Adv Syst Sci Appl 2020; **04**:60–69
Published online at https://ijassa.ipu.ru.

Multi-Criteria Estimation of Input Parameters in Natural Gas Quality Analysis

Ivan Brokarev[1]*, Sergei Vaskovskii[2]

[1] *National University of Oil and Gas Gubkin University, Moscow, Russia. E-mail: brokarev.i@gubkin.ru*

[2] *V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia. E-mail: v63v@yandex.ru*

Abstract: In this paper, we present a method for assessing input parameters in a statistical model used for the quality analysis of the natural gas. The analysis is done by measuring physical quantities of natural gas through the hierarchy analysis, compromise programming, correlation analysis and assessing the practical possibility of measuring selected input physical quantities by available means. The problems arising when selecting input parameters are also considered. The results are compared with the previously obtained results of the model input parameters correlation analysis in order to improve the developed system for natural gas quality analysis. The proposed method is applied to assess input parameters for certain samples of natural gas based mixtures. Based on the results of the multicriteria assessment, the speed of sound, thermal conductivity and concentration of carbon dioxide were selected as input parameters for the developed natural gas quality analysis system.

Keywords: multi-criteria estimation, analytic hierarchy process, correlation analysis, natural gas quality analysis, neural network analysis

1. INTRODUCTION

To analyze the quality of natural gas is to determine the energy characteristics out of the measured physical parameters of the gas. At present, this problem is usually solved in the gas industry by traditional methods, but there are alternative approaches [1]. The essence of alternative methods and systems based on them is to determine the required indicators by a statistical model, where the measured physical parameters of natural gas are input parameters. These models, mostly neural networks, seem to be the most effective means to solve the problem in comparison with other predictive models [2]. However, if we use artificial intelligence methods, we encounter a number of problems that need to be solved for the developed method to succeed in analyzing the quality of natural gas. The main problems of the approach are high computational costs, a large number of input physical parameters to measure, no general algorithm to choose the architecture and parameters of the model, and the model's decrease in accuracy, when the parameters deviate from standard values. Some of the listed problems are associated with suboptimal choice of input parameters for the models. For instance, if the number of input variables is excessive, the time spent on the analysis procedure increases, up to the point when it is impossible to analyze in real time. In addition, some of the input parameters are uninformative, so the model becomes less accurate.

*Corresponding author: brokarev.i@gubkin.ru

In practice, we choose the input parameters for a statistical model by heuristic methods and a complete enumeration of the available parameters. However, there are algorithms aimed to optimize the set of input parameters and to choose the parameters more efficient.

This article discusses our methodology to select input parameters for a natural gas quality analysis system; the methodology is based on a combination of correlation analysis and a multi-criteria estimation method.

2. METHODOLOGY FOR MULTI-CRITERIA ESTIMATION OF INPUT PARAMETERS

The authors of [2, 3] chose the measured physical parameters out of a relationship between the parameter under consideration and the component composition of the gas. In addition, the authors took into account if it is possible to measure this parameter with commercially available and relatively inexpensive measuring instruments. To determine the relationship between the parameters and the component composition of natural gas, the authors conducted a correlation analysis between the physical parameters and simulated samples of natural gas of various component composition. Based on the calculated Pearson correlation coefficients for the simulated sample, the authors figured out if they chose the input parameters correctly. If the parameter had a high correlation (the value of the correlation coefficient in the range 0.7 - 1 modulo) with certain gas components and at the same time low correlation (0 - 0.3 modulo) with other input parameters, then this parameter was added to a set of model input parameters.

We propose to marry the described approach with multi-criteria estimation methods. The first step of the proposed methodology is to select criteria to estimate the input parameters for the model. The selected criteria are divided into three groups, the first of which is the availability of technology to measure the parameter (0 if absent, 1 if available), high correlation with the model output parameters (component composition), and low correlation with other input physical parameters of the gas. The second group includes three criteria: correlations with the main components of the equivalent pseudogas - methane (CH_4), propane (C_3H_8), and nitrogen (N_2).

The above criteria are not equivalent. The most important criterion is the availability of technology to measure the parameter. The second is the criterion for high correlation with the model output parameters. Correlation criteria are expressed as Pearson correlation coefficients r. The least important criterion is the one of low correlation between the input parameters; this is necessary mainly to eliminate the possible multicollinearity of the input parameters.

Based on the results of the previous studies and analysis [4-8], we choose the following physical parameters as the input parameters: speed of sound (c), thermal conductivity (χ), concentration of carbon dioxide (CO_2), dynamic viscosity (η), and dielectric permittivity (ε).

To form an optimal set of input physical parameters of natural gas, we propose to use the analytic hierarchy process (AHP) [9]. This method allows you to find such an alternative, that is best consistent with how you understand the essence of the problem and the problem requirements. The main advantage of the analytic hierarchy process is high versatility - the method can be used to solve a wide variety of problems, regardless of the field of application. The disadvantage of the analytic hierarchy process is a large amount of information needed to be obtained from experts. However, we compensate for this disadvantage if we calculate the correlation coefficients and ask an expert to assess only an availability of technology to measure the investigated parameter.

It should be noted that we could represent our system hierarchically to describe how changing priorities at higher levels influences the priorities of lower level elements. Hierarchies provide more detailed information about the structure and function of the system at the lower levels. To satisfy the constraints on level elements, it is best to reproduce the constrains at the next higher level. Hierarchically structured natural systems are much more efficient. Hierarchies are stable in the sense that small changes have little effect, and flexible in the sense that if we add elements to a well-structured hierarchy, we do not destroy the hierarchy's characteristics.

The first stage of the AHP is to build a hierarchical structure that includes the goal, criteria, alternatives, and other relevant factors. The structure to choose the optimal set is shown in Fig. 1.

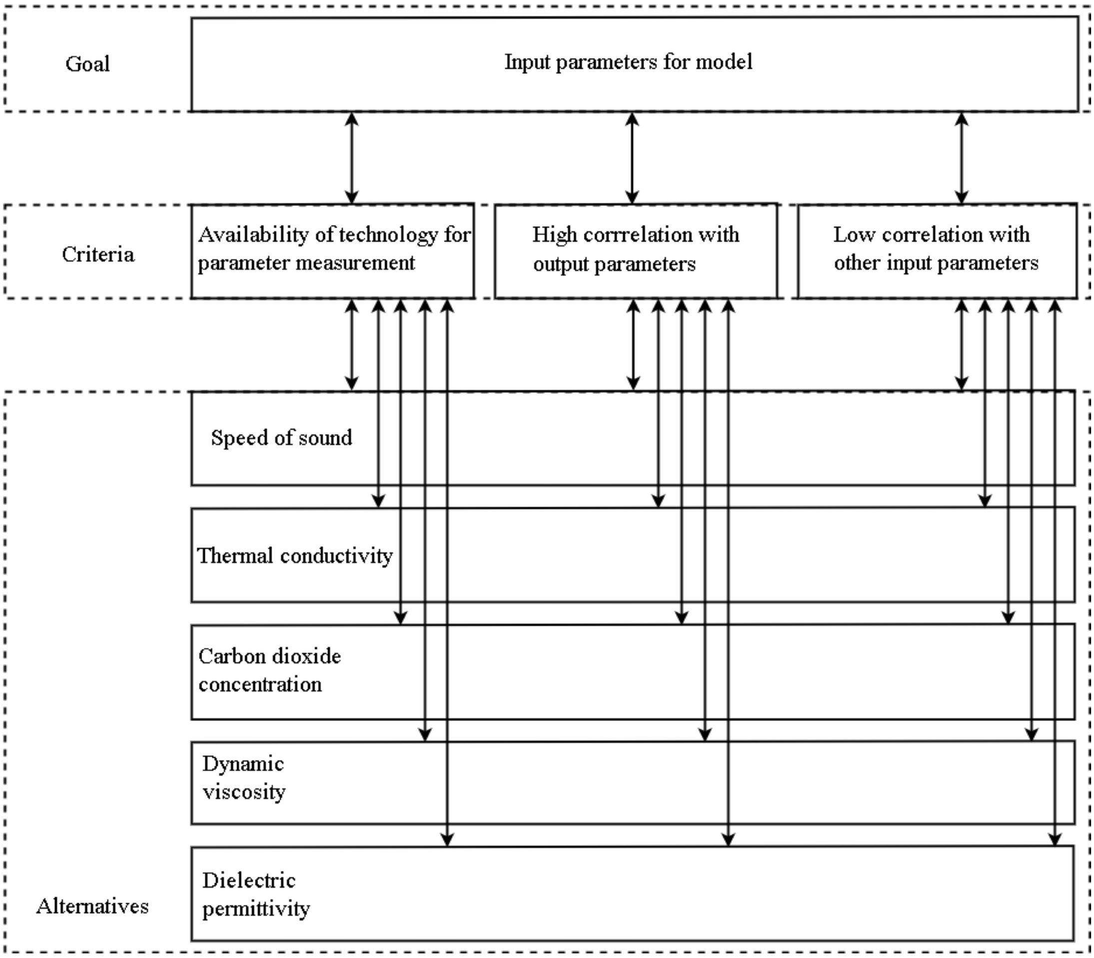

Fig. 2.1. The structure of how the optimal set of the model input parameters is formed as we analyze the hierarchies.

The next step is to build a matrix of pairwise criteria comparisons. We compare each criterion relative to all others. To conduct paired comparisons, T. Saaty [9] developed a scale of relative importance, presented in table 1.

Table 2.1. Scale of relative importance

Level of importance index	Level of importance	Essence of the level of importance
0	Incomparable	Difficult to compare criteria
1	Equal importance	Two criteria contribute equally
3	Moderate superiority	One criteria has moderate advantage over the other
5	Strong advantage	One criteria has strong advantage over the other
7	Substantial advantage	One criteria is so much superior over the other, that the advantage is substantial
9	Very strong advantage	One criteria is obviously superior over the other
2,4,6,8	Intermediate solutions between two importance criteria	Applied in edge cases where it is difficult to determine what criteria wins

According to the list of important criteria given in the introduction, we pairwise-compared the criteria on a qualitative scale and subsequently converted the criteria to points. The matrix of pairwise comparisons is compiled as we calculate the coefficients a_{ij} - the ratio of the weight of criterion i to criterion j. The matrix of pairwise comparisons of criteria is shown in Table 2.

Table 2.2. Matrix of pair-wise criteria comparisons

a_{ij}	Tekh.	r(CH_4)	r(C_3H_8)	r(N_2)	r(c)	r(χ)	r(CO_2)	r(η)	r(ε)
Tekh.	1	2	2	2	3	3	3	3	3
r(CH_4)	1/2	1	1	1	2	2	2	2	2
r(C_3H_8)	1/2	1	1	1	2	2	2	2	2
r(N_2)	1/2	1	1	1	2	2	2	2	2
r(c)	1/3	1/2	1/2	1/2	1	1	1	1	1
r(χ)	1/3	1/2	1/2	1/2	1	1	1	1	1
r(CO_2)	1/3	1/2	1/2	1/2	1	1	1	1	1
r(η)	1/3	1/2	1/2	1/2	1	1	1	1	1
r(ε)	1/3	1/2	1/2	1/2	1	1	1	1	1

Then the matrix of pairwise criteria comparisons is normalized. For this, the new element of the matrix A_{ij} is calculated as the ratio of the original element a_{ij} to the sum of the elements of the corresponding column. The average values for each row (Avg_i) are found as well; the value is called the criterion weighted column for the goal. The normalized matrix of pairwise comparisons of criteria is shown in Table 3.

Table 2.3. Normalized matrix of pair-wise criteria comparisons

A_{ij}	Tekh.	r(CH_4)	r(C_3H_8)	r(N_2)	r(c)	r(χ)	r(CO_2)	r(η)	r(ε)	Avg_i
Tekh.	0.240	0.267	0.267	0.267	0.214	0.214	0.214	0.214	0.214	0.235
r(CH_4)	0.120	0.133	0.133	0.133	0.143	0.143	0.143	0.143	0.143	0.136
r(C_3H_8)	0.120	0.133	0.133	0.133	0.143	0.143	0.143	0.143	0.143	0.136
r(N_2)	0.120	0.133	0.133	0.133	0.143	0.143	0.143	0.143	0.143	0.136
r(c)	0.080	0.067	0.067	0.067	0.071	0.071	0.071	0.071	0.071	0.070
r(χ)	0.080	0.067	0.067	0.067	0.071	0.071	0.071	0.071	0.071	0.070
r(CO_2)	0.080	0.067	0.067	0.067	0.071	0.071	0.071	0.071	0.071	0.070
r(η)	0.080	0.067	0.067	0.067	0.071	0.071	0.071	0.071	0.071	0.070
r(ε)	0.080	0.067	0.067	0.067	0.071	0.071	0.071	0.071	0.071	0.070

Ones we determined the weights of the criteria, we can proceed to the next step – to determine the weights of the alternatives. For this, it is proposed to use the method of compromise programming [10]. This is a multi-criteria optimization method, where a solution is determined to minimize the distance from the target point to a set of effective solutions. The first step in this method is to calculate the value of the criteria for each alternative. The value of the criterion for the presence of the technology to measure a parameter is equal to zero if the technology is absent and to unity if the technology is present. The rest of the criteria values were calculated as Pearson's correlation coefficients modulo for a sample of natural gas with the following concentration ranges for each component: 85 - 100% for methane, 0 - 5% for nitrogen, carbon dioxide and propane. Table 4 shows the values of the criteria for each alternative.

Table 2.4. Values of criteria for each alternative

O_{ki}	c	χ	CO_2	η	ε
Tekh.	1	1	1	0	0
r(CH_4)	0.899	0.848	0.628	0.785	0.117
r(C_3H_8)	0.477	0.557	0.053	0.339	0.906
r(N_2)	0.787	0.764	0.083	0.750	0.129
r(c)	1	0.992	0.233	0.629	0.451
r(χ)	0.992	1	0.141	0.535	0.554
r(CO_2)	0.233	0.141	1	0.541	0.445
r(η)	0.629	0.535	0.541	1	0.403
r(ε)	0.451	0.554	0.445	0.403	1

At the next step, the proximity degree values of the criteria are calculated. For this, the criteria are divided into two categories for minimum and maximum. In this problem, it is necessary both to maximize the availability of measurement technology and correlation with the output parameters, and to minimize the correlation between the input parameters. For the minimum criteria, the proximity degree of the each alternative according to the each criterion o_{ki} is calculated by the formula:

$$o_{ki} = \frac{\max(O_{ki}) - O_{ki}}{\max(O_{ki}) - \min(O_{ki})},\ i = 1...n \tag{2.1}$$

where O_{ki} is an estimated value of alternative i over criteria k, min and max are min and max operations, n is the number of alternatives.

For the max criteria, we compute the proximity degree of alternatives as:

$$o_{ki} = \frac{O_{ki} - \min(O_{ki})}{\max(O_{ki}) - \min(O_{ki})},\ i = 1...n \tag{2.2}$$

The computation results for the proximity degree of alternatives over the criteria are shown in Table 5.

Table 2.5. Proximity degree of alternatives over the criteria

o_{ki}	c	χ	CO_2	η	ε
Tekh.	1	1	1	0	0
r(CH_4)	1	0.935	0.653	0.854	0
r(C_3H_8)	0.497	0.591	0	0.335	1
r(N_2)	1	0.967	0	0.947	0.065
r(c)	0	0.010	1	0.484	0.716
r(χ)	0.009	0	1	0.541	0.519
r(CO_2)	0.893	1	0	0.534	0.646
r(η)	0.621	0.779	0.769	0	1
r(ε)	0.920	0.747	0.930	1	0

After we determined the proximity degree of alternatives, it is necessary to determine how to assess each alternative according to the criteria. We need to take into account the criteria' significance calculated above. For this, the set of criteria G for each pair of alternatives is divided into two subsets: G_1 is a subset of criteria, where the considered i-th alternative is superior to the j-th; G_2 is a subset of criteria, where the considered i-th alternative is inferior to the j-th one. For example, the first alternative (the speed of sound c) exceeds the fifth (dielectric permittivity ε) according to the first, second, fourth, fifth, sixth, and eighth criteria, but is inferior in all others. The general formula to calculate β_{ij} estimates is as follows:

$$\beta_{ij} = \frac{\sum\limits_{k \in G_1} Avg_k * (o_{ki} - o_{kj})}{\sum\limits_{k \in G_2} Avg_k * (o_{kj} - o_{ki})} \tag{2.3}$$

Here Avg_k are the values of criteria weights from Table 2; o_{ki} and o_{kj} -are the degrees of proximity of i and j alternatives over k-th criteria from Table 4.

Since β_{ii} = 1, the assessment matrix for the alternative is the one in Table 6.

Table 2.6. The assessment matrix for the alternatives

β_{ij}	c	χ	CO_2	η	ε
c	1	1.277	6.391	5.207	4.046
χ	0.783	1	5.667	4.230	4.391
CO_2	0.156	0.176	1	0.850	1.452
η	0.192	0.236	1.177	1	2.036
ε	0.247	0.228	0.688	0.491	1

The final step of methodology proposed is to calculate the eigenvector of the assessment matrix for the alternatives and subsequently normalize the matrix to determine the required vector of weights for the alternatives. The resulting vector looks like this: [0.412; 0.351; 0.074; 0.095; 0.068]. Based on the obtained vector, it can be concluded that the first (speed of sound) and second (thermal conductivity) alternatives are superior.

Next, it is necessary to evaluate the data obtained for consistency of the alternatives assessment matrix and to reassess the alternatives. For this, the following parameters were calculated: the maximum (main) eigenvalue λ_{max}, the concordance index (CI) and the concordance ratio (CR).

We calculate the maximum eigenvalue. Then we use the matrix for evaluating alternatives as follows: the sum of each column of the matrix is multiplied by the corresponding element of the weight vector of the alternatives, and then the resulting numbers are summed up. The closer the maximum eigenvalue is to the number of alternatives (the dimension of the alternative assessment matrix) n, the more consistent the result is. For the obtained assessment matrix for the alternatives and the vector of alternatives' weights λ_{max} is 5.106 for five alternatives.

The results are then compared with the randomly scored data for a scoring matrix of the same dimension. For this, the consistency index is calculated using the formula:

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{2.4}$$

The consistency index for the received data is 0.026.

The consistency index of a randomly generated inversely symmetric matrix with the corresponding reciprocal values of the elements is called the random index (RI), and the value of the random index increases as the order of the matrix increases. Table 7 shows the order of the matrix and the average values of the random index, determined for random samples [9].

Table 2.7. Dependence of the random index on the order of the matrix.

Matrix order	1	2	3	4	5	6	7	8	9	10
Average RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

For the analyzed matrix dimension, the random index is 1.12 according to Table 6.

The ratio of the consistency index to the mean random index for a matrix of the same order is called the consistency ratio.

$$CR = \frac{CI}{RI} \tag{2.5}$$

A value of this parameter that is less than or equal to 0.1 is considered acceptable. If the consistency ratio is 0, it means that the data is completely consistent. If the concordance ratio exceeds 0.1, then we need to revise the estimates and recalculate the concordance parameters. In the case study, this is not required, since the agreement ratio is 0.024. Based on the calculated characteristics, it follows that the selected estimates of alternatives are consistent.

3. CONFIRMATION OF THE MULTI-CRITERIA ESTIMATION RESULTS BY NEURAL NETWORK ANALYSIS

To confirm the results of the multi-criteria estimation of the input parameters for the neural network model, we performed a neural network analysis. For our experiment, we simulated a neural network model on the calculated data. The developed recurrent neural network [11] was trained, tested, and simulated with different sets of input parameters to determine the required component composition of gas mixtures. For five sets of input parameters, ten sets of parameters are possible, three in each. All possible sets of input parameters for this case are shown in Table 8.

Table 3.8. Possible sets of input parameters for the considered case

Set number	1	2	3	4	5	6	7	8	9	10
Input parameters	c, χ, CO_2	c, χ, η	c, χ, ε	c, CO_2, η	c, CO_2, ε	c, η, ε	χ, CO_2, η	χ, CO_2, ε	χ, η, ε	CO_2, η, ε

For each option from the set, the accuracy characteristics to determine the required parameters were calculated: maximum absolute deviation (MaxAE), mean absolute deviation (MAE), maximum relative deviation (MaxAPE), mean relative deviation (MAPE) according to the formulas presented below.

Table 3.9. Simulation results of the developed neural network model for various sets of input parameters

Component	Characteristic	Set number									
		1	2	3	4	5	6	7	8	9	10
Methane	MaxAE, %	0.425	0.696	0.751	0.516	0.567	0.832	0.524	0.535	0.556	0.987
	MAE, %	0.008	0.011	0.012	0.009	0.010	0.015	0.009	0.009	0.010	0.018
	MaxAPE, %	0.451	0.714	0.789	0.562	0.579	0.875	0.542	0.551	0.561	0.991
	MAPE, %	0.009	0.011	0.012	0.010	0.011	0.016	0.010	0.011	0.010	0.018
Nitrogen	MaxAE, %	0.285	0.482	0.621	0.378	0.415	0.781	0.394	0.419	0.402	0.871
	MAE, %	0.011	0.013	0.016	0.012	0.013	0.018	0.013	0.014	0.013	0.018
	MaxAPE, %	0.308	0.482	0.568	0.376	0.421	0.790	0.382	0.407	0.399	0.862
	MAPE, %	0.012	0.017	0.018	0.012	0.013	0.017	0.012	0.012	0.012	0.018
Propane	MaxAE, %	0.256	0.521	0.535	0.279	0.293	0.541	0.311	0.321	0.361	0.671
	MAE, %	0.007	0.012	0.012	0.007	0.007	0.012	0.008	0.008	0.009	0.014
	MaxAPE, %	0.235	0.539	0.546	0.291	0.318	0.567	0.309	0.311	0.354	0.649
	MAPE, %	0.006	0.011	0.011	0.007	0.007	0.012	0.007	0.007	0.008	0.013

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |Y_{output} - Y_{target}| \tag{3.6}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n} |\frac{Y_{output} - Y_{target}}{Y_{target}}| \tag{3.7}$$

$$MaxAE = \max |Y_{output} - Y_{target}| \tag{3.8}$$

$$MaxAPE = 100\% \max |\frac{Y_{output} - Y_{target}}{Y_{target}}| \tag{3.9}$$

Here Y_{output} – the value obtained from the statistical model, Y_{target} – the initial values of the target parameters, n – sample size, max – operator to compute maximum value.

The proposed calculation was performed in Matlab 2019b [12] with a plug-in for NIST REFPROP [13]. The data included 150000 gas mixtures with the ranges described above and the calculated physical input parameters under study. The simulation results of the developed neural network model are shown in Table 9. The accuracy of carbon dioxide characteristics are not shown, since the concentration of this component is an input parameter and is considered known.

From the results of the neural network analysis, it can be concluded that the best accuracy is in the first set; the set contains the alternatives prevailing from the multi-criteria estimation. At the same time, the worst accuracy is in the tenth set with the least weighty alternatives. The complete ranking of all sets of input parameters is shown in Table 10. This procedure was carried out using expert estimations to clarify the calculated accuracy characteristics. Based on the results obtained, it follows that the results of the multi-criteria estimation and neural network analysis coincide.

Table 3.10. Ranged sets of input parameters

Set number	1	2	3	4	5	6	7	8	9	10
Set rank	I	VII	VIII	II	V	IX	III	IV	VI	X

4. CONCLUSION

This paper investigates how to choose input parameters for the statistical model to analyze the quality of natural gas. It is shown that the proposed technique based on multi-criteria estimation methods allows us to choose the quantitative and qualitative values of the criteria for each input physical parameter. Also, the proposed methodology makes it possible to choose the parameters before training the statistical model. This significantly reduces the computational and time costs; this is extremely useful when we analyze hierarchies in real time.

Note, that our technique showed the speed of sound and thermal conductivity to dominate over the rest of the input parameters. These results partially coincide with previous studies, where only the correlation analysis of input parameters was used [2, 3].

It was also found that the concentration of carbon dioxide, inferior to most of the considered alternatives, might not be used in further studies as an input parameter of the statistical model, although it is needed to reduce the amount of unknown concentrations of the components of the analyzed gas.

The results of our multi-criteria estimation of the input parameters are confirmed by the results of our neural network analysis carried out on various sets of input parameters.

To investigate this area further, we need to apply our system to analyze the quality of natural gas [14] in laboratory and industrial conditions. We plan to use a recurrent neural network model as the main model to determine the required quality characteristics of natural gas by analogy with the neural network model of a multilayer perceptron, which was used earlier [15].

ACKNOWLEDGEMENTS

The authors would like to thank the international cooperation projects between BRISK II TA and Erasmus + 2017-1-SE01-KA107-034292 Staff Mobility for the opportunity to conduct this study.

REFERENCES

1. Dorr H., Koturbash T., Kutcherov V. (2019) Review of impacts of gas qualities with regard to quality determination and energy metering of natural gas. *Measurement Science and Technology*, **30**(2), 1—20.
2. Vaskovskii S.V., Brokarev I.A. (2020) Comparative analysis of statistical models for the task of natural gas composition analysis. *Journal of Information Technologies and Computing Systems*, **1**, 34—43.
3. Koturbash T.T., Brokarev I.A. (2018) Metod opredeleniya svoystv i sostava prirodnogo gaza po izmereniyam ego fizicheskikh parametrov [The method of determination of the properties and the composition of natural gas by measuring of its physical parameters]. *Sensors and Systems*, **1**, 43—50, [in Russian].
4. Koturbash T., Bicz A., Bicz W. (2016) New instrument for measuring velocity of sound and quantitative characterization of binary gas mixtures composition. *Measurement Automation Monitoring*, **1**, 254—258.
5. Löfqvist T., Delsing J., Sokas K. (2003) Speed of sound measurements in gas — mixtures at varying composition using an ultrasonic gas flow meter with silicon based transducers. *International Conference on Flow Measurement*, Groningen, Netherlands.
6. *Xensor Integration Website* [Online]. Available: http://www.xensor.nl
7. *Dynament Infrared Gas Sensors* [Online]. Available: https://www.dynament.com
8. *Bright Sensors BlueEye* [Online]. Available: https://www.brightsensors. com
9. Saati T. L. (1989) Prinyatie reshenij. Metod analiza ierarhij. *Radio i svyaz'*, 316 p.
10. Ringuest J.L. (1992) Compromise Programming. In: Multiobjective Optimization: Behavioral and Computational Considerations. *Springer*, Boston, MA, 51—59.
11. Vaskovskii S.V., Brokarev I.A. (2019) Distributed data gathering system for natural gas composition analysis. *Proceedings of the 22nd International Scientific Conference on Distributed Computer and Communication Networks: Control, Computation, Communications (DCCN-2019, Moscow)*, 381—388.
12. *Matlab 2018b Software* [Online]. Available: https://www.mathworks.com.
13. *REFPROP Software* [Online]. Available: https://www.nist.gov/srd/refprop.
14. Brokarev I.A., Vaskovskii S.V. (2019) Distributed data gathering system to analyze natural gas composition. *Advances in Systems Science and Applications*, **19**(4), 14—24.
15. Brokarev I.A. Iskusstvennyie neyronnyie seti dlya resheniya zadachi analiza komponentnogo sostava gazovyih smesey [Artificial neural networks for solving the problem of analyzing the composition of gas mixtures]. *Upravlenie bolshimi sistemami [Large-scale Systems Control]*, **80**, 98—115.

