risk of fire, explosion or pollution (hereafter dangerous activities). This risk is multiplied if the above-mentioned activities are executed in the environment with a high level of fire danger.

Based on the valid legislation and in terms of company's regulations, in such cases the so-called *fire assistance* is executed. It comprises a group of experts in the field of fire protection. The assistance consists of the preventive part, i.e. inspection of the environment, prohibition of entry for unauthorized personnel, permanent supervision of activities, potential prohibition of an activity during increased endangerment, safety supervision over the object after the execution of works, etc. and the repressive part, i.e. immediate action during fire, prevention of spreading fire, suppression activities, immediate call a firefighter unit, coordination of rescue operations, initiation of evacuation, etc. With respect to

the nature of executed activities, it is necessary to determine the range, staff and equipment of the assistance unit with the necessary technology.

Managing and supplying the fire assistance in the company was assigned to employees with a university degree. The individual units of the company who intend to carry out activities with an increased danger (based on their subjective risk assessment) ask for staffing the assistance with members of firefighter unit (so-called professional assistance) or they execute the assistance by means of their own fire patrol.

In their request, they determine the needed staff, range, and equipment of the assistance unit (with respect to their own evaluation). The authorized employee of the firefighter unit will determine, based on the given request or after the consultation and/or environment inspection, the suitable range, staff, and equipment of the unit. Consequently, the unit will execute the assistance in the time needed.

A critical factor of such system of organization and supervision of the assistance is the evaluation of the risk of potentially dangerous activities by the coordinating units. To avoid under or over-estimation of the risk of an unwanted activity, there was elaborated a probabilistic model for the risk estimation during dangerous activities. For an adequacy judgment of the model, the method of expert assessment was used.

The company's request to make the model simple and lucid brought two technical restrictions for the model:

- Based on the long-term experience with the usage of the Failure Mode and Effects Critical Analysis – FMEA), there was a request for a numeric range of the risk extent from 1 to 1000.
- The interpolating table of the meaning of the estimated risks and the way of execution of the fire assistance was appointed in advance. It is denoted as follows (Table 1).

Table 1 – Interpreting Table of the Meaning of the Risks Estimated by the Model

Risk estimated by the model	Way of execution of the fire assistance
0 – 100	Without fire assistance
101 – 300	Fire assistance - FA
301 – 500	FA, more members with precautions, i.e. more mobile fire extinguishers + hydrant
501 – 750	FA + professional firefighter unit (PFU) + ability to provide the firefighter technology based on the decision of the chief of the firefighter unit
751 – 1000	FA + PFU + firefighter technology + ability to render special actions based on the decision of the chief of the firefighter unit

From Table 1 results the necessity of professional assistance in case the risk estimated by the model has a value greater than 500. The procedure of the model creation exceeds the scope of this article and is described in depth in (Turisová, 2004).

Formally, we can represent the examined model of the risk calculation by this formula:

$$R = X \cdot L \cdot C \cdot (M + D) \cdot (k \cdot (Z + N + H) + O) + \varepsilon.$$

The meaning of the individual variables and their relevant values is described in Table 2. Based on computer aided simulation the k value with the required precision was determined by a group of experts. The verification of adequacy of the model resulted from the application of the expert assessment method.

Table 2 – Meaning and Importance of Individual Variables in the Assistance Adequacy Model

Label	Meaning	
L	Inflammable or flame supporting gases – under pressure	L1=10
	Inflammable or flame supporting gases	L2=9
	Critical technological appliances containing inflammable liquids of all fire danger rating levels	L3=9
	Explosive powder	L4=9
	Inflammable liquids I. and II. fire danger class – under pressure	L5=8
	Inflammable liquids I. and II. fire danger class	L6=7
	Inflammable liquids III. and IV. fire danger class – under pressure	L7=6
	Inflammable liquids III. and IV. fire danger class	L8=5
	Solids – high inflammation	L9=4
	Solids – medium inflammation	L10=3
	Solids – low inflammation	L11=2
	Almost fireproof	L12=1
C	Burning	C1=5
	Welding	C2=4
	Grinding	C3=3
	Gluing	C4=2
	Others	C5=1
M	Place with an increased fire danger	M1=10
	Under the ground (under the ground level)	M2=9
	In cable channels	M3=8
	In heights	M4=7
	Pipe bridges, conveyor bridges	M5=6
	Above the ground level in an enclosed area	M6=5
	In an open space	M7=4
D	In the storage area	D1=10
	In the piping	D2=8
	In the technological appliances	D3=6
	In others	D4=4
H	Negligible	H1=1
	Up to 5 million Sk (165,970 €)	H2=4

Label	Meaning	
	Up to 10 million Sk (331,939 €)	H3=6
	Up to 20 million Sk (663,878 €)	H4=8
	Over 20 million Sk (663,878 €)	H5=10
N	Immediately	N1=2
	Up to one week	N2=4
	Up to one month	N3=6
	Up to one year	N4=8
	Irreparable	N5=10
Z	High	Z1=100
	Medium	Z2=80
	Low	Z3=60
	Other appliances	Z4=10
O	None	O1=2
	1 person	O2=6
	Up to 5 people	O3=8
	Over 5 people	O4=10
T	None	T1=1
	Fire	T2=2
	Explosion	T3=3
	Pollution	T4=4

Legend: *L* – Substances occurring in the place of assistance, *C* – Activity with an increased danger of fire, *M* – Place of assistance execution, *D* – Type of workplace storage of the dangerous substance, *H* – Direct primary damages, *N* – Indirect primary damages - repairable, *Z* – Secondary damages on technology (domino effect), *O* – Endangerment of a human life, *T* - Type of endangerment.

5 THE PROCEDURE OF A PRACTICAL VERIFICATION OF THE MODEL ADEQUACY

- In cooperation with the purchaser of the assistance there was elaborated a list of dangerous works (activities). It includes both the activities that were executed in the past and activities that might potentially occur in the future. Various locations (types of works) were taken into consideration, so that they created a sample set of possible assistance orders, i.e. they could cover the potential range and size of endangerment.
- 120 activities were processed.
- From empirical data, which were acquired from real assistances, 30 of them were selected as calibrating variables and were included into the database. Calibrated, realized and empirically verified assistances were selected on the basis of relevant activities, so that they could cover the whole range of possible risk assessment.
- There was created a group of 7 experts from various relevant fields (representatives of customers, creators of the model and other fire specialists).
- The task of the group of experts was to evaluate the risk (the whole database of 150 activities) based on the model of assistance adequacy, but from the

client’s point of view. Thus, it was not a strict determination of the risk, but estimation - how can a trained amateur (representative of the purchaser) proceed with the help of the risk evaluation model.

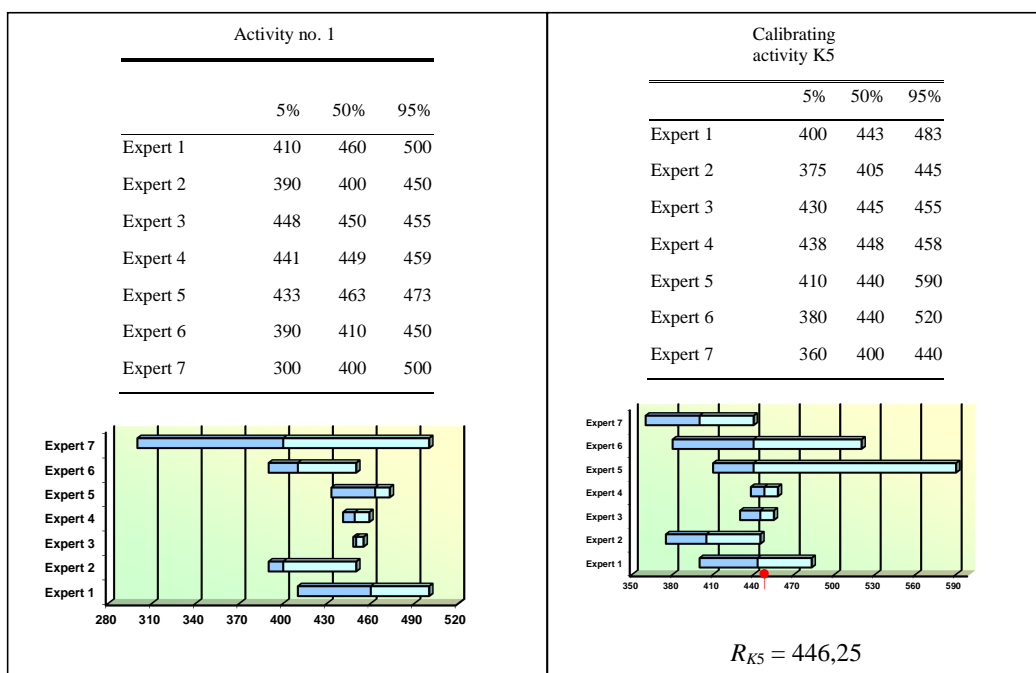
Experts carried out an interval estimation of the risk for every activity from the database, hence for calibrating activities, taking into consideration the following percentiles 5%, 50% a 95% (Table 3):

$q_{5\%}$ - risk estimation – lower bound of the estimation (lower estimation of the risk from the customer’s side is not very probable - max 5%),

$q_{50\%}$ - risk estimation – middle estimation that is evidently the most common one, it is the value that will the most frequently represent actually calculated value of the risk by the customer,

$q_{95\%}$ - risk estimation – upper bound of the estimation (higher estimation of the risk from the customer’s side is not very probable - max 5%).

Table 3 – Sample of Interval Estimations of the Risk Determined by Individual Experts for Activity no. 1 and the Calibrating Activity K5 (Value R_{K5} represents the actually known resulting value)



- The assessment of all activities in the database was executed once again independently by the team of the model’s creators, who made a point estimation R_i for every activity i from the database. Based on this model, all activities were stratified into four categories:

- Category I.: Activity i with estimate R_i from 0 to 200.
- Category II.: Activity i with estimate R_i from 201 to 400.
- Category III.: Activity i with estimate R_i from 401 to 600.
- Category IV.: Activity i with estimate R_i from 601 to 1000.
- the informativeness and calibration of each expert was calculated for every category (Table 4). For every activity, there was determined the empirical density of the distribution of the risk value by the expert assessment based on presented weights.

Table 4 – Weights of Experts for Individual Activities

Expert	Calibration $C(e)$	Informativeness $I(S_p)$	Order	Weights $w(e)$
1	0.1050	0.01928	3	0.2
2	0.0010	0.07949	-	0
3	0.4025	0.29030	1	0.2
4	0.2134	0.32010	2	0.2
5	0.0910	0.40010	4	0.2
6	0.0120	0.12400	-	0
7	0.0050	0.29710	-	0
Combined expert	0.0877	0.02132	5	0.2

Calibration and informativeness was determined for each expert. In order to assign weights, the marginal success value α was chosen. For each selection of the marginal value the weights were changed (because for higher values of α , more experts are excluded, and the weights are more focused on the remaining experts). Similarly, a *combined expert*, which was created as a combination of other experts, is dependent upon α . For the combined expert, we can also enumerate 5%, 50% and 95% quantiles and eventually calibration and informativeness.

In the model, we made a selection α with respect to the weight of this expert in a way that the combined expert was the last one, who met the criteria for getting into the group of experts.

- From the empirical distribution function of expert assessment, there was determined a mean M_i for every activity i .
- For every activity i the *adequacy index* was calculated:

$$V_i = M_i - R_i$$

For every category of activities, there was created a histogram of adequacy indexes, mean and sample standard deviation (Figure 1).

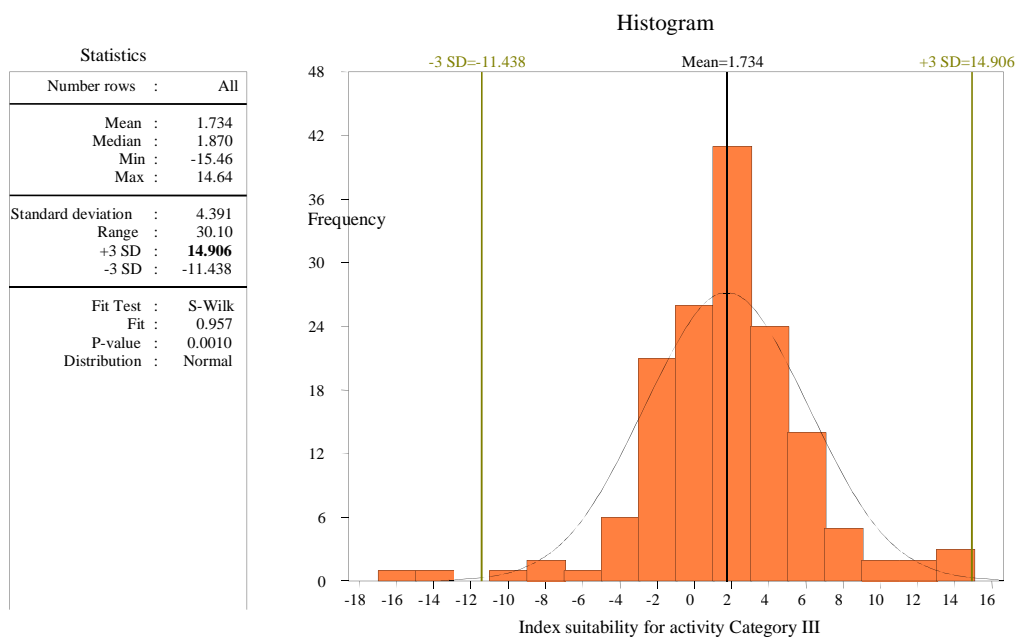


Figure 1 – The histogram of the estimation error for Category III activities (Estimation error $\varepsilon = 14.906$)

- Based on the mean of adequacy indexes extended by the \pm triple of the sample standard deviation, there was determined the so-called estimation error interval for every category of activities.
- The expert team decided (wanting rather to provide any assistance than not provide the appropriate professional assistance) to consider the estimation error ε to be the absolute value of the upper bound of the interval of an error estimated for Category III (Figure 1). (It is a logical decision, because Category III includes the “marginal” activities, when it is necessary to make a decision about the potential professional assistance of the firefighter unit.)
- The resulting model formula for the risk calculation has a form:

$$R = X \cdot L \cdot C \cdot (M + D) \cdot (k \cdot (Z + N + H) + O) + \varepsilon,$$

where $X = 0.0394$, $k = 0.125$, and $\varepsilon = 14.906$.

The meaning of other variables is determined in Table 2, whereas parameter X was adjusted, so that we can meet the upper bound of the range of possible risk $R = 1000$. Such an adjusted model was once again empirically verified on a database sample of actually executed assistances.

6 CONCLUSION

In company practice we often encounter the problem of qualified estimation of some important characteristics necessary for the process of decision making of the top management. A typical example is risk evaluation, which is based, in addition to other factors, on the probability of occurrence of an unwanted phenomenon. The estimation of the probability of occurrence is commonly a reason for a big faultiness of the above-mentioned estimates. It is relatively difficult to estimate an occurrence of any given phenomenon, which is very improbable, i.e. the occurrence rate is a very small number, especially if the assessor has no experience with the assessing of the given phenomenon. On the other hand, if we want to make a rational decision resulting from quantitative characteristics, the precision of above-mentioned estimation is a very important factor of a good, rational decision. Implementation of the probabilistic model as an aid for managers for selection of a suitable type of fire assistance is an example of a procedure, by which it is possible to decrease a risk of error occurrence in the risk evaluation systematically. The adequacy of the theoretically determined model was verified by the method of expert assessment, based on which were appointed the balancing variables, so that the model could meet all the required criteria and at the same time remain simple, easily interpretable and trustworthy. The methodology of the expert assessment that was used for the verification of the given model has proven to be usable. Moreover, it seems that in other areas of managerial decision making it can be understood as one of the fundamental effective methods for reaching a consensus.

AKNOWLEDGEMENT

This contribution is the result of the project implementation: Centre for research of control of technical, environmental and human risks for permanent development of production and products in mechanical engineering (ITMS: 26220120060) supported by the Research & Development Operational Program funded by the ERDF and project VEGA No. 1/0102/11 - Methods and techniques of experimental modelling of in-house manufacturing and non-production processes.

REFERENCES

- COOKE, R.M. (1991), *Experts in Uncertainty*. Oxford University Press, 1991.
- GOOSSENS, L. H. J., et al. (1998), "Evaluation of weighting schemes for expert judgment studies", in *Probabilistic Safety Assessment and Management*, Springer, London.

TKÁČ, M. (2000), “Matematická štatistika a teória pravdepodobnosti,” in *Projekt Tempus - Phare IB_JEP-13406-98*, SjF, Technická univerzita v Košiciach, ISBN 80-7099-536-X.

TURISOVÁ, R. (2004), “*Ekonomické aspekty odhadu rizík v integrovaných systémoch riadenia podnikov*,” PhD. Thesis, Technical University of Kosice, Slovak Republic.

ABOUT THE AUTHORS

Ing. **Renáta Turisová**, PhD., Technical University of Košice, Faculty of Mechanical Engineering, Department of Industrial Engineering and Management, Němcovej 32, 04200 Košice, Slovak Republic
e-mail: renata.turisova@tuke.sk

Dr.h.c. mult. prof. Ing. **Jozef Mihok**, PhD., President of the Office for Standards, Metrology and Testing Slovak Republic, Štefanovičova 3, P. O. Box 76, 810 05 Bratislava 15, Slovak Republic
e-mail: predseda@normoff.gov.sk

doc. Ing. **Jaroslava Kádárová**, PhD., Technical University of Košice, Faculty of Mechanical Engineering, Department of Industrial Engineering and Management, Němcovej 32, 04200 Košice, Slovak Republic
e-mail: jaroslava.kadarova@tuke.sk