Science of the Total Environment 654 (2019) 1000–1009

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Virtual water quality monitoring at inactive monitoring sites using Monte Carlo optimized artificial neural networks: A case study of Danube River (Serbia)

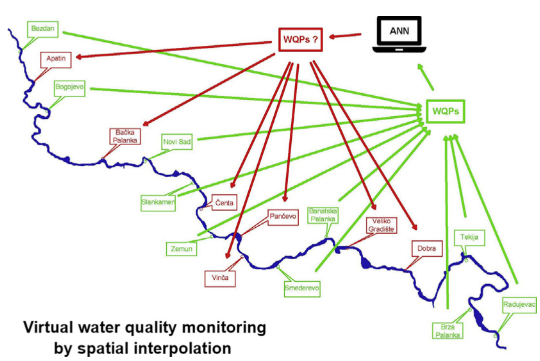
Tatjana Mitrović [a], Davor Antanasijević [b,*], Saša Lazović [c], Aleksandra Perić-Grujić [d], Mirjana Ristić [d]

[a] Jaroslav Cerni Institute for Development of Water Resources, Jaroslava Cernog 80, 11226 Belgrade, Serbia
[b] Innovation Center of the Faculty of Technology and Metallurgy, Karnegijeva 4, 11120 Belgrade, Serbia
[c] Institute of Physics Belgrade, University of Belgrade, Pregrevica 118, 11080 Belgrade, Serbia
[d] University of Belgrade, Faculty of Technology and Metallurgy, Karnegijeva 4, 11120 Belgrade, Serbia

HIGHLIGHTS

- WQPs at inactive monitoring sites were estimated by ANN models.
- The selection of input sites for spatial interpolation was done by similarity index.
- MCS routine for the selection of inputs was modified to fit simultaneous models.
- The best modeling strategy for study area was determined.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 15 August 2018
Received in revised form 6 November 2018
Accepted 13 November 2018
Available online 14 November 2018

Editor: Damia Barcelo

Keywords:
ANNs
Monte Carlo simulations
River water monitoring
Water quality prediction
Inactive monitoring sites

ABSTRACT

Rationalization of water quality monitoring stations nowadays is applied in many countries. In some cases, missing data from abandoned/inactive stations, spatial and temporal, could be very important, hence the use of artificial neural networks (ANNs) for virtual water quality monitoring at inactive monitoring sites was investigated. The aim was to develop single-output and simultaneous ANNs for the spatial interpolation of 18 water quality parameters at single- and multi-inactive monitoring sites on Danube River course through Serbia. Those different modeling approaches were considered in order to determine the most suitable combination of models. The variable selection and sensitivity analysis in the case of simultaneous models were performed using a modified procedure based on Monte Carlo Simulations (MCS). In general, the multi-target models tend to be more accurate than single target ones, while single output models outperform the simultaneous ones. Hence, for particular monitoring network and set of water quality parameters the optimal combination of models must be defined based on model's accuracy and computational effort needed. The MCS selection procedure has proved to be efficient only in the case of simultaneous multi-target model. MCS based analysis of input-output interactions has shown all significant interactions in the case of simultaneous single-target are grouped as a complex cluster of interactions, where majority of inputs influence on several outputs. In the case multi-target model those interactions were portioned in five separate clusters, there majority of them mimic the input-output interactions that are present in single output models. The modeling strategy for study area was proposed on the basis of the performance of created models (mean average percentage error < 10%): simultaneous multi-target model for pH,

* Corresponding author.
E-mail address: dantanasijevic@tmf.bg.ac.rs (D. Antanasijević).

https://doi.org/10.1016/j.scitotenv.2018.11.189

alkalinity, conductivity, hardness, dissolved oxygen, HCO_3^-, SO_4^{2-} and Ca, single-output multi-target models for temperature and Cl^-, simultaneous single-target models for Mg and CO_2, single output single target models for NO_3^-.

1. Introduction

River water quality is on one hand influenced by natural as well as anthropogenic factors, like population growth and industrialization, but on another hand it strongly determines the use of fresh water, aquatic ecosystem status, and even human health. An inevitable step in ensuring good quality of river water is monitoring. In this way, continuous collection of data on status of surface water is ensured, as well as taking measures for the elimination of potential hazards. One of the potential problems, linked to monitoring stations, is missing data, caused by different reasons.

The current tendency is to reduce the number of monitoring sites wherever possible, in order to reduce costs (Chapman et al., 2016). Typically, monitoring stations that have been observed to have similar impacts like one chosen as representative, or have a similar trend in data analysis are abolished (abandoned). In this case, studies and research were focused on number of water quality monitoring stations optimization. For example, Chapman et al. (2016) have used combined cluster and discriminant analysis (Kovács et al., 2014) to estimate the efficiency of monitoring network at Austrian and Hungarian section of the Danube River, while Antanasijević et al. (2018) have proposed self-organizing network based similarity index for the optimization of sampling locations in an existing river water quality monitoring network of the River Danube on its stretch through Serbia.

Excluding some stations from monitoring program could have significant implications in future, for different reasons: necessity of activation of some inactive station and comparison with past data; the emergence of a new source of pollution near inactive monitoring station, indicating serious pollution, the occurrence of invasive species, or their extinction.

Regarding the water quality parameters (WQPs) prediction, artificial neural networks (ANNs) (Peleato et al., 2018) have been successfully applied for the estimation of temperature (Sahoo et al., 2009), chloride (Salami and Ehteshami, 2015; Barzegar et al., 2016), fluoride (Barzegar et al., 2017), electrical conductivity (Barzegar et al., 2018), alkalinity (Salami and Ehteshami, 2015), total hardness (Salami and Ehteshami, 2015), salinity (Huang and Foo, 2002; Salami Shahid and Ehteshami, 2016; Barzegar and Moghaddam, 2016), total dissolved solids (Salami et al., 2016), sodium adsorption ratio (Salami et al., 2016), ammonia nitrogen (Wang et al., 2013), bicarbonate (Salami et al., 2016), chemical and biological oxygen demand (COD and BOD) (Ay and Kisi, 2014; Dogan et al., 2009; Salami et al., 2016; Salami Shahid and Ehteshami, 2016; Verma and Singh, 2013), dissolved oxygen (DO) (Antanasijević et al., 2014; Keshtegar and Heddam, 2017; Salami et al., 2016; Salami and Ehteshami, 2015; Salami Shahid and Ehteshami, 2016; Wang et al., 2013), DO percentage (Salami and Ehteshami, 2015), etc.

An ANN can be described as an information process system which consists of many nonlinear and densely interconnected processing units. With this parallel-distributed processing architecture, ANNs have been proven to be an efficient alternative to traditional methods for hydrological modeling (Chang et al., 2007). As Rigol et al. (2001) have noted, the advantage of ANNs for spatial interpolation is that the input variables are not assumed necessarily to be linearly related with the data being interpolated, and that combinative effects are taken into account during modeling. ANNs have been effectively applied for various spatial interpolation tasks, e.g. surface air temperatures (Li et al., 2004; Snell et al., 2000), solar radiation (Li et al., 2004), wind speed (Philippopoulos and Deligiorgi, 2012), soil salinity (Shahabi et al., 2017) etc. Even an integrated ANN-kriging approach was recently proposed for spatial prediction of saline and sodic soils in rice–shrimp farming land (Dinh et al., 2017).

The focus of this study is on a monitoring network of Danube River course through Serbia, namely on the prediction of common water quality parameters (WQPs) at monitoring sites (MSs) that have become non-operational (inactive) since 2012 network rationalization (Antanasijević et al., 2018). This virtual monitoring can be performed by spatial interpolation using measured WQPs data from neighboring MSs. For this task, ANNs was selected, since simplicity and robustness of their application are more important than an accurate description of the various internal sub-processes (Lima et al., 2016). Also, it was proven that ANNs can provide similar or better performance models in comparison with alternative techniques, such as various kriging methods (universal, ordinary, etc.) or partial thin plate splines (Rigol, 2003).

The novelty of this work lies is the fact that single output and simultaneous ANNs for the spatial interpolation of common 18 WQPs at single inactive MS on Danube River course through Serbia, as well as at multi inactive MSs were developed by the variable selection and sensitivity analysis performed using procedure based on Monte Carlo Simulations (MCS) that was previously applied for single output ANN models (Gao et al., 2018; Šiljić et al., 2015), and which is modified in this work to fit simultaneous models.

2. Materials and methods

2.1. Study area and water quality data

The Danube River with its length of 2857 km (588 km through Serbia) is the second largest watercourse in Europe and thus very important International River. About 16% of its total drainage basin (817,000 km^2) belongs to the Serbian territory. It is the main source for domestic and industrial water supply and irrigation in Serbia. In addition, it serves as an international waterway and as receiving waters for wastewater effluents. The major municipal pollution sources come from the cities of Belgrade (1.7 million inhabitants) and Novi Sad (300,000 inhabitants), which do not have satisfying wastewater purification treatment plants. These untreated wastewaters, which are discharged into the river directly, are sources of significant organic and nutrient pollution (The International Commission for the Protection of the Danube River (ICPDR), 2018).

Two hydroelectric power plants (Iron Gate I and II) with their reservoir systems are located at the Serbian territory. These reservoirs change the Danube regime and trap millions of tons of sediment per year which are considerable deposit for nutrients and hazardous pollutants originating upstream of the dam. As a result, the water residence time and temperature increase thermal stratification changes, primary production in situ enhance etc. Considering these facts, there is a big impact of the dams on the river aquatic life as well as on the environment at all (Mitrović et al., 2010).

The dataset used in this study was generated through monitoring of the water quality of Danube River (Serbia). Monthly and semi-monthly

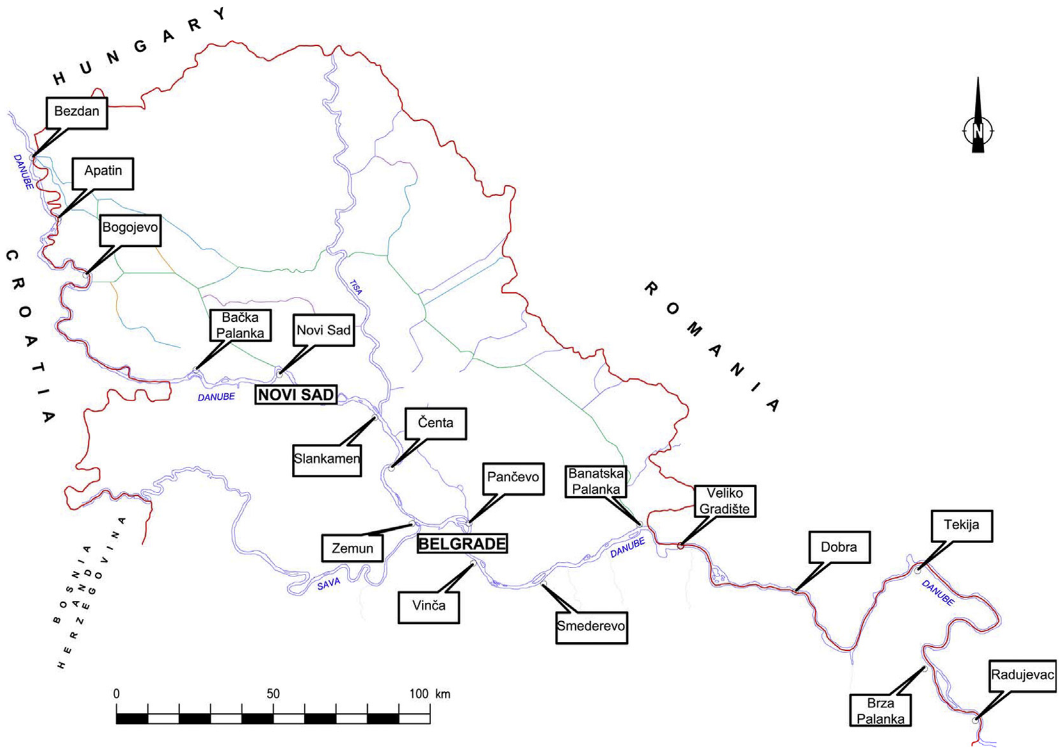


Fig. 1. Danube River course through The Republic of Serbia.

sampling was carried out during ten years (2002–2011) at 17 monitoring sites (Fig. 1). A dataset consisted of 18 WQPs:

a) 8 non-specific WQPs, namely temperature (T), pH, total suspended solids (TSS), hardness, alkalinity, electrical conductivity, biological oxygen demand (BOD), chemical oxygen demand (COD), and
b) 10 specific WQPs, whereby

 i. two gaseous WQPs, i.e. dissolved oxygen (DO) and CO_2,
 ii. two cations, i.e. Ca and Mg,
 iii. five anions, namely HCO_3^-, NO_3^-, PO_4^{3-}, Cl^-, SO_4^{2-}, and
 iv. total phosphorus (P).

For model generation, available dataset was split into three sub-sets: training, validation and testing in ratio 8:1:1, respectively.

2.2. Modeling approaches

If a monitoring network of few dozen monitoring sites is considered, then its rationalization will yield at least several inactive monitoring sites. This allows the development of two types of ANN prediction models (Fig. 2), which have been already compared in studies related to the modeling of water quality monitoring data (e.g. (Nevers and Whitman, 2011)):

1) single-target (ST) models, i.e. for each inactive monitoring site (target) separate model can be created, and
2) multi-target (MT) model, i.e. data for all inactive monitoring sites are combined to create a single prediction model. The MT model discussed in this section should not be confused with multi-output regression (Borchani et al., 2015) that is often labeled in the same way.

Although ST models are more frequently used, MT models can be suitable alternative because of the obvious reduction of computational

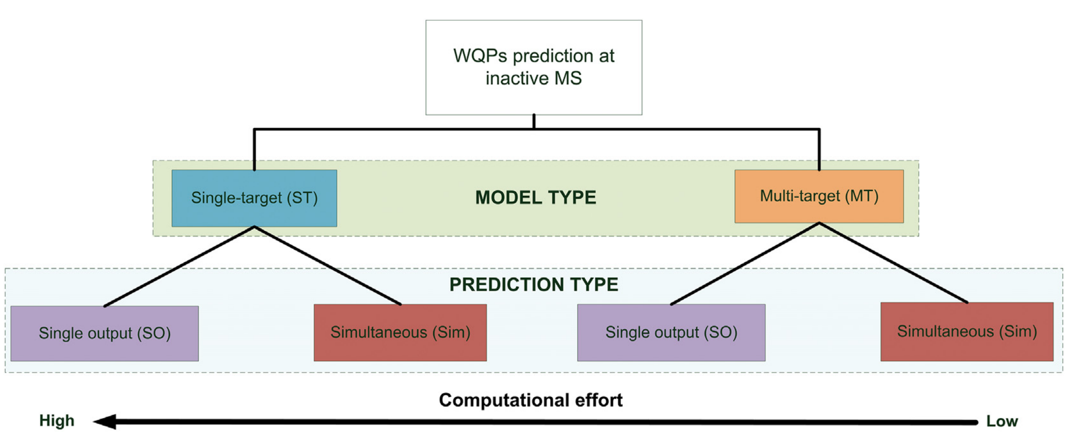


Fig. 2. Modeling approaches.

cost. Also a MT model is smaller than the total size of the ST models and it explains dependencies between different targets (Kocev et al., 2009).

Further, ANN models can be produced to be single output (SO) and simultaneous (Sim), which defines not only the computational efficiency, but also their accuracy, since simultaneous model is obtained by a compromising optimization of all outputs, hence there is no guarantee that minimum error for each output is reached (Chang et al., 2007). While SO ANN models are the most commonly used, several studies have shown that ANNs can simultaneously predict several outputs with desired accuracy, e.g. simultaneous prediction of: five traffic-related pollutants at the national level (Antanasijević et al., 2017), seven meteorological parameters in a weather station (Raza and Jothiprakash, 2014), physical and chemical properties prediction (Ghaedi, 2015), as well as multi-output time series forecasting of electricity prices (Gareta et al., 2006) and demand (An et al., 2013).

The issue related to the computational effort, which can be expressed in the number of models, is more pronounced in the current study, concerning that the number of WQPs that should be predicted for a single inactive monitoring site can be very large (≥20). As can be observed in Fig. 2a, the creation of theoretically most accurate SO-ST models actually means that up to several hundred models are needed to cover each inactive monitoring site (MS), since its number depends both on the number of inactive MS, as well as on the number of WQPs that should be predicted. Therefore, other modeling approaches appear to be more practical since they demand the creation of model(s) which number ranges from 1 to up to the number of WQPs that are subject of prediction (Fig. S1 in the Supplement). Regarding that their accuracy is questionable, it should be empirically verified and benchmarked using SO-ST models, which is one of the aims of current study.

2.3. ANN modeling

ANNs are data-driven methods capable to fit highly nonlinear relations between several input and output variables, which is typically achieved by training performed using iterative algorithms. They are consisted of layers of neurons, usually three, and their training implies (i) random initialization of the weights, (ii) iterative adjustment of the weights, and (iii) the determination of the optimal value of weights based on external criterion, e.g. sum-of-squares error or mean squared error.

In this study, three-layered feed-forward neural network was used for the creation of prediction models. The BFGS algorithm, a quasi-Newton iterative method proposed independently by Broyden–Fletcher–Goldfarb–Shanno (Borsato et al., 2011; Zounemat-Kermani et al., 2016), was used for ANN training, since it is regarded as one of the most powerful methods to solve unconstrained optimization problem (Dai, 2013). Although, BFGS has high memory requirements, due to storage of the Hessian matrix, its fast convergence makes it more efficient in comparison with standard back-propagation algorithm (Nawi et al., 2006). Different types of activation functions (Identity, Logistic sigmoid, Hyperbolic tangent and Exponential) were tested in the hidden and output layers to achieve the best model setup. The complexity of models was determined empirically by testing models with predefined lower and upper limits of hidden neurons. Overtraining has been prevented by stopping the network training at the point where errors for the validation set started to increase. The generalization capability of final models is evaluated on their performance on testing set. All ANN models were generated using STATISTICA Automated Neural Networks module (TIBCO Software Inc., 2017).

As it was already stated, two types of models were constructed: single- and multi-target (Fig. 3). The ST models are generated by coupling historical monitoring data from three monitoring sites: upper neighboring active monitoring site (NAMS) and down NAMS, where both serve as “input” sites, and inactive monitoring site. The selection of representing input sites is discussed in Section 2.5.

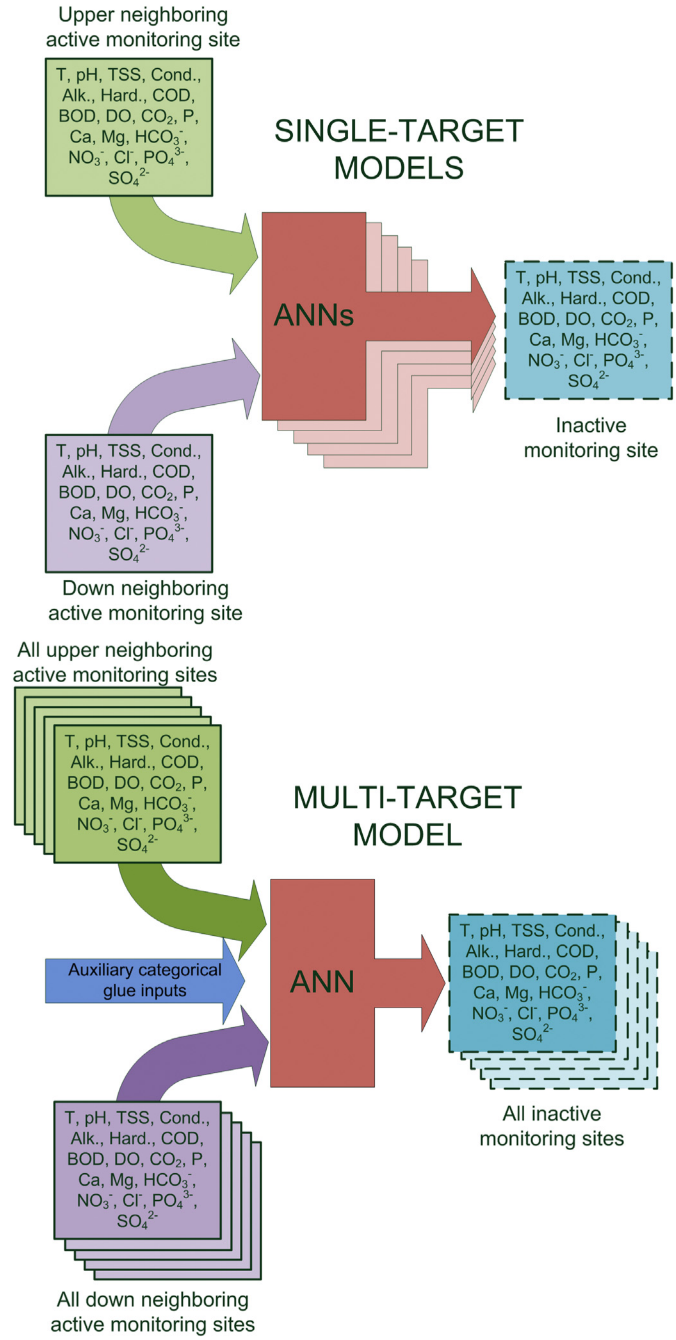


Fig. 3. Single- and multi-target ANN models.

Since seven monitoring sites are not operational on Danube section through Serbia, based on the principles of data fusion (Fosdick et al., 2016), the monitoring data for all those sites are gathered in a single MT model using auxiliary (categorical) “glue” inputs. In this case, three glue inputs are needed, each labeling one of sites that are coupled: upper glue variable (UGV) that marks upper NAMS, down glue variable (DGV) that marks down NAMS, and inactive glue variable (IGV) that is related to the inactive MS.

In the case of single output models, only corresponding WQP from upper and down NAMS are used as inputs, e.g. temperature (T) at inactive MS is predicted using only T measure at upper and down NAMS. In the case of simultaneous models, all available input WQPs are used for the prediction of corresponding WQPs at inactive MS. But concerning that the selection of the best subset of measured input variables is vital for the performance of an ANN model, additional (optimized) simultaneous models are created based on MCS input selection procedure (Šiljić et al., 2015). The reduction of the number of inputs should reduce the number of free parameters in the model, hence improving its generalization and computational efficiency (Fernando et al., 2005). This particular selection procedure was used since as a model-based approach provide the real influence of inputs on the output results, and also it allows analysis for all outputs in a single run of initial model.

2.4. MCS input selection for simultaneous models

MCS input selection procedure has been previously applied for the selection of the best subset of inputs for single output model, yielding model with better performance which has used 25% less inputs (Šiljić et al., 2015). This procedure comprises of several steps:

1. estimation of probability density functions (PDFs) for each input,
2. selection of the most significant PDF based on the Kolmogorov-Smirnov non-parametric test,
3. re-sampling of inputs according to the selected PDFs,
4. construction of MCS dataset that contains blocks of n patterns per input, where each block had one input with re-sampled values, while other inputs were set to the median of measured values,
5. application of initial ANN model to MCS dataset,
6. quantification of a specific input significance based on difference between maximum and minimum predicted output value (Δ) for each block in the MCS dataset,
7. generation of new ANN model with selected inputs and it's evaluation, and
8. sensitivity analysis based on Δ values obtained for final model.

In the case of simultaneous models, additional operations in step 6 are required:

6.1. comparison of particular input significance for different output variables based on normalized Δ (Δ_{norm})
6.2. selection of most significant inputs for each output variable based on predefined number of significance levels (e.g. 1st and 2nd) and/or Δ_{norm} threshold (e.g. $\Delta_{norm} \geq 0.90$).

Application of this procedure is presented in details in Section 3.1.

2.5. Created models

The selection of the pair of active MSs, which WQPs data were used for the prediction of WQPs at particular inactive MS, was performed using two criterions: geographical position and statistical similarity of measured WQPs at target and input sites.

In the first step, active MS are classified into upper and down MS based on their geographical position relative to the particular inactive MS. Namely, all active MS located in the section presiding the target inactive MS are labeled as upper MS, while others are labeled as down MS. This is done under the assumption that MS at Danube river entry and exit point will remain active after rationalization of monitoring network.

In the second step, the one active MS that have highest WQPs pattern similarity with target MS, from both groups, is selected and used for the creation of model. In the current study, similarity between two MSs is determined using location similarity index (LSI) (Antanasijević et al., 2017), which is based on the self-organizing network classification. The LSI values ranges from 0 to 100%, where higher value indicate higher similarity. The LSI for studied monitoring network was published in our previous study (Antanasijević et al., 2017).

The selected pairs of active MSs for each of seven inactive MS with corresponding LSI values are presented in Table 1. As it can be observed,

Table 1
Combination of MS used for model creation.

Inactive MS	IGV[a]	LSI (upper NAMS)	UGV[b]	LSI (down NAMS)	DGV[c]
Apatin	2	89% (Bezdan)	1	95% (Bogojevo)	3
Bačka Palanka	4	88% (Bezdan)	1	84% (Novi Sad)	5
Čenta	7	95% (Slankamen)	6	73% (Banatska Palanka)	12
Pančevo	9	91% (Slankamen)	6	74% (Banatska Palanka)	12
Vinča	10	81% (Zemun)	8	87% (Smederevo)	11
Veliko Gradište	13	81% (Smederevo)	11	90% (Brza Palanka)	16
Dobra	14	83% (Smederevo)	11	91% (Tekija)	15

[a] Upper glue variable.
[b] Down glue variable.
[c] Inactive glue variable.

the LSI values in the majority of cases were higher than 80%, indicating high patterns similarity, and supporting the creation of ANN models with satisfactory performance.

To reduce the computational effort needed, the SO-ST and Sim-ST models were created only for one inactive MS, i.e. Apatin (Table S1 in the Supplement). Concerning that 18 WQPs are measured at each site, 18 separate SO- and two Sim-ST (initial and MCS optimized) were created and evaluated.

In the case of MT models, monitoring data for all inactive MS are combined yielding dataset of 466 patterns (Table S2 in the Supplement). Again, 18 separate SO- and two Sim-MT were created and evaluated.

3. Results and discussion

3.1. Optimization and performance of models

In case of single target modeling, i.e. prediction of WQPs at Apatin using Bezdan and Bogojevo data, the models selected based on error obtained on validation data with their topology and testing performance are presented in Table S3 in the Supplement, while performance metrics are defined in Table 2. It can be observed that single output models (SO-ST) were superior (overall NSE = 0.64) in comparison with the initial simultaneous (Sim-ST) model (overall NSE = 0.41), as well as that this simultaneous model is a good starting point for further optimization, concerning that 2/3 of WQPs are predicted with satisfactory accuracy (Fig. 4), according to NSE ratings (Table 2).

In further step, MCS procedure for the selection of inputs was applied. Table S4 in the Supplement shows the PDFs obtained for 36 inputs used in Sim-ST model and the values of Kolmogorov-Smirnov test, which was used for the selection of the best fitting PDF, based on its significance (p). The examples of fitted PDFs are presented in Fig. S2 in the Supplement. The MCS dataset was assembled of blocks of inputs with 100 re-sampled values.

The Δ values obtained for the each output of Sim-ST model are presented in the Supplement (see Table S5). In addition, the Δ_{norm} values were calculated by scaling Δ values in the range 0 to 1 (see Table S6 in the Supplement), in order to allow the comparison of the significance of each input-output combination. Finally, the selection of most significant inputs for each output variable, based on 1st and 2nd significance level and Δ_{norm} threshold of 0.90, was performed (Fig. 5a). Hence, the number of inputs has been reduced for 50%, from 36 to only 18, and new simultaneous ST model (labeled as MCS-Sim-ST) was created. Its testing performance is given in Table S3, and it can be noted that prediction result has not been enhanced by MCS, regarding the low NSE (Fig. 4) and overall NSE values (0.30).

It can be concluded that limited number of data point impedes the development of accurate simultaneous single target model. It seems that only possibility for the accurate simultaneous prediction at single location is the reduction of the number of outputs that are

Table 2
Performance metrics with guides on their values, where n is the number of cases and Y_o, Y_p and Y_m are observed, predicted and mean observed output values, respectively.

Metrics	Calculation	Ratings	Short description (Moriasi et al. 2007)
Coefficient of determination (R^2)		Acceptable $R^2 > 0.50$	It describes the proportion of the observed variance explained by the model. R^2 ranges from 0 to 1, with higher values indicating less error variance.
Mean absolute error (MAE)	$\frac{1}{n}\sum\lvert Y_o - Y_p\rvert$		RMSE and MAE are frequently reported because they indicate error in the unit of outputs. Their values of 0 indicate a perfect fit.
Root mean squared error (RMSE)	$\sqrt{\frac{1}{n}\sum(Y_o - Y_p)^2}$		
Nash-Sutcliffe efficiency (NSE)	$1 - \frac{\sum(Y_o - Y_p)^2}{\sum(Y_o - Y_m)^2}$	Very good[a] $0.75 < NSE \le 1$ Good $0.65 < NSE \le 0.75$ Satisfactory $0.50 < NSE \le 0.65$ Unsatisfactory $NSE \le 0.50$	Normalized metrics that determines the relative magnitude of the residual variance (noise) compared to the measured data variance (information). It ranges between - ∞ and 1.0, where 1 is the optimal value, while values < 0.0 indicates that the mean observed value is a better predictor than the model.
Mean absolute percentage error (MAPE)	$\frac{100\%}{n}\sum\frac{\lvert Y_o - Y_p\rvert}{Y_o}$	Highly accurate[b] $MAPE \le 10\%$ Good $10 < MAPE \le 20\%$ Reasonable $20 < MAPE \le 50\%$ Inaccurate $MAPE > 50\%$	Frequently used metrics that gives overall relative error, where low values are preferable.

[a]Indicative ratings recommended for a monthly time step (Moriasi et al., 2007).
[b]According to Lewis interpretation the MAPE results (Pao et al., 2012).

simultaneously modeled, and finding their most efficient combination, which is out of the scope of this study.

xThe selected multi-target models with their topology and testing performance are presented in Table S7 in the Supplement. Again, the single output models (SO-MT) were superior (overall NSE = 0.68) in comparison with the initial simultaneous (Sim-MT) model (overall NSE = 0.48). Also, those multi-target models have shown better performance in comparison with single target ones (Fig. 4). This confirms the benefits of data fusion that yields significantly higher number of data patterns available for model development.

Data related to the optimization of Sim-MT models are presented in Table S4 in the Supplement (the best fitting PDFs), Table S8 in the Supplement (the obtained Δ values), Table S9 in the Supplement (the Δ_{norm} values) and Fig. 5b (highly significant inputs with their interactions).

The MCS-Sim-MT has been created with only 19 inputs (plus three glue ones) which makes a reduction of 44% in comparison with the

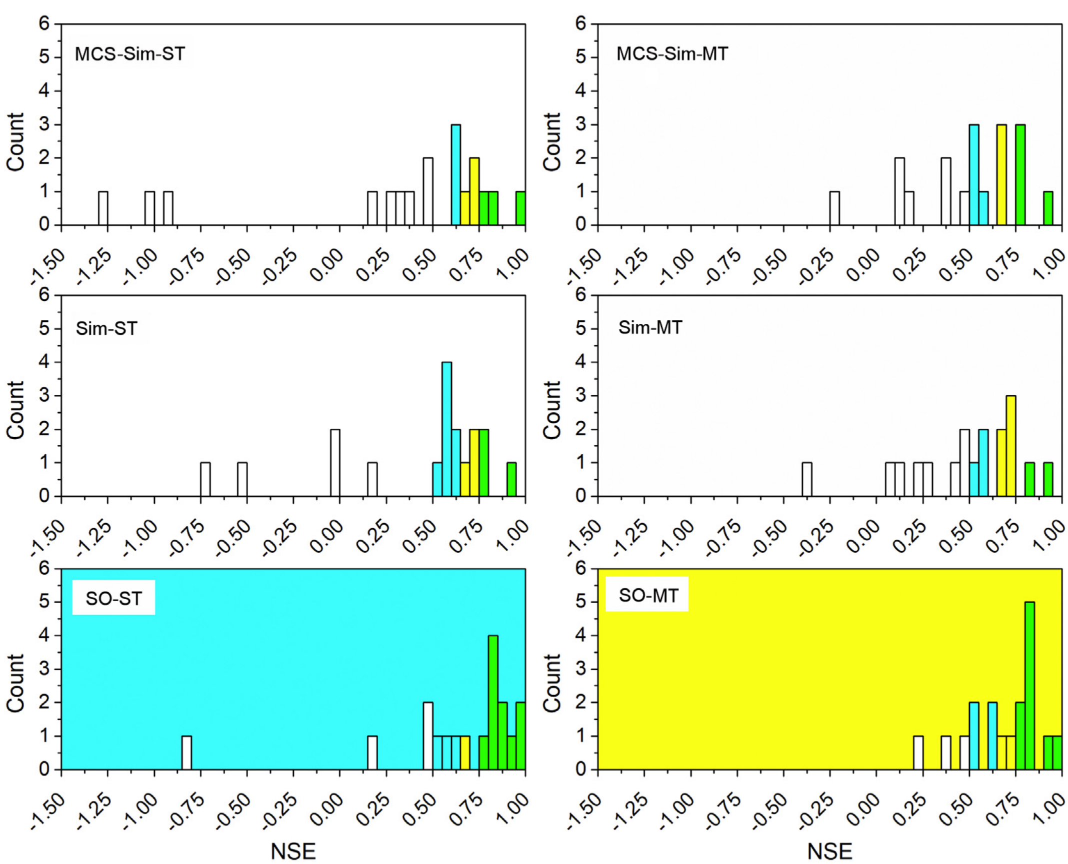


Fig. 4. Histograms of NSE values (for colors see Table 2).

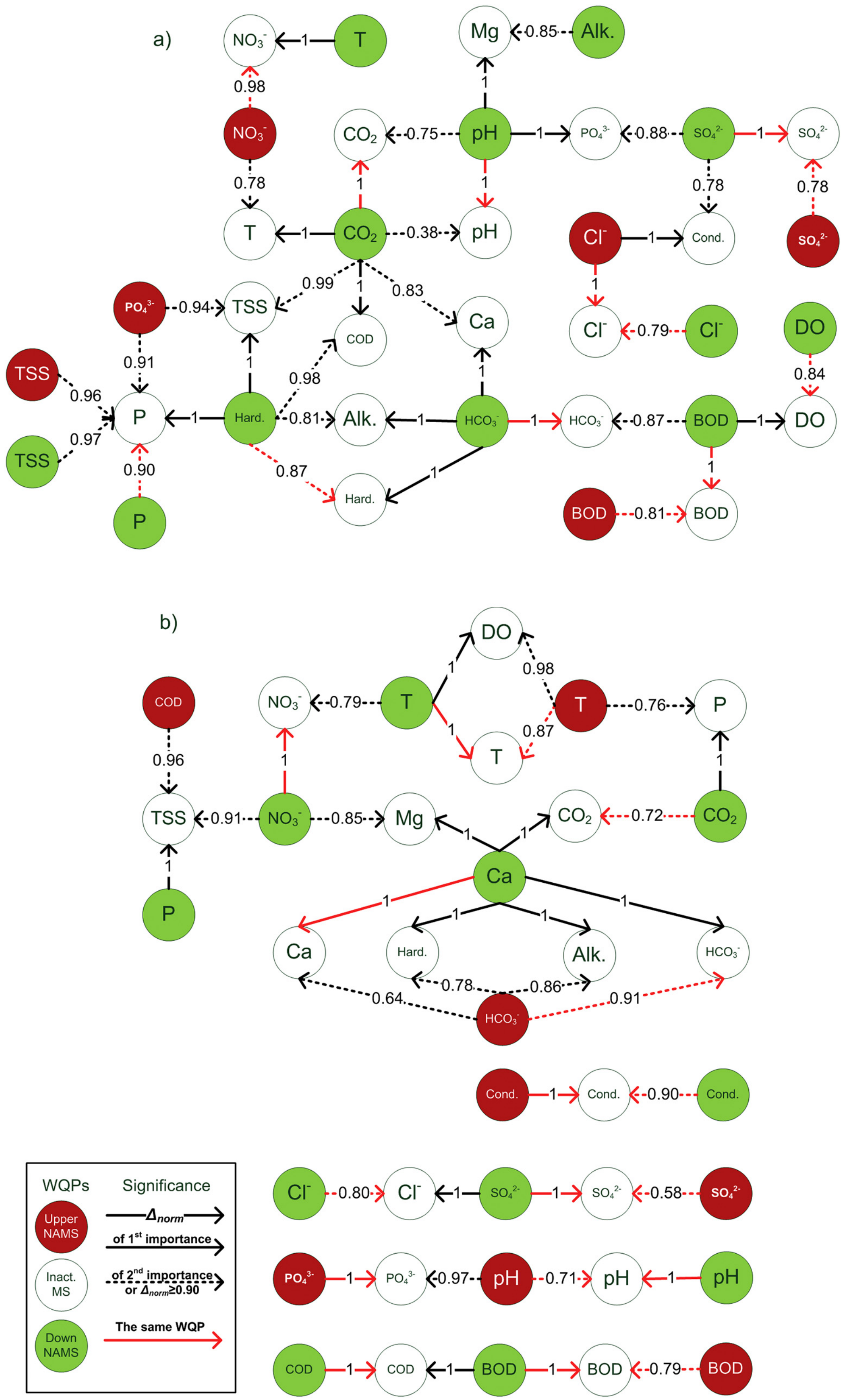


Fig. 5. Highly significant interaction between WQPs for a) Sim-ST and b) Sim-MT.

initial Sim-MT model. The optimized model performance was slightly better with overall NSE = 0.50, which is a threshold for satisfactory performance (Table 2), but still, it had poor performance in comparison with corresponding single output models.

3.2. Overview of input-output interactions

From the above results it can be observed that best performance is obtained when spatial interpolation of particular WQP was performed

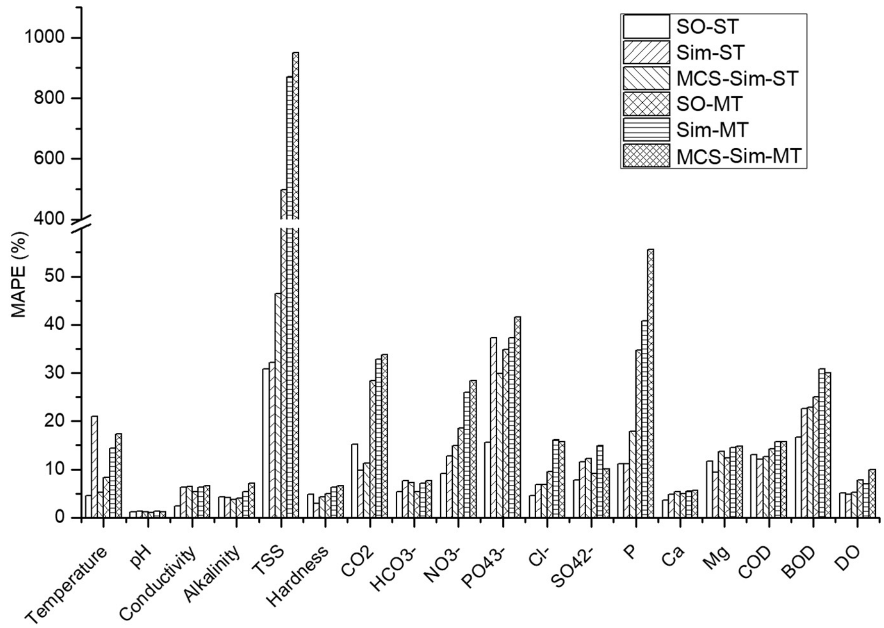


Fig. 6. MAPE depending modeling approach.

with the same WQP measured at similar neighboring locations. This is the case in all single target models, where by default input-output interactions at 1st and 2nd significance level are made by the same WQPs. In the case of simultaneous single-target model (Fig. 5a) one can be observed that BOD, Cl^- and SO_4^{2-} were only WQPs determined at 1st and 2nd significance level determined only by the same input WQPs, while in the case of pH, CO_2, P, DO, Hardness, NO_3^- and HCO_3^- one of the 1st or 2nd significance level was determined by the same WQP. Moreover, all significant interactions in this case present the complex cluster of interactions, where the majority of inputs has the influence on several outputs. In order to quantify the suitability of input-output interactions, to each input-output interaction made by the same WQPs, the weight 1 was given, while the others had 0. For the single output models, the overall suitability index has value that equals two times the number of inputs, i.e. 36, while Sim-ST has index value of only 13.

In the case of simultaneous multi-target model this suitability index is higher (17), and it should be noted that Sim-MT (Table S3) had better performance than Sim-ST (Table S6). From Fig. 5b it can be noted that in the case of Sim-MT(i) T, Conductivity, pH, BOD, SO_4^{2-}were determined only by the same WQPs at the 1st and 2nd significance level, while (ii) COD, CO_2, NO_3^-, Ca, HCO_3^-, Cl^-, PO_4^{3-} were determined by the same WQP at one of the 1st or 2nd significance level. More important, the single cluster of interactions that was observed in the case of single-target model, was portioned in five separate clusters (Fig. 5b) and majority of them mimic the input-output interactions that are present in single output models.

Table 3
Proposed modeling strategy for study area.

Relative error	Model type	Parameters	Number of models
Single-target			
MAPE < 10%	Single output	NO_3^-	7
	Simultaneous	Mg, CO_2	7
Multi-target			
MAPE < 10%	Single output	Temperature, Cl^-	2
	Simultaneous	pH, Alkalinity, Conductivity, Hardness, DO, HCO_3^-, SO_4^{2-}, Ca	1
10% < MAPE < 30% (SO-ST)		TSS, COD, BOD, P, PO_4^{3-}	

3.3. Modeling strategy for studied area

In order to resolve two (opposite) goals, i.e. high accuracy and low computational cost, suitable modeling strategy for studied area has been determined based on MAPE for each modeled WQP. The aim was to define a set of models that will give highly accurate predictions (MAPE ≤ 10%) with lowest possible computational cost. After MAPE for each modeling approach and WQP was assessed (Fig. 6), the combination of models was determined (Table 3). These results indicate that TSS, COD, BOD, P and PO_4^{3-} cannot be predicted with such high accuracy, while the concentration of most other WQPs, all except NO_3^-, Mg, CO_2, can be obtained using multi-target models.

4. Conclusion

The prediction of 18 common water quality parameters (WQPs) on inactive monitoring station/stations on the Danube River at the territory of Republic of Serbia was performed by developing and testing of single output and simultaneous artificial neural network (ANNs) for spatial interpolation of these WQPs at single- and multi-inactive monitoring sites. Monthly and semimonthly data collected during ten years (2002–2011) were used for models development and testing. Monte Carlo Simulation was applied for the variable selection and the analysis of input-output interactions.

Results presented in this study have shown that single output models have outperformed simultaneous ones in the case of majority WQPs. The benefit of data fusion in the case of multi-target models has been observed, concerning that overall Nash-Sutcliffe efficiency (NSE) has increased in comparison with single target models, i.e. from 0.64 to 0.68 in case of single output models and from 0.41 to 0.48 in the case of simultaneous ones. Also, only in the case of simultaneous multi-target model the Monte Carlo Simulation based selection of inputs was successful in providing model with enhanced performance.

After the performance of all type of ANN models had been analyzed, it was determined that the majority of studied WQPs (13/18) were predicted with relative error <10%, which makes the virtual monitoring highly accurate alternative to the field measurements of those WQPs. Moreover, ¾ of studied WQPs can be predicted with desired accuracy using multi-target models which significantly reduce the computational effort and time.

Acknowledgements

The authors are grateful to the Ministry of Education, Science and Technological Development of the Republic of Serbia for financial support (project number 172007).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2018.11.189.

References

An, N., Zhao, W., Wang, J., Shang, D., Zhao, E., 2013. Using multi-output feedforward neural network with empirical mode decomposition based signal filtering for electricity demand forecasting. Energy 49, 279–288. https://doi.org/10.1016/j.energy.2012.10.035.

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. J. Hydrol. 519, 1895–1907. https://doi.org/10.1016/j.jhydrol.2014.10.009.

Antanasijević, D., Pocajt, V., Perić-Grujić, A., Ristić, M., 2017. Multiple-input-multiple-output general regression neural networks model for the simultaneous estimation of traffic-related air pollutant emissions. Atmos. Pollut. Res. 9, 388–397. https://doi.org/10.1016/j.apr.2017.10.011.

Antanasijević, D., Pocajt, V., Antanasijević, J., Perić-Grujić, A., Ristić, M., 2018. A novel SON2-based similarity index and its application for the rationalization of river water quality monitoring network. River Res. Appl. 34, 144–152. https://doi.org/10.1002/rra.3231.

Ay, M., Kisi, O., 2014. Modelling of chemical oxygen demand by using ANNs, ANFIS and k-means clustering techniques. J. Hydrol. 511, 279–289. https://doi.org/10.1016/j.jhydrol.2014.01.054.

Barzegar, R., Moghaddam, A.A., 2016. Combining the advantages of neural networks using the concept of committee machine in the groundwater salinity prediction. Model. Earth Syst. Environ. 2, 26.

Barzegar, R., Adamowski, J., Moghaddam, A.A., 2016. Application of wavelet-artificial intelligence hybrid models for water quality prediction: a case study in Aji-Chay River, Iran. Stoch. Env. Res. Risk A. 30, 1797–1819.

Barzegar, R., Moghaddam, A.A., Adamowski, J., Fijani, E., 2017. Comparison of machine learning models for predicting fluoride contamination in groundwater. Stoch. Env. Res. Risk A. 31, 2705–2718.

Barzegar, R., Moghaddam, A.A., Adamowski, J., Ozga-Zielinski, B., 2018. Multi-step water quality forecasting using a boosting ensemble multi-wavelet extreme learning machine model. Stoch. Env. Res. Risk A. 32, 799–813.

Borchani, H., Varando, G., Bielza, C., Larrañaga, P., 2015. A survey on multi-output regression. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 5, 216–233. https://doi.org/10.1002/widm.1157.

Borsato, D., Pina, M.V.R., Spacino, K.R., dos Santos Scholz, M.B., Filho, A.A., 2011. Application of artificial neural networks in the geographical identification of coffee samples. Eur. Food Res. Technol. 233, 533–543. https://doi.org/10.1007/s00217-011-1548-z.

Chang, F.J., Chiang, Y.M., Chang, L.C., 2007. Multi-step-ahead neural networks for flood forecasting. Hydrol. Sci. J. 52, 114–130. https://doi.org/10.1623/hysj.52.1.114.

Chapman, D.V., Bradley, C., Gettel, G.M., Hatvani, I.G., Hein, T., Kovács, J., Liska, I., Oliver, D.M., Tanos, P., Trásy, B., Várbíró, G., 2016. Developments in water quality monitoring and management in large river catchments using the Danube River as an example. Environ. Sci. Pol. 64, 141–154. https://doi.org/10.1016/j.envsci.2016.06.015.

Dai, Y.-H., 2013. A perfect example for the BFGS method. Math. Program. 138, 501–530. https://doi.org/10.1007/s10107-012-0522-2.

Dinh, Q.T., Liang, D., Thi Anh Thu, T., Le, T.D.H., Dinh Vuong, N., Pham, V.T., 2017. Spatial prediction of saline and sodic soils in rice–shrimp farming land by using integrated artificial neural network/regression model and kriging. Arch. Agron. Soil Sci. https://doi.org/10.1080/03650340.2017.1352088.

Dogan, E., Sengorur, B., Koklu, R., 2009. Modeling biological oxygen demand of the Melen River in Turkey using an artificial neural network technique. J. Environ. Manag. 90, 1229–1235. https://doi.org/10.1016/j.jenvman.2008.06.004.

Fernando, T.M.K.G., Maier, H.R., Dandy, G.C., May, R., 2005. Efficient selection of inputs for artificial neural network models. In: Zerger, A., Argent, R.M. (Eds.), Modsim 2005: International Congress on Modelling and Simulation: Advances and Applications for Management and Decision Making: Advances and Applications for Management and Decision Making. MSSANZ, Melbourne, Victoria, pp. 1806–1812.

Fosdick, B.K., Deyoreo, M., Reiter, J.P., 2016. Categorical data fusion using auxiliary information. Ann. Appl. Stat. 10, 1907–1929.

Gao, M., Yin, L., Ning, J., 2018. Artificial neural network model for ozone concentration estimation and Monte Carlo analysis. Atmos. Environ. https://doi.org/10.1016/j.atmosenv.2018.03.027.

Gareta, R., Romeo, L.M., Gil, A., 2006. Forecasting of electricity prices with neural networks. Energy Convers. Manag. 47, 1770–1778. https://doi.org/10.1016/j.enconman.2005.10.010.

Ghaedi, A., 2015. Simultaneous prediction of the thermodynamic properties of aqueous solution of ethylene glycol monoethyl ether using artificial neural network. J. Mol. Liq. 207, 327–333. https://doi.org/10.1016/j.molliq.2015.04.015.

Huang, W., Foo, S., 2002. Neural network modeling of salinity variation in Apalachicola River. Water Res. 36, 356–362. https://doi.org/10.1016/S0043-1354(01)00195-6.

Keshtegar, B., Heddam, S., 2017. Modeling daily dissolved oxygen concentration using modified response surface method and artificial neural network: a comparative study. Neural Comput. Appl., 1–12 https://doi.org/10.1007/s00521-017-2917-8.

Kocev, D., Džeroski, S., White, M.D., Newell, G.R., Griffioen, P., 2009. Using single- and multi-target regression trees and ensembles to model a compound index of vegetation condition. Ecol. Model. 220, 1159–1168. https://doi.org/10.1016/j.ecolmodel.2009.01.037.

Kovács, J., Kovács, S., Magyar, N., Tanos, P., Hatvani, I.G., Anda, A., 2014. Classification into homogeneous groups using combined cluster and discriminant analysis. Environ. Model. Softw. 57, 52–59. https://doi.org/10.1016/j.envsoft.2014.01.010.

Li, B., McClendon, R.W., Hoogenboom, G., 2004. Spatial interpolation of weather variables for single locations using artificial neural networks. Trans. ASAE 47, 629–637.

Lima, A.R., Cannon, A.J., Hsieh, W.W., 2016. Forecasting daily streamflow using online sequential extreme learning machines. J. Hydrol. 537, 431–443. https://doi.org/10.1016/j.jhydrol.2016.03.017.

Mitrović, T., Obradović, V., Golobočanin, D., Ogrinc, N., Miljević, N., 2010. Spatial and temporal variability of stable isotopes and biological parameters for the Danube River in Serbia. Isot. Environ. Health Stud. 46, 166–179.

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50, 885–900.

Nawi, N.M., Ransing, M.R., Ransing, R.S., 2006. An improved learning algorithm based on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method for back propagation neural networks. Proceedings of the Sixth International Conference on Intelligent Systems Design and Applications (ISDA'06). IEEE, Jinan, China, pp. 152–157 https://doi.org/10.1109/ISDA.2006.95.

Nevers, M.B., Whitman, R.L., 2011. Efficacy of monitoring and empirical predictive modeling at improving public health protection at Chicago beaches. Water Res. 45, 1659–1668. https://doi.org/10.1016/j.watres.2010.12.010.

Pao, H.-T., Fu, H.-C., Tseng, C.-L., 2012. Forecasting of CO_2 emissions, energy consumption and economic growth in China using an improved grey model. Energy 40, 400–409. https://doi.org/10.1016/j.energy.2012.01.037.

Peleato, N.M., Legge, R.L., Andrews, R.C., 2018. Neural networks for dimensionality reduction of fluorescence spectra and prediction of drinking water disinfection by-products. Water Res. 136, 84–94. https://doi.org/10.1016/j.watres.2018.02.052.

Philippopoulos, K., Deligiorgi, D., 2012. Application of artificial neural networks for the spatial estimation of wind speed in a coastal region with complex topography. Renew. Energy 38, 75–82. https://doi.org/10.1016/j.renene.2011.07.007.

Raza, K., Jothiprakash, V., 2014. Multi-output ANN model for prediction of seven meteorological parameters in a weather station. J. Inst. Eng. Ser. A 95, 221–229. https://doi.org/10.1007/s40030-014-0092-9.

Rigol, J.P., 2003. Neural networks for spatial interpolation of meteorological data. 3rd Conference on Artificial Intelligence Applications to the Environmental Science. American Meteorological Society, Boulder, Colorado.

Rigol, J.P., Jarvis, C.H., Stuart, N., 2001. Artificial neural networks as a tool for spatial interpolation. Int. J. Geogr. Inf. Sci. 15, 323–343. https://doi.org/10.1080/13658810110038951.

Sahoo, G.B., Schladow, S.G., Reuter, J.E., 2009. Forecasting stream water temperature using regression analysis, artificial neural network, and chaotic non-linear dynamic models. J. Hydrol. 378, 325–342. https://doi.org/10.1016/j.jhydrol.2009.09.037.

Salami, E.S., Ehteshami, M., 2015. Simulation, evaluation and prediction modeling of river water quality properties (case study: Ireland Rivers). Int. J. Environ. Sci. Technol. 12, 3235–3242. https://doi.org/10.1007/s13762-015-0800-7.

Salami Shahid, E., Ehteshami, M., 2016. Application of artificial neural networks to estimating DO and salinity in San Joaquin River basin. Desalin. Water Treat. 57, 4888–4897. https://doi.org/10.1080/19443994.2014.995713.

Salami, E.S., Salari, M., Ehteshami, M., Bidokhti, N.T., Ghadimi, H., 2016. Application of artificial neural networks and mathematical modeling for the prediction of water quality variables (case study: southwest of Iran). Desalin. Water Treat. 3994, 1–12. https://doi.org/10.1080/19443994.2016.1167624.

Shahabi, M., Jafarzadeh, A.A., Neyshabouri, M.R., Ghorbani, M.A., Valizadeh Kamran, K., 2017. Spatial modeling of soil salinity using multiple linear regression, ordinary kriging and artificial neural network methods. Arch. Agron. Soil Sci. 63, 151–160. https://doi.org/10.1080/03650340.2016.1193162.

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. Environ. Sci. Pollut. Res. Int. 22, 4230–4241. https://doi.org/10.1007/s11356-014-3669-y.

Snell, S.E., Gopal, S., Kaufmann, R.K., 2000. Spatial interpolation of surface air õ-temperatures using artificial neural networks: evaluating their use for downscaling GCMs. J. Clim. 13, 886–895. https://doi.org/10.1175/1520-0442(2000)013<0886:SIOSAT>2.0.CO;2.
The International Commission for the Protection of the Danube River (ICPDR), 2018. River Basin [WWW Document]. URL. http://www.icpdr.org/.
TIBCO Software Inc, 2017. STATISTICA Automated Neural Networks (SANN) - Neural Networks: An Overview [WWW Document]. URL. http://documentation.statsoft.com/STATISTICAHelp.aspx?path=SANN/Overview/SANNNeuralNetworksAnOverview.
Verma, a.K., Singh, T.N., 2013. Prediction of water quality from simple field parameters. Environ. Earth Sci. 69, 821–829. https://doi.org/10.1007/s12665-012-1967-6.
Wang, Y., Zheng, T., Zhao, Y., Jiang, J., Wang, Y., Guo, L., Wang, P., 2013. Monthly water quality forecasting and uncertainty assessment via bootstrapped wavelet neural networks under missing data for Harbin, China. Environ. Sci. Pollut. Res. Int. 20, 8909–8923. https://doi.org/10.1007/s11356-013-1874-8.
Zounemat-Kermani, M., Kişi, Ö., Adamowski, J., Ramezani-Charmahineh, A., 2016. Evaluation of data driven models for river suspended sediment concentration modeling. J. Hydrol. 535, 457–472. https://doi.org/10.1016/j.jhydrol.2016.02.012.