USE OF DIFFERENT STATISTICAL APPROACHES IN PREDICTION OF METABOLIZABLE ENERGY OF DIETS FOR BROILERS

Lato L. Pezo[1], Radmilo R. Čolović[2]*, Dragan V. Palić[2], Aca P. Jovanović[1]

[1]University of Belgrade, Institute of General and Physical Chemistry,
Studentski Trg 12 - 16, 11000 Belgrade, Serbia

[2]University of Novi Sad, Institute of Food Technology,
Bulevar cara Lazara 1, 21000 Novi Sad, Serbia

*Corresponding author:
Phone: +381214853796
Fax: +38121450725
E-mail address: radmilo.colovic@fins.uns.ac.rs

ABSTRACT: Energy value of diets has importance for feed producers and farmers. Methods for *in vivo* determination of metabolisable energy have high accuracy, but they are time and cost consuming. The aim of this study was to investigate the effect of enzymatic digestible organic matter and values of proximate chemical analysis on prediction of the nitrogen corrected true metabolisable energy (TME_n) of diets for broilers. The performance of Artificial Neural Network was compared with the performance of first order polynomial model, as well as with experimental data in order to develop rapid and accurate method for prediction of TME_n content. Analysis of variance and post-hoc Tukey's HSD test at 95% confidence limit have been calculated to show significant differences between different samples. Response Surface Method has been applied for evaluation of TME_n. First order polynomial model showed high coefficients of determination ($r^2 = 0.859$). Artificial Neural Network model also showed high prediction accuracy ($r^2 = 0.992$). Principal Component Analysis was successfully used in prediction of TME_n.

Key words: *true metabolisable energy, broiler diets, FOP, ANN*

Abbreviations: ANN - Artificial Neural Network, BFGS - Broyden–Fletcher–Goldfarb–Shanno, CA - Crude Ash, CFa - Crude Fat, CFi - Crude Fibre, CP - Crude Protein, DM - Dry Matter, EDOM - Enzymatic Digestible Organic Matter, FOP- First Order Polynomial model, ME - Metabolisable Energy, MLP – Multi-layer Perceptron model, OM - Organic Matter, PCA - Principal Component Analysis, r^2 - Coefficient of determination, RSM - Response Surface Methodology, SD - Standard Deviation, SOS - Sum of Squares, TME_n - True Metabolisable Energy corrected for nitrogen.

INTRODUCTION

Energy value of diets has importance for feed producers and farmers. Since methods for *in vivo* metabolizable energy (ME) determination require the use of live animals they can be considered as most accurate. On the other hand, these methods are often time and cost consuming (Elkin, 1987; Mohamed et al., 1984; Palić and Leeuw, 2009; Pojić et al., 2008). There has been a considerable interest to find accurate methods for ME prediction, which will be also rapid and inexpensive (Robbins and Firman, 2005; Zhang et al., 1994).

Recently, mathematical modelling has been increasingly used for the study of the given systems. Developed empirical models show a reasonable fit to experimental data and successfully predict ME (Perai et al., 2010). Nonlinear models are found to be more suitable for real process simulation. First order polynomial (FOP), using Response Surface Methodology (RSM) and Artificial Neural Network (ANN) models have gained momentum for modelling and control of processes (Khuri and Mukhopadhyay, 2010; Priddy and Keller, 2005).

ANN models are recognized as a good modelling tool since they provide the empirical solution to the problems from a set of experimental data, and are capable of handling complex systems with non-linearities and interactions between decision variables (Almeida, 2002). The specific objective of this study was to investtigate the effect of EDOM and values of proximate chemical analysis on nitrogen corrected true metabolisable energy (TME_n) content of diets for broilers. The performance of ANN was compared with the performance of FOP, as well as to experimental data, in order to develop rapid and accurate method for prediction of TME_n.

MATERIAL AND METHODS

Feed and assays

Twenty one diets for broilers were used in the study. Proximate chemical composition of the diets was determined according to AOAC standard methods (AOAC, 2000). Estimation of the enzymatic digestibility of organic matter (EDOM) was performed by use of modified method of Boisen and Fernandez (1997). *In vivo* TME_n content of the diets was determined using the assay described by McNab and Blair (1988).

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a mathematical procedure used as a central tool in exploratory data analysis (Brlek et al., 2013). It is a multivariate technique in which the data are transformed into orthogonal components that are linear combinations of the original variables. PCA is performed by Eigenvalue decomposition of a data correlation matrix (Abdi and Williams, 2010). This transformation is defined in such a way that the first component has the largest possible variance. This analysis is used to achieve maximum separation among clusters of parameters (Pezo et al., 2013). This approach, evidencing spatial relationship between processsing parameters, enabled a differentiation between the different samples.

First order polynomial (FOP) model

According to general recommendations, prior to ANN modelling, analysis of variance (ANOVA) was performed, in order to check the significant effect of the input variables over the output, as well as to justify the later use of ANN model by coefficient of determination (r^2). Analysis and mathematical modelling was performed using StatSoft Statistica 10.0 software (Statistica, 2010).

The FOP model was used for estimation of the main effect of the process variables on responses. The independent variables used for modelling were dry matter (DM), crude protein (CP), crude fibre (CFi), crude fat (CFa), crude ash (CA), organic matter (OM) and enzymatic digestible organic matter (EDOM), while TME_n was response variable. FOP model was fitted to data collected by experimental measurements:

$$Y = \beta_0 + \sum_{i=1}^{7} \beta_i X_i \quad (1)$$

where: β_0 and β_i are constant regression coefficients, Y is response variable, while X_i are independent variables. The significant terms in the model were found using ANOVA for each dependent variable.

Artificial Neural Network (ANN) modelling

The database for ANN was randomly divided to: training data (60%), cross-validation (20%) and testing data (20%). The cross-validation data set was used to test the performance of the network, while training was in progress as an indicator of the level of generalization and the time at which the network has begun to overtrain.

Testing data set was used to examine the network generalization capability.

To improve the behaviour of the ANN, both input and output data were normalized. In order to obtain good network behaviour, it is necessary to make a trial and error procedure and also to choose the number of hidden layers, and the number of neurons in hidden layer(s). A multi-layer perceptron model (MLP) consisted of three layers (input, hidden and output). Such a model has been proven as a quite capable of approximating nonlinear functions (Hu and Weng, 2009) giving the reason for choosing it in this study. In this work the number of hidden neurons for optimal network was ten. Broyden-Fletche-Goldfarb-Shanno (BFGS) algorithm was used for ANN modelling.

After defining the architecture of ANN, the training step was initiated. The training process was repeated several times in order to get the best performance of the ANN, due to a high degree of variability of parameters. It was accepted that the successful training was achieved when learning and cross-validation curves (Sum of Squares vs. training cycles) approached zero. Testing was carried out with the best weights stored during the training step. Coefficient of determination (r^2) and SOS were used as parameters to check the performance (i.e. the accuracy) of the obtained ANNs.

After the best behaved ANN was chosen, the model was implemented using an algebraic system of equations to predict TME_n content of studied diets.

The goodness of fit

The goodness of fit for developed models (FOP and ANN) were evaluated using the coefficient of determination (r^2), the mean relative percent error (P), the root mean square error (RMSE) and the reduced chi-square (χ^2). The higher the values of r^2 and the lower the values of P, RMSE and χ^2, the better is the goodness of the fit. These parameters can be calculated as follows:

$$P = \frac{100}{N} \cdot \sum_{i=1}^{n} \frac{\left| Y_{\exp,i} - Y_{pre,i} \right|}{Y_{\exp,i}},$$

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^{n} \left(Y_{\exp,i} - Y_{pre,i} \right)^2 \right]^{1/2},$$

$$\chi^2 = \frac{\sum_{i=1}^{n} \left(Y_{\exp,i} - Y_{pre,i} \right)^2}{N - n}$$

where $Y_{\exp,i}$ is the ith experimentally observed response Y, $Y_{pre,i}$ is the ith predicted Y, N is the number of observations and n is the number of constants.

Sensitivity analysis

Sensitivity analysis is a sophisticated technique which is necessary to use for studying the effects of observed input variables and also the uncertainties in obtained models and general network behaviour. Neural networks were tested using sensitivity analysis, to determine whether and under what circumstances obtained model might result in an ill-conditioned system (Taylor, 2006). On the basis of developed ANN model, sensitivity analysis was performed in order to more precisely define the influence of processing variables on the observed outputs. The infinitesimal amount (+0.0001%) has been added to each input variable, in 10 equally spaced individual points encompassed by the minimum and maximum of the train data. These signals were normally distributed with a constant intensity and frequency. It was used to test the model sensitivity and measurement errors.

RESULTS AND DISCUSSION

Results of proximate chemical analysis, EDOM and TME_n content of broiler diets are presented using descriptive statistics in Table 1. DM, CP, CFa, CFi, CA, OM, EDOM, and TME_n varied significantly, implying that fitting of the experimental data can be performed using FOP and ANN modelling.

Principal component analysis (PCA)

Preliminary performed calculation for estimation of effects, using RSM of experimental data, showed that only EDOM, CFa, CFi, and CA variables influenced TME_n at statistically significant level ($p<0.05$). Therefore DM, CP and OM were excluded from further calculation.

The PCA applied to the given data set has shown a differentiation between the samples according to used process parameters, and it was used as a tool in exploratory data analysis to characterize and differentiate neural network input parameters (Figure 1). As it can be seen, there is a neat separation of the observed samples according to used assays. Quality results show that the first two principal components, accounting for 81.53% of the total variability for TME_n, can be considered sufficient for data representation. CFi content, CA, TME_n and EDOM had been more influential for the first factor coordinate calculation (accounting 25.5, 29.9, 22.9 and 20.1% contribution, respectively), while CFa content had been more influential for the second factor coordinate calculation (67.3%, respectively). PCA (Figure 1) showed quite good discrimination between samples. Starter samples (1 and 2) are grouped at the centre of the graph, while LP and HP finisher samples (3 and 4) are grouped at the lower side of the graph (with increased CFa). Finisher and grower feed samples (5 and 6) are located on the right side of the graph (with increased CFi and CA). Starter samples (7-12) are located on the left side of the graph, with increased EDOM and TME_n. Grower samples (13-16) are located in the central area of the graph. Starter samples containing by-products of biodiesel production (17-21) are located at the top of graph, with the lowest values of CFa, CFi, CA, TME_n and EDOM. Position of proximate chemical analysis parameters EDOM

Table 1.
Results of proximate chemical analysis, EDOM and TME_n content of complete diets for broilers

	DM (%)	CP* (%)	CFi* (%)	CFa* (%)	CA* (%)	OM* (%)	TME_n (MJ/Kg DM)	EDOM (%)
Average	88.42	23.53	3.36	6.68	6.18	93.82	15.58	80.52
SD	1.06	3.09	1.73	3.43	1.55	1.55	1.01	3.42
Min.	85.30	16.63	1.90	2.79	3.81	88.61	12.60	71.50
Max.	89.85	27.37	8.87	14.88	11.39	96.19	17.18	86.76
Var.	1.13	9.55	3.00	11.80	2.41	2.41	1.02	11.71

*Results are expressed on dry matter basis; SD - Standard deviation; Min. – Minimum; Max. – Maximum; Var. – Variance
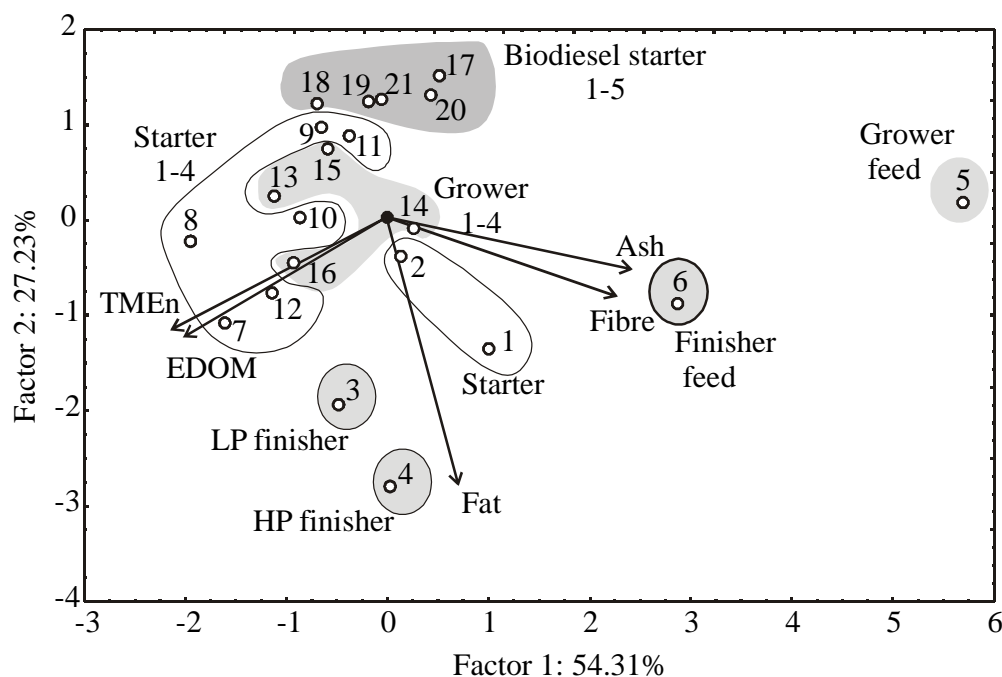

Figure 1. Biplot graph of broiler diets with the results of proximate analysis, EDOM and TME_n content

and TME_n is showing that EDOM and TME_n are positively correlated, while CFi and TME_n are negatively correlated, meaning that when CFi of the diet is decreasing and EDOM is increasing, TME_n will increase.

Analysis of variance and FOP model

Analysis of variance (ANOVA) was conducted for obtained FOP model, and output were tested against the impact of input variables (Table 2). Analysis revealed that linear terms considerably influenced forming of FOP model for TME_n calculation.

According to ANOVA results, TME_n was mostly affected by CA, which was statistically significant at $p < 0.001$ level. CFa was also very influential, at $p < 0.001$ level as well. CFi and EDOM were also very influential, statistically significant at $p < 0.05$ level. DM, CP and OM were excluded from this calculation, as explained before. The coefficient of determination (r^2) for FOP model was rather low (0.859), which leads to the conclusion that ANN model developing would improve the validity of the model.

The residual variance, marked as 'Error' in Table 2, presents the model disagreement with the experimental values i.e. contributions of terms that are not described in the FOP model. The residual analysis showed that the mean of residuals was equal to zero, and the standard deviation was 0.379 (with minimum of -0.527 and the maximum of 0.902). These results showed a good approximation to a normal distribution around zero with a probability of 95% (2×SD), which means a good generalization ability of the developed model for the range of observed experimental values (the skewness parameter showed minimal deviations from normal distribution, 0.546, while the Kurtoisis parameter showed almost negligible difference in "peakedness" compared to normal distribution, -0.254). Developed model showed statistically insignificant deviation from the experimental values of the model, which confirmed its suitability. Evaluated values of coefficient of determination (0.859), the mean relative percent error (2.072), the root mean square error (0.370) and the reduced chi-square (0.169) confirmed that obtained FOP model was statistically significant and in agreement with experimental results.

Neurons in the ANN hidden layer

All variables considered in the RSM, were also used for the ANN modelling. Determination of the appropriate number of hidden layers and number of hidden neurons in each layer is one of the most critical tasks in ANN design. The number of neurons in a hidden layer depends on the complexity of the relationship between inputs and outputs. As this relationship becomes more complex, more neurons should be added (Ćurčić et al., 2014).

The optimum number of hidden neurons was chosen upon minimizing the difference between predicted ANN values and desired outputs, using Sum of Squares (SOS) during testing as performance indicator. Used multi-layer perceptron models (MLPs) were marked according to StatSoft Statistica's notation. MLP was followed by number of inputs, number of neurons in the hidden layer, and the number of outputs.

Table 2.
Analysis of variance (ANOVA) of TME_n in diets for broilers

Factor	df	SS	F	p
CFi	1	1.30*	7.25	0.02
CFa	1	5.88[+]	32.70	<0.001
CA	1	9.70[+]	54.00	<0.001
EDOM	1	0.84*	4.67	0.46
Error	16	2.88		
r^2		0.859		

[+]Significant at $p < 0.001$ level, *Significant at $p < 0.05$ level, 95% confidence limit; df - Degrees of freedom, SS - Sum of squares, F - F-test value

Table 3.
Performance of the optimal ANN

Network name	MLP 4-10-1	
Training	Testing	Validation
$r^2=0.992$	$r^2=0.894$	$r^2=0.999$
Training error	Testing error	Validation error
0.0005	0.0038	0.0012
Training algorithm: BFGS 30		
Error function: SOS		
Hidden activation: Exponential		
Output activation: Tanh		
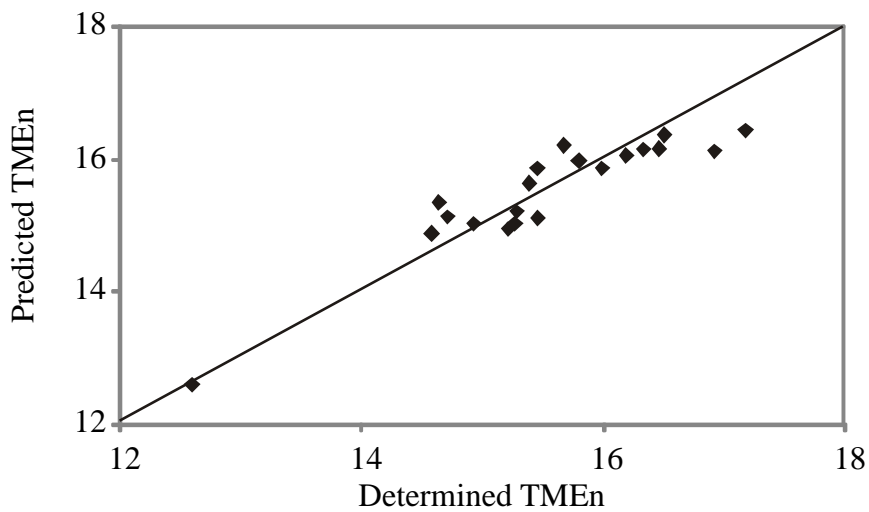
Figure 2. Comparison of experimentally obtained TME_n with ANN and FOP predicted values
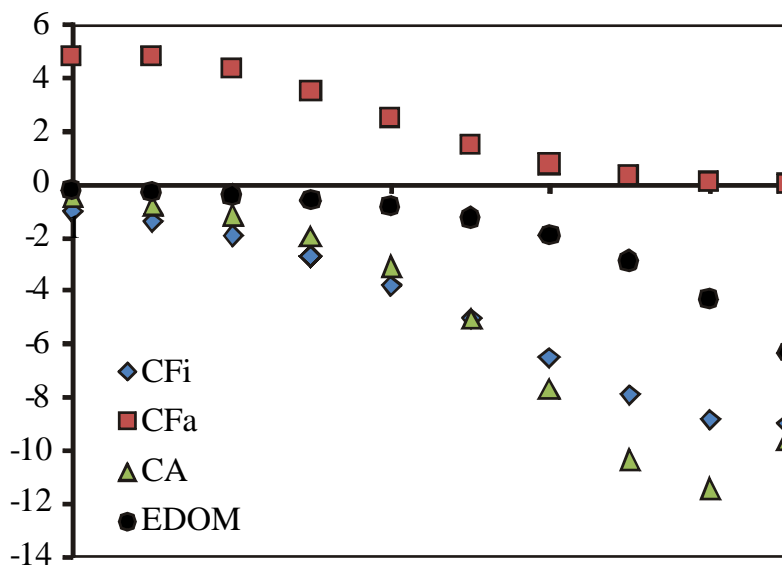

Figure 3. Sensitivity analysis - the influence of the input over the output variables

According to ANN performance (Table 3), it was noticed that the optimal number of neurons in the hidden layer for TME_n calculation was 9 (network MLP 4-9-1),

when obtaining high values of r^2 (0.992 for ANN during training period, compared to 0.859 for FOP model) and low values of SOS.

Simulation of the ANN

Optimal network, used for prediction of TME_n was able to predict reasonably well the output for a broad range of the process variables (coefficients of determination reached 0.992 for TME_n prediction). The predicted values were very close to the experimental (target) values in most cases, in terms of r^2 value for both FOP and ANN models.

It can be seen that the r^2 value for ANN model is higher than that associated with the FOP model. This is in line with findings of Perai et al. (2010) who compared different statistical approaches for prediction of ME of meat and bone meal. Generally, ANN model is more complex (55 weights-biases for ME calculation) than FOP, and it has performed better fitting of experimental data due to the high nonlinearity of the developed system (Karlović et al., 2013; Chattopadhyaya and Rangarajana, 2014).

The mean and the standard deviation of residuals have also been analysed. The mean of residuals for ANN model was equal to 0.014, and the standard deviation was 0.110 (with minimum of -0.586 and the maximum of 0.701). The skewness parameter showed minimal deviations from normal distribution (-0.050), while the Kurtoisis parameter showed almost negligible difference in "peakedness" compared to normal distribution (0.439). Evaluated values of coefficient of determination (0.859), the mean relative percent error (2.072), the root mean square error (0.370) and the reduced chi-square (0.169) are confirming that obtained FOP model was statistically significant and in agreement with experimental results.

Sensitivity analysis

In order to assess the effect of changes in the outputs due to the changes in the inputs, a sensitivity analysis was performed. The greater effect observed in the output implies that greater sensitivity is presented with respect to the input (Pezo et al., 2013). Sensitivity analysis has been performed to test an infinitesimal change in an input value in 10 equally spaced individual points, ranged by the minimum and maximum of the observed assay, in order to explore the changes in observed outputs. It is also used to test the model sensitivity and measurement errors.

The influence of the input over the output variables, i.e. calculated changes of output variables for infinitesimal changes in input variables, is shown on Figure 3. Obtained values corresponded to the level of experimental errors, and also showed the CFa, CFi, CA and EDOM influence on TME_n.

Sensitivity analysis is used to show the influence of the inputs, but it also shows the importance of an input variable at a given point in the input space (Saltelli and Annoni, 2010). As it can be seen on Figure 3, TME_n was affected more strongly by the infinitesimal changes of CFa, CFi, CA, and EDOM at the extreme values of the input range. These findings are in accordance with PCA and ANOVA analysis, as well as with experimental measurements.

CONCLUSIONS

This paper presented different statistical approaches for prediction of *in vivo* TME_n content in complete diets for broilers using the results of proximate chemical analysis and EDOM. FOP and ANN-based models were developed for prediction of TME_n for a wide range of input variables. Both models are easy to implement and could be effectively used for predictive purposes, modelling and optimization. As compared to RSM, ANN model yielded a better fit of experimental data. Taking into account that a considerable amount and wide variety of data were used in the present work to obtain the ANN model, and considering that the model turned out to yield a sufficiently good representation of the experimental results, it can be expected that it will be useful in practice.

REFERENCES

1. Abdi, H., Williams, L. J. (2010). Principal component analysis, *Wiley Interdisciplinary Reviews: Computational Statistics*, *2 (4),* 433-459.

2. Almeida, J. S. (2002). Predictive non-linear modeling of complex data by artificial neural networks, *Current Opinion in Biotechnology*, *13 (1),* 72-76.

3. AOAC (2000). Official Methods of Analysis, 17th Edition, AOAC INTERNATIONAL, Gaithersburg, MD 20877-2417, USA.

4. Boisen, S., Fernandez, J.A. (1997). Prediction of the total tract digestibility of energy in feedstuffs and feed diets by *in vitro* analyses. *Animal Feed Science and Technology, 68,* 277-286.

5. Brlek, T., Pezo, L., Voća, N., Krička, T., Vukmirović, Đ., Čolović, R., Bodroža-Solarov, M. (2013). Chemometric approach for assessing the quality of olive cake pellets, *Fuel Processing Technology, 116,* 250-256.

6. Ćurčić, B., Pezo, L. L., Filipović, V. S., Nićetin, M. R., Knežević, V. (2014). Osmotic treatment of fish in two different solutions-artificial neural network model, Journal of Food Processing and Preservation, doi:10.1111/jfpp.12275

7. Chattopadhyaya, P.B., Rangarajana, R. (2014). Application of ANN in sketching spatial nonlinearity of unconfined aquifer in agricultural basin, *Agricultural Water Management, 133*, 81-91.

8. Elkin, R. G. (1987). A review of duck nutrition research, *World's Poultry Science Journal, 43,* 84–106.

9. Hu, X., Weng, Q. (2009). Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multilayer perceptron neural networks, *Remote Sensing of Environment, 113 (10),* 2089-2102.

10. Karlović, S., Bosiljkov, T., Brnčić, M., Jezek, D., Tripalo, B., Dujmić, F., Dzineva, I., Skupnjak, A. (2013). Comparison of artificial neural network and mathematical models for drying of apple slices pretreated with high intensity ultrasound, *Bulgarian Journal of Agricultural Sciences, 19,* 1372-1377.

11. Khuri, A. I., Mukhopadhyay, S. (2010). Response surface methodology, *Wiley Interdisciplinary Reviews: Computational Statistics, 2 (2),* 128-149.

12. McNab, J.M., Blair, J.C. (1988). Modified assay for true and apparent metabolizable energy based on tube feeding. *British Poultry Science, 29,* 697-707.

13. Mohamed, K., Leclercq, B., Anwar, A., El-Alaily, H., Soliman, H. (1984). A comparative study of metabolizable energy in ducklings and domestic chicks, *Animal Feed Science and Technology, 11,* 199–209.

14. Palić, D., Leeuw, K. J. (2009). Comparison of three in vitro methods for determining and predicting the organic matter digestibility of complete diets for ruminants, *Acta Periodica Technologica, 40,* 79-86.

15. Perai, A. H., Moghaddam, H. N., Asadpour, S., Bahrampour, J., Mansoori, Gh. (2010). A comparison of artificial neural networks with other statistical approaches for the prediction of true metabolizable energy of meat and bone meal, *Poultry Science, 89 (7),* 1562-1568.

16. Pezo L., Ćurčić, B., Filipović, V., Nićetin, M., Koprivica, G., Mišljenović, N., Lević, Lj. (2013). Artificial neural network model of pork meat cubes osmotic dehydration, *Hemijska industrija, 67,* 465-475.

17. Pojić, M., Palić, D., Mastilović, J., Janić-Hajnal, E. (2008). The introduction of a method for determination of organic matter digestibility in feeds into routine laboratory practice, *Food processing, quality and safety, 35 (3),* 151-154.

18. Priddy, K. L., Keller, P. E. (2005). Artificial neural networks: an introduction. SPIE Press, Washington.

19. Robbins, D. H., Firman, J. D. (2005). Evaluation of metabolizable energy of meat and bone meal for chickens and turkeys by various methods, *International Journal of Poultry Science, 4 (9),* 633-638.

20. Saltelli, A., Annoni, P. (2010). How to avoid a perfunctory sensitivity analysis, *Environmental Modelling & Software, 25 (12),* 1508-1517.

21. STATISTICA (Data Analysis Software System). (2010). v. 10., Stat-Soft, Inc., USA, (www.statsoft.com).

22. Taylor, B. J. (2006). Methods and procedures for the verification and validation of artificial neural networks. Springer Science+Business Media, New York.

23. Zhang, W. J., Campbell, L. D., Stothers, S. C. (1994). An investigation of the feasibility of predicting nitrogen-corrected true metabolizable energy (TME_n) content in barley from chemical composition and physical characteristics*, Canadian Journal of Animal Science, 74 (2),* 355-360.

УПОТРЕБА РАЗЛИЧИТИХ СТАТИСТИЧКИХ ПРИСТУПА ЗА ПРЕДВИЂАЊЕ МЕТАБОЛИЧКЕ ЕНЕРГИЈЕ СМЕСА ЗА ИСХРАНУ БРОЈЛЕРА

Лато Л. Пезо[1], Радмило Р. Чоловић*[2], Драган В. Палић[2], Аца П. Јовановић[1]

[1]Универзитет у Београду, Институт за општу и физичку хемију,
Студентски трг 12-16, 11000 Београд, Србија
[2]Универзитет у Новом Саду, Научни институт за прехрамбене технологије у Новом Саду,
Булевар цара Лазара 1, 21000 Нови Сад, Србија

Сажетак: Енергетска вредност смеса за исхрану животиња је за произвођаче хране за животиње, као и за сточаре, један од најбитнијих параметара нутритивног квалитета. *In vivo* методе за одређивање метаболичке енергије имају велику тачност, али су временски заметне и скупе. Циљ овог истраживања био је да се испита утицај ензимске сварљивости органске материје, као и основног хемијског састава смеса, на предвиђање стварне метаболичке енергије смеса за исхрану бројлера. Резултати добијени употребом вештачких неуронских мрежа, поређени су са резултатима добијеним применом полинома првог степена, као и са експерименталним резултатима, у циљу добијања брзог и тачног поступка за предвиђање стварне метаболичке енергије. За испитивање значајности разлика између вредности, коришћени су анализа варијансе и post-hoc Tukey's HSD тест са 95%-ним интервалом поверења. Метода одзивне површине је коришћена за евалуацију стварне метаболичке енергије. Применом полинома првог степена и вештачких неуронских мрежа, добијене су високе вредности коефицијента детерминације ($r^2 = 0.859$ односно $r^2 = 0.992$). Метод главних компоненти је такође успешно искоришћен за предвиђање стварне метаболичке енергије.
Кључне речи: *стварна метаболичка енергија, смесе за исхрану бројлера,* FOP, ANN