

Sequential dynamic artificial neural network modeling of a full-scale coking wastewater treatment plant with fluidized bed reactors

Hua-Se Ou¹ · Chao-Hai Wei² · Hai-Zhen Wu³ · Ce-Hui Mo¹ · Bao-Yan He¹

Received: 25 November 2014 / Accepted: 7 May 2015 / Published online: 7 June 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract This study proposed a sequential modeling approach using an artificial neural network (ANN) to develop four independent models which were able to predict biotreatment effluent variables of a full-scale coking wastewater treatment plant (CWWTP). Suitable structure and transfer function of ANN were optimized by genetic algorithm. The sequential approach, which included two parts, an influent estimator and an effluent predictor, was used to develop dynamic models. The former parts of models estimated the variations of influent COD, volatile phenol, cyanide, and NH_4^+ -N. The later parts of models predicted effluent COD, volatile phenol, cyanide, and NH_4^+ -N using the estimated values and other parameters. The performance of these models was evaluated by statistical parameters (such as coefficient of determination (R^2), etc.). Obtained results indicated that the estimator developed dynamic models for influent COD ($R^2=0.871$), volatile phenol ($R^2=0.904$), cyanide ($R^2=0.846$), and NH_4^+ -N.

Responsible editor: Angeles Blanco

Electronic supplementary material The online version of this article (doi:10.1007/s11356-015-4676-3) contains supplementary material, which is available to authorized users.

✉ Hua-Se Ou
touhuase@jnu.edu.com

¹ Key Laboratory of Water/Soil Toxic Pollutants Control and Bioremediation of Guangdong Higher Education Institutions, School of Environment, JiNan University, The Second Science and Technology Building, 601 Huangpu Avenue West, Guangzhou, China P.C. 510632

² Key Lab of Pollution Control and Ecosystem Restoration in Industry Clusters, Ministry of Education, College of Environment and Energy, South China University of Technology, Guangzhou 510006, China

³ School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, China

N ($R^2=0.777$), while the predictor developed feasible models for effluent COD ($R^2=0.852$) and cyanide ($R^2=0.844$), with slightly worse models for effluent volatile phenol ($R^2=0.752$) and NH_4^+ -N ($R^2=0.764$). Thus, the proposed modeling processes can be used as a tool for the prediction of CWWTP performance.

Keywords Artificial neural network · Modeling · Coking wastewater · Cyanide · Phenol

Abbreviations

CWW	Coking wastewater
CWWTPs	Coking wastewater treatment plants
COD	Chemical oxygen demand
BOD	Biochemical oxygen demand
ADM1	Anaerobic digestion model no.1
ANN	Artificial neural network
BP-ANN	Back-propagation-ANN
GA	Genetic algorithm
DO	Dissolved oxygen
NH_4^+ -N	Ammonium nitrogen
GDX	Gradient descent with momentum and adaptive learning rate back propagation
BR	Bayesian regularization
RMSE	Root mean squared error
MAE	Mean absolute error
R^2	Coefficient of determination

Introduction

Coking wastewater (CWW) is a heterogeneous complex generated during the coal coking, coal gas purification, and by-product recovery processes in coke factory which provide

important raw materials for steelmaking (Pal and Kumar 2014). It contains considerable amounts of phenols, polycyclic aromatic hydrocarbons, heterocyclic compounds, cyanide, and sulfide compounds. Some of these contaminants are listed as US-EPA and EU priority pollutants (Angelino and Gennaro 1997; Zhang et al. 2012). Thus, CWW has been considered the most toxic wastewater (Kim et al. 2007). The purification and harmless disposal of CWW are important to insure water safety of receiving water bodies.

The conventional activated sludge process was the most widely used treatment technology for CWW (Pal and Kumar 2014). However, the biofilm process was increasingly applied for CWW treatment (Lai et al. 2008), especially the fluidized bed biofilm reactor (Jing et al. 2009). Recently, several novel types of processing, for example, anaerobic–aerobic–hydrolysis–aerobic (A/O/H/O) combined with fluidized bed biofilm reactor, were successfully employed in CWW treatment plants (CWWTPs) (Zhang et al. 2012). Since a strong nonlinear relationship exists between the operating parameters and water quality variables in biological processes, the operational difficulties are frequently encountered in CWWTPs. Furthermore, typical CWW contained high level cyanide ($>50 \text{ mg L}^{-1}$) and volatile phenol ($>200 \text{ mg L}^{-1}$), which had significant biotoxic to normal microorganism. Therefore, the biological processes in CWWTPs may be susceptible to disturbances and toxic loadings (Kim et al. 2009). The unsteady operation of CWWTPs caused failure to achieve the effluent standards.

On-line sensors were used in municipal WWTPs, which provided normal parameters, such as dissolved oxygen (DO), COD, and ammonium nitrogen ($\text{NH}_4^+ \text{-N}$), for routine monitoring and operation. However, the lack of accurate on-line sensors for typical toxic variables in WWTPs, such as cyanide and volatile phenol, resulted in delayed monitoring of effluent. Even though these variables can be measured via offline (laboratory) analyses, an inevitable time delay in a range of few hours or days would result in the discharge of effluent containing over-standard toxic contaminants, which aggravated the potential risk of toxic effluent disposal (Hong et al. 2007). Therefore, a reliable simulating model for CWWTPs was essential to predict its performance and controlling its operation.

The most popular model was the International Water Association Anaerobic Digestion Model No.1 (ADM1) (Blumensaat and Keller 2005). However, the estimated parameters of ADM1 were generally case specific and difficult to adapt for system modification (Hong et al. 2003), and the contaminant composition of CWW was different from normal municipal wastewater. Therefore, the direct application of ADM1 for CWWTPs might be infeasible. Alternatively, artificial neural network (ANN) has been increasingly used to model wastewater treatment processes in recent years. The advantages of ANN included the following: (1) learning of nonlinear data processing, (2) capability of generalization,

and (3) tolerance of failure or incomplete data. Therefore, the simulation using ANN has been extensively studied for municipal WWTPs (Bongards 1999; Dürrenmatt and Gujer 2012; Güçlü and Dursun 2010). Furthermore, the operations of WWTPs for industrial wastewater, such as paper-making wastewater (Huang et al. 2012) and oily wastewater (Pendashteh et al. 2011), were predicted using ANN. For CWW, only a few research developed ANN for the prediction of the performance of single treatment process, such as Fenton oxidation (Zhu et al. 2011).

To the best of our knowledge, there is no literature that reported the study about ANN simulation of CWWTPs. The current study is the first report about the modeling of the variables in a full-scale CWWTP using BP-ANN combined with GA. The BP-ANN models were developed to successfully simulate the variation tendencies of COD, volatile phenol, cyanide, and $\text{NH}_4^+ \text{-N}_{\text{in}}$, the influent and effluent of biotreatment processes. The prediction performances of the models under dynamical circumstances have been evaluated and compared using statistical parameters. This study can be used as a guide for the use of routinely on-line measured parameters, such as pH, COD, DO, and $\text{NH}_4^+ \text{-N}$ as input data, to predict the performance of CWWTPs.

Materials and methods

CWWTP and data set

The data sets were collected from No.1 Songshan CWWTP, which was located in Shaoguan Steel Company Ltd, Guangdong province, China, with a designed treatment capacity of $1680 \text{ m}^3 \text{ day}^{-1}$. The continuous distilled ammonia wastewater and discontinuous desulfurization wastewater contributed the influents. Table 1 presents the statistical information of the monitoring variables of these influents, and Fig. 1 shows the schematic diagram of CWWTP.

Three treatment processes, including pretreatment, biotreatment, and posttreatment, were conducted during daily operation. The pretreatment included degreasing and sedimentation. The influent was first degreased, following by a coagulation sedimentation using ferrous sulfate and polyacrylamide to remove most particle materials and part of organic contaminants. Then, liquid effluent went through the biotreatment processes. An anoxic–aerobic₁–aerobic₂ (A/O₁/O₂) system coupled with internal-loop biological fluidized beds was applied to degrade organic contaminants. These three fluidized beds have volumes of 2316, 3280, and 3920 m^3 . The biological effluent then went through secondary clarifiers, where the bacteria and remaining particles were coagulated using poly-aluminium chloride. Finally, the secondary clarifier effluent was discharged or reused. The data sets contained daily variables of the biotreatment influent and

Table 1 Statistical characteristic of measured variables in no. 1 Songshan CWWTP

Parameter	Distilled ammonia wastewater			Desulfurization wastewater			Total influent			Biotreatment influent			Biotreatment effluent		
	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean
pH	11.4	8.3	10.6	11.1	4.9	9.6	11.1	9.1	10.2	10.0	9.1	9.3	9.4	5.9	7.5
COD (mg L ⁻¹)	4526.6	2465.4	3276.1	11250.2	6216.4	8364.8	4899.5	2323.4	3499.8	3370.1	898.2	2104.7	419	136	233
Volatile phenol (mg L ⁻¹)	1326.54	264.65	596.64	526.64	216.64	346.75	1225.87	312.45	653.39	852.14	146.32	397.26	0.68	0.04	0.23
Cyanide (mg L ⁻¹)	41.89	4.64	17.98	125.46	51.28	72.69	46.81	5.32	19.09	8.87	0.94	3.48	2.58	0.02	0.84
Sulfides	216.6	14.3	51.6	315.5	182.7	214.2	239.3	14.5	56.4	53.8	3.5	23.8	32.5	2.4	12.2
NH ₄ ⁺ -N (mg L ⁻¹)	164.3	4.3	30.7	1762.6	423.6	726.9	179.4	5.4	33.6	119.1	4.2	21.7	40.2	0.7	4.3
Mean flow rate (m ³ h ⁻¹)	60			Variable			~ 60			~ 60			~ 60		
Inflow period	24 h continuous inflow			Variable			-			-			-		

effluent from 1 January 2012 to 31 December 2012, with a total of 358 data points (8 data were absent during holiday).

Analytical methods

The operational data of influent and effluent of biotreatment process were used to develop models. The analyzed variables included COD, volatile phenols, NH₄⁺-N, total sulfides, cyanide, pH, flow rate, and DO. The COD was measured by online COD_{max}-Plus monitors (Hach, USA), and NH₄⁺-N was obtained using Amtax Compact analyzers (Hach, USA). The concentration of volatile phenols, sulfide, and cyanide was determined according to Standard Methods for the Examination of Water and Wastewater (APHA et al. 1998). The pH value was determined by on-line GLI pH/ORP monitors (Hach, USA). Furthermore, the DO concentrations in O₁ and O₂ fluidized beds were monitored using on-line LDO monitors (Hach, USA). The flow rate was determined by FXP4000 electromagnetic flow meters (ABB, Switzerland).

BP-ANN model and genetic algorithm

A multilayer perceptron neural network model trained by the BP algorithm, which was a widely and successfully applied ANN for prediction purposes (Show et al. 2013), was used to

predict the variables in CWWTP. The BP-ANN structure consisted of one input layer, a minimum of one intermediate or hidden layer, and one output layer. Each layer had a number of neurons, which were connected linearly by weights to the neurons in the neighboring layers. The structure of the ANN, which had significant effects on the prediction, should be carefully selected (Lee et al. 2011). A genetic algorithm was used as the selected strategy to determine the number of hidden layers, the number of neurons in each layer, and the transfer function type of each neuron (Goldberg 1989). All simulations were conducted using a genetic adapted neural network program that has been developed by Hong (Hong YS 1998). In this study, the number of generations in genetic algorithm was 100, and each generation contained 250 candidate BP-ANN structures (population). The probability of selection, crossover, and mutation was 45, 40, and 15 %, respectively. Three typical transfer functions, including logistic sigmoid, hyperbolic tangent, and linear functions, were simultaneously tested.

Training algorithm

The training algorithm had effects on the performance of BP-ANN models. Gradient descent with momentum and adaptive learning rate back propagation (GDX) algorithm was robust

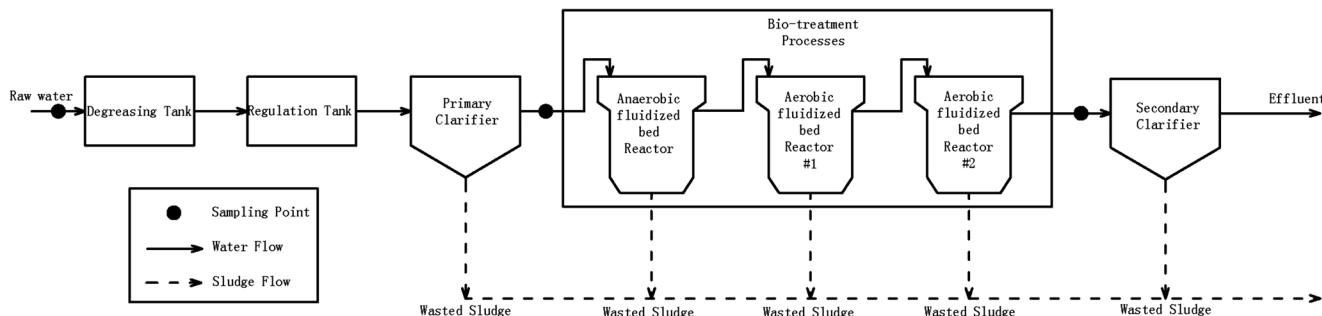


Fig. 1 Schematic diagram of no.1 Songshan coking wastewater treatment plant

and widely used for BP-ANN modeling (Riedmiller 1994). Recently, the BP-ANN model trained by Bayesian regularization (BR) was confirmed to provide satisfactory generalization ability (Aggarwal et al. 2005). Therefore, GDX and BR were used to train the BP-ANN models. Their prediction performance was investigated by statistic parameters. In order to avoid over fitting of model, training and testing process were used (Schulze et al. 2005). The available raw data set (358 daily water quality data) was randomly partitioned into three data sets: training set, validating set, and testing set. The number proportion of [training set: (validating set+testing set)] was set at 2:1. The training set (233 daily data, ~65 %) was used to generate the models. The validating set (72 daily data, ~20 %), which was not involved in the developments of models, was used to optimize the network geometry and model parameters. Furthermore, the testing set (53 daily data, ~15 %) was used to evaluate the applicability of the model. In the present study, training was stopped at each iteration (epochs), for every candidate model that was tested at each step.

Sequential training processes

Most previous research applied to WWTPs aimed at the prediction of effluent quality based on the influent quality. However, some process variables in the influent cannot be directly obtained due to several reasons, such as a lack of reliable online measurement systems and the delay of offline measurement. In order to synchronously predict the variables of influent and effluent, Lee et al. developed a sequential modeling process, which was composed of back-propagation-ANN (BP-ANN) and genetic algorithm (GA) (Lee et al. 2011). In current study, similar sequential modeling of BP-ANN was conducted to develop four independent models for the predictions of COD, volatile phenol, cyanide, and NH_4^+ -N removal in biotreatment processes. The sequential modeling approach was composed of two parts: (1) a process disturbance estimator and (2) a process behavior predictor (Lee et al. 2011). The architecture of this modeling approach is presented in Fig. 2. In the case of COD, the first part (process disturbance estimator) received seven measurable input variables (flow rate_{in}, NH_4^+ -N_{in}, pH_{in}, volatile phenol_{in}, cyanide_{in}, sulfides_{in}, and 1 day before COD_{in}) and provided an estimation of influent COD (COD_{in}). The output of the estimator subsequently formed part of the inputs for the second part. The second part predicted the effluent COD (COD_{ef}) based on nine inputted variables, including estimated COD_{in}, DO₁ in O₁ fluidized bed, DO₂ in O₂ fluidized bed, measured flow rate_{in}, NH₄⁺-N_{in}, pH_{in}, volatile phenol_{in}, cyanide_{in}, and sulfides_{in}. The sequential modeling approaches of volatile phenol, cyanide, and NH_4^+ -N all followed similar processes. More details can be found in Lee et al. (2011).

Evaluation of model performance

The performance of models was evaluated in terms of root mean squared error (RMSE, Eq. 1), mean absolute error (MAE, Eq. 2), and coefficient of determination (R^2 , Eq. 3) between the predicted values and the measured values in training, validating, and testing data sets.

$$MAE = \frac{1}{N} \left(\sum_{i=1}^N |y_i - y'_i| \right) \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - y'_i)^2}{N}} \quad (2)$$

$$R^2 = \left(\frac{\sum_{i=1}^N (y_i - y_{i,\text{mean}})(y'_i - y'_{i,\text{mean}})}{\sqrt{\sum_{i=1}^N (y_i - y_{i,\text{mean}})} \sqrt{\sum_{i=1}^N (y'_i - y'_{i,\text{mean}})}} \right)^2 \quad (3)$$

where N was the number of data points, y_i and y'_i were the observed and predicted values, respectively. Furthermore, $y_{i,\text{mean}}$ and $y'_{i,\text{mean}}$ were the mean observed values and mean predicted values, respectively

Results and discussion

Comparison of GDX and BR algorithms

BP-ANN models were developed to estimate the influent variables and predict the effluent variables of biotreatment. GDX and BR algorithms were used to train all models until the best fitting network architectures were obtained by GA. For each variable (Table 2), the network structure obtained using GDX algorithm was more complicated than the structure obtained using BR algorithm. The training iterations using GDX were also more than those using BR. The corresponding model performance statistics for the best network architectures using GDX and BR are listed in Table 3. All MAEs and RMSEs of training and validating data sets using GDX algorithm were lower than those using BR algorithm, indicating that GDX algorithm trained the better fitting models for target-known data set. However, the MAEs and RMSEs of testing data sets using GDX algorithm were significantly higher than those using BR algorithm, suggesting that the BR-trained models had better generalizing capacity for target-unknown data set. The results of R^2 also showed that BR had better performance than GDX. Consequently, BR was the better training algorithm which had high generalizing capacity for the prediction of biological influent and effluent variables in current CWWTP. The best model for each predicted variable using BR algorithm was chosen and used for subsequent modeling.

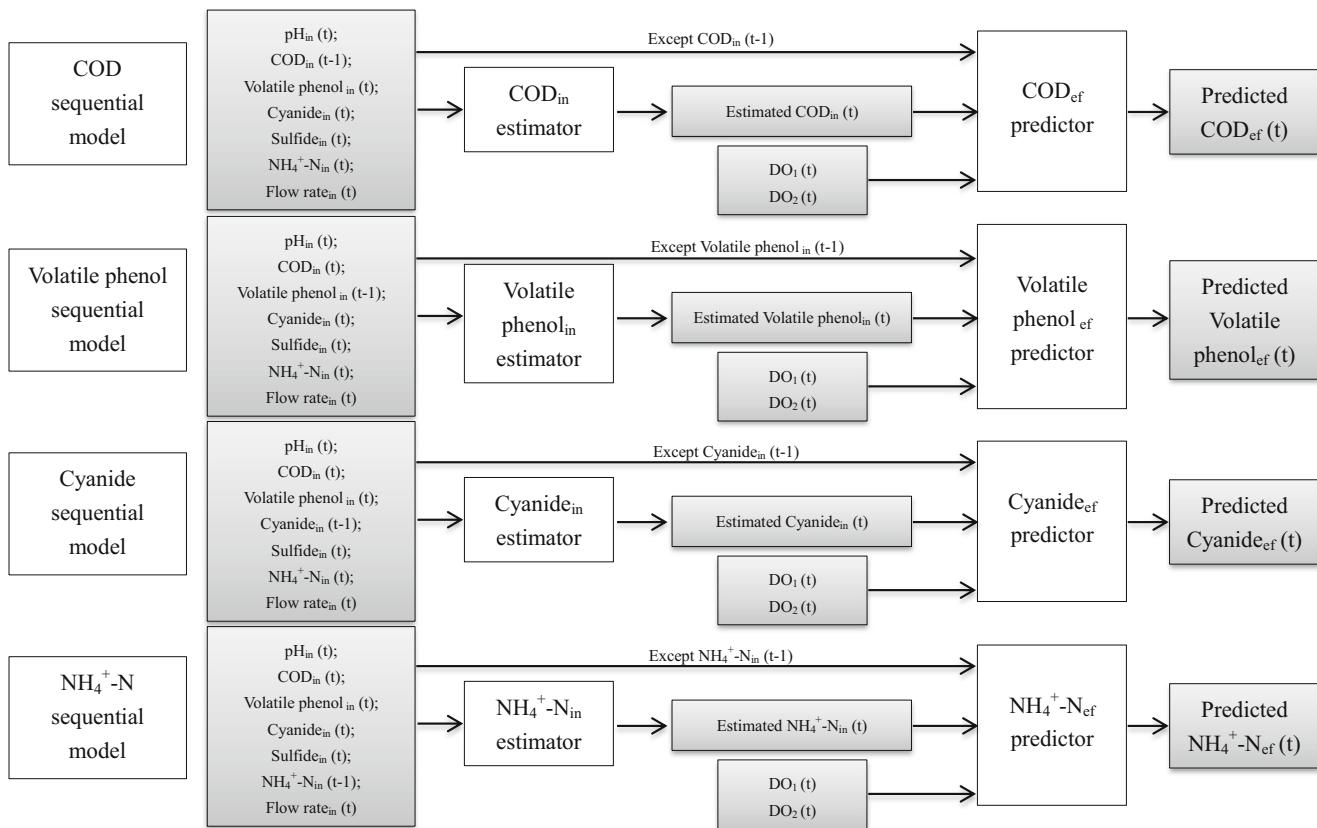


Fig. 2 Schematic diagram of the sequential ANN modeling processes

Estimation of biotreatment influent variables

Four independent BP-ANN models were developed to estimate the influent COD, volatile phenol, cyanide, and $\text{NH}_4^+ \text{-N}$

(Fig. 2). The structure of the network (number of layers, number of hidden nodes, learning rate, momentum values) was optimized during the training phase using GA. The validating data sets were used to guarantee good generalization ability of

Table 2 Optimized architecture of model network using GDX and BR algorithms

Predicted parameters	Training algorithm	Best network architecture	Iteration	Testing node numbers for hidden layer
COD_{in}	GDX	5:11:1	1659	2-40
	BR	5:6:1	672	
$\text{Volatile phenol}_{\text{in}}$	GDX	5:7:1	1974	2-40
	BR	4:4:1	364	
$\text{Cyanide}_{\text{in}}$	GDX	4:6:1	2561	2-40
	BR	3:4:1	921	
$\text{NH}_4^+ \text{-N}_{\text{in}}$	GDX	5:7:1	6359	2-40
	BR	5:6:1	568	
COD_{ef}	GDX	7:13:1	3164	2-50
	BR	8:10:1	1263	
$\text{Volatile phenol}_{\text{ef}}$	GDX	8:10:1	2469	2-50
	BR	7:10:1	684	
$\text{Cyanide}_{\text{ef}}$	GDX	6:10:1	1897	2-50
	BR	6:6:1	2415	
$\text{NH}_4^+ \text{-N}_{\text{ef}}$	GDX	8:12:1	6845	2-50
	BR	7:8:1	3695	

Two BP-ANN training algorithms, including gradient descent with adaptive learning rate (GDX) algorithm and Bayesian regularization (BR) algorithm, were used to train all models. The models were trained and tested until the best fitting network architectures were obtained by genetic algorithm

Table 3 ANN model performance statistics

Predicted parameters	Training algorithm	R^2	Training		Validating		Testing	
			MAE	RMSE	MAE	RMSE	MAE	RMSE
COD_{in}	GDX	0.716	92.6	112.6	93.2	110.9	216.7	243.8
	BR	0.871	105.8	127.7	119.4	140.4	126.6	148.2
$\text{Volatile phenol}_{\text{in}}$	GDX	0.813	21.64	25.14	22.16	24.64	41.62	58.62
	BR	0.904	22.94	28.62	23.16	25.16	27.64	32.37
$\text{Cyanide}_{\text{in}}$	GDX	0.842	0.16	0.24	0.21	0.27	0.73	0.82
	BR	0.846	0.21	0.26	0.24	0.31	0.31	0.35
$\text{NH}_4^+ \text{-N}_{\text{in}}$	GDX	0.679	3.9	4.3	4.6	5.1	12.6	14.3
	BR	0.777	4.8	5.5	5.1	5.9	5.9	6.6
COD_{ef}	GDX	0.816	16.9	18.4	17.9	19.8	49.6	59.9
	BR	0.852	18.1	19.6	20.4	21.7	23.6	26.9
$\text{Volatile phenol}_{\text{ef}}$	GDX	0.649	0.02	0.02	0.03	0.03	0.23	0.31
	BR	0.752	0.02	0.03	0.03	0.04	0.04	0.04
$\text{Cyanide}_{\text{ef}}$	GDX	0.791	0.08	0.08	0.10	0.12	1.42	1.62
	BR	0.844	0.09	0.10	0.12	0.13	0.14	0.16
$\text{NH}_4^+ \text{-N}_{\text{ef}}$	GDX	0.729	0.9	1.0	1.0	1.3	4.6	5.7
	BR	0.764	1.0	1.1	1.3	1.4	1.7	1.8

Two BP-ANN training algorithms, including gradient descent with adaptive learning rate (GDX) algorithm and Bayesian regularization (BR) algorithm, were used to train all models. The models were trained and tested until the best fitting network architectures were obtained by genetic algorithm. MAE and RMSE both had the unit of mg L^{-1} for COD, volatile phenol, cyanide, and $\text{NH}_4^+ \text{-N}$

the model. Then, the results in testing data sets were compared with evaluating the model performance. The best suited ANN models were optimized as 7:5:6:1 for COD_{in} , 7:4:4:1 for volatile phenol_{in}, 7:3:4:1 for cyanide_{in}, and 7:5:6:1 for $\text{NH}_4^+ \text{-N}_{\text{in}}$ corresponding node number of input, 1st hidden, 2nd hidden, and output layers, respectively (Fig. 2). The weight matrices of developed ANN models are listed in Tables S1–S4. The relative importance of variables was calculated according to Soleymani et al. (2011), and the results are presented in Fig. S1.

The best ANN modeling results for COD are shown in Figs. 3a and 4a. The optimized model for COD_{in} had RMSEs of 127.7 (6 %), 140.4 (7 %), and 148.2 mg L^{-1} (7 %) on the training, validating, and testing data sets, respectively (R^2 of 0.871 for overall data sets). The ANN results in Fig. 3a showed that the variation tendency of the measured concentrations can be well estimated, which indicated that the optimized COD_{in} estimator provided a satisfactory estimating performance.

The performance of the optimized model for volatile phenol_{in} was as good as that of COD_{in} . The RMSEs of the optimized volatile phenol_{in} estimation model for training, validating, and testing data sets were 28.62 (7 %), 25.16 (6 %), and 32.37 mg L^{-1} (8 %), respectively, and the R^2 value for overall data sets was 0.904. As shown in Fig. 3b, the estimated volatile phenol_{in} values were able to track the highly fluctuated observed volatile phenol_{in} values. Thus, statistical measures of accuracy (RMSEs and R^2 values) and visual

inspection indicated that volatile phenol_{in} concentration can also be well estimated. In addition, this simple model for the volatile phenol_{in} estimation also had good generalization capability without over-fitting, as demonstrated by low level of RMSEs (<10 %) between the observed and estimated testing data sets.

The same simulation procedure for the COD_{in} and volatile phenol_{in} estimations was applied to develop a process model for cyanide_{in} estimation. More specifically, it was impossible to measure real-time cyanide_{in} concentration because of lack of on-line sensors. The best ANN modeling result for cyanide_{in} is shown in Figs. 3c and 4c. Good correlation was obtained between the observed and estimated values with RMSEs of 0.26 (7 %), 0.31 (9 %), and 0.35 mg L^{-1} (10 %) for the training, validating, and testing data sets, respectively (R^2 value of 0.846 for overall data sets). Of note, the highly fluctuated observed cyanide_{in} values in data points 0–53 (Fig. 3c) and the continuous high values in data points 95–115 (Fig. 3c) were successfully estimated by this model.

The best ANN modeling results for $\text{NH}_4^+ \text{-N}$ are shown in Figs. 3d and 4d. Compared with COD_{in} , volatile phenol_{in}, and cyanide_{in}, the $\text{NH}_4^+ \text{-N}$ estimation model had lowest R^2 value at 0.777 for overall data sets. The optimized BP-ANN model for $\text{NH}_4^+ \text{-N}$ had RMSEs of 5.5 (10 %), 5.9 (11 %), and 6.6 mg L^{-1} (12 %) for training, validating, and testing data sets, respectively. The total variation tendency of observed $\text{NH}_4^+ \text{-N}_{\text{in}}$ was predicted by optimized ANN model (Fig. 3d).

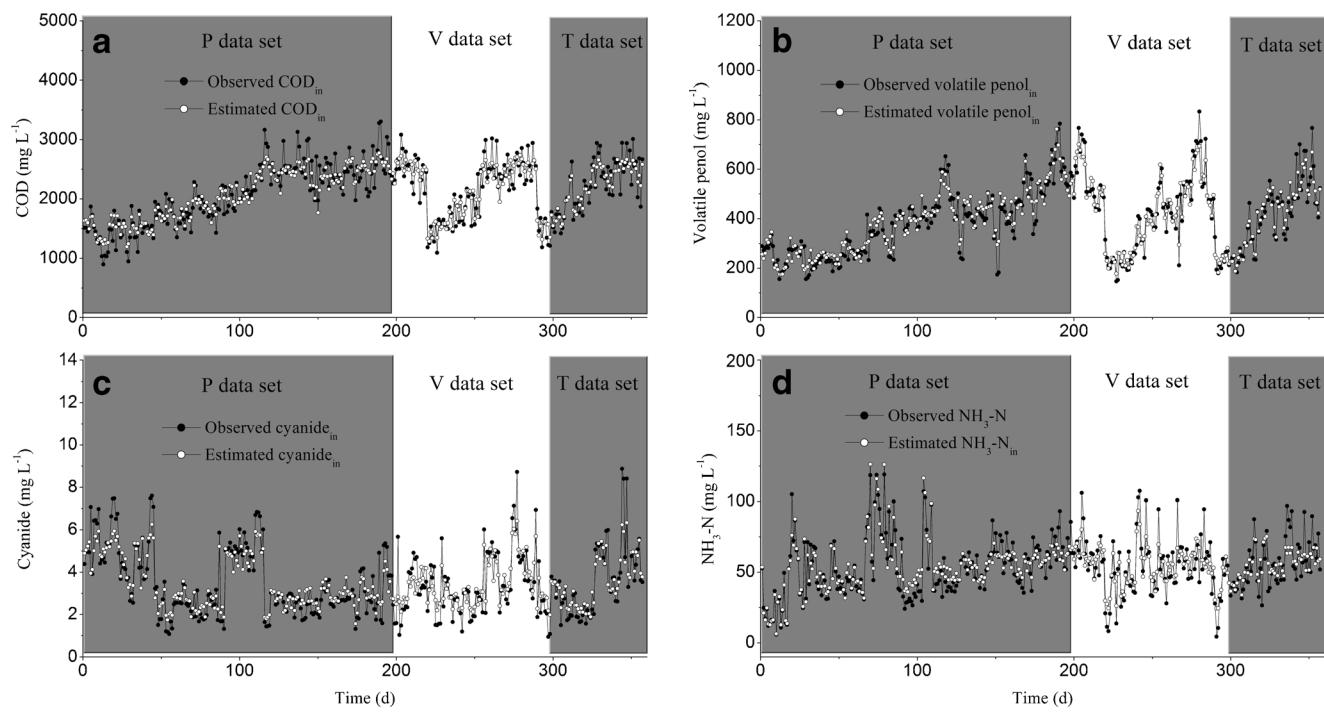


Fig. 3 Estimated and observed values of multiple parameters **a** COD_{in}, **b** volatile phenol_{in}, **c** cyanide_{in}, and **d** NH₄⁺-N_{in}

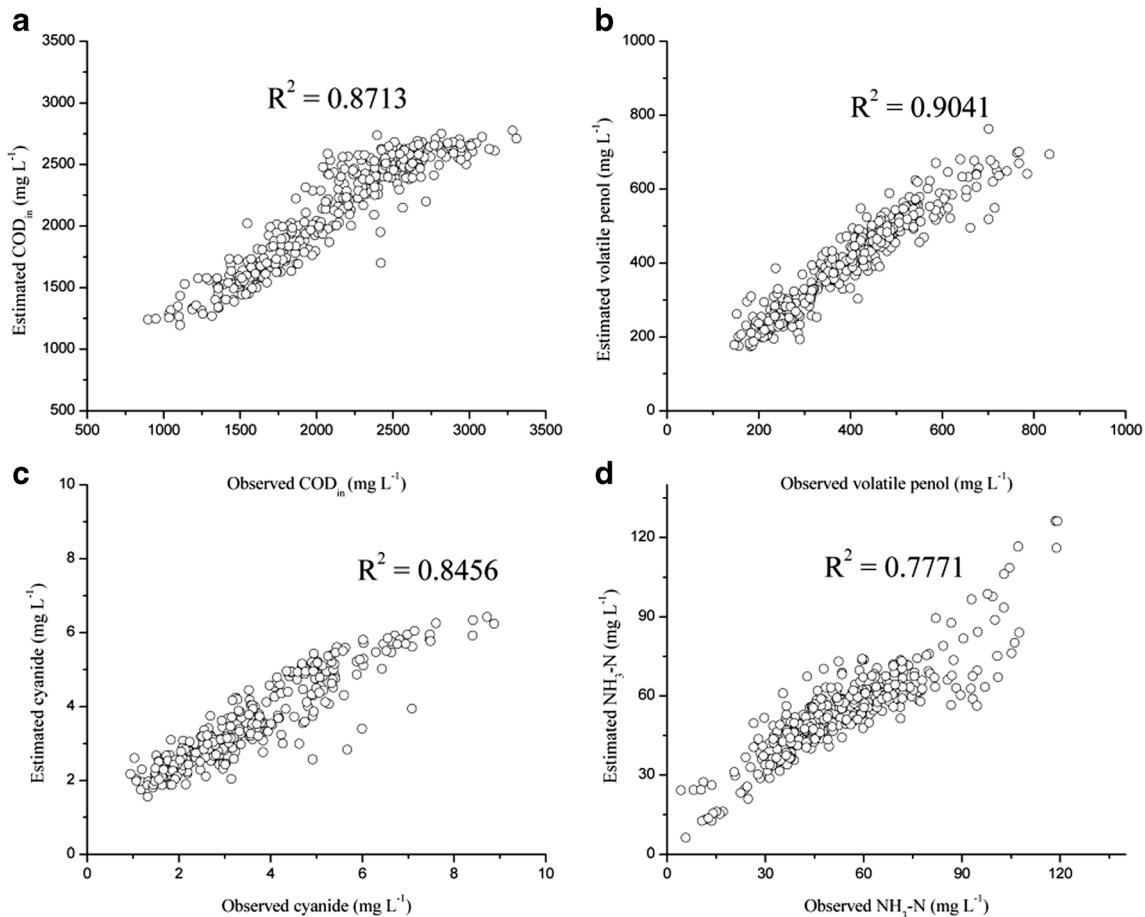


Fig. 4 Regression plots of estimated and observed values of multiple parameters **a** COD_{in}, **b** volatile phenol_{in}, **c** cyanide_{in}, and **d** NH₄⁺-N_{in}

However, the suddenly fluctuated values were not well estimated, resulting in high RMSEs and low R^2 .

Prediction of biotreatment effluent variables

The estimated concentrations of COD_{in} , $\text{cyanide}_{\text{in}}$, volatile phenol_{in}, and $\text{NH}_4^+ \text{-N}_{\text{in}}$, which were the outputs of process disturbance estimator, were subsequently fed into the process behavior predictor as part of the input with other variables to predict the concentrations of COD_{ef} , $\text{cyanide}_{\text{ef}}$, volatile phenol_{ef}, and $\text{NH}_4^+ \text{-N}_{\text{ef}}$ respectively (Fig. 2). The best suited model was optimized as 9:8:10:1 for COD_{ef} , 9:7:10:1 for volatile phenol_{ef}, 9:6:6:1 for $\text{cyanide}_{\text{ef}}$, and 9:7:8:1 for $\text{NH}_4^+ \text{-N}_{\text{ef}}$ corresponding node number of input, 1st hidden, 2nd hidden, and output layers, respectively. The weight matrices of developed ANN models are listed in Tables S5–S8, and the relative importance of variables is presented in Fig. S2. Furthermore, the residual distribution of the predicted values is presented in Fig. S3.

The observed and predicted values of COD_{ef} had RMSEs of 19.6 (8 %), 21.7 (9 %), and 26.9 mg L⁻¹ (11 %) for training, validating, and testing data sets (Fig. 5a). This model performed satisfactorily in terms of predicting COD_{ef} based on the estimated COD_{in} over the full data range (R^2 value of 0.852), as shown in Fig. 6a. This model showed slightly worse prediction performance (lower R^2 values) compared with the COD_{in} estimation model. However, the fluctuated COD

values were well estimated, indicating that this model can predict the COD_{ef} .

The lack of on-line monitoring sensor for toxic cyanide might induce high risk of CWWTP effluent. Therefore, the prediction of $\text{cyanide}_{\text{ef}}$ was important to guarantee the safety of effluent received water bodies. The performance of the BP-ANN model for $\text{cyanide}_{\text{ef}}$ prediction gave an RMSE of 0.10 mg L⁻¹ (8 %) on the training data set. The RMSEs for validating and testing data sets were 0.13 (11 %) and 0.16 mg L⁻¹ (13 %) (Fig. 5). The observed and predicted $\text{cyanide}_{\text{ef}}$ values on all three data sets are presented in Fig. 5c, and the regression plots of predicted and observed values are presented in Fig. 6c (R^2 of 0.844). The average differences between observed and predicted $\text{cyanide}_{\text{ef}}$ were low, indicating that this prediction model can predict $\text{cyanide}_{\text{ef}}$.

The lack of on-line volatile phenol monitoring sensor in CWWTP also resulted in risk of effluent. In volatile phenol_{ef} prediction model, the overall prediction performance was slightly worse (R^2 value of 0.752), and some data points were unpredicted compared with the observed volatile phenol_{ef} (data points 60–75, Fig. 5b). The performance of BP-ANN model for volatile phenol_{ef} prediction gave an RMSE of 0.116 mg L⁻¹ on the training data set. The RMSEs for validating and testing data sets were 0.04 (22 %) and 0.04 mg L⁻¹ (22 %) (Fig. 5). The differences between the estimated and measured data in volatile phenol_{ef} models were higher compared with COD_{ef} models (lower R^2 value), but still reflected the general

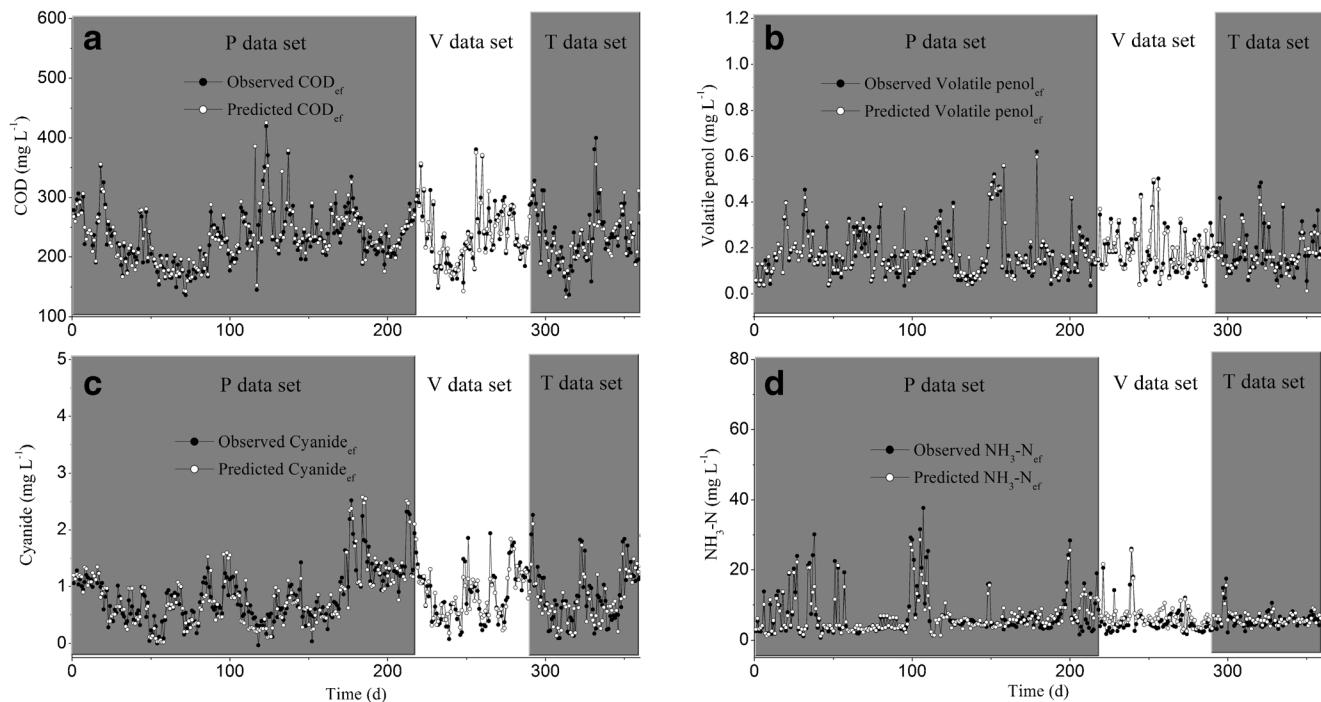


Fig. 5 Predicted and observed values of multiple parameters. **a** COD_{ef} , **b** volatile phenol_{ef}, **c** cyanide_{ef}, and **d** $\text{NH}_4^+ \text{-N}_{\text{ef}}$

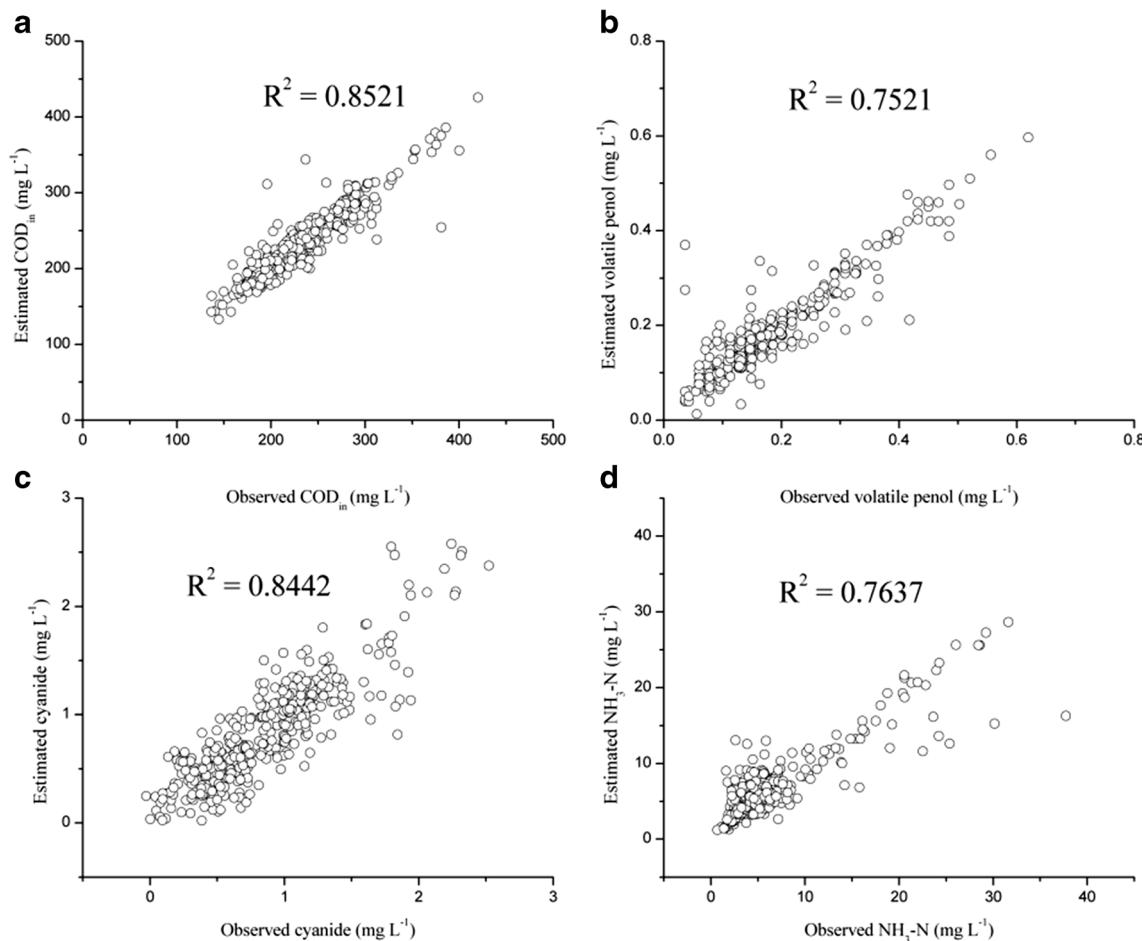


Fig. 6 Regression plots of predicted and observed values of multiple parameters. **a** COD_{ef}, **b** volatile phenol_{ef}, **c** cyanide_{ef} and **d** NH₄⁺-N_{ef}

variation tendency of volatile phenol_{ef} values. The poor fit of simulated and observed volatile phenol_{ef} can be attributed to the relative lower concentration (mean value at 0.23 mg L⁻¹) than COD (mean COD at 233 mg L⁻¹). Furthermore, the dissolved volatile phenols in CWW included phenol, cresol, and other phenol derived matters, which were characterized by high volatility. The atmospheric pressure, water temperature, and the aeration rate of fluidized bed might have effects on the removal of volatile phenols. Therefore, further improvement of the volatile phenol_{ef} prediction capability could be achieved using more detailed information about these operation parameters of fluidized bed.

Similarly, the optimized BP-ANN model for the prediction of NH₄⁺-N_{ef} was less accurate than those of COD_{ef} and cyanide_{ef}. The R^2 value for the overall data sets was 0.764. The optimized BP-ANN model for NH₄⁺-N_{ef} had RMSEs of 1.1 (17 %), 1.4 (24 %), and 1.8 mg L⁻¹ (27 %) on the training, validating, and testing data sets, respectively. The poor prediction of NH₄⁺-N_{ef} might be due to several reasons. The first one might be the intermittent incoming desulfurization wastewater, which contained higher dosage cyanide than continuous

distilled ammonia wastewater. The sudden inflow of desulfurization wastewater induced highly fluctuated cyanide_{in} values in data points 0–53 (Fig. 3c) and the continuous high level (4–8 mg L⁻¹) in data points 95–115 (Fig. 3c).

It was reported that cyanide had significant inhibitory effects on both ammonia-oxidizing bacteria and nitrite-oxidizing bacteria (Kim et al. 2011). Second, the high NH₄⁺-N content of influent might also be a possible reason to the poorer fit of NH₄⁺-N_{ef} model. It was observed that the NH₄⁺-N_{in} value was high in the range of data points 65–110 (NH₄⁺-N_{in} dosage >50 mg L⁻¹). The fluctuated cyanide and hydraulic load had impacts on the microorganism, such as nitrifying bacteria, in fluidized beds. NH₄⁺-N_{ef} dosage was high in the range of data points 0–55 and 95–110 (NH₄⁺-N dosage in the range of 10–40 mg L⁻¹), which reflected an inhibition of ammonia-oxidizing process. Of note, the predicted NH₄⁺-N_{ef} could not follow the observed values in these data ranges (Fig. 5d). Thus, the NH₄⁺-N_{ef} prediction accuracy could be improved using more detailed information about intermittent incoming desulfurization wastewater such as inflow period, flow rate, and the concentrations of cyanide_{in}.

Conclusions

In this study, a BP-ANN combined with sequential modeling approach and GA was used to develop dynamic models for the prediction of COD, volatile phenol, cyanide, and NH_4^+ -N removal in a full-scale CWWTP. The former parts of these models estimated the variations of influent COD, volatile phenol, cyanide, and NH_4^+ -N. The estimated values, with other measured parameters, were fed into the later parts of models which predicted the variations of effluent COD, volatile phenol, cyanide, and NH_4^+ -N. The results indicated that the influent estimator successfully developed dynamic models for COD, volatile phenol, cyanide and NH_4^+ -N, and the effluent predictor developed feasible models for effluent COD and cyanide prediction. Further research would be focused on the improvement of prediction capability for effluent volatile phenol and NH_4^+ -N by using more relevant variables.

Acknowledgments This work was supported by the State Key Program of National Natural Science Foundation of China (No. 21037001), the National Natural Science Foundation of China (No. 51308224), and the National Natural Science Foundation of China (No. 21377040).

References

- Aggarwal K, Singh Y, Chandra P, Puri M (2005) Bayesian regularization in a neural network model to estimate lines of code using function points. *J Comput Sci* 1:505–509
- Angelino S, Gennaro M (1997) An ion-interaction RP-HPLC method for the determination of the eleven EPA priority pollutant phenols. *Anal Chim Acta* 346:61–71
- APHA, AWWA, WEF, (1998) Standard methods for the examination of water and wastewater. American Public Health Association, Washington, DC
- Blumensaat F, Keller J (2005) Modelling of two-stage anaerobic digestion using the IWA Anaerobic Digestion Model No. 1 (ADM1). *Water Res* 39:171–183
- Bongards M (1999) Controlling biological wastewater treatment plants using fuzzy control and neural networks, Computational Intelligence. Springer, 142–150
- Dürrenmatt DJ, Gujer W (2012) Data-driven modeling approaches to support wastewater treatment plant operation. *Environ Model Softw* 30:47–56
- Goldberg DE (1989) Genetic algorithms in search, optimization, and machine learning. ReadingAddison, -Wesley
- Güclü D, Dursun S (2010) Artificial neural network modelling of a large-scale wastewater treatment plant operation. *Bioprocess Biosyst Eng* 33:1051–1058
- Hong YS BS, Charles T (1998) A genetic adapted neural network analysis of performance of the nutrient removal plant at Rotorua. The Institution of Professional Engineers New Zealand Annual Conference, Auckland
- Hong Y-ST, Rosen MR, Bhamidimarri R (2003) Analysis of a municipal wastewater treatment plant using a neural network-based pattern analysis. *Water Res* 37:1608–1618
- Hong SH, Lee MW, Lee DS, Park JM (2007) Monitoring of sequencing batch reactor for nitrogen and phosphorus removal using neural networks. *Biochem Eng J* 35:365–370
- Huang M, Wan J, Wang Y, Ma Y, Zhang H, Liu H, Hu Z, Yoo C (2012) Modeling of a paper-making wastewater treatment process using a fuzzy neural network. *Korean J Chem Eng* 29:636–643
- Jing JY, Feng J, Li WY (2009) Carrier effects on oxygen mass transfer behavior in a moving-bed biofilm reactor. *Asia Pac J Chem Eng* 4: 618–623
- Kim YM, Park D, Lee DS, Park JM (2007) Instability of biological nitrogen removal in a cokes wastewater treatment facility during summer. *J Hazard Mater* 141:27–32
- Kim YM, Park D, Lee DS, Jung KA, Park JM (2009) Sudden failure of biological nitrogen and carbon removal in the full-scale pre-denitrification process treating cokes wastewater. *Bioresour Technol* 100: 4340–4347
- Kim YM, Lee DS, Park C, Park D, Park JM (2011) Effects of free cyanide on microbial communities and biological carbon and nitrogen removal performance in the industrial activated sludge process. *Water Res* 45:1267–1279
- Lai P, Zhao H, Ye Z, Ni J (2008) Assessing the effectiveness of treating coking effluents using anaerobic and aerobic biofilms. *Process Biochem* 43:229–237
- Lee J-W, Suh C, Hong Y-ST, Shin H-S (2011) Sequential modelling of a full-scale wastewater treatment plant using an artificial neural network. *Bioprocess Biosyst Eng* 34:963–973
- Pal P, Kumar R (2014) Treatment of coke wastewater: a critical review for developing sustainable management strategies. *Separat Purif Rev* 43:89–123
- Pendashteh AR, Fakhru'l-Razi A, Chaibakhsh N, Abdullah LC, Madaeni SS, Abidin ZZ (2011) Modeling of membrane bioreactor treating hypersaline oily wastewater by artificial neural network. *J Hazard Mater* 192:568–575
- Riedmiller M (1994) Advanced supervised learning in multi-layer perceptrons—from backpropagation to adaptive learning algorithms. *Comput Standard Interf* 16:265–278
- Schulze F, Wolf H, Jansen H, Veer P (2005) Applications of artificial neural networks in integrated water management: fiction or future? *Water Sci Technol* 52:21–31
- Show K-Y, Lee D-J, Pan X (2013) Simultaneous biological removal of nitrogen–sulfur–carbon: recent advances and challenges. *Biotechnol Adv* 31:409–420
- Soleymani AR, Saen J, Bayat H (2011) Artificial neural networks developed for prediction of dye decolorization efficiency with UV/ $\text{K}_2\text{S}_2\text{O}_8$ process. *Chem Eng J* 170:29–35
- Zhang W, Wei C, Chai X, He J, Cai Y, Ren M, Yan B, Peng P, Fu J (2012) The behaviors and fate of polycyclic aromatic hydrocarbons (PAHs) in a coking wastewater treatment plant. *Chemosphere* 88:174–182
- Zhu X, Tian J, Liu R, Chen L (2011) Optimization of Fenton and electro-Fenton oxidation of biologically treated coking wastewater using response surface methodology. *Sep Purif Technol* 81:444–450