Strojnícky časopis – *Journal of* MECHANICAL ENGINEERING,
VOL 75 (2025), NO 1, 61 – 70

COMPARISON OF THE REGRESSION METHOD AND THE NEURAL NETWORK METHOD IN SPECIFIC CASES OF ENGINEERING PRACTICE

KRIŽAN Peter[*1], SVÁTEK Michal[1], MATÚŠ Miloš[1], HANKO Lukáš[1]

[1]*Slovak University of Technology in Bratislava, Faculty of Mechanical Engineering, Institute of Production Engineering and Quality Management, Nám. Slobody 17, 812 31 Bratislava, Slovakia,*
[*]*e – mail: peter.krizan@stuba.sk*

Abstract: The presented article presents a summary of research findings based on a comparison of two statistical methods, both of which were applied to the same problem. These were data that were measured during the biomass compaction process, with softwood used as the input material. The aim of the analysis was to create models for both selected statistical methods that would sufficiently describe the given process and subsequently compare them. The second very important step was to compare the applicability of these methods from the point of view of limiting criteria and to define their advantages and disadvantages in practical use. For the purposes of the experiment itself, the biomass pressing process was used using a uniaxial press. In the case of this process, parameters were defined that significantly affect the quality of the resulting press, which is represented by the measured quantity - the press density. The parameters already mentioned, which by their nature significantly affect the quality of the press, are the pressing pressure, pressing temperature, humidity of the pressed material and the size of the pressed material fraction. The process itself and the subsequent experimental obtaining of results from the given research took place in laboratory conditions using an experimental uniaxial press. Subsequently, through the steps of statistical analysis, we carried out estimates of effects, individual tests of hypotheses about the significance of the model and effects and, last but not least, also the predictive ability of the models thus obtained. The obtained models, which were created using the two methods already mentioned, were subsequently compared based on their predictive ability, specifically through the so-called predictive ability of the model, which is represented by the coefficient of determination R2, which is defined as the ratio of the sum of squares SSM explained by the model to the total sum of squares SST. As a result, it expresses the degree of agreement of the observed values with the model. Based on the experimental knowledge thus obtained, the aim is to point out the significance and importance of statistical methods and methodologies that are useful in processing data of various nature and scope. It is precisely the modelling of a process, which is often complex in nature or demanding on the data obtained, that needs to be differentiated in the statistical methods used. The research results presented in the paper demonstrate their significance.

KEYWORDS: statistical methods, neural networks, regression analysis, biomass, densification, briquettes density, comparation

1 Introduction

Statistical methods play an increasingly important role in the analysis and evaluation of tasks in all industries [1]. The same is true in the case of engineering practice. In many cases, thanks to an appropriately used statistical method, it is possible to perfectly describe the given (e.g. production) process with regard to the monitored (required) output [2]. In the given case, we are talking about revealing the dependence between the input variables and the output variables of the given process [3].

The mentioned methods differ according to their applicability with regard to the nature of the given variable and their amount (number) [4 - 9]. According to these parameters, we can divide the methods leading to the detection of the dependence of variables according to Figure

DOI: 10.2478/scjme-2025-0007, Print ISSN 0039-2472, On-line ISSN 2450-5471

1. As can be seen from the given figure, some methods can be used for multiple types of variables. Of course, all the mentioned methods have their limitations and conditions of use. Much depends on the nature of the variables, their quantity, etc. In our article, we therefore decided to compare the selected method from the above list with a method that is still less known in some industries. This is the use of artificial neural networks.

The use of artificial neural networks is a relatively young method that is gradually finding its application in several industries [2, 6, 7]. This method, unlike other statistical approaches that depend on the assumption of a certain statistical distribution of classified data, does not assume such a distribution, or he doesn't expect him. In the case of classical statistical methods, it is assumed that each class in the variable space will have some n-dimensional distribution [5]. However, in such a case, the problem with the statistical approach is that the assumed model may not fit the type of data in n-dimensional space [8]. As a result of this fact, statistical approaches can be considered limiting precisely because of the strict assumption of a certain data model [3].

Independent variable (variables)		**Dependent variable (variables)**			
		One		More than one	
		Quantitative	Qualitative	Quantitative	Qualitative
One	Quantitative	Regression analysis	Logistic regression Discriminant analysis	Canonical correlation analysis	Multi-group discriminant analysis
	Qualitative	t - test	Discreet discriminant analysis	MANOVA	Discreet multi-group discriminant analysis
More than one	Quantitative	**Multiple regression analysis**	Logistic regression Discriminant analysis	Canonical correlation analysis	Multi-group discriminant analysis
	Qualitative	ANOVA	Discreet discriminant analysis Conjoint analysis	MANOVA	Discreet multi-group discriminant analysis

Fig. 1 Overview of the applicability of individual methods for detecting dependencies

2 Materials and methods

The basis for the comparison of the mentioned methods were the results of the experiment, which was carried out in laboratory conditions and whose target was to describe the biomass compaction process by a statistical model.

The main target of the first phase was the analysis of the measured data in the statistical software environment. In this case, it was a matter of dividing the analysis into two parts - a part devoted to analysis through the classical method (see Fig. 1) and a part devoted to analysis through an artificial neural network.

In our case, we focused on regression analysis, as it is the most frequently used method in statistical practice [10]. Specifically, with regard to the processed data, we devoted ourselves to multiple regression analysis. The reason was that in the given case we had several input (independent) variables and one output (dependent) variable [11]. As an alternative solution (for comparison), we chose the method of artificial neural networks.

As Figure 2 shows, the individual variables were of different character, some were continuous and some were discrete. In the case of pressing pressure and temperature, it was a continuous variable that can take on values in the range (63kPa - 191kPa) [12]. The pressing temperature also has the character of a continuous variable and its values are in the range (55°C - 130°C) [12-13]. Discrete variables include moisture content of the compacted material (%) and fraction size (mm) [11-12]. These variables can acquire previously known values from the given range. The mentioned four variables belong to the group of independent variables [13]. The dependent variable in this case is the density of the resulting compact (kg.dm^3) and represents the only dependent variable.

Pressure - p [kPa]	Temperature - T [°C]	Humidity - w [%]	Faction - L [mm]	Measured density [kg.dm3]
95	85	8	1	1,135
159	85	8	1	1,157
95	115	8	4	1,167
159	115	8	4	1,206
95	85	12	1	0,8
159	85	12	1	1,007
95	115	12	4	1,128
159	115	12	4	1,135
95	85	8	4	1,089
159	85	8	4	1,081
95	115	8	1	1,191
159	115	8	1	1,236
95	85	12	4	0,755
159	85	12	4	0,96
95	115	12	1	1,174
159	115	12	1	1,236
63	100	10	2	1,023
191	100	10	2	1,101
127	55	10	2	0,972
127	130	10	2	1,146
127	100	5	2	1,186
127	100	15	2	1,053
127	100	10	0,5	1,024
127	100	10	4	1,012
127	100	10	2	1,016
127	115	15	0,5	1,159
191	115	15	0,5	1,187
191	130	5	0,5	1,192
127	130	5	0,5	1,206
95	130	5	0,5	1,201

Fig. 2 Table of input values

3 Processing the measured data using an artificial neural network

In the next step, using the SAS JMP 17 software, we prepared a proposal for processing the measured data using an artificial neural network with three neurons in a layer. This was the basic setting we used for the given case. A schematic representation of the mentioned proposal is shown in Figure No. 1.

As we mentioned in the introduction, the principle of data processing by neural networks is their learning and subsequent evaluation of the result with a repeated verification process. As Figure 4 shows, the result of the analysis in JMP 17 statistical software are two result values.

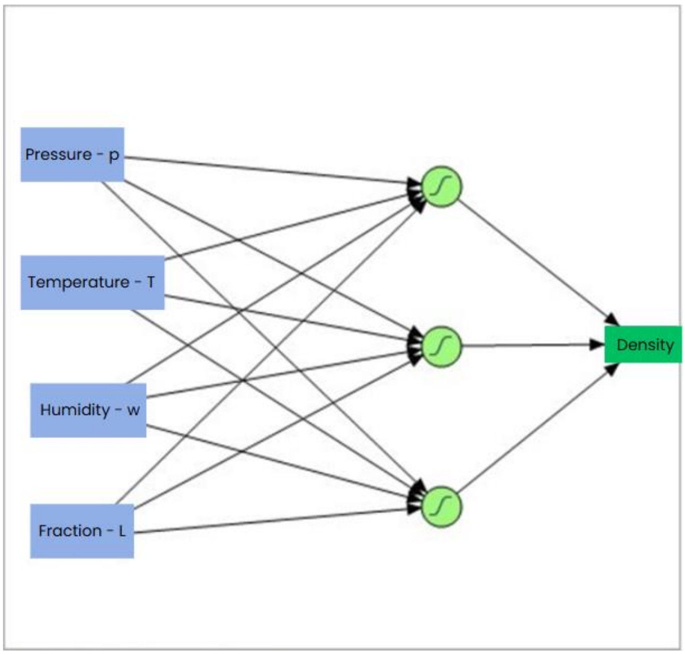

Fig. 3 Diagram – neural network with three neurons

Training – Density: Measures	Value	Validation – Density: Measures	Value
RSquare	0,9943059	RSquare	0,8922261
RASE	0,0082411	RASE	0,0455831
Mean Abs Dev	0,0069252	Mean Abs Dev	0,0409353
-LogLikelihood	-81,11245	-LogLikelihood	-10,01567
SSE	0,00163	SSE	0,0124669
Sum Freq	24	Sum Freq	6

Fig. 4 Summary of fit for artificial neural network – 3 neurons

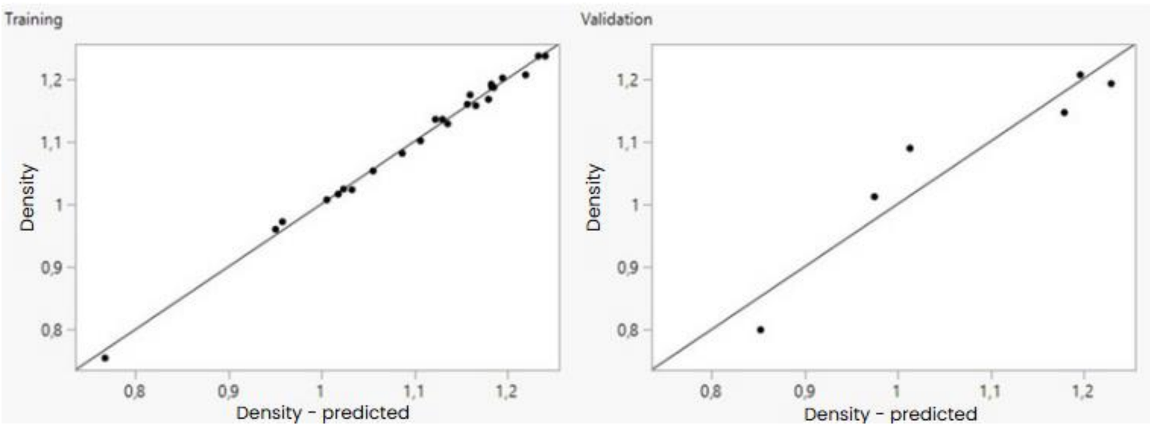

Fig. 5 Actual by predicted plot – 3 neurons

The first column in Figure 4 shows the results of the model that was generated by the selected "training data". In this case, it is the data through which the model trains, or optimizes its values up to the best achieved criteria parameters. The second column contains the validation model, i.e. a model that serves to validate the model created on the training data. The validation model is created on a group of data other than the training data, but from the given range. As the results show, the achieved R^2 value is more than 99% for the training data. However, in the case of the validation model, it is only 89%.

Figure 5 shows the interpolation of a linear straight line in the case of training data, or of the model created by training data and interpolation by validation data (validation model). As can be seen from the graphic display, in the case of training data, it is a better overlap with a straight line, or the points gets closer to the given straight line. In the case of validation data, it is exactly the opposite.

In the same way as with the design of three neurons on one layer, we also proceeded in the case of five neurons in a layer. Figure 6 shows a schematic representation of the proposal. The procedure itself for analyzing the artificial neural network designed in this way is the same as in the previous case.

In the case of the second design of the artificial neural network, we proceeded in the same way as in the previous case, figure 6.

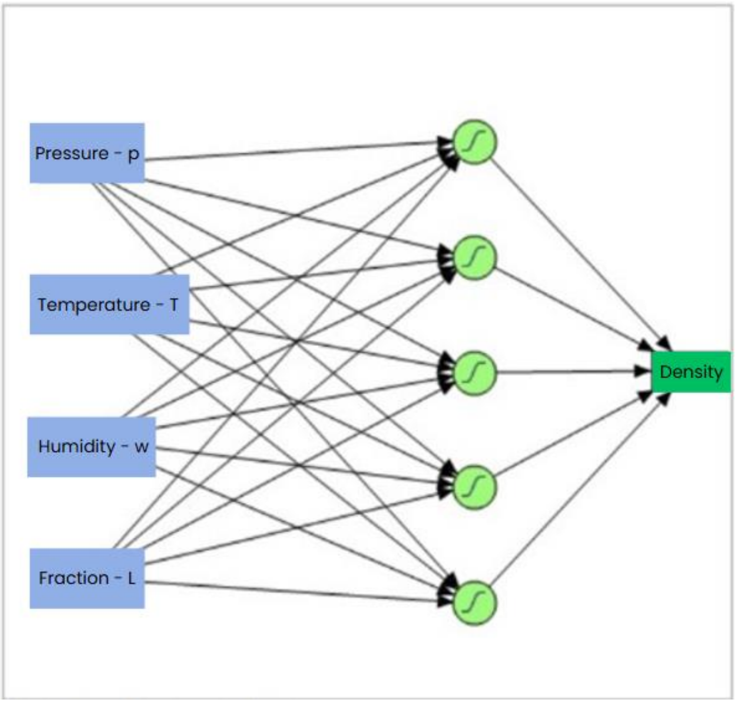

Fig. 6 Diagram – neural network with five neurons

The result in the second case is the data in Figure 7. The first column contains the results of the model, which was implemented on "training data", and the second contains the validation model. As the results of the second case show, the achieved value of R^2 is more than 75% in the case of training data. In the case of the validation model, it is more than 99%.

Training		Validation	
Density		Density	
Measures	Value	Measures	Value
RSquare	0,7550548	RSquare	0,9999959
RASE	0,0505712	RASE	0,0003198
Mean Abs Dev	0,0223079	Mean Abs Dev	0,0002284
-LogLikelihood	-37,57043	-LogLikelihood	-39,77377
SSE	0,0613787	SSE	6,1353e-7
Sum Freq	24	Sum Freq	6

Fig. 7 Summary of fit for artificial neural network – 5 neurons

Figure 8 shows the interpolation of a linear line by points in the case of training data and also validation data. As can also be seen from the graph, in the case of training data there is a worse overlap with a straight line than in the case of validation data, where it is exactly the opposite.

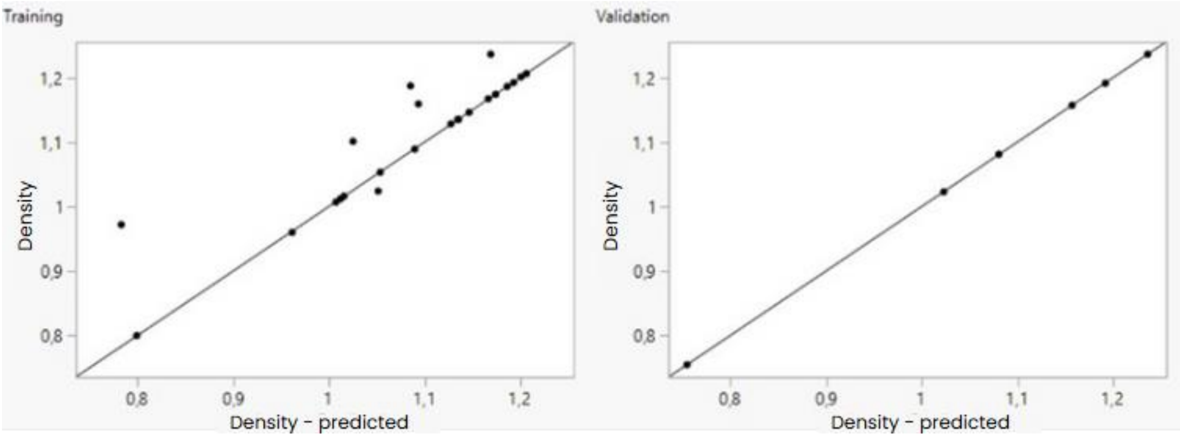

Fig. 8 Actual by predicted plot – 5 neurons

4 Processing the measured data using a regression analysis

The second, compared, method is regression analysis. As with artificial neural networks, we used the same SAS JMP 17 statistical software.

In the first step, we estimated the parameters of the regression model. Figure No. 11 shows estimate of individual parameters of the model, variables and also their mutual interactions.

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	0,7994324	0,105552	7,57	<,0001*
Pressure	0,0008498	0,000363	2,34	0,0303*
Temperature	0,0043225	0,000726	5,95	<,0001*
Humidity	-0,02196	0,005023	-4,37	0,0003*
Fraction	-0,014995	0,009119	-1,64	0,1166
(Pressure-130,2)x(Temperature-103,5)	-2,955e-5	2,676e-5	-1,10	0,2833
(Pressure-130,2)x(Humidity-9,8333)	0,0001117	0,000137	0,81	0,4256
(Pressure-130,2)x(Fraction-2,0333)	-7,168e-5	0,000292	-0,25	0,8085
(Temperature-103,5)x(Humidity-9,833)	0,0012285	0,000381	3,23	0,0044*
(Temperature-103,5)x(Fraction-2,033)	0,0001651	0,000606	0,27	0,7882
(Humidity-9,8333)x(Fraction-2,0333)	-0,000255	0,004429	-0,06	0,9548

Fig. 9 Parameter estimates

The next step was to determine the effects of individual parameters. In picture no. 10, the mentioned effects are shown together with the order according to the size of the effect from the largest to the smallest. The graph also graphically displays the limit of the effectiveness of the given effect.

Effect Summary

Source	Logworth	PValue
Temperature	5,005	0,00001
Humidity	3,484	0,00033
Temperature x Humidity	2,354	0,00443
Pressure	1,519	0,03030
Fraction	0,933	0,11655
Pressure x Temperature	0,548	0,28333
Pressure x Humidity	0,371	0,42557
Temperature x Fraction	0,103	0,78818
Pressure x Fraction	0,092	0,80850
Humidity x Fraction	0,020	0,95477

Fig. 10 Effect summary results

Last but not least, we performed an analysis using the ANOVA method (analysis of variance) for the entire model. As the results in Figure 11 show, the criterion value is less than the value of $p < 0.05$, which means that the given model is statistically significant [14].

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	10	0,32725180	0,032725	7,8823
Error	19	0,07888237	0,004152	**Prob > F**
C. Total	29	0,40613417		<,0001*

Fig. 11 ANOVA for regression model

The last step was to evaluate the quality of the given model through the evaluation criterion, which in our case was the value of the coefficient of determination R^2.

Summary of Fit

RSquare	0,805773	AICc	-50,7406
RSquare Adj	0,703548	BIC	-52,2792
Root Mean Square Error	0,064434		
Mean of Response	1,097833		
Observations (or Sum Wgts)	30		

Fig. 12 Summary of fit for regression model

As Figure 12 shows, the value of R^2 in the case of the regression model was 81%, which represents a value at the limit of acceptability of the model (>80%). However, the value of R^2 can increase as the number of effects in the model increases, because the model can better fit its curvature to the observed values, even if this does not correspond to the true dependence [13-14]. Therefore, the adjusted coefficient of determination R^2_{adj} is used to express the degree of agreement of the model with the observed values. The value of the criterion R^2_{adj} in this case is 70.3%, which represents a very low value.

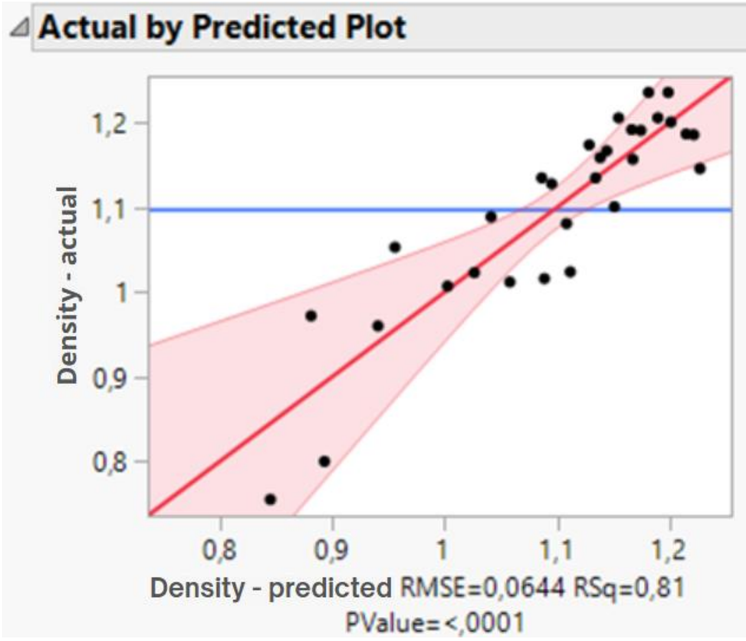

Fig. 13 Actual by predicted plot

Figure 13 shows a graph of predicted and measured values. As can be seen, the values deviate considerably from the straight line and are also outside the prediction range. This only confirms the quality of the given model in its absolute value, represented by the R^2 criterion.

In the case of the comparison of the results of both methods and the subsequently performed analyzes by these methods, we were not concerned with the absolute value of the R^2 criterion, but with the comparison of this value in the given methods. We did not evaluate the quality of the model itself, but the quality of the model by one method and another. The reason for the quality of the given model can be measured data (outliers, etc.) or other influences.

In general, however, we can conclude that the use of artificial neural networks gave us a better % value of the evaluation criterion R^2, i.e. this method in this case gave us a more accurate model than the regression analysis method. Added value is the conditions under which neural networks can be used.

CONCLUSION

By means of the mentioned research, the use of two selected statistical methods was compared on a selected case of analysis of measured values. The main conclusions we can state as a result of this study are:

- The advantage of artificial neural networks is that they are independent of the statistical distribution of the data, the normality of the data is not required, nor the sample size, and no particular model is expected.

- The accuracy of the model obtained through the use of artificial neural networks was significantly higher than the model obtained by regression analysis. The value of the criterion R^2 was in the case of the regression analysis R^2 = 81% and in the case of the artificial neural network R^2 = 99%.

- In the same way as when using other statistical methods, it can be stated that even when artificial neural networks are used, the so-called relearning the network, which in practice means that fewer neurons in a layer can bring a better result of the quality of the resulting model than a higher number of neurons. The R^2 value was R^2 = 99% for a network with 3 neurons and R^2 = 76% for a network with 5 neurons.

- The advantage of the application of artificial neural networks is the use of the so-called training data and subsequent verification by validation on selected data from the scope of the given selection.
- A very positive aspect of the use of artificial neural networks for data classification is the relatively high classification success achieved.
- The advantage of artificial neural networks is that their output can be discrete, continuous, vector or combined.
- Disadvantages of artificial neural networks are relatively long training time and demanding, less transparent internal processing (so-called "black-box").

ACKNOWLEDGMENT

This paper is a part of the research done within the project VEGA 1/0533/23 „Research of technological and structural parameters of the pressing process of composite biofuels from alternative raw materials“, funded by the by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences, within the project APVV-19-0607 “Optimized progressive shapes and unconventional composite raw materials of high-grade biofuels” funded by the Slovak Research and Development Agency.

REFERENCES

[1] Dung, H. T., et al. "Comparison of Multi-Criteria Decision Making Methods Using The Same Data Standardization Method" Strojnícky časopis – Journal of Mechanical Engineering 72 (2), pp. 57 – 72, **2022**. DOI: 10.2478/scjme-2022-0016

[2] Albrecht, R. F., Reeves, C. R., Steele, N. C. "Artificial Neural Nets and Genetic Algorithms", Springer Science & Business Media. 737 pp., **2012**. ISBN 9783709175330

[3] Ustynenko, O., et al. "Mathematical Model and Optimization Algorithm By Mass of Tracked Load-Carrier/Prime Mover MT-LB Transmission", Strojnícky časopis – Journal of Mechanical Engineering 72 (1), pp. 125 – 134, **2022**. DOI: 10.2478/scjme-2022-0012

[4] Anderson, F. “MANOVA by Example: Hands on Approach Using SAS”, Independently Published. 40 pp., **2019**. ISBN 978-1-674-89137-8

[5] Fahrmeir, L., Kneib, T., Lang, S., Marx, B. D. "Regression: Models, Methods and Applications", Springer Berlin Heidelberg. 746 pp., **2022**. ISBN 978-3-662-63881-1

[6] Gregor, M. "Umelá inteligencia 1", CEIT, a.s. Žilina, 164 pp. **2014**. ISBN 978-80-971684-1-4, (in Slovak)

[7] Gregor, M., Nemec, D., Hruboš, M., Spalek, J. "Umelá inteligencia 2, Hlboké učenie", CEIT, a.s. Žilina, 368 pp., **2018**. ISBN 978-80-89865-03-1 (in Slovak)

[8] Iavobucci, D. "Analysis of Variance (ANOVA)", CreateSpace Independent Publishing Platform, 302 pp., **2016**. ISBN 978-1-530-33202-1

[9] Kleinbaum, D. G., Kupper, L. L., Nizam, A., Rosenberg, E. S. "Applied Regression Analysis and Other Multivariable Methods", Cengage Learning 1072 pp., **2013**. ISBN 978-1-285-96375-4

[10] Križan, P., Šooš, Ľ., Matúš, M., Svátek, M., Vukelič, D. "Evaluation of measured data from research of parameters impact on final briquettes density", In: Aplimat - Journal of Applied Mathematics 3 (3), ISSN 1337-6365, **2010**.

[11] Križan, P., Svátek, M. "Navrhovanie experimentu pre výskum procesu zhutňovania", In: ERIN 2008. Education, Research, Innovation: the 2nd International Conference for Young researchers and PhD. Students, Bratislava, 23.-24.4. 2008, Bratislava, STU v Bratislave, **2008**. ISBN 978-80-227-2849-2 (in Slovak)

[12] Križan, P., Svátek, M. "Plán experimentálneho výskumu procesu zhutňovania drevnej biomasy a metodika vyhodnocovania nameraných údajov", In: Mechanical Engineering 2007, the 11th International Scientific Conference. Bratislava, November 29-30. 2007, Bratislava, STU v Bratislave, **2007**. ISBN 978-80-227-2768-6 (in Slovak)

[13] Meloun, M., Militký, J. "Statistická analýza experimentálních dat", ACADEMIA, 953 pp., **2004**. ISBN 80-200-1254-0 (in Czech)

[14] Šoltés, E. "Regresná a korelačná analýza s aplikáciami", Bratislava IURA Edition, 287 pp., **2008**. ISBN 978-80-8078-163-7