Prediction of nitrate concentration in Danube River water by using artificial neural networks

Lidija J. Stamenković, Sanja Mrazovac Kurilić and Vladanka Presburger Ulniković

ABSTRACT

This paper describes the development of a model based on artificial neural networks (ANN) which aims to predict the concentration of nitrates in river water. Another 26 water quality parameters were also monitored and used as input parameters. The models were trained and tested with data from ten monitoring stations on the Danube River, located in its course through Serbia, for the period from 2011 to 2016. Multilayer perceptron (MLP), standard three-layer network is used to develop models and two input variable selection techniques are used to reduce the number of input variables. The obtained results have shown the ability of ANN to predict the nitrate concentration in both developed models with a value of mean absolute error of 0.53 and 0.42 mg/L for the test data. Also, the application of IVS has contributed to reduce the number of input variables and to increase the performance of the model, especially in the case of variance inflation factor (VIF) analysis where the estimation of multicollinearity among variables and the elimination of excessive variables significantly influenced the prediction abilities of the ANN model, r – 0.91.

Key words | ANN, Danube, MLP, nitrate, WQP

Lidija J. Stamenković
The Academy of Applied Technical and Preschool
 Studies,
Section Vranje,
Filipa Filipovića 20, Vranje 17500,
Serbia

Sanja Mrazovac Kurilić (corresponding author)
Vladanka Presburger Ulniković
Faculty of Ecology and Environmental Protection,
University 'Union-Nikola Tesla',
Cara Dušana 62-64, Belgrade 11000,
Serbia
E-mail: mrazovac@gmail.com

HIGHLIGHTS

- Represents the application of specific parameters to a neural network for estimating nitrate content.
- Analyzes in a new way a part of the flow of the Danube River through Serbia.
- Expands the range of applications of modeling and neural networks in the analysis of the environmental state.

INTRODUCTION

River pollution is one of the most widespread environmental problems nowadays (Dulo 2008; Abowei & Ekubo 2011; Hossain 2019). In recent decades, growing industrial, economic, as well as agricultural growth and development have greatly contributed to river pollution in many regions. The most important sources of river pollution are of anthropogenic origin: unhygienic living, industry, agriculture, sewage leaks, landfills, and application of wastewater for irrigation (Mateo-Sagasta et al. 2017). At the international level, many efforts are being made to preserve and improve water quality

and protect the status of aquatic ecosystems. The European Parliament adopted the 2000 Water Framework Directive establishing an integrated and coordinated water management system to safeguard their quality (Council Directive 2000/60/EC 2000). An integral part of the Water Framework Directive is the Nitrate Directive, the adoption of which aims to reduce water pollution by nitrates caused by agricultural activities (European Commission 1991). Member States are required to implement the Directive into their legislation and to continuously monitor and submit reports on the

ecological and chemical status of water resources, including national and transboundary catchments. Therefore, gathering accurate and reliable data on all water quality parameters is of the utmost importance. In order to obtain as reliable data as possible, it is very important to have alternative models for estimating water quality parameters, which would give a more complete insight into the status of surface water. Water quality indicators to monitor and determine its pollution status are defined through biological and physico-chemical parameters.

Nitrate is generally found in nature and it is the end product of the aerobic decomposition of organic nitrogenous matter and organic micro-organisms. Non-polluted natural water contains a small amount of nitrate. Surface water seldom contains more than 5 mg/L and often less than 1 mg/L of nitrate (Davies *et al.* 2008). According to claims of the Global Environment Monitoring System database, the concentration of nitrate in most rivers in populated regions today is approximately 70 mg/L while the healthy water quality standard suggested by the World Health Organization is 50 mg/L of nitrate as NO_3 or 10 mg/L nitrogene (NO_3-N). Nitrate has both positive and negative effects. A positive effect is that nitrate is an essential plant nutrient that is important for protein synthesis, growth of plants and nitrogen fixation. A negative effect is that nitrate contamination in drinking water causes health problems such as blue baby syndrome, thyroid disease and colorectal cancer (Ward *et al.* 2018). Nitrate N is mobile and it can be lost from the soil by leaching. Additional transport of nitrate N to surface water occurs through subsurface drainage or base flow. A very small amount of nitrate N is lost from the soil over the surface runoff (Jackson *et al.* 1973).

Nitrate content in European rivers originates predominantly from the agricultural sector, while contributions from the other sectors are of minor importance (Grizzetti *et al.* 2008; European Environment Agency (EEA) 2015). David *et al.* (1997) concluded that high soil mineralization rates, fertilization and tile drainage contribute notably to nitrate transport to rivers.

Water quality modeling has increased in importance in the last few decades. Water quality modeling is used to predict trends in water quality and pollutant concentrations (Najah *et al.* 2011; Antanasijević *et al.* 2014). Surface water models make significant contributions and give guidance on how to make decisions. Two main approaches to surface water quality modeling are process-based models and statistical models based on historical data. Artificial neural networks (ANN) are the result of attempts to apply the least complex models for forecasting. The ANN models need less data for forecasting and do not require a complex mathematical description of the process (Nayak *et al.* 2005). This type of modeling is very useful for the modeling of non-linear problems.

ANN represent a sophisticated and very useful modeling technique, which in recent years has had important applications in all areas of science. Artificial neural networks find application for classification and prediction in complex data where the connection between variables is complex and non-linear. Networks notice and learn data patterns and can produce results based on those patterns (Singh *et al.* 2009; Singh *et al.* 2012). From the aspect of this paper, their application is particularly important in the field of environmental protection, but first of all in the area of water protection and control of water quality (Xu & Liu 2013; Stamenković *et al.* 2017; Antanasijević *et al.* 2018). In past years, in order to achieve the best results, neural networks have been improved, and accordingly, today we have different types of networks. However, the literature data show that the multilayer perceptron (MLP) is the most frequently used network for solving various types of problems (Haykin 1994; Sarkar & Pandey 2015; Cabaneros *et al.* 2019). The MLP network consists of three layers of artificial neurons, or process units: input, one or more hidden layers, and one output layer. In input layer of the network the input data are presented, in the hidden layer the data are processed, i.e. the patterns among the presented data are represented, and the results are produced through the output layer. In this paper, a standard three-layer MLP neural network with one hidden layer is applied. In a standard MLP neural network each layer in the network architecture is connected to the previous layer. The number of neurons in the input and output layer depends of the number of input and output variables. Namely, the number of neurons in the input is equal to the number of input variables, while the number of neurons in the output layer represents the number of output variables. On the other side, the number of neurons in the hidden layer can be determined by a trial-and-error method, or may be estimated by applying a formula. The crucial step in model development is the training process. The training process means the adjustment of the connection

weights, i.e. the adjustment accuracy between the predicted and measured value expressed through model error after processing all presented data. During the training process there are different network parameters, such as learning rate, learning algorithm, activation function and number of neurons in the hidden layer, whose role in the process is to regulate the connection weights until the model gives results with an acceptable value of error. More detailed information on ANN functions can be found in the relevant literature (Kalogirou 2003; Wu *et al.* 2014).

Artificial neural networks are a very useful technique of many scientific disciplines such as ecology, analytical chemistry, and water quality (Alizadeh & Kavianpour 2015; Shamshirband *et al.* 2019). Literature on ANN models is very widespread (Diamantopoulou *et al.* 2005; Najah *et al.* 2009).

Application of neural networks for water quality modeling has been considered in different studies (Antanasijević *et al.* 2013; Šiljić *et al.* 2015). The results of those studies showed that ANN models provide good and reliable predictions and perform considerably better compared with the linear MLR model. Benzer & Benzer (2018) estimate the changes in the amount of nitrate in surface water using ANN and predict the nitrate value for the years 2020 and 2030. Raj Shrestha & Rode (2007) applied ANN for the simulation of streamflow nitrate concentration. Foddis *et al.* (2019) apply an MLP-ANN-based approach for assessing nitrate contamination of groundwater in Sardinia.

Taking into account that studies based on modeling nitrates in river by using ANN and water quality parameters as inputs are rare (Anmala & Venkateshwarlu 2019; Tiyasha *et al.* 2020), in this paper the application of ANN for the prediction of nitrate concentration in rivers is presented. The use of artificial neural networks to determine the concentration of nitrate provides another opportunity to obtain as reliable data as possible. The existence of alternative models, including the model based on neural networks, is

Figure 1 | Locations of measuring points (1-Bezdan; 2-Bogojevo; 3-Novi Sad; 4-Slankamen; 5-Zemun; 6-Smederevo; 7-Banatska Palanka; 8-Tekija; 9-Brza Palanka; 10-Radujevac).

presented in this paper as one of the contributions to this common goal of gaining a better insight into the quality of river water. Nitrates were taken as the output variable since they are very important in terms of maintaining water quality, and they are one of the basic chemical indicators of the presence of nutrients in water. For this purpose, a standard three-layer neural network was used, using two different approaches for selecting the most important input variables for the development of an ANN model.

METHODS

Study area

The Danube River is of immense importance for Europe and the countries through which it flows (Mihic & Andrejevic 2012). So, monitoring of its quality is very important. According to literature data, most pollutants reach the Danube from wastewater from its tributaries, the agricultural sector and from industrial and public sewage systems along its course (SEPA 2016). It has a navigable length of 2,415 km, and 588 km lies within the Republic of Serbia, from Bezdan to Prahovo (Mihic et al. 2011). Data used in this work were collected from ten locations on the Danube River through Serbia. Locations of measuring points are shown in Figure 1.

Source of data

Data used in the modeling process were taken from the monitoring network under the jurisdiction of the Agency for Environmental Protection of the Republic of Serbia (http://www.sepa.gov.rs/).

For that purpose, available data of the Danube River water quality in its flow through Serbia were used. The monitoring program has been adopted by aligning Serbia's legislation with the EU Water Framework Directive. In this regard, measuring stations, ten in total, have been designated so as to provide a comprehensive overview of the water quality of the Danube River. Data from ten monitoring stations and 27 water quality parameters (WQP), which are listed in the Table 1, were used in this study. The values of 27 parameters were measured monthly (12 times per year) for the period from 2011 to 2016. Part of the data, from 2011

Table 1 │ List of input/output WQP

WQP	Unit
Hardness	mg/L
Total Suspended Solids (TSS)	mg/L
Dissolved Oxygen (DO)	mg/L
% Saturation of Dissolved Oxygen (SO)	%
Alkalinity	mmol/L
Total Hardness (TWH)	mg/L
CO_2	mg/L
CO_3^{2-}	mg/L
HCO_3^-	mg/L
Total Alkalinity (TA)	mg/L
pH	/
Conductivity	mS/cm
Total Dissolved Salts (TDS)	mg/L
Ammonia	mg/L
Nitrite	mg/L
Organic Nitrogen	mg/L
Total Nitrogen	mg/L
PO_4^{3-}	mg/L
Phosphorus, total	mg/L
Ca	mg/L
Mg	mg/L
Cl	mg/L
SO_4^{2-}	mg/L
Chemical Oxygen Demand (COD)	mg/L
Biochemical Oxygen Demand (BOD)	mg/L
Total Organic Carbon (TOC)	mg/L
Nitrate as NO_3-N	mg/L

to 2015, was used for training networks, while data for 2016 were used for network testing.

Sampling and analytical methods of nitrate

Sampling was carried out from the midstream of the Danube River, according to the standard procedure. The sampling balloons were rinsed three times with the river water, then immersed at 30–40 cm below the level of the water surface and filled with water without bubbles (moving the balloon opposite to the river course).

NO_3-N nitrates were determined spectrophotometrically at a wavelength of 400 nm using a NitraVer reagent.

MLP model development parameters

In this paper, a standard three-layer MLP neural network with one hidden layer is applied. The crucial step in model development is the training process. MLP model development and architecture parameters are the same for all created MLP models in this study. One of the important steps in ANN model development is data pre-processing, because large values of some input variables could 'overshadow' the impact of the input variables with smaller values. According to that, a rescaling method used to ensure that all inputs fall in a similar range is standard. The hyperbolic tangent function was used as a neuron activation function and the conjugate gradient algorithm was applied as a weight optimization algorithm. The validation dataset is extracted randomly by choosing 30% of data patterns from the total number in the training dataset.

Input variable selection (IVS) techniques

Many literature data have shown that in order to obtain an ANN model with satisfactory performance it is desirable to choose those input parameters for the development of models that have the greatest influence on the dependent variable (Hu *et al.* 2007; Li *et al.* 2015; Tran *et al.* 2015). Smaller or larger numbers of input parameters than the optimal number for solving the problem implies that the network does not 'learn' good enough relationships among the variables, and thus makes bad predictions. On the other hand, if the number of inputs is significantly high, it affects the complexity of the network, and that problem is called over-training (Bowden *et al.* 2005). To solve this problem, different methods are used that can generally be classified into model-based and model-free methods. The main approaches that have been included in those two methods for IVS in ANN modeling can be generally classified into five selection procedures: methods based on *a priori* knowledge of variables and the problem, methods based on linear correlation, methods based on data-mining techniques, methods based on forward selection, stepwise selection and backward elimination and sensitivity analysis using trained ANN. The simplest and at the same time the most commonly used method is correlation analysis, in which the selection is based on the correlation between the selected inputs and the output variables. The values of the

Pearson correlation coefficient indicate which potential input variables are in significant correlation with the dependent variable. Based on these values, it can be seen whether the input variables are adequate for model development. In this paper, in order to assess the influence of individual input parameters on the concentration of nitrate in water, the correlation analysis is applied.

As mentioned earlier, the selection of the most important and uncorrelated input variables is essential for good prediction by artificial neural network models. For this purpose, two techniques for selecting the most important input variables have been applied in this paper: model-based (MB) and model-free techniques.

Table 2 | Skewness and kurtosis values of all WQP

| WQP	Skewness	Kurtosis
Hardness	0.089	−1.284
Total Suspended Solids	2.006	5.894
Dissolved Oxygen	0.087	−0.963
Percent Saturation of Dissolved Oxygen	1.329	4.677
Alkalinity	*18.791*	*419.660*
Total Water Hardness	0.572	0.275
CO_2	1.955	7.398
CO_3^{2-}	*2.946*	*10.496*
HCO_3^-	0.415	0.458
Total Alkalinity	0.710	0.454
pH	−0.059	0.855
Conductivity	0.744	0.337
Total Dissolved Salts	0.746	0.330
Ammonia	*6.717*	*96.397*
Nitrite	*6.657*	*67.721*
Organic Nitrogen	1.362	2.198
Total Nitrogen	1.217	2.953
PO_4^{3-}	*6.750*	*68.466*
Phosphorus, total	*5.277*	*38.061*
Ca	0.342	0.155
Mg	0.823	1.788
Cl	0.567	0.600
SO_4^{2-}	0.678	−0.149
COD	*4.461*	*42.123*
BOD	1.572	4.369
TOC	*8.436*	*109.020*
Nitrate as NO_3-N	0.615	−0.097
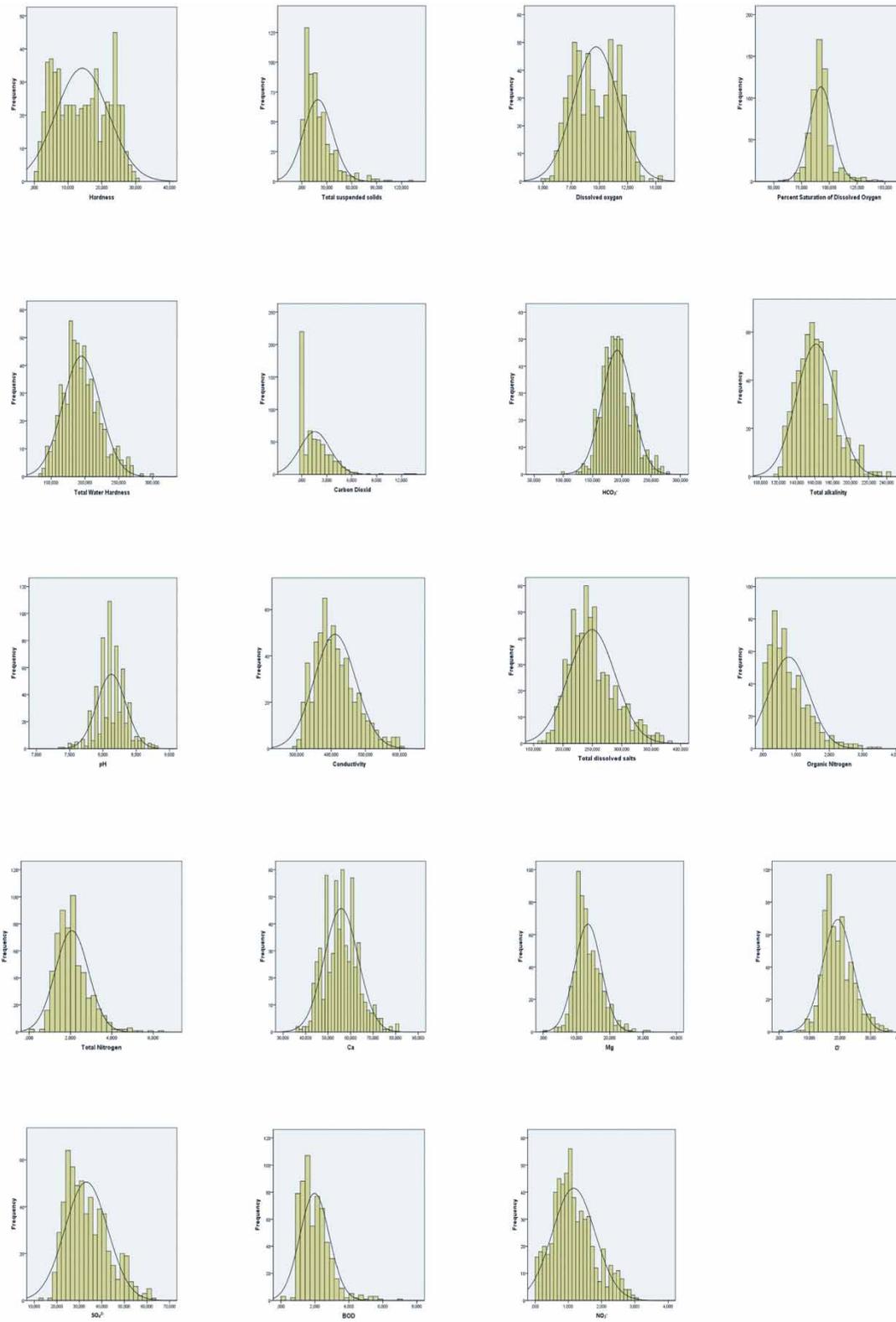
Figure 2 │ Histograms of near-normal distribution.

In the first applied technique for selecting input variables, the model-based technique, an ANN model was created with each individual potential input variable, and the results of the created ANN models were used to compute the determination coefficient values (R^2). The determination coefficient is defined as the proportion of the variance in the dependent variable that is explained by the independent variable. R^2 values were taken as indicators of the significance of the input variables.

The other approach to select the most important input variables, the model-free technique, is based on values of the variance inflation factor (VIF). Namely, the VIF statistics allow us to estimate the multicollinearity of the observed variables, which is significant, since the multicollinearity of the input variables significantly reduces the prediction capacity of the model, in this case the ANN model. The value of the VIF is based on the linear relationship between the input variables (Alin 2010). Since different values of VIF are suggested in the literature to indicate significant multicollinearity between the input parameters, in this paper the input parameters for the creation of the final model are those whose value of VIF is less than 10.

Model performance indicators

In order to determine the level or success of the created model prediction, different indicators of model performance can be found in the literature (Olyaie *et al.* 2015; Nabavi-Pelesaraei *et al.* 2019). Applied indicators mostly compare true or measured values with the values of the corresponding variables obtained by the prediction of the created models. In this paper, in order to evaluate the performances of both MLP models, the indicators that are most often found in the relevant literature, and which are recommended for the assessment of hydrological models, were used (Wang *et al.* 2009): the root mean square error (RMSE), the mean absolute error (MAE), the correlation coefficient (r) and the determination coefficient (R^2).

RESULTS AND DISCUSSION

Statistical analysis of data

The statistical techniques used in this paper to investigate the relationship between variables imply the fulfilment of a
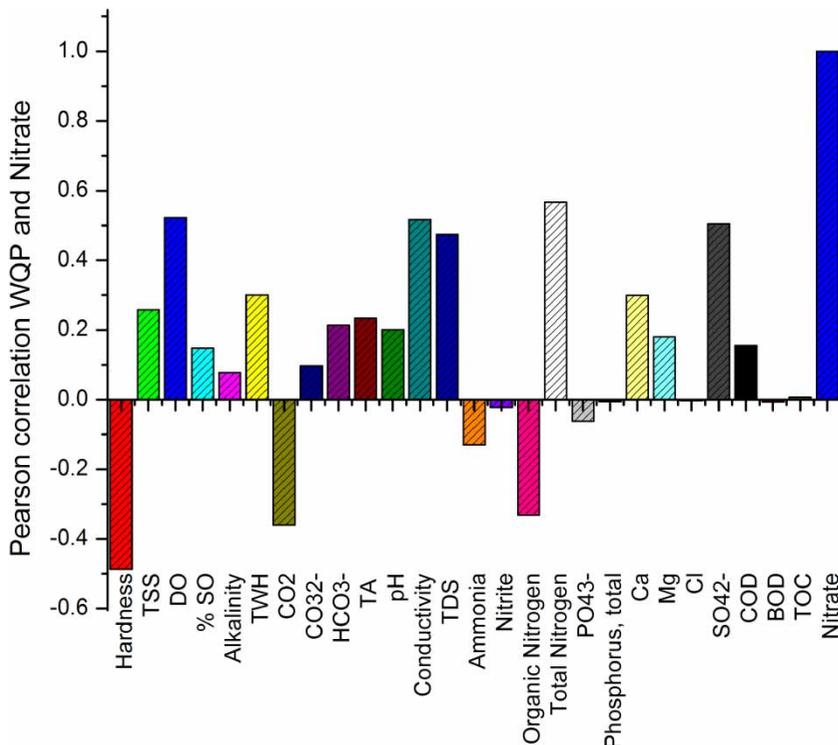

Figure 3 | Correlation coefficients between the input variables and the dependent variable.

Table 3 │ Values of R^2 for ANN models created with each input variable individually

WQP-MLP model	R^2
Hardness	0.167
Total Suspended Solids	0.024
Dissolved Oxygen	*0.315*
Percent Saturation of Dissolved Oxygen	0.003
Alkalinity	*0.214*
Total Water Hardness	0.031
CO_2	*0.209*
CO_3^{2-}	0.011
HCO_3^-	0.157
Total Alkalinity	0.028
pH	0.059
Conductivity	*0.217*
Total Dissolved Salts	*0.337*
Ammonia	0.07
Nitrite	0.069
Organic Nitrogen	0.02
Total Nitrogen	*0.524*
PO_4^{3-}	0.007
Phosphorus, total	0.184
Ca	0.058
Mg	0.005
Cl	0.059
SO_4^{2-}	*0.279*
COD	0.034
BOD	0.000
TOC	0.037

certain condition in terms of data distribution. Namely, the performance of the created models depends on the fulfilment of these data assumptions. Statistics of all water

quality parameters (WQP) with values of statistical indicators of the respective distribution are shown in Table 2. Histograms of parameters that show approximately normal distribution are shown in Figure 2. The values of skewness and kurtosis were taken to assess whether the data have a normal distribution. According to the literature data, the values of these statistical indicators in the range of −2 to +2 indicate that the data have a normal distribution (Wu *et al.* 2010; Barzegar *et al.* 2016; Šiljić Tomić *et al.* 2018). As can be seen, most of the parameters involved in this study show a normal or approximately normal distribution, i.e. among the 27 parameters, only eight show a significant deviation from the skewness and kurtosis limit values. These results indicate that the model may have reduced prediction accuracy.

Input variable selection and results of created ANN models

First, in order to examine the relationships between all chosen input parameters and nitrate concentration as output or dependent variable, correlation analysis was used. The existence of a correlation between input/output parameters and their potential convenience to be used in ANN model development are considered in relation to the value of the correlation coefficient. Input parameters which have statistically significant correlation with the model output in this case are those with the value of correlation coefficient above 0.2. Figure 3 shows the values of the correlation coefficient for each of the input parameters for model development. As can be seen, out of a total of 26 water quality parameters, 15 of them show that they are in significant correlation with the concentration of nitrates, i.e. with the

Table 4 │ Detailed information on created MLP models

Model	IVS	Critical value[a]	Variable inputs	Architecture[b]
MLP-MB	Model based	$R^2 < 0.2$	Dissolved Oxygen, Alkalinity, CO_2, Conductivity, Total Dissolved Salts, Total Nitrogen, SO_4^{2-}	$7 - 20 - 1$
MLP-VIF	VIF analysis	VIF >10	Total Suspended Solids, Percent Saturation of Dissolved Oxygen, Alkalinity, CO_2, CO_3^{2-}, pH, Conductivity, Total Dissolved Salts, Ammonia, Nitrite, Organic Nitrogen, Total Nitrogen, PO_4^{3-}, Phosphorus, total, Ca, Mg, Cl, SO_4^{2-}, COD, BOD, TOC	$21 - 20 - 1$

[a]For input removal.
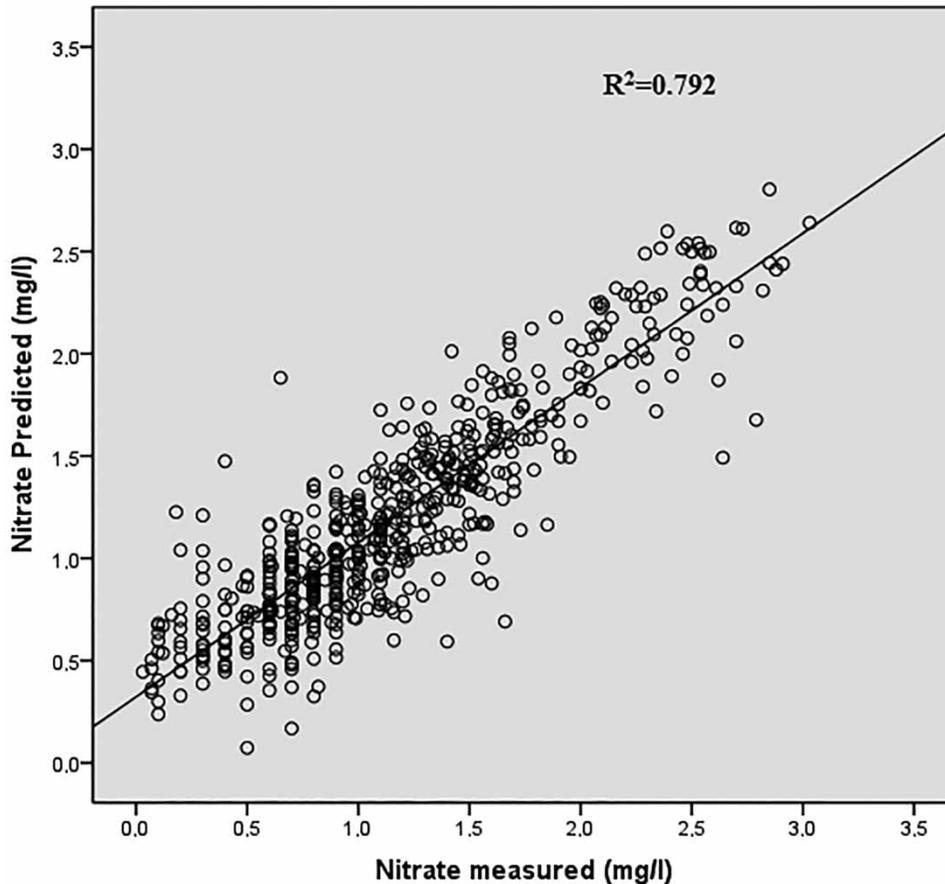[b]The number of neurons per layer.

Figure 4 │ Results of MLP-MB model prediction (training data).

output, dependent variable. Based on the obtained results of the correlation analysis, it can be concluded that the observed water quality parameters can be applied as input variables for the development of the ANN model.

In order to select the most significant inputs, the first IVS technique was applied. ANN models were created with each input variable individually. The results of the created ANN models are shown in Table 3. All input variables which indicate that the independent variable provides information about the output variable in the created models (R^2 higher than 0.2) are considered as significant for the model. The other variables show notably lower values of determination coefficient values. As can be seen, the values of R^2 indicate that seven input variables (italics marked in the table) are significant for the prediction of nitrate concentration. Among the other inputs, total nitrogen as input parameter showing the good prediction ability of the individual MLP model, which could explain the lower

values of R^2 for some of the water quality parameters (ammonia, nitrite) which are expected to have significant correlation with nitrate concentration. Based on these results, the seven input variables were selected for creating the ANN model.

The number of layer neurons for the study's created models are presented in Table 4. The results of the created model for the training data are shown in Figure 4. The values of model performance indicators for the test data are shown in Table 5.

Table 5 │ Performances of created MLP models – test data

| | **Performance indicators** | | |
Model	*RMSE* [mg/L]	*MAE* [mg/L]	*r*
MLP-MB	0.76	0.53	0.85
MLP-VIF	0.68	0.42	0.91

Table 6 │ VIF analysis

WQP	VIF
Hardness	*14.196*
Total Suspended Solids	1.292
Dissolved Oxygen	*19.03*
Percent Saturation of Dissolved Oxygen	5.218
Alkalinity	1.182
Total Water Hardness	*19.928*
CO_2	2.015
CO_3^{2-}	5.081
HCO_3^-	*30.973*
Total Alkalinity	*28.249*
pH	2.832
Conductivity	8.480
Total Dissolved Salts	6.932
Ammonia	1.470
Nitrite	1.069
Organic Nitrogen	1.908
Total Nitrogen	2.734
PO_4^{3-}	1.014
Phosphorus, total	1.113
Ca	9.663
Mg	7.366
Cl	2.735
SO_4^{2-}	2.690
COD	1.220
BOD	1.309
TOC	1.057

In order to select the significant input variable set to create an optimal ANN model with good performance, the second IVS approach based on VIF analysis was applied. The obtained results of VIF analysis are presented in Table 6. The input variables considered as significant to ANN model creation are those with lower VIF values. In other words, from the basic set of input parameters, a total of 26, those whose VIF value was greater than ten were removed, and the final ANN model was created with 21 inputs (Table 6).

The obtained results of the created model, the training data, are shown in Figure 5, while the performance indicator values, for the validation set of data, are shown in Table 5.

Therefore, based on the applied IVS techniques for selecting the most important input variables, the final ANN

models were created: MLP-MB and MLP-VIF. Detailed information on the created MLP models is shown in Table 4.

The performance indicators of the created models indicate that both applied techniques for selecting input parameters give satisfactory prediction results. Application of the model-based technique for selecting input parameters gave results with prediction errors of the models RMSE – 0.76 mg/L and MAE – 0.53 mg/L, that are satisfactory accuracy results, taking into account that the number of inputs is significantly reduced from 26 to only seven input parameters. On the other hand, after the VIF analysis, five parameters were removed from the initial set of inputs, and in this way, the MLP-VIF model with 21 inputs provided prediction results slightly better than the MLP-MB model, with errors predicting RMSE – 0.68 mg/L and MAE – 0.42 mg/L.

If we observe the results of the correlation analysis and afterwards applied IVS technique, in the case of the model-based approach it can be clearly seen that those parameters that are in good correlation with the output variable also give good results in the prediction of individual MLP models. In the case of the VIF analysis, due to the presence of multicollinearities between the inputs and their removal from the initial set, there was an improvement in the prediction of the created MLP-VIF model, *r* – 0.91. Namely, it is obvious that there are input water quality parameters that are in significant correlation, such as: hardness of water and total hardness, dissolved oxygen and saturation of dissolved oxygen, alkalinity of water and total alkalinity; which can reduce the ability of the model to notice and recognize the input–output relations and patterns, and can reduce the performance of the created models. In other words, high VIF values of individual input variables indicate that one of the highly correlated independent variables had to be removed from the model.

In this paper, discussion about data out of ranges is not included due to the extensiveness of the research. In the MLP model development part a rescaling method is listed that is used to ensure that all inputs fall in a similar range.

CONCLUSIONS

In this paper, models based on artificial neural networks have been used to predict the concentration of nitrates in
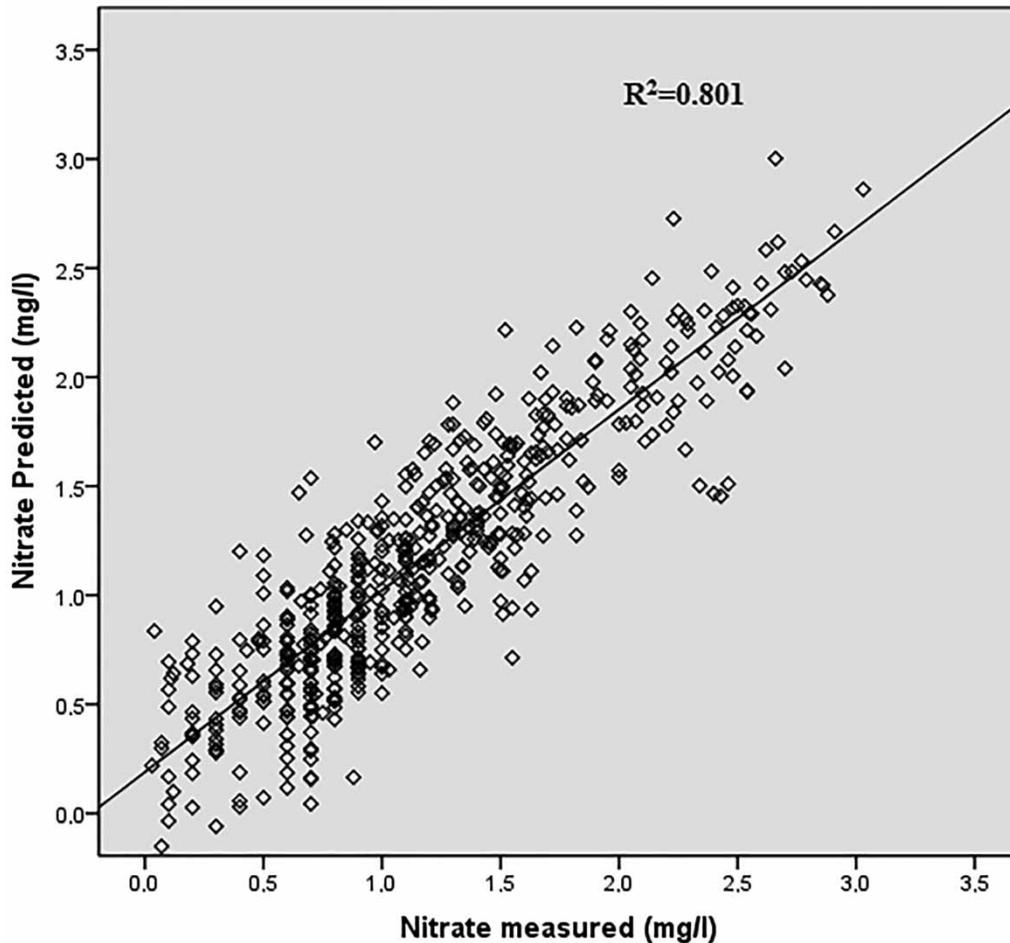
Figure 5 | Comparison of actual and MLP results (training data).

the Danube River. The models were developed using data of water quality parameters of the Danube River on its course through the Republic of Serbia. They were based on monitoring data from ten measuring stations for the period from 2011 to 2016, including 27 water quality parameters. For the development of the model, water quality parameters (as the input parameters) and standard three-layer neural networks were used. By applying the correlation analysis, the degree of correlation between nitrate concentration in the river water and other water quality parameters was determined. Results showed a significant relationship between individual WQPs and nitrate concentration, as confirmed by MLP models created with each input parameter individually. The application of the VIF analysis contributed to the reduction of multicollinearity between the input

parameters, resulting in very satisfactory prediction results on the validation set of data, with $r – 0.91$ and errors RMSE – 0.68 mg/L and MAE – 0.42 mg/L.

From the obtained results it can be concluded that neural networks can be used to predict the concentration of nitrates in water, based on their relationship with other water quality parameters. Nitrates are one of the most important water pollutants, especially in agricultural areas, and their monitoring by different models is significant.

Therefore, the use of neural networks as an alternative model can be significant, especially in developing countries such as Serbia, in order to obtain as complete and reliable data on water quality parameters as possible. Results obtained in this research can serve as a starting point for the development of new ANN models in order to predict

other water quality parameters. In addition, future research may be directed toward testing other input parameters to improve the performance of the model developed in this paper, as well as comparing the results obtained with those of linear models.

ACKNOWLEDGEMENT

REFERENCES

Abowei, J. F. N. & Ekubo, A. T. 2011 Aspects of aquatic pollution in Nigeria. *Research Journal of Environmental and Earth Sciences* 3 (6), 673–693. Retrieved from: https://maxwellsci. com/print/rjees/v3-673-693.pdf.

Alin, A. 2010 Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics* 2 (3), 370–374. http://doi.org/10. 1002/wics.84.

Alizadeh, M. J. & Kavianpour, M. R. 2015 Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, Pacific Ocean. *Marine Pollution Bulletin* 98 (1–2), 171–178. http://doi.org/10.1016/j.marpolbul.2015.06.052.

Anmala, J. & Venkateshwarlu, T. 2019 Statistical assessment and neural network modeling of stream water quality observations of Green River watershed, KY, USA. *Water Science and Technology: Water Supply* 19 (6), 1831–1840. http://doi.org/10.2166/ws.2019.058.

Antanasijević, D., Pocajt, V., Povrenović, D., Perić-Grujić, A. & Ristić, M. 2013 Modelling of dissolved oxygen content using artificial neural networks: Danube River, North Serbia, case study. *Environmental Science and Pollution Research* 20 (12), 9006–9013. http://doi.org/10.1007/s11356-013-1876-6.

Antanasijević, D., Pocajt, V., Perić-Grujić, A. & Ristić, M. 2014 Modelling of dissolved oxygen in the Danube River using artificial neural networks and Monte Carlo Simulation uncertainty analysis. *Journal of Hydrology* 519, 1895–1907. http://doi.org/10.1016/j.jhydrol.2014.10.009.

Antanasijević, D., Pocajt, V., Antanasijević, J., Perić-Grujić, A. & Ristić, M. 2018 A novel SON^2-based similarity index and its application for the rationalization of river water quality monitoring network. *River Research and Applications* 34 (2), 144–152. http://doi.org/10.1002/rra.3231.

Barzegar, R., Asghari Moghaddam, A. & Tziritis, E. 2016 Assessing the hydrogeochemistry and water quality of the Aji-Chay River, northwest of Iran. *Environmental Earth Sciences* 75 (23), 1486. http://doi.org/10.1007/s12665-016-6302-1.

Benzer, S. & Benzer, R. 2018 Modelling nitrate prediction of groundwater and surface water using artificial neural networks. *Journal of Polytechnic* 21 (2), 321–325. http://doi. org/10.2339/politeknik.385434.

Bowden, G. J., Dandy, G. C. & Maier, H. R. 2005 Input determination for neural network models in water resources applications. Part 1 – background and methodology. *Journal of Hydrology* 301 (1–4), 75–92. http://doi.org/10.1016/ j.jhydrol.2004.06.021.

Cabaneros, S. M., Calautit, J. K. & Hughes, B. R. 2019 A review of artificial neural network models for ambient air pollution prediction. *Environmental Modelling & Software* 119, 285–304. http://doi.org/10.1016/j.envsoft. 2019.06.014.

Council Directive 2000/60/EC 2000 Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. *Official Journal of the European Parliament* L327, 1–73.

David, M. B., Gentry, L. E., Kovacic, D. A. & Smith, K. M. 1997 Nitrogen balance in and export from an agricultural watershed. *Journal of Environmental Quality* 26 (4), 1038–1048. http://doi.org/10.2134/jeq1997. 00472425002600040015x.

Davies, O. A., Abolude, D. S. & Ugwumba, A. A. A. 2008 Phytoplankton of the lower reaches of Okpoka Creek, Port Harcourt, Nigeria. *Journal of Fisheries International* 3, 83–90. Retrieved from: http://medwelljournals.com/abstract/amp. php?doi=jfish.2008.83.90.

Diamantopoulou, M. J., Antonopoulos, V. Z. & Papamichail, D. M. 2005 The use of a neural network technique for the prediction of water quality parameters of Axios River in Northern Greece. *European Water* 11/12, 55–62. Retrieved from: http://www.ewra.net/ew/pdf/EW_2005_11-12_07.pdf.

Dulo, S. O. 2008 Determination of some physico-chemical parameters of the Nairobi River, Kenya. *Journal of Applied Sciences and Environmental Management* 12 (1), 57–62. Retrieved from: www.bioline.org.br/ja.

European Commission 1991 Council directive 91/676/EEC (Nitrates directive). *Official Journal of the European Communities* L 375 (L), 1–8.

European Environment Agency (EEA) 2015 *Nutrient Trend*. Available from: https://www.eea.europa.eu/data-and-maps/ daviz/rivers-nutrient-trend#tab-googlechartid_chart_11 (retrieved August 22, 2019).

Foddis, M. L., Montisci, A., Trabelsi, F. & Uras, G. 2019 An MLP-ANN-based approach for assessing nitrate contamination. *Water Science and Technology: Water Supply* 19 (7), 1911–1917. http://doi.org/10.2166/ws.2019.066.

Grizzetti, B., Bouraoui, F. & De Marsily, G. 2008 Assessing nitrogen pressures on European surface water. *Global Biogeochemical Cycles* 22 (4), GB4023, http://doi.org/10. 1029/2007GB003085.

Haykin, S. 1994 *Neural Networks: A Comprehensive Foundation*. Macmillan College Publishing, New York, USA.

Hossain, M. F. 2019 Water pollution. In: *Sustainable Development for Mass Urbanization*, Elsevier, Amsterdam, The Netherlands, pp. 151–267. http://doi.org/10.1016/B978-0-12-817690-0.00010-5.

Hu, T., Wu, F. & Zhang, X. 2007 Rainfall–runoff modeling using principal component analysis and neural network. *Nordic Hydrology* **38** (3), 235–248. http://doi.org/10.2166/nh.2007.010.

Jackson, W. A., Asmussen, L. E., Hauser, E. W. & White, A. W. 1973 Nitrate in surface and subsurface flow from a small agricultural watershed. *Journal of Environmental Quality* **2** (4), 480–482. http://doi.org/10.2134/jeq1973.00472425000200040017x.

Kalogirou, S. A. 2003 Artificial intelligence for the modeling and control of combustion processes: a review. *Progress in Energy and Combustion Science* **29**, 515–566. http://doi.org/10.1016/S0360-1285(03)00058-3.

Li, X., Maier, H. R. & Zecchin, A. C. 2015 Improved PMI-based input variable selection approach for artificial neural network and other data driven environmental and water resource models. *Environmental Modelling & Software* **65**, 15–29. http://doi.org/10.1016/j.envsoft.2014.11.028.

Mateo-Sagasta, J., Zadeh, S. M. & Turral, H. 2017 *Water Pollution from Agriculture: A Global Review*. FAO, Rome, Italy and IWMI, Colombo, Sri Lanka. http://doi.org/http://www.fao.org/3/a-i7754e.pdf.

Mihic, S. D. & Andrejevic, A. 2012 European policy for the promotion of inland waterway transport – a case study of the Danube River. In: *Sustainable Development* (C. Ghenai, ed.), InTech, Rijeka, Croatia, pp. 23–40. Rijeka: IntechOpen. http://doi.org/10.5772/26716.

Mihic, S., Golusin, M. & Mihajlovic, M. 2011 Policy and promotion of sustainable inland waterway transport in Europe – Danube River. *Renewable and Sustainable Energy Reviews* **15** (4), 1801–1809. http://doi.org/10.1016/j.rser.2010.11.033.

Nabavi-Pelesaraei, A., Rafiee, S., Mohtasebi, S. S., Hosseinzadeh-Bandbafha, H. & Chau, K. 2019 Comprehensive model of energy, environmental impacts and economic in rice milling factories by coupling adaptive neuro-fuzzy inference system and life cycle assessment. *Journal of Cleaner Production* **217**, 742–756. http://doi.org/10.1016/j.jclepro.2019.01.228.

Najah, A., Elshafie, A., Karim, O. A. & Jaffar, O. 2009 Prediction of Johor River water quality parameters using artificial neural networks. *European Journal of Scientific Research* **28** (3), 422–435.

Najah, A., El-Shafie, A., Karim, O. A., Jaafar, O. & El-Shafie, A. H. 2011 An application of different artificial intelligences techniques for water quality prediction. *International Journal of the Physical Sciences* **6** (22), 5298–5308.

Nayak, P. C., Sudheer, K. P. & Ramasastri, K. S. 2005 Fuzzy computing based rainfall–runoff model for real time flood forecasting. *Hydrological Processes* **19** (4), 955–968. http://doi.org/10.1002/hyp.5553.

Olyaie, E., Banejad, H., Chau, K. W. & Melesse, A. M. 2015 A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river

systems: a case study in United States. *Environmental Monitoring and Assessment* **187** (4), 189. http://doi.org/10.1007/s10661-015-4381-1.

Raj Shrestha, R. & Rode, M. 2007 Modelling catchment-scale nitrate transport using a combined process-based and data-driven approach. In: *Water Quality and Sediment Behaviour of the Future: Predictions for the 21st Century* (B. W. Webb & D. De Boer, eds), IAHS Publication 314, IAHS Press, Wallingford, UK, pp. 125–133. Retrieved from: https://iahs.info/uploads/dms/14031.18-125-134-314-34-Shrestha--Rode.pdf.

Sarkar, A. & Pandey, P. 2015 River water quality modelling using artificial neural network technique. *Aquatic Procedia* **4**, 1070–1077. http://doi.org/10.1016/J.AQPRO.2015.02.135.

SEPA 2016 *Agencija za zaštitu životne sredine – Ministarstvo poljoprivrede i zaštite životne sredine*. http://www.sepa.gov.rs/index.php (retrieved January 28, 2016).

Shamshirband, S., Jafari Nodoushan, E., Adolf, J. E., Abdul Manaf, A., Mosavi, A. & Chau, K. 2019 Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll *a* concentration in coastal waters. *Engineering Applications of Computational Fluid Mechanics* **13** (1), 91–101. http://doi.org/10.1080/19942060.2018.1553742.

Šiljić, A., Antanasijević, D., Perić-Grujić, A., Ristić, M. & Pocajt, V. 2015 Artificial neural network modelling of biological oxygen demand in rivers at the national level with input selection based on Monte Carlo simulations. *Environmental Science and Pollution Research* **22** (6), 4230–4241. http://doi.org/10.1007/s11356-014-3669-y.

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. http://doi.org/10.1016/J.SCITOTENV.2017.08.192.

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. http://doi.org/10.1016/j.ecolmodel.2009.01.004.

Singh, K. P., Gupta, S., Kumar, A. & Shukla, S. P. 2012 Linear and nonlinear modeling approaches for urban air quality prediction. *Science of The Total Environment* **426**, 244–255. http://doi.org/10.1016/j.scitotenv.2012.03.076.

Stamenković, L. J., Antanasijević, D. Z., Ristić, M. Đ., Perić-Grujić, A. A. & Pocajt, V. V. 2017 Prediction of nitrogen oxides emissions at the national level based on optimized artificial neural network model. *Air Quality, Atmosphere & Health* **10** (1), 15–23. http://doi.org/10.1007/s11869-016-0403-6.

Tiyasha, Tung, T. M. & Yaseen, Z. M. 2020 A survey on river water quality modelling using artificial intelligence models: 2000–2020. *Journal of Hydrology* **585**, 124670. http://doi.org/10.1016/j.jhydrol.2020.124670.

Tran, H. D., Muttil, N. & Perera, B. J. C. 2015 Selection of significant input variables for time series forecasting. *Environmental Modelling & Software* **64**, 156–163. http://doi.org/10.1016/j.envsoft.2014.11.018.

Wang, W. C., Chau, K. W., Cheng, C. T. & Qiu, L. 2009 A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology* **374** (3–4), 294–306. http://doi.org/10.1016/j.jhydrol.2009.06.019.

Ward, M. H., Jones, R. R., Brender, J. D., de Kok, T. M., Weyer, P. J., Nolan, B. T., Villanueva, C. M. & van Breda, S. G. 2018 Drinking water nitrate and human health: an updated review. *International Journal of Environmental Research and Public Health* **15** (7), 1557. http://doi.org/10.3390/ijerph15071557.

Wu, M.-L., Wang, Y.-S., Sun, C.-C., Wang, H., Dong, J.-D., Yin, J.-P. & Han, S.-H. 2010 Identification of coastal water quality by statistical analysis methods in Daya Bay, South China Sea. *Marine Pollution Bulletin* **60** (6), 852–860. http://doi.org/10.1016/j.marpolbul.2010.01.007.

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 & Software* **54**, 108–127. http://doi.org/10.1016/j.envsoft.2013.12.016.

Xu, L. & Liu, S. 2013 Study of short-term water quality prediction model based on wavelet neural network. *Mathematical and Computer Modelling* **58** (3–4), 807–813. http://doi.org/10.1016/j.mcm.2012.12.023.