

Modeling total phosphorus removal in an aquatic environment restoring horizontal subsurface flow constructed wetland based on artificial neural networks

Wei Li¹ · Yan Zhang^{1,2} · Lijuan Cui¹ · Manyin Zhang¹ · Yifei Wang¹

Received: 30 December 2014 / Accepted: 9 April 2015 / Published online: 23 April 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract A horizontal subsurface flow constructed wetland (HSSF-CW) was designed to improve the water quality of an artificial lake in Beijing Wildlife Rescue and Rehabilitation Center, Beijing, China. Artificial neural networks (ANNs), including multilayer perceptron (MLP) and radial basis function (RBF), were used to model the removal of total phosphorus (TP). Four variables were selected as the input parameters based on the principal component analysis: the influent TP concentration, water temperature, flow rate, and porosity. In order to improve model accuracy, alternative ANNs were developed by incorporating meteorological variables, including precipitation, air humidity, evapotranspiration, solar heat flux, and barometric pressure. A genetic algorithm and cross-validation were used to find the optimal network architectures for the ANNs. Comparison of the observed data and the model predictions indicated that, with careful variable selection, ANNs appeared to be an efficient and robust tool for predicting TP removal in the HSSF-CW. Comparison of the accuracy and efficiency of MLP and RBF for predicting TP removal showed that

the RBF with additional meteorological variables produced the most accurate results, indicating a high potentiality for modeling TP removal in the HSSF-CW.

Keywords Aquatic environment restoration · Horizontal subsurface constructed wetland · Total phosphorus · Artificial neural networks · Model

Introduction

As the economy in China has developed, increased discharges of nitrogen- and phosphorus-rich wastewater into water bodies such as lakes, estuaries, and reservoirs have led to a great number of ecological problems. Constructed wetlands (CWs) are now recognized as a cost-effective ecotechnology, owing to their many advantages, including low maintenance, their capacity to absorb sudden increased nutrient loads, and their low energy consumption (Kadlec and Wallace 2009). The most common types of CWs are free water surface and horizontal subsurface flow CWs (HSSF-CWs), both of which have been used for the treatment of types of wastewater such as industrial wastewater, agricultural runoff, municipal wastewater, domestic wastewater, and stormwater (Guo et al. 2014; Huang et al. 2015; Khan et al. 2009; Kipasika et al. 2014; Vymazal and Biezinova 2015). However, few researches about the use of HSSF-CWs for the restoration of waterfowl-contaminated aquatic environment have been reported.

Contaminants in wastewater are removed by different components, and treatment performances are related to the design and operation of individual CWs. Complicated physical, biological, and chemical processes are responsible for the removal of different kinds of contaminants in CWs. Macrophytes, the CW substrate, water flow, and microorganisms provide the

Responsible editor: Marcus Schulz

Electronic supplementary material The online version of this article (doi:10.1007/s11356-015-4527-2) contains supplementary material, which is available to authorized users.

✉ Lijuan Cui
lkyclj@126.com

- ¹ Institute of Wetland Research, Chinese Academy of Forestry, Haidian District, Beijing 100091, China
- ² Key Laboratory of Water and Sediment Sciences, Ministry of Education, Department of Environmental Engineering, Peking University, Beijing 100871, China

media for these processes. The small localized environments of CWs, combined with various physicochemical reaction conditions, contribute to the variation in the removal of certain contaminants.

Currently, efforts to quantify the physicochemical interactions that are related to contaminant removal processes rely largely on explicit mathematical models. Effective models for these processes would make the operation mechanisms more visible and can provide scientific support for improved management of CWs. The first-order $k-C^*$ model, proposed by Kadlec and Knight (1996), is the most commonly used model in describing contaminant treatment processes. Many efforts have been made to improve the accuracy of different kinetic models, and details of several modified models, such as the CW2D and Monod models, have been reported (Akhbari et al. 2012; Wynn and Liehr 2001). Incorporating additional parameters increases the accuracy of model predictions. However, models may become increasingly complicated as more parameters are included. Meanwhile, incompatibility with the assumption of the normal distribution adds constraints to the applicability of models. Contaminant removal processes and interactions are nonlinear (Kadlec and Knight 1996; Rousseau et al. 2004), and datasets may not be either normally or exponentially distributed. Therefore, models without data linear correlation and certain distribution requirements have been increasingly documented (Akratos et al. 2008; Guo et al. 2014; Song et al. 2013).

Artificial neural networks (ANNs), including multilayer perceptron (MLP) and radial basis function (RBF), can be used to describe complex internal relationships without any data distribution requirements. ANNs can also effectively solve nonlinear prediction problems with non-normal distributions. To date, ANNs have been widely used to forecast water demand in urban areas (Liu et al. 2003), to forecast daily stream flow (Wang et al. 2006), to simulate the lake level fluctuations (Talebizadeh and Moridnejad 2011), and to find the deterministic factors influencing algal blooms (Wilson and Recknagel 2001). ANN-based predictions of the effluent contaminant concentrations, such as chemical oxygen demand, biochemical oxygen demand, total nitrogen, total suspended solids, and organic matter, have also been studied (Akratos et al. 2009; Naz et al. 2009; Pastor et al. 2003; Tomenko et al. 2007). However, few published data are available on the application of ANNs to simulation of total phosphorus (TP) removal in waterfowl-contaminated aquatic environment, which places restrictions to the wide application of HSSF-CWs.

With this in mind, the main objectives of the current study are therefore (1) to examine whether ANNs can be used to predict the removal of TP in a HSSF-CW used for aquatic environment restoration, and (2) to exploit additional variables for which data are either available or easily measured in the field to improve the TP removal.

Materials and methods

Site description

The HSSF-CW is located in Beijing Wildlife Rescue and Rehabilitation Center, Beijing, China (Fig. 1). An artificial lake (~1 ha surface area) that is a major habitat for waterfowl falls inside the center. Excessive nutrient inputs, such as nitrogen and phosphorus mainly from the frequent activities of waterfowl, caused eutrophication problems in the lake. The artificial lake was built without any water outlets or facilities for the improvement of hydrodynamics. Besides, high-density polyethylene (HDPE)-impermeable membrane (tensile strength=>17 MPa, thickness=1.5–3.0 mm, permeability coefficient=< 1.0×10^{-13} g/(cm/s/Pa)) was used for preventing infiltration processes of the lake water into the groundwater environment. With continuous inputs of nutrient-enriched wastewater, the artificial lake operated with a poor water recycling capacity.

To improve the water quality of the artificial lake, a HSSF-CW system was constructed and has been in operation since 2008. *Iris tectorum* was planted at a density of 3 stems/m². The HSSF-CW bed was filled with gravel (5–30 mm in diameter) to a depth of 0.80 m. Oxygenation processes were enhanced by introducing oxygen through ventilation pipes (10 cm in diameter and 0.50 m in length). Wastewater in the lake was pumped to the first pond, with an influent hydraulic loading of 200 m³/day, and then flew sequentially to the second and third ponds by gravity. Treated water at the outlet of the third pond flew back to the lake. Plants were allowed to grow naturally for about 2 months before sampling for water quality. The HSSF-CW froze over and so was shut down through winter. Dead leaves and branches of aboveground plants were not harvested.

Sample collection and analysis

Treatment performance of the HSSF-CW system has been monitored since August 2008. To determine the water quality, triplicate water samples were collected at each sampling site every 2 weeks from August 2008 to November 2012. A YSI 6-series sonde (YSI Inc., Yellow Springs, OH, USA) was used to record information on a range of variables, including total suspended solid, water temperature, dissolved oxygen, pH, conductivity, and turbidity. Total nitrogen and TP concentrations in 5-mL water samples were determined using a SmartChem 200 discrete chemistry analyzer (WESTCO, USA), following digestion with H₂SO₄–HClO₄. Each subsample was filtered through a Millipore membrane filter (0.45 μm) and was immediately analyzed for concentrations of phosphate, nitrate, and ammonia. Chemical oxygen demand was measured using colorimetric procedures (0–1500 mg/L; Hach Corp., Loveland, CO, USA). Water depth

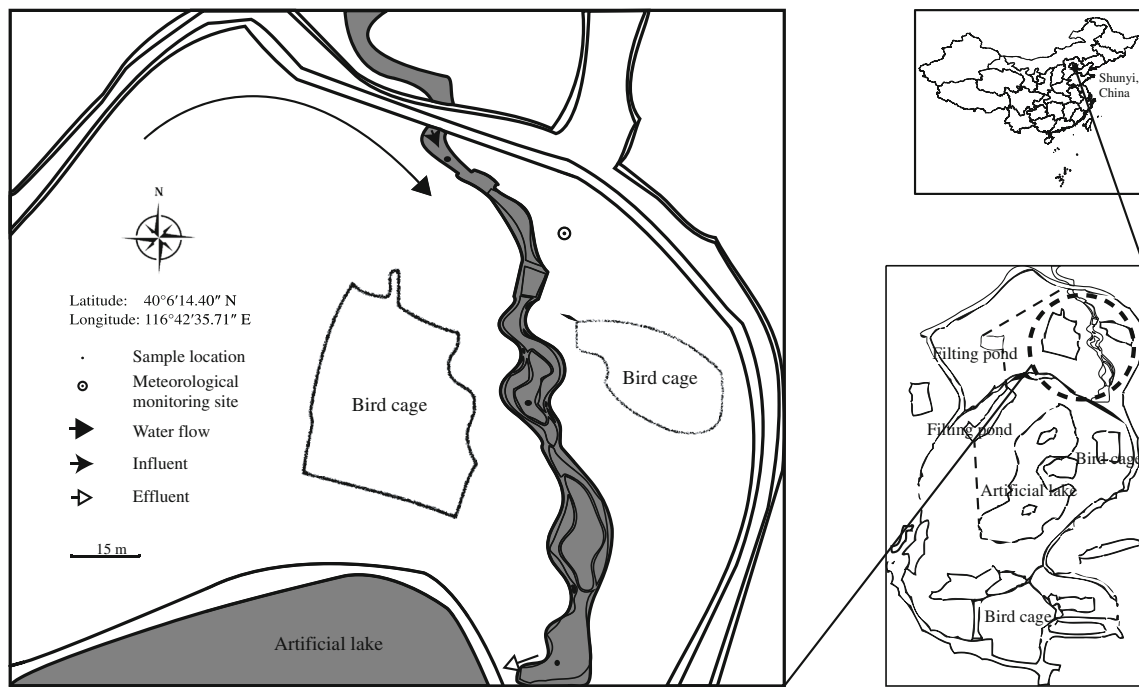


Fig. 1 Diagram of the horizontal subsurface flow constructed wetland system

and flow velocity at each sampling location were measured by FlowTracker (YSI Inc., Yellow Springs, OH, USA), and triplicate measurements were recorded. Meteorological variables, including solar heat flux, barometric pressure, precipitation, evapotranspiration, and humidity, were recorded hourly by WeatherHawk (CampSci, USA). Although the samples were analyzed for a range of contaminants, only TP values were included in the model. The resulting physicochemical data were summarized in Table S1.

Statistical analysis

SPSS 19.0 (IBM, USA) was used to determine the effects of sampling dates and locations on water quality. Relationships between TP removals and different environmental variables were determined using SPSS 19.0.

Principal component analysis

Data standardization and extraction are crucial for function approximations. Principal component analysis (PCA) can be used to decrease the complexity of high-dimensional functions and, therefore, could increase the accuracy of model predictions (Jain and Dubes 1988). Using dimensional reduction, PCA orthogonalizes the components of different variables and ranks the resulting principal components in order. As a result, components for which the greatest degree of variance is explained come first, while those that contribute least to the total variance of the dataset are eliminated.

In the present study, we used the nonlinear iterative partial least squares iteration algorithm, which has a proven higher accuracy than other methods (StatSoft 1998). *N*-fold cross-validation was performed to maximize the chance of finding significant relationships among different variables and to ensure the general validity of models. In order to eliminate issues caused by using various scales and units for different variables, data were rescaled to [−1, 1], using a normalization method based on the following equation (Sarle 2001):

$$x_t = \frac{2(x - x_{\min})}{x_{\max} - x_{\min}} - 1 \tag{1}$$

where x_t is the transformed data, and x_{\max} and x_{\min} are the maximum and minimum values, respectively.

Results of the PCA indicated that variables differed in their importance, and these variables were then considered as the inputs for the development of ANNs according to their availability in the field.

ANN

The generalization capability of ANNs refers to the capability of the optimized networks to choose the correct responses to various learning samples. Topology and connection weights of neurons are the factors that have the most influence on the generalization capability of a neural network (Sarle 2001). MLP and RBF, with inherent advantages and disadvantages for each function, are the most common types of ANNs. The RBF network, with a single hidden layer of radial neurons, is a feedforward learning network. It can minimize the prediction

error of outputs by modeling a Gaussian response surface for each neuron. In addition, RBF can be trained more quickly than MLP. However, the large number of neurons required for the RBF fitting processes increases the dimensionality of the input space. Therefore, both RBF and MLP were used in this study in order to select the most suitable networks.

As the removal processes of different contaminants and the interactions among different variables proved to be close to the pattern of continuous functions (Reed et al. 1995), a typical three-forward neural network, which includes three layers (input layer, hidden layer, and output layer), was used to approximate the continuous functions of the TP removal. Based on the PCA results, the inputs for ANNs were the variables that influenced the TP removal mostly, while the TP removal was considered as the single output layer. The hidden layer was set to one, and the number of neurons in the hidden layer was determined by the Vapnik-Chervonenkis (VC) dimension. The network generalization was optimized with a lower tolerance than a given value ε at a high degree of confidence as follows:

$$\varepsilon \leq \frac{VC}{M} \ln \frac{M}{VC} \quad (2)$$

For the neural network with a single hidden layer

$$VC = I \times H + H \times O \quad (3)$$

where I , H , and O are the number of neurons in the input layer, the hidden layer, and the output layer, respectively. Training was processed to choose a combination of M and H to make the tolerance error smallest.

A genetic algorithm was introduced to the learning processes in order to calculate the connection weights among different neurons. A number of hybridizations and a constant probability mutation were used by the genetic operator. As a result, the neural network tended to minimize the total system error by constantly upgrading the initial connection weights. The activation function for the hidden layer used in this study was the scaled conjugate gradient algorithm. As this algorithm was independent of specified parameters and time consuming, line searches could be avoided (Møller 1993). The algorithm determined the minimum number of points in the weight space by approximating the quadratic errors. By avoiding line searching at each iteration, the speed of convergence increased.

In order to estimate the generalization error of the different networks, leave-one-out cross-validation was used for training processes (Sarle 2001). Merits of the fitting processes (F) were calculated using the following function:

$$F = \frac{1 + \alpha \left(1 - \frac{H}{H_{\max}} \right)}{SSE} \quad (4)$$

where α is the impact coefficient for the number of neurons in the hidden layer, and SSE represents the sum of squares for error.

Based on hybridization of the classical gradient algorithm, constant probability mutation operation was adopted by the mutation operators (Krishna and Murty 1999). Datasets compiled for model development were randomly partitioned into learning sets and test sets by stratified sampling. As a result, about 70 and 30 % of the data were used for training and testing, respectively. We only used learning sets for training in the model development, whereas test sets were used to assess the overall performance of ANNs.

Results and discussion

Relationships between TP removal and environmental variables

To prepare for PCA, a correlation analysis was carried out between candidate input variables and the coefficients were calculated using the following equation:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5)$$

where \bar{x} and \bar{y} are the mean values of x_i and y_i , respectively.

As -1 and $+1$ are equally important when considering the significance of environmental variables and 0 indicates no correlation in PCA, absolute values of the correlation coefficients were used in the following analysis. In this study, 330 of 582 data points were used to identify the correlations between TP removal and different environmental variables.

Consequently, the absolute coefficients ranged from 0 to 0.84 (Table 1). The influent TP concentration (absolute r value=0.62, $p=0.013$), flow rate (absolute r value=0.84, $p=0.004$), temperature (absolute r value=0.73, $p=0.029$), and porosity (absolute r value=0.65, $p=0.032$) were strongly correlated to TP removal. The flow rate was correlated with water temperature, but the coefficient value was low (absolute r value=0.28; $p=0.057$), indicating that these two variables were independent. However, certain physicochemical variables, such as DO, turbidity, and conductivity, were correlated with TP removal with low absolute r values ranging from 0.39 to 0.42. The four factors of area, inlet–outlet distance, bulk density, and porosity, were strongly correlated with each other (absolute $r=0.57$ – 0.73 , $p<0.05$). However, with the exception of porosity, these variables had little impact on TP removal. There were no significant relationships between the meteorological variables and TP removal (absolute $r=0.01$ – 0.20 , $p>0.05$).

Table 1 Pearson correlation coefficients between TP removal and different environmental variables from the HSSF-CW

Variable	Physicochemical factor					Hydrological factor		Wetland unit factor				Meteorological factor				
	TP _i	DO	Turb	Cnd	T	D	Flow	A	Dis	Den	Por	Sol	Bar	Pre	ET	Hum
TP _i	1															
DO	0.62	1														
Turb	0.56	0.47	1													
Cnd	0.61	0.67	0.42	1												
T	0.63	0.67	0.48	0.8	1											
D	0.07	0.03	0.12	0.05	0.06	1										
Flow	0.55	0.44	0.36	0.39	0.28	0	1									
Aera	0.08	0.02	0.07	0	0	-0.03	0.01	1								
Dis	0.03	-0.04	0.13	0.06	0.02	-0.07	0.01	0.64*	1							
Den	-0.05	0.04	0.06	0.01	0.03	0.01	-0.07	-0.63*	-0.62*	1						
Por	-0.01	-0.01	0.06	0.03	-0.02	-0.04	0.04	0.73*	0.63*	-0.57*	1					
Sol	0.27	-0.3	0.17	0.22	0.33	0	0.01	0.02	0.03	0.1	0.03	1				
Bar	0.02	0.10	0	0.11	0.17	0	0	0.12	0	0.06	0.01	0.11	1			
Pre	0.32	-0.27	0.26	0.19	0.15	0.12	0.1	0.03	0.01	0.02	0.07	0.19	0.01	1		
ET	-0.11	0.14	0	0.13	0.23	0	-0.18	0.01	0	-0.02	0.06	0.24	0.09	-0.13	1	
Hum	0.03	-0.10	0.01	0.20	-0.14	0.01	0	0.05	0.01	0.07	-0.05	0.02	-0.01	0.22	0.15	1
Removal	0.62*	0.39	0.40	0.42	0.73*	-0.06	0.84**	0.31	-0.12	0.2	-0.65*	0.15	0.01	-0.17	0.20	0.09

TP_i influent total phosphorus concentration, DO dissolved oxygen, Turb turbidity, Cnd conductivity, T water temperature, D water depth, Flow flow rate, Dis inlet–outlet distance, Den soil density, Por porosity, Sol solar heat flux, Bar barometric pressure, Pre precipitation, ET evapotranspiration, Hum air humidity

p*<0.05; *p*<0.01

Principal component analysis

PCA was performed with 302 iterations and converged in the criterion of 0.0001. Sevenfold cross-validation was used in this study. Contributions of the estimated principal components to the total variance were calculated (Fig. 2a). Principal components that contributed less than 5 % to the total variance were eliminated, resulting in a three-dimensional dataset, which explained the 91.45 % of the total variance (Fig. 2b).

The three principal components were rotated using the varimax rotation method to facilitate interpretation. Results indicated that the first principal component, contributing 38.29 % to the total variance, represented the influence of the influent TP concentration, temperature, dissolved oxygen, turbidity, and conductivity. The intercepts of the extracted variables were 0.981, 0.977, 0.936, 0.961, and 0.929, respectively. Out of these variables, the influent TP concentration and temperature were selected as the inputs for the ANNs. The first principal component therefore represented physicochemical effects.

The second principal component accounted for 32.02 % of the total variance and primarily reflected the interaction between TP removal and wetland unit variables. The intercepts for area, distance, soil density, and porosity were 0.926, 0.961, 0.937, and 0.980, respectively. In this study, we used porosity

as an input variable because using this variable makes it easier to compare effects of components such as plants and substrates on the removal of contaminants in future research.

The third principal component explained 21.14 % of the total variance and was related to the flow rate and water depth. The loadings of these two variables were 0.981 and 0.973, respectively. Compared to water depth, the flow rate was selected as an input variable based on its strong correlation with TP removal.

As a result, the input variables, including the influent TP concentration, temperature, porosity, and flow rate, represented physicochemical characteristics, wetland unit features, and hydraulic dynamics and were selected for the development of ANNs.

ANNs

In this study, two kinds of ANNs were developed. The first kind of ANNs was developed from the four input parameters that were extracted by PCA, as detailed above. Improved ANNs that included meteorological variables, such as solar heat flux, barometric pressure, precipitation, evapotranspiration, and humidity, were also developed in order to grasp seasonal effects on the TP removal.

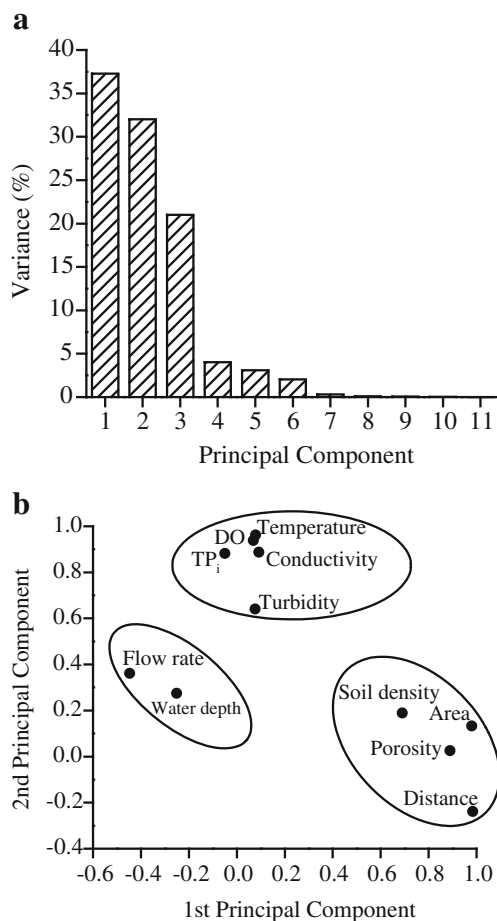


Fig. 2 Variance in total phosphorus removals explained by extracted principal components (a) and principal component ordinations for different environmental variables (b). TP_i influent concentration of total phosphorus, DO dissolved oxygen, $Distance$ inlet–outlet distance

ANNs with four input parameters

The number of neurons in the MLP input layer was set as 4, corresponding with the PCA results. The output layer contained a single TP removal neuron with a linear activation function. Generally, ANNs make training easily fall into a local minimal architecture and searching is usually limited to a subspace of the entire structure (Murata et al. 1994). In order to solve this problem, a genetic algorithm was incorporated into the training processes. The genetic algorithm upgrades the initial network weights by constant genetic mutation, so that the networks tend to minimize the total system error (Krishna and Murty 1999). Large learning rates produced oscillations in ε estimates, while small ones increased the number of epochs. In order to achieve a satisfactory level of accuracy, training was carried out by setting the constant training epoch value to 1000 and the constant learning rate to 0.15 (Murata et al. 1994). As training progressed, the number of neurons in the hidden layer was optimized, while other parameters were maintained constant. The performance of the MLP

functions was optimized according to the ε estimates. After training, the ε values for each model were estimated and the general performance of different architectures was evaluated by averaging the individual ε estimates (Fig. 3a). As a result, the MLP network was completely fine-tuned for 276 iterations with a learning rate equal to 0.21. The optimal network architecture for a minimum ε value of 0.055 was 4–13–1 for the number of neurons in the input, hidden, and output layers, respectively.

The MLP input variables were used for RBF training. The number of neurons in the input layer coincided with the number of input variables, and the only TP removal neuron in the output layer was exploited by the linear activation function. Gaussian function was used to generate output signals for the number of neurons in the hidden layer. Using the leave-one-out cross-validation method, the ε estimates were calculated for each model, and the performance of the trained models for each network architecture was tested by averaging the individual VS ε estimates (Fig. 3a). As a result, the algorithm that was initialized with the estimated optimal architecture converged in 18 epochs and the best architecture achieved was

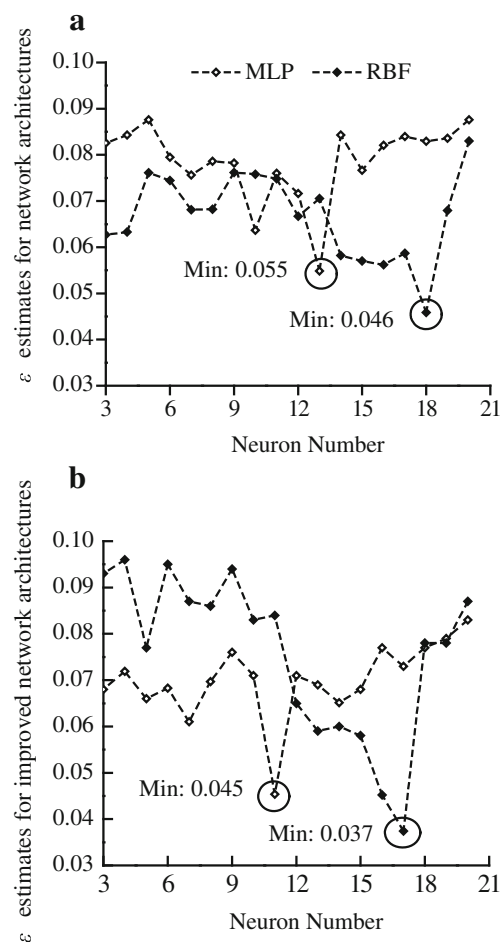


Fig. 3 Estimated ε values for different network topological architectures with different numbers of neurons in the hidden layer. **a** ANNs with four input parameters. **b** ANNs with additional parameters

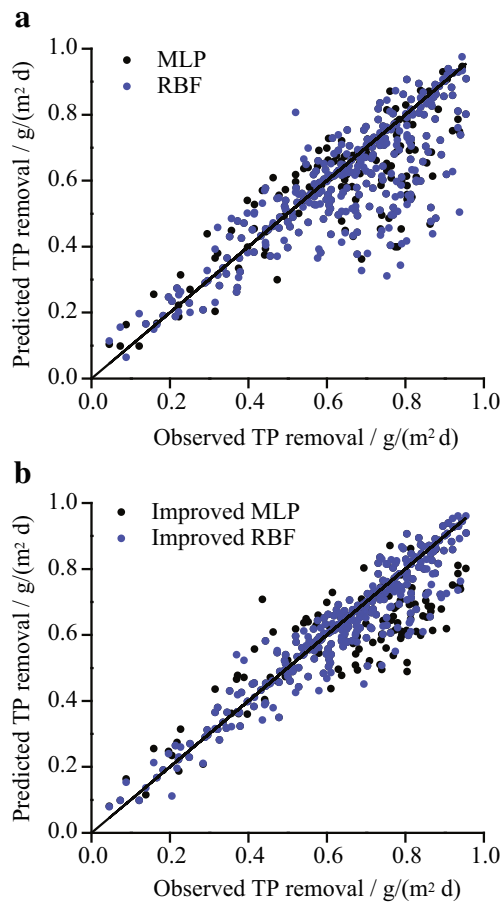


Fig. 4 Experimental and model-predicted removals of total phosphorus. **a** ANNs with four input parameters. **b** ANNs with additional parameters

4–18–1 for the number of neurons in the input, hidden, and output layers, respectively.

Validation of the MLP and RBF networks was performed by comparing the observed and predicted TP removals (Fig. 4a). Quantitative statistics, including determination coefficients (R^2) and mean absolute errors (MAEs) between the actual and simulated outputs, were considered as indicators of the model performance. Modeling results suggested a strong correlation between the observed and predicted TP removals, indicated by the high R^2 values in Table 2. Model predictions were close to the observed values when TP removal was smaller than 0.6 g/(m²/day). The MAEs for the networks were

somewhat lower than those reported in previous studies (Bowes et al. 2010; Nayak et al. 2006; Steer et al. 2002). Both of the MLP and RBF networks were underpredicted when the removal was higher than 0.6 g/(m²/day). Most of data points were within the range from 0.20 to 0.89 g/(m²/day), with the exception of some outliers that may have been caused by extreme hydraulic characteristics and temperature values. Extreme meteorological variables might have also contributed to the outliers.

Improved ANNs with meteorological variables

Previous research demonstrated that rainfall has an impact on the evapotranspiration rate and can introduce a certain amount of phosphorus to the local environments (Kadlec and Wallace 2009; Sharpley and Kleinman 2003). Meanwhile, metabolism processes, which can transport rhizosphere water containing a certain content of phosphate (0.60–1.31 g/(m²/day)) to different organs and tissues (Kadlec and Wallace 2009), were influenced by variable meteorological conditions. In this study, five additional variables (solar heat flux, barometric pressure, precipitation, evapotranspiration, and humidity) were introduced to the development of ANNs. Even though meteorological variables and TP removal were not highly correlated, these variables were included in order to improve the model accuracy.

Both MLP and RBF were trained for nine input variables. Setups for the improved ANNs were similar to those used for the networks earlier. As a result, the optimal numbers of neurons in the hidden layer for MLP and RBF were 11 and 17, respectively (Fig. 3b). Modeling results indicated that the improved ANNs ($R^2_{LS}=0.887-0.912$, $R^2_{TS}=0.739-0.801$) performed better than the ANNs ($R^2_{LS}=0.739-0.851$, $R^2_{TS}=0.651-0.717$) developed from the four input variables. Most of the predictions were close to the observed values, and the number of outliers was greatly reduced (Fig. 4b), indicating improved model performance of the improved ANNs. As both the predictions and the input variables were averaged over extended time periods, the fit of the RBF, with its high R^2 and low MAE, was considered to be more accurate for predicting TP removal in the HSSF-CW system. In addition, training algorithm for the RBF network resulted in a smaller

Table 2 Statistical parameters for the two kinds of ANNs

	R^2_{LS}	p_{LS}	MSE_{LS}	MAE_{LS}	R^2_{TS}	p_{TS}	MSE_{TS}	MAE_{TS}	Epoch ^a
MLP	0.739	0.004	0.055	0.049	0.651	0.013	0.093	0.060	370
RBF	0.851	0.002	0.046	0.037	0.717	0.004	0.082	0.053	45
Improved MLP	0.887	0.002	0.045	0.033	0.739	0.007	0.080	0.051	301
Improved RBF	0.912	0.000	0.037	0.016	0.801	0.004	0.057	0.029	15

R^2 determination coefficient, p p value, MSE mean-squared error, MAE mean absolute error

^a Number of training epochs

number of epochs (Table 2), which is an important property of adaptive systems (Kadlec and Wallace 2009).

Conclusions

ANNs with meteorological variables included are efficient alternatives for modeling TP removal in the HSSF-CW system. In this study, the improved RBF network gave good results for TP removal. Comparison of the observed and predicted TP removals showed that it is useful to evaluate the model performance, as by doing this, the important role of meteorological effects was highlighted. Additional variables should be studied in combination with interactions. Stochastic models could be further improved by incorporating variables related to explicit removal mechanisms for different contaminants.

Acknowledgments This study was funded by the National Nonprofit Institute Research Grant of the Chinese Academy of Forestry entitled “Dynamic mechanisms of phosphorus removal in subsurface constructed wetlands” (CAFINT2013C13) and the Fundamental Research Funds for the Central Non-profit Research Institution of CAF “N removal mechanism of subsurface constructed wetland in health wetland” (CAFYBB2014QA029). We are grateful to all the members of the research team for their helpful comments and advice.

References

- Akhbari A, Zinatizadeh A, Mohammadi P, Mansouri Y, Irandoust M, Isa M (2012) Kinetic modeling of carbon and nutrients removal in an integrated rotating biological contactor-activated sludge system. *Int J Environ Sci Technol* 9:371–378
- Akratos CS, Papaspyros JN, Tsihrintzis VA (2008) An artificial neural network model and design equations for BOD and COD removal prediction in horizontal subsurface flow constructed wetlands. *Chem Eng J* 143:96–110
- Akratos CS, Papaspyros JNE, Tsihrintzis VA (2009) Total nitrogen and ammonia removal prediction in horizontal subsurface flow constructed wetlands: use of artificial neural networks and development of a design equation. *Bioresour Technol* 100:586–596
- Bowes MJ, Neal C, Jarvie HP, Smith JT, Davies HN (2010) Predicting phosphorus concentrations in British rivers resulting from the introduction of improved phosphorus removal from sewage effluent. *Sci Total Environ* 408:4239–4250
- Guo YM, Liu YG, Zeng GM, Hu XJ, Xu WH, Liu YQ, Liu SM, Sun HS, Ye J, Huang HJ (2014) An integrated treatment of domestic wastewater using sequencing batch biofilm reactor combined with vertical flow constructed wetland and its artificial neural network simulation study. *Ecol Eng* 64:18–26
- Huang JJ, Gao X, Balch G, Wootton B, Jørgensen SE, Anderson B (2015) Modelling of vertical subsurface flow constructed wetlands for treatment of domestic sewage and stormwater runoff by subwet 2.0. *Ecol Eng* 74:8–12
- Jain AK, Dubes RC (1988) Algorithms for clustering data. Prentice-Hall, Inc., USA
- Kadlec RH, Knight RL (1996) Treatment wetlands. Lewis, Boca Raton
- Kadlec RH, Wallace SD (2009) Treatment wetlands. CRC, USA
- Khan S, Ahmad I, Shah MT, Rehman S, Khaliq A (2009) Use of constructed wetland for the removal of heavy metals from industrial wastewater. *J Environ Manag* 90:3451–3457
- Kipasika HJ, Buza J, Lyimo B, Miller WA, Njau KN (2014) Efficiency of a constructed wetland in removing microbial contaminants from pre-treated municipal wastewater. *Phys Chem Earth* 72–75:68–72
- Krishna K, Murty MN (1999) Genetic K-means algorithm. *Syst Man Cybern C* 29:433–439
- Liu JG, Savenije HG, Xu JX (2003) Forecast of water demand in Weinan City in China using WDF-ANN model. *Phys Chem Earth* 28:219–224
- Møller MF (1993) A scaled conjugate gradient algorithm for fast supervised learning. *Neural Netw* 6:525–533
- Murata N, Yoshizawa S, Amari S (1994) Network information criterion-determining the number of hidden units for an artificial neural network model. *IEEE Trans Neural Netw* 5:865–872
- Nayak PC, Rao YS, Sudheer K (2006) Groundwater level forecasting in a shallow aquifer using artificial neural network approach. *Water Resour Manag* 20:77–90
- Naz M, Uyanik S, Yesilnacar MI, Sahinkaya E (2009) Side-by-side comparison of horizontal subsurface flow and free water surface flow constructed wetlands and artificial neural network (ANN) modelling approach. *Ecol Eng* 35:1255–1263
- Pastor R, Benqlilou C, Paz D, Cardenas G, Espuña A, Puigjaner L (2003) Design optimisation of constructed wetlands for wastewater treatment. *Resour Conserv Recycl* 37:193–204
- Reed SC, Crites RW, Middlebrooks EJ (1995) Natural systems for waste management and treatment. McGraw-Hill, Inc., USA
- Rousseau DP, Vanrolleghem PA, Pauw ND (2004) Constructed wetlands in Flanders: a performance analysis. *Ecol Eng* 23:151–163
- Sarle WS (2001) Neural network FAQ, part 2 of 7: learning, periodic posting to the usenet newsgroup comp. ai. neural-nets. *Neural Netw* 23:551–566
- Sharpley A, Kleinman P (2003) Effect of rainfall simulator and plot scale on overland flow and phosphorus transport. *J Environ Qual* 32:2172–2179
- Song K, Park Y-S, Zheng F, Kang H (2013) The application of artificial neural network (ANN) model to the simulation of denitrification rates in mesocosm-scale wetlands. *Ecol Inform* 16:10–16
- StatSoft I (1998) Statistical neural network. StatSoft Inc., USA
- Steer D, Fraser L, Boddy J, Seibert B (2002) Efficiency of small constructed wetlands for subsurface treatment of single-family domestic effluent. *Ecol Eng* 18:429–440
- Talebzadeh M, Moridnejad A (2011) Uncertainty analysis for the forecast of lake level fluctuations using ensembles of ANN and ANFIS models. *Expert Syst Appl* 38:4126–4135
- Tomenko V, Ahmed S, Popov V (2007) Modelling constructed wetland treatment system performance. *Ecol Model* 205:355–364
- Vymazal J, Biezinova T (2015) The use of constructed wetlands for removal of pesticides from agricultural runoff and drainage: a review. *Environ Int* 75:11–20
- Wang W, Gelder PHV, Vrijling JK, Ma J (2006) Forecasting daily streamflow using hybrid ANN models. *J Hydrol* 324:383–399
- Wilson H, Recknagel F (2001) Towards a generic artificial neural network model for dynamic predictions of algal abundance in freshwater lakes. *Ecol Model* 146:69–84
- Wynn TM, Liehr SK (2001) Development of a constructed subsurface flow wetland simulation model. *Ecol Eng* 16:519–536