PREDICTION THE AVERAGE ANNUAL CONCENTRATION OF PHOSPHATE IN RIVERS USING ANN AND MLR MODELS

Lidija Stamenković
College of Applied Professional Studies, Vranje, Serbia, lidija.stm@gmail.com

Abstract: Clean and quality water is one of the main aims in achieving the seventeen adopted Sustainable Development Goals. That means taking legislative measures at national and international level for the preservation of water quality, prevention of pollution and continuous monitoring. According to global and regional cooperation which is stated by numerous legal regulations, all member countries have obligation to monitoring water quality and submit annual report. For that reason it is very important to obtain accurate and precise data of all monitored water pollutants. In developing countries, because of poor and underdeveloped environmental infrastructure, obtained data for pollutants could have a certain percentage of uncertainty. In order to avoid that problem it is very important the existence of a larger number of models to estimate concentration of water pollutants, including concentration of phosphate. Due to the increasing influence of anthropogenic sources on the content of phosphorus, first in the soil and then in water sources, and negative effects caused by the eutrophication process, it is crucial to monitor the state of water in terms of the presence of phosphate. In the present study, in order to obtain one optimal model to predict the average annual concentration of phosphate in rivers was used two different approaches. In the first approach, was used artificial neural network (ANN), two different ANN architecture: Multilayer Perceptron (MLP) and Radial Basis Function (RBF). In the second approach were used two type of Multiple Linear Regression (MLR) technique, simultaneous and stepwise MLR. Available data used in this study were obtained from Eurostat-the statistical office of the European Union, for 19 European countries for the period from 2004 to 2012. For all created models was used the same dataset. As input variables were used sustainable development indicators, in total 10. All selected inputs were chosen on the basis of their influence on phosphate concentration in rivers and also reviewing of adequate literature. As model performance indicators in this study was used Mean Absolute Error (MAE) and coefficient of determination (R^2). Results of all created models first at all showed that comparing two ANN models, better performances on training and test dataset shows MLP model. On test dataset, MLP model gives satisfactory prediction with value of coefficient of determination 0,657. On the other side, developed MLR models show significantly poorer results. Taking into account performances of created models, it is clearly that MLP model shows the best predictive results. On the basis of results of MLP model, could be recognized that it can be an alternative model for prediction annual concentration of phosphate in rivers, and significant tool in water pollutant monitoring.
Keywords: ANN, MLR, phosphate in rivers, water quality.

1. INTRODUCTION

Among the others parts of environment, one of significant segment represent water. In the last decades, clean and satisfactory water quality, represents the challenge that society is facing. According to that, international legislation in the area of environmental protection is directed towards the protection and preservation of water quality. One of newest international legislation is adopted by United Nations in 2015, and represents Sustainable Development Agenda with 17 the main Sustainable Development Goals and 169 particularly defined goals, which includes water quality and all the rest environmental area (European Environment Agency, 2018; United Nations, 2015).

Bearing in mind that the rivers in terms of pollution are endangered, both, due to the direct introduction of contaminants and pollution caused by the dissolution of soil contaminants compounds which are came in it by anthropogenic influence, it is very important to continual monitoring all pollutants. Phosphorus and nitrogen are the two main elements responsible for the eutrophication process in aquatic system. This process causes aquatic ecosystem degradation and also, different problems in terms of water quality and water uses (Boeykens, Piol, Samudio Legal, Saralegui, & Vázquez, 2017). Although phosphate occurs naturally in water; from the aspect of water pollution more significant is high level of phosphorus concentration reached by human activities (Marcos, 2007). Human activities related to higher production of phosphate in water are production of domestic and different industrial wastes, production and consumption detergents, cosmetics and pharmaceutical products as well as use of fertilizers in agricultural sector (Hashim et al., 2019; L. Stamenković, Jevtić, & Mrazovac Kurilić, 2018).

Because of nature of rivers pollution which is not located only on one State, legal measures on national and global level obligate all Member States, developed and developing, to reporting about water quality and their progress in that direction. According to that, it is crucial to obtain precise an accurate data about all water quality parameters including phosphate concentration. In the case of developed countries, continuous monitoring is available because of developed environmental infrastructure i.e. high number of water quality monitoring stations. That leads to precise estimation of concentration of pollutants. However, in developing countries because of poorer environmental infrastructure, ability for accurate estimation of concentration of pollutants is more reduced or unavailable. Because of that, one or more alternative models can contribute to resolve this lack.

This study was focused on models based on Artificial Neural Networks (ANN) and Multiple Linear Regression (MLR). By using two different approaches an optimal model was developed for prediction of average annual concentration of phosphate in rivers.

2. MATERIALS AND METHODS

Source of data and input and output variables

For creating an optimal model for prediction of phosphate concentration in rivers, in the present work was used Eurostat database (European Commission, 2015). As input data for model creation were used sustainable development indicators: Total waste (TW), Population connected to at least secondary waste water treatment (SWT), Resource productivity and domestic material consumption-chain linked volumes-2010 (DMC), Real Gross Domestic Product (GDP), Shares of environmental and labour taxes in total tax revenues (ET), Recycling rate of municipal waste (RMW), Waste Manufacture of textiles, wearing apparel, leather and related products (WT), Waste Manufacture of wood and of products of wood and cork, except furniture; manufacture of articles of straw and plaiting materials (WW), Waste Manufacture of chemical, pharmaceutical, rubber and plastic products (WCh) and Final Energy Consumption (FEC). Those indicators were chosen, in total 10, on the basis of their influence on the production of phosphate. Although, the majority of data are available, for some indicators data are not available. Because of that, in this paper were used only data for 19 European countries, for period from 2004 to 2012. List of chosen input and output variables and their units are presented in Table 1. All models are created with the same dataset.

Table 1. List of variables and their units

Variable	TW	SWT	DMC	GDP	ET	RMW	WT	WW	WCh	FEC	Phosphate in rivers
Unit	kg **p**er **c**apita	% of population	€ per kg	€ per capita	% of T.Tax	% of TW	kg pc	kg pc	kg pc	million of toe	mg PO4/l

Modeling techniques

Artificial neural networks (ANN) represent one of the most commonly used technique in Ecological modeling for the last two decades. Application of ANN on the environmental field is based on their possibility to resolve often non-linear relationships between different parameters. During the last years, advance in ANN architectures as well as training algorithms, contribute to improve performance of ANN models. Detailed information about basic ANN modeling can be found in appropriate literature (Ozesmi, Tan, & Ozesmi, 2006; Wu, Dandy, & Maier, 2014). Several previous study have shown that ANN achieve satisfactory prediction results in water quality modeling (Singh, Basant, Malik, & Jain, 2009; Šiljić Tomić, Antanasijević, Ristić, Perić-Grujić, & Pocajt, 2018). In the present study were used two different ANN architecture: Multilayer Perceptron (MLP), which is most used type of ANN and Radial Basis Function (RBF). The basic difference between two applied networks architectures is the way of network learning process as well as the role of the hidden layer.

Multiple linear regression (MLR) is statistical technique of finding linear equation-model between the dependent variable and appropriate set of independent variables. The aim of this technique is to find linear combination of independent variables which can be able to determine dependent variable with smaller error in prediction. The mathematical form of MLR prediction model is presented in form of regression equation. This type of modeling includes many types of approaches. In the present work were used two MLR approaches: simultaneous and stepwise MLR. Detailed information about MLR modeling can be found in literature (Moghaddasi & Noorian-Bidgoli, 2018; Shaw, 2003).

In order to evaluate the performance of created models in this paper was used two statistical metrics: coefficient of determination (R^2) and Mean Absolute Error (MAE). The two parameters are calculated by using equations presented in paper (L. J. Stamenković, Antanasijević, Ristić, Perić-Grujić, & Pocajt, 2015).

3. RESULTS AND DISCUSSION

In order to obtain one optimal model to predict the average annual concentration of phosphate in rivers, in this study as mentioned above, was used two different approaches. First model was developed by using two types of artificial neural networks: RBF and standard tree layers network-MLP. With the aim to estimate prediction capability of ANN model, multiple linear regression models (MLR) were developed and compared with ANN model with better performances. Details of created ANN models and their performances on training data are presented in Table 2.

Table 2. Model summary (ANN-training data)

Model	Number of neurons in layer	AF[5]-H	AF-O	R^2	MAE
MLP	10-5-1[6]	Hyperbolic tangent	Sigmoid	***0,769***	***0,0213***
RBF	10-9-1	Softmax	Identity	0,554	0,0317

As it can be seen, according to MAE values for both created ANN models, MLP model shows significant smaller values of error in prediction MAE-0,0213, compared to RBF model where MAE value is 0,0317, i.e. smaller deviation between actual and MLP model estimated data of phosphate concentration. Also, the values of the coefficient of determination R^2-0,769 indicate that MLP model shows better correlation between measured and predicted values. After training phase, the new, unknown data set were presented to the models in order to estimate their real capability to predict concentration of phosphate in rivers. According to obtained results it can be concluded that MLP model shows better performances on test dataset comparing to the RBF results (Figure 1.) Also, results showed that performances of both models applied on unknown data set are slightly worse comparing to the results in training phase. Those discrepancies could be because of insufficiently training time of models. However, prediction results of ANN models are satisfactory.

With the aim to getting an optimal model, in the present study were also developed MLR model. Two different regression approaches were applied on existed dataset: first one was standard or simultaneous MLR and second one was stepwise MLR where independents variables were gradually involved in regression equation. Details of created MLR models and their performances on training data are presented in Table 3. As it can be seen, MLR simultaneous approach shows slightly better performances comparing to MLR stepwise model with R^2 values 0,384 and 0,363. Also, the values of models error show small advantages of MLR simultaneous model. Although the number of input variable included in stepwise MLR model is smaller, in total six. However, taking into account all created models and their performances, it is clearly recognized that MLP model i.e. model based on artificial neural network shows the best predictive results.

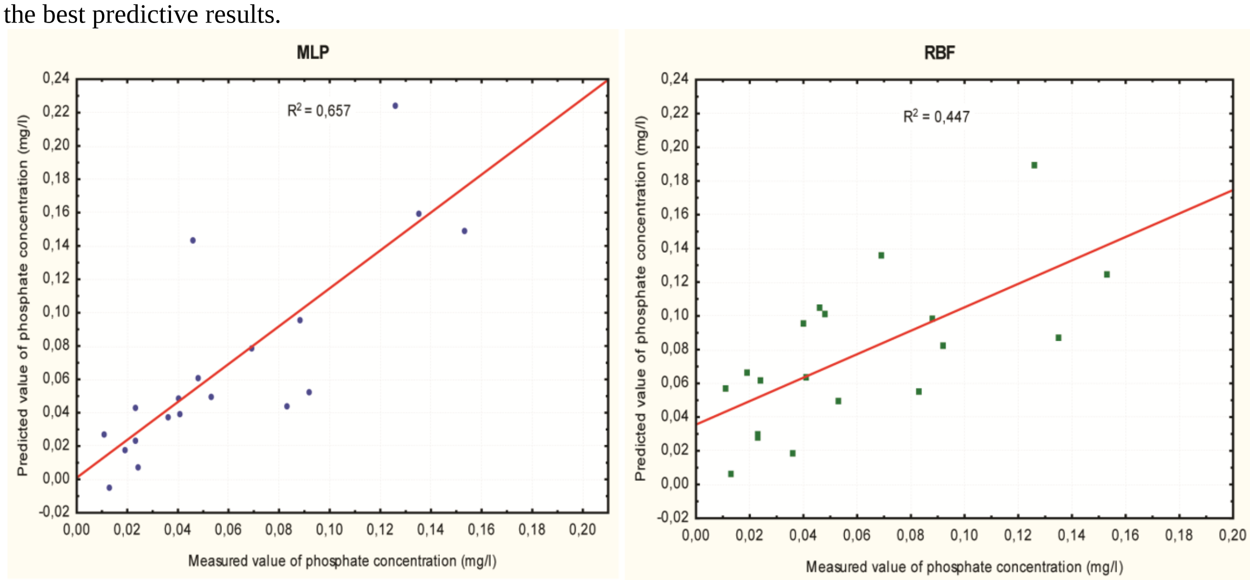

Figure 1. Comparison of the measured and ANN predicted values (MLP and RBF-test data)

[5] Activation Function
[6] Input–Hidden-Output layers (I-H-O)

Table 3. Model summary (MLR-training data)

		MLR simultaneous	MLR stepwise
Model equation	Constant	0,007435053	-0,001826844
	TW	0,000004171	0,000003951
	SWT	-0,000132995	
	DMC	0,063380518	0,065290352
	GDP	-0,000001306	-0,000001157
	ET	-0,001383781	
	RMW	0,000277177	
	WT	0,001812755	0,001792011
	WW	-0,000076479	-0,000071498
	WC	0,000039269	
	FEC	0,000181679	0,000212283
Model performance	R^2	***0,384***	0,363
	MAE	***0,0380***	0,0389

Prediction results of MLP and MLR models for every countries included in this study are showed in Figure 2. The results obtained confirm previously discussed values of model performance metrics that the best prediction was made from MLP model. Although MLP model shows good prediction in most cases, in some countries (Italy and Bulgaria) deviations between predicted and measured values are significant. Variation for some input parameters i.e. the quality of training dataset for those countries could be a reason for poorer prediction results.

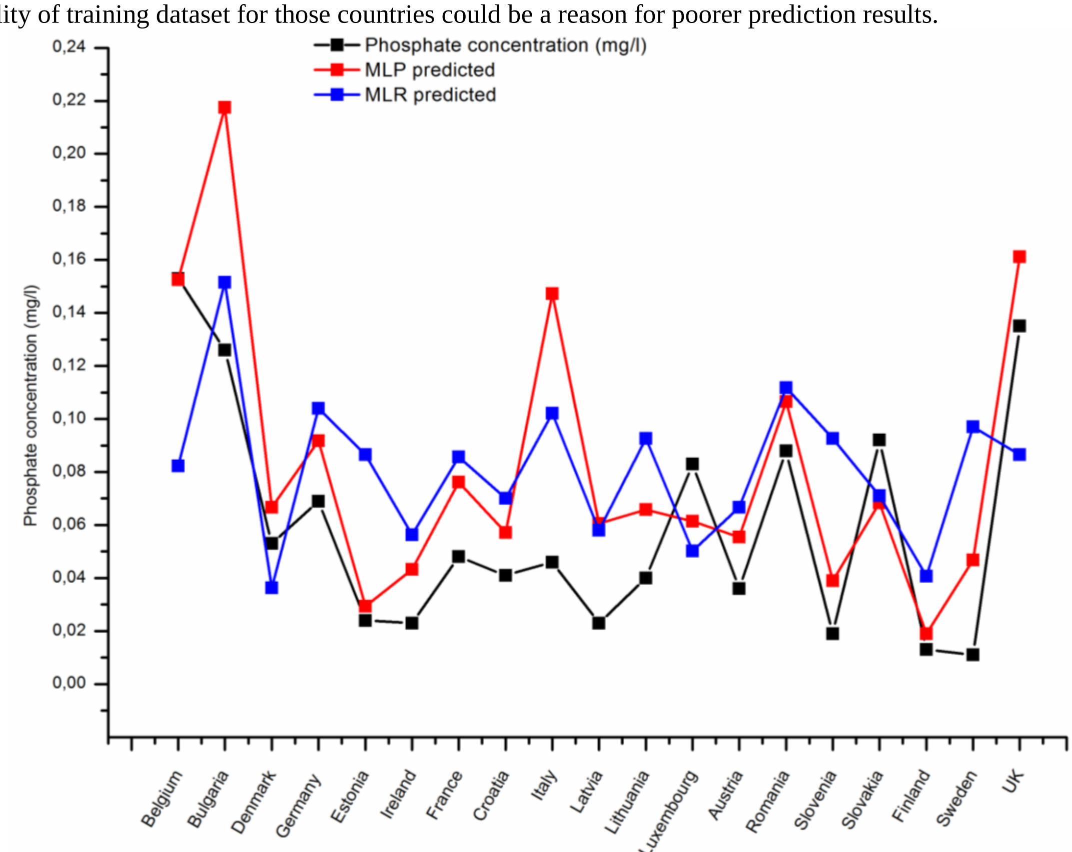

Figure 2. Prediction results of MLP and MLR models for every countries included in this study (test data)

4. CONCLUSION

This study investigated two different approaches, Artificial Neural Network and Multiple Linear Regression modeling techniques, for prediction annual concentration of phosphate in rivers. For this purpose were used data for 19 European countries and 10 sustainable development indicators as input variables. The results obtained showed that Multilayer Perceptron gives satisfactory prediction, better than other created models. In general it can be concluded that ANN modeling technique can be one useful tool and alternative model for estimation concentration of phosphate in rivers. As an auxiliary model, ANN is much more important for developing countries, because of their poorer environmental infrastructure. On the other hand, results of ANN model indicated that although it shows good predictions, the model shows some deviations. In this sense, future research can be focused on optimizing the model. It means examination multicollinearity between the input variables, removing the redundant or introduction the new inputs to model and optimization other ANN parameters.

REFERENCES

[1] Boeykens, S. P., Piol, M. N., Samudio Legal, L., Saralegui, A. B., & Vázquez, C. (2017). Eutrophication decrease: Phosphate adsorption processes in presence of nitrates. *Journal of Environmental Management*, *203*, 888–895.

[2] European Commission. (2015). Indicators - Eurostat. Retrieved August 28, 2018, from https://ec.europa.eu/eurostat/web/sdi/indicators

[3] European Environment Agency. (2018). *Water is life*. Retrieved from https://www.eea.europa.eu/publications/eea-signals-2018-water-is-life

[4] Hashim, K. S., Khaddar, R. Al, Jasim, N., Shaw, A., Phipps, D., Kot, P., … Alawsh, R. (2019). Electrocoagulation as a green technology for phosphate removal from river water. *Separation and Purification Technology*, *210*, 135–144.

[5] Marcos, V. S. (2007). *Wastewater Characteristics, Treatment and Disposal :* . PB - IWA Publishing. Retrieved from http://www.oapen.org/record/640137

[6] Moghaddasi, M. R., & Noorian-Bidgoli, M. (2018). ICA-ANN, ANN and multiple regression models for prediction of surface settlement caused by tunneling. *Tunnelling and Underground Space Technology*, *79*, 197–209.

[7] Ozesmi, S. L., Tan, C. O., & Ozesmi, U. (2006). Methodological issues in building, training, and testing artificial neural networks in ecological applications. *Ecological Modelling*, *195*(1-2), 83–93.

[8] Shaw, P. J. A. (2003). *Multivariate statistics for the Environmental Sciences*. *New York*. New York.

[9] Singh, K. P., Basant, A., Malik, A., & Jain, G. (2009). Artificial neural network modeling of the river water quality-A case study. *Ecological Modelling*, *220*(6), 888–895.

[10] Stamenković, L. J., Antanasijević, D. Z., Ristić, M. Đ., Perić-Grujić, A. A., & Pocajt, V. V. (2015). Modeling of methane emissions using artificial neural network approach. *Journal of the Serbian Chemical Society*, *80*(3), 421–433.

[11] Stamenković, L., Jevtić, P., & Mrazovac Kurilić, S. (2018). Application of artificial neural networks to predict phosphate in rivers. In *Conference VIVUS 2018* (pp. 124–131). Naklo, Slovenia: Biotechnical Centre Naklo, Strahinj 2018.

[12] Šiljić Tomić, A., Antanasijević, D., Ristić, M., Perić-Grujić, A., & Pocajt, V. (2018). A linear and non-linear polynomial neural network modeling of dissolved oxygen content in surface water: Inter- and extrapolation performance with inputs' significance analysis. *Science of The Total Environment*, *610-611*, 1038–1046.

[13] United Nations. (2015). *United Nations Millennium Development Goals*. New York. Retrieved from http://www.un.org/millenniumgoals/

[14] Wu, W., Dandy, G. C., & Maier, H. R. (2014). Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. *Environmental Modelling and Software*, *54*, 108–127.