



## Process Systems Engineering and Process Safety

A data-derived soft-sensor method for monitoring effluent total phosphorus<sup>☆</sup>Shuguang Zhu<sup>1,2,3,4</sup>, Honggui Han<sup>1,2,\*</sup>, Min Guo<sup>1,3,4</sup>, Junfei Qiao<sup>1,2</sup><sup>1</sup> College of Electronic Information & Control Engineering, Beijing University of Technology, Beijing 100124, China<sup>2</sup> Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124, China<sup>3</sup> Engineering Research Center of Digital Community, Ministry of Education, Beijing 100124, China<sup>4</sup> Beijing Laboratory for Urban Mass Transit, Beijing 100124, China

## ARTICLE INFO

## Article history:

Received 10 November 2016

Received in revised form 7 April 2017

Accepted 12 June 2017

Available online 1 August 2017

## Keywords:

Data-derived soft-sensor

Effluent total phosphorus

Wastewater treatment process

Radial basis function neural network

Partial least square method

## ABSTRACT

The effluent total phosphorus (ETP) is an important parameter to evaluate the performance of wastewater treatment process (WWTP). In this study, a novel method, using a data-derived soft-sensor method, is proposed to obtain the reliable values of ETP online. First, a partial least square (PLS) method is introduced to select the related secondary variables of ETP based on the experimental data. Second, a radial basis function neural network (RBFNN) is developed to identify the relationship between the related secondary variables and ETP. This RBFNN easily optimizes the model parameters to improve the generalization ability of the soft-sensor. Finally, a monitoring system, based on the above PLS and RBFNN, named PLS-RBFNN-based soft-sensor system, is developed and tested in a real WWTP. Experimental results show that the proposed monitoring system can obtain the values of ETP online and own better predicting performance than some existing methods.

© 2017 The Chemical Industry and Engineering Society of China, and Chemical Industry Press. All rights reserved.

## 1. Introduction

During the recent decades, the increasing awareness about the negative impact of eutrophication to the qualities of water bodies has led to more stringent wastewater treatment requirements and regulations [1]. In wastewater treatment process (WWTP), Phosphorus (P) is of particular interest because of the major concern for elevated levels of P in many natural waterways. Along with nitrogen, excessive P concentrations will promote eutrophication and the associated detrimental environmental impacts. To control eutrophication, a major requirement for improving the performance relies on the online measurements of effluent total phosphorus (ETP) [2,3]. However, the online values of ETP are hard to measure [4,5].

To measure the values of ETP online, the physicochemical detection methods have been developed [6]. For example, an ion chromatography was developed to detect the values of ETP in the oil wastewater [7]. A compact portable flow analysis system, including an ultra-violet photo-reactor and a spectrophotometry method, was introduced for the rapid measurement of ETP in [8]. Moreover, a microfabricated cobalt electrode was designed for detecting the values of ETP in [9]. For these above methods [7–9], the results show that they could obtain the values

of ETP online with suitable accuracy in the laboratory conditions or some special detecting stations [10]. However, there are several limitations of these methods, such as expensive capital and maintenance costs, as well as time-delayed responses [11,12].

In recent years, since the large amount of process data have been routinely measured and collected in modern WWTPs, the data-derived soft-sensor method has been considered as an interesting alternative for measuring the values of water qualities online [13]. A data-derived soft-sensor method is an input–output model, where the inputs usually consist of easy-to-measure variables in the form of plant's signals. For example, a self-validating soft-sensor method was proposed to predict the values of biological oxygen demand (BOD) with the capability of performing self-diagnostics, self-reconstruction and online uncertainty measurement [14]. In addition, a soft-sensor, based on the multi-sensor water quality monitoring system, was used to monitor the chemical oxygen demand (COD) and total suspended solids (TSS) concentrations of the effluents in a real WWTP [15]. Due to the unknown input–output relationship, neural networks, originally inspired by the abilities of human beings to perform many complicated tasks with ease, have been usually used to model complex relationships between inputs and outputs [16]. For example, an adaptive network-based fuzzy inference system was developed to predict suspended solids (SS) and effluent COD of a paper mill wastewater treatment process [17]. A soft-sensor, combining the multi-layer neural network and the multi-way principal components analysis, was developed for predicting the values of ETP in a full-scale WWTP [18]. The results show that this soft-sensor can improve the predicting capability. However, in most real WWTPs, most of the input variables, such as the

<sup>☆</sup> Supported by the National Science Foundation of China (61622301, 61533002), Beijing Natural Science Foundation (4172005), and Major National Science and Technology Project (2017ZX07104).

\* Corresponding author at: College of Electronic Information & Control Engineering, Beijing University of Technology, Beijing 100124, China.

E-mail addresses: Zhushuguang@emails.bjut.edu.cn (S. Zhu), recharhdhan@bjut.edu.cn (H. Han), guomin@bjut.edu.cn (M. Guo).

influent TSS, influent BOD, and influent total phosphorous, cannot be measured online. To predict the values of ETP with some easy-to-measure variables, a soft-sensor, based on a feed-forward back-propagation neural network (FBNN), was designed and applied to a small scale municipal WWTP [19]. The results demonstrate that the proposed soft-sensor method can significantly improve the estimation accuracy than some existing methods. However, since the soft-sensor method combines the FBNN with auto-regressive exogenous (ARX), it is time-consuming to adjust the parameters of the soft-sensor method [20]. Although neural networks have been successfully used in predicting the effluent qualities of WWTPs [21–23], the parameters adjustment of the soft-sensor methods is still an open problem [24,25].

In this study, a data-derived soft-sensor method, including a partial least square (PLS) method and a radial basis function neural network (RBFNN), is investigated to measure the values of ETP online. The PLS method is used to select the most suitable process variables for the data-driven soft-sensor method to rise the accuracy and reduce computing time. Meanwhile, the RBFNN algorithm is developed to model complex relationships between the process variables and ETP. Moreover, based on the soft-sensor method, an online monitoring system is designed and tested in a real WWTP. The results demonstrate that the proposed online monitoring system owns better performance than the results in [7,8,26].

## 2. Methods

### 2.1. Hardware setup

The experimental setup is shown in Fig. 1 based on a real anaerobic anoxic oxic ( $A^2/O$ ) WWTP (about  $120 \text{ m}^3 \cdot \text{d}^{-1}$ ). It is located at the R&D center of Gaobeidian WWTP which is one of the largest municipal WWTP in China. The  $A^2/O$  WWTP consists of two anaerobic tanks, two anoxic compartments, and four oxic compartments. Meanwhile, the mechanical mixers are installed in anaerobic and anoxic tanks to ensure well-mixed conditions. The influent tank is used to subside the content of suspended solids by gravity, and the settler tank is designed to remove the flocs of biological growth. Moreover, the sludge will be removed from the bottom of settler tank if the effluent water is discharged from the surface of the same tank.

In this study, the online monitoring system consists of two parts: the online instruments and the soft-sensor method. The online instruments are used to measure the easy-to-measure variables from different tanks of the  $A^2/O$  process, including the dissolved oxygen concentration (DO), oxidation-reduction potential (ORP), TSS, pH, ammonium nitrogen ( $\text{NH}_4\text{-N}$ ), nitrate nitrogen ( $\text{NO}_3\text{-N}$ ), and temperature ( $T$ ). All the instruments are operated in a continuous measurement mode to ensure the frequency of data collection, and the measured data are stored in their own memories. Since importing data through the SD-Card or the USB ports is time-consuming and low efficiency, a data recorder is installed to record the data from the online instruments and directly import the data into the computer automatically. In addition, since there is no direct contact between the wastewater treatment process and the data recorder, the transmission error can be ignored and the maintenance cost is acceptable. Further, following the data recorder, a soft-sensor mode, embedding the PLS method and the RBFNN together, is designed for predicting the values of ETP.

### 2.2. Data-derived soft-sensor

The proposed data-derived soft-sensor can be described as an input–output process model, in which the inputs consist of easy-to-measure variables and the output is the online predicting values of ETP. In this soft-sensor model, the online instruments will consistently collect and transmit new data to the soft-sensor method. Then the soft-sensor will update its parameters to follow the change of data online. The steps and methodologies of the data-derived soft-sensor method are shown in Fig. 2, and the main techniques are briefly introduced in the following section.

#### 2.2.1. Data collection and data transmission

The first step in the proposed data-derived soft-sensor method is the data collection, which is to obtain the process data. To ensure the consistency and authenticity of the input data, the process data from this  $A^2/O$  WWTP are routinely recorded and stored in the database.

In this data-derived soft-sensor system, a hardware system, including the online instruments and the transmission component, is set up (refer to Fig. 1). Nine process variables (as well as their collecting points, main apparatus and instruments, and sampling frequency), measured from the  $A^2/O$  WWTP, are listed in Table 1. The sensors of

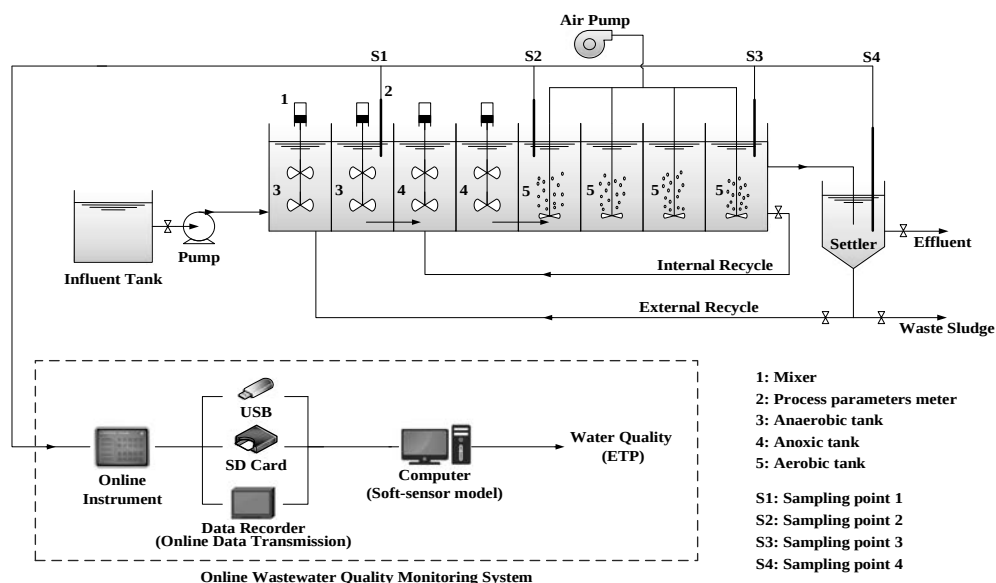


Fig. 1. Schematic diagram of the monitoring system.

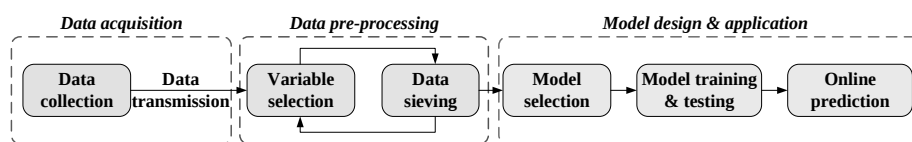


Fig. 2. The typical steps and methodologies of the data-derived soft-sensor.

Table 1

The information of the process variables in the A<sup>2</sup>/O WWTP

| Parameters         | Unit               | Collecting points          | Collecting sensors | Sampling frequency |
|--------------------|--------------------|----------------------------|--------------------|--------------------|
| Temperature        | °C                 | Anaerobic tank (S1)        | SensoLyt700IQ      | 1 min              |
| ORP <sub>1</sub>   | mV                 | Anaerobic tank (S1)        | SensoLyt700IQ      | 5 min              |
| DO <sub>1</sub>    | mg·L <sup>-1</sup> | Oxic tank (first tank, S2) | TriOxmatic700IQ    | 5 min              |
| DO <sub>2</sub>    | mg·L <sup>-1</sup> | Oxic tank (last tank, S3)  | TriOxmatic700IQ    | 5 min              |
| TSS                | g·L <sup>-1</sup>  | Oxic tank (last tank, S3)  | ViSolid700IQ       | 5 min              |
| ORP <sub>2</sub>   | mV                 | Settler (S4)               | SensoLyt700IQ      | 5 min              |
| pH                 | -                  | Settler (S4)               | SensoLyt700IQ      | 1 min              |
| NH <sub>4</sub> -N | mg·L <sup>-1</sup> | Settler (S4)               | AmmoLyt700IQ       | 5 min              |
| NO <sub>3</sub> -N | mg·L <sup>-1</sup> | Settler (S4)               | NitraLyt700IQ      | 5 min              |

the process variables are operated in a continuous/online measurement mode in this study. Meanwhile, the real values of ETP are collected by the laboratory analysis, and the sampling frequency of ETP is 5 min. Moreover, a data transmitting system has been designed to transmit the data directly between the sensors and the PC-computer.

## 2.2.2. Variable selection

In WWTPs, the data of the plants are redundant and possibly insignificant to calculate the measurements of the hard-to-measure variables. Moreover, in many situations, the computational capability of the soft-sensor methods cannot handle too many inputs. The quantity of the data with high-dimensionality restricts the development of soft-sensor methods. Therefore, it is necessary to pre-process the data before they are processed by a soft-sensor. A possible approach to overcome the dimensionality problem is variable selection, which is used to choose the easy-to-measure secondary variables that are most informative for the process, as well as the variables that provide high generalization ability.

The variable selection is to eliminate the useless process variables and choose the important variables, which is a crucial stage for soft sensor methods since a model with too many inputs may lead to over fitting and time consuming. In fact, the most commonly used techniques for the variable selection are multivariate linear techniques [27]. For example, the principal component analysis (PCA) [28,29] and PLS [30,31] have been widely applied to obtain a low-dimensional subspace. PCA can extract feature information from the original data source, and form a new set of uncorrelated data with lower dimension. However, the PCA technique fails to extract the explanatory information of the input variables corresponding to the model output. On the other hand, PLS can extract principle components from inputs to form a new matrix (score matrix) that are maximally uncorrelated among themselves and maximally linear dependent between the input and output variables [32]. In addition, the output can be represented by the original input variables after the essential components are chosen. Thus, the PLS technique is utilized to select the secondary variables for predicting the values of ETP in the proposed data-derived soft-sensor method.

Suppose that there is a data set  $\{X, y\}$ , where  $X$  is an  $n \times \alpha$  data matrix with  $\alpha$  being the number of independent variables and  $n$  the number of samples, and  $y$  is the corresponding dependent variable vector with size  $n \times 1$ . The underlying assumption of PLS is that the observed data are generated by a system or process that is driven by a small number of

latent variables. In the following, PLS adopts a two-step strategy to generate a functional relationship between  $X$  and  $y$ :

$$X = TP^T + E = \sum_{i=1}^{\alpha} t_i p_i^T + E, \quad (1)$$

$$y = UQ^T + F = \sum_{i=1}^{\alpha} u_i q_i^T + F, \quad (2)$$

where  $T$ ,  $P$  and  $E$  are the score matrix, loading matrix and residual matrix of  $X$  block.  $U$ ,  $Q$  and  $F$  are the score matrix, loading matrix and residual matrix of  $y$ ,  $t_i$ ,  $p_i$ ,  $u_i$  and  $q_i$  ( $i = 1, 2, \dots, \alpha$ ) are the corresponding vectors of  $T$ ,  $P$ ,  $U$  and  $Q$ .  $\alpha$  is the total number of available secondary variables.

The above two equations formulate a PLS outer model. A least squares regression is subsequently performed on the extracted orthogonal latent variables. The input and output score vectors are then related by:

$$\hat{u}_i = b_i t_i, \quad (3)$$

where

$$b_i = \frac{u_i^T t_i}{t_i^T t_i}, \quad (4)$$

$b_i$  is the  $i$ th regression coefficient, and the vector of the regression coefficients is  $b = [b_1, b_2, \dots, b_{\alpha}]^T$ . Then, the predictors for the output variable can be selected in descending order of the magnitude of  $b$ :

$$R_{\text{select}} = \frac{\|b_{\text{select}}\|}{\|b\|}, \quad (5)$$

where  $b_{\text{select}}$  is the vector of the regression coefficients of the selected variables,  $R_{\text{select}}$  is the importance of the selected variables. More predictors will be selected when the threshold of  $R_{\text{select}}$  is larger.

## 2.2.3. Model design

In the model design process, the model structure will fit the specific application task, and the model parameters will determine the generalization abilities of the data-derived soft-sensor. Due to their simple topological structure and universal approximation ability, the RBFNN algorithm has been widely used in nonlinear system modeling and

control [33]. Moreover, the RBFNN algorithm does not need any assumptions about the functional relationship between the dependent and independent variables [34]. Besides, the RBFNN algorithm has been proved with high accuracy, robustness to noise, as well as high computational efficiency [35,36]. Thus, a RBFNN is used to design the soft-sensor model in this study.

As shown in Fig. 3, the structure of the basic RBFNN consists of one input layer, one output layer, and one hidden layer. To better illustrate the predicting method, a multi-input single-output (MISO) RBFNN model is used. A single output RBFNN with  $K$  hidden neurons can be described as follows

$$\hat{y}(t) = \sum_{k=1}^K w_k(t) \theta_k(\mathbf{x}(t)), \quad (6)$$

where  $\mathbf{x}$  and  $\hat{y}$  are the input and output of RBFNN respectively,  $w_k$  is the connecting weight between the  $k$ th hidden neuron and the output neuron, and  $\theta_k$  is the output function of the  $k$ th hidden neuron, which can be described as

$$\theta_k(\mathbf{x}(t)) = e^{-\|\mathbf{x}(t) - \boldsymbol{\mu}_k(t)\| / \sigma_k^2}. \quad (7)$$

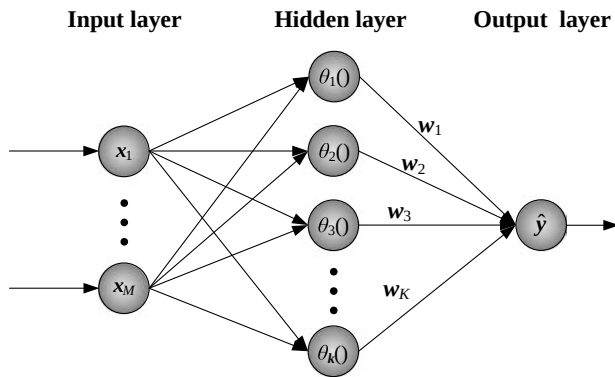


Fig. 3. The structure of a RBFNN.

In the above formula,  $\|\mathbf{x}(t) - \boldsymbol{\mu}_k(t)\|$  should be changed to  $\|\mathbf{x}(t) - \boldsymbol{\mu}_k(t)\|^2$  where  $\boldsymbol{\mu}_k$  is the center vector of the  $k$ th neuron, and  $\|\mathbf{x}(t) - \boldsymbol{\mu}_k(t)\|$  indicates the Euclidean distance between  $\mathbf{x}$  and  $\boldsymbol{\mu}_k$ ,  $\sigma_k$  is the width of the  $k$ th neuron,  $e(t)$  is the current predicting error of RBFNN

$$e(t) = \hat{y}(t) - y(t), \quad (8)$$

$\hat{y}(t)$  is the output of the neural network and  $y(t)$  is the system output for the current input sample  $\mathbf{x}$  at time  $t$ . The training algorithm for  $\boldsymbol{\mu} = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_K]^T$ ,  $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \dots, \sigma_K]^T$ , and  $\mathbf{w} = [w_1, w_2, \dots, w_K]^T$  is similar to the description in [37], and is omitted here.

The performance of the soft-sensor method is evaluated by the root mean-square -error (RMSE), which is defined as

$$RMSE(T) = \sqrt{\frac{1}{T} \sum_{t=1}^T (y(t) - \hat{y}(t))^2}, \quad (9)$$

where  $T$  is the number of samples for the test set. The predicting performance of RBFNN will be discussed in the next section.

Meanwhile, the percentage predicting accuracy is another key criterion to evaluate the performance of the soft-sensor method

$$p(T) = \frac{1}{T} \sum_{t=1}^T (1 - e(t)/y(t)) \times 100\%, \quad (10)$$

where  $p(T)$  is the predicting accuracy and  $e(t)$  is the error between the real output  $y(t)$  and the output of RBFNN  $\hat{y}(t)$ .

### 2.3. Experimental procedures

In summary, the procedure of the proposed PLS-RBFNN-based soft-sensor method includes the following steps:

- Step 1: In the beginning of the procedure, routinely acquire and store the historical and online process data in the data acquisition system of the plant.
- Step 2: Preprocess the data to deal with missing values, and scale each variable to have zero mean and unit variance.
- Step 3: Select the secondary variables by the PLS technique.
- Step 4: Divide the data into the training samples and testing samples.
- Step 5: Set the initial parameters of RBFNN randomly. Train and test the RBFNN-based model by using the adaptive computation algorithm in [35] with different sets of data to find the best RBFNN-based model and ensure its performance.
- Step 6: Complete the model design of the PLS-RBFNN-based soft-sensor.
- Step 7: Apply the proposed PLS-RBFNN-based soft-sensor to a real WWTP to predict the ETP values.
- Step 8: Update the PLS-RBFNN-based soft-sensor regularly until reach the stop situation.

Following the above procedures, the proposed data-derived soft-sensor method is then applied to predict the values of ETP in a real WWTP. For our proposed PLS-RBFNN-based soft-sensor, the convergence of RBFNN with respect to the parameter adjusting is an important issue and needs careful investigation. When the proposed PLS-RBFNN-based method is applied to a real WWTP, new data of ETP will be acquired through time, and the parameter adjusting strategy would be crucial for a successful application. As shown in [37], the convergence of the proposed RBFNN can be demonstrated. This means that the learning algorithm proposed in this paper can guarantee the convergence of the PLS-RBFNN-based soft-sensor when new data of ETP are collected and used for model updating.

### 3. Experiment Studies and Results

This section presents and compares the results of the variable selection and the predicting performance in two cases: one with all the secondary variables and the other with the selected secondary variables. Moreover, the performance of the proposed PLS-RBFNN-based soft-sensor method is compared with some other existing methods.

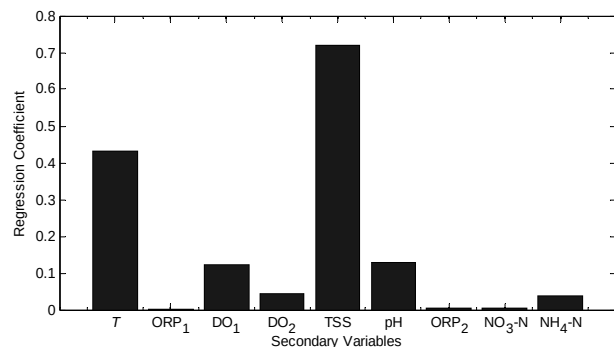


Fig. 4. The regression coefficient of each parameter shown in Table 1.

**Table 2**

The range of values of each process variables

| Variables       | $T/^{\circ}\text{C}$ | ORP <sub>1</sub> /mV | DO <sub>1</sub> /mg·L <sup>-1</sup> | DO <sub>2</sub> /mg·L <sup>-1</sup> | TSS/g·L <sup>-1</sup> | pH    | ORP <sub>2</sub> /mV | NO <sub>3</sub> -N/mg·L <sup>-1</sup> | NH <sub>4</sub> -N/mg·L <sup>-1</sup> |
|-----------------|----------------------|----------------------|-------------------------------------|-------------------------------------|-----------------------|-------|----------------------|---------------------------------------|---------------------------------------|
| Range of values | 22–26                | (-100)–(-310)        | 0.5–1.7                             | 1–3                                 | 2–3.5                 | 6.2–7 | (-150)–50            | 2–6                                   | 2–5                                   |

### 3.1. Variable selection results

As shown in Table 1, there are total nine process variables whose values were collected directly by the online instruments. After the abnormal data were removed, a dataset containing 800 samples, corresponding to the real process operation (from 1st September 2014 to 3rd September 2014), is selected.

In order to unify the time series of the online and laboratory data, the sampling frequency of all parameters is 5 min. The dataset is then separated into a training set with 720 samples and a test set with 80 samples. The process variables that are used as the inputs of the soft-sensor are selected according to the PLS technique. Fig. 4 shows the results of the regression coefficients of the nine process variables. Moreover, the value ranges/dispersions of the process variables are shown in Table 2. As seen in Fig. 4, the regression coefficients of  $T$ , TSS, pH, DO<sub>1</sub> and DO<sub>2</sub> are larger than the other the regression coefficients of ORP<sub>1</sub>, ORP<sub>2</sub>, NH<sub>4</sub>-N, and NO<sub>3</sub>-N. To show the regression coefficients more clearly, the details are listed in Table 3.

**Table 3**

Regression coefficient of every process variables

| Variables              | $T$          | ORP <sub>1</sub> | DO <sub>1</sub> | DO <sub>2</sub> | TSS          | pH           | ORP <sub>2</sub> | NO <sub>3</sub> -N | NH <sub>4</sub> -N |
|------------------------|--------------|------------------|-----------------|-----------------|--------------|--------------|------------------|--------------------|--------------------|
| Regression coefficient | <b>0.435</b> | 0.013            | <b>0.120</b>    | 0.046           | <b>0.719</b> | <b>0.127</b> | 0.019            | 0.022              | 0.033              |

Based on the analysis results in [31], the threshold of  $R_{\text{select}}$  is generally set between 0 and 1. Based on the features of the process data and the characteristics of WWTP, the value of  $R_{\text{select}}$  is set as 0.85 in this paper. As a result, the selected secondary variables are:  $T$ , TSS, pH, and DO<sub>1</sub>. Therefore, four process variables are selected as the secondary variables for predicting the values of ETP in the proposed soft-sensor method.

### 3.2. Predicting performance

In this work, the predicting performance is evaluated and compared by two cases: one with all the secondary variables (Case 1) and the other with the selected secondary variables (Case 2). In Case 1, a 9-20-1 RBFNN is built for predicting the values of ETP, and in Case 2, a 4-10-

1 RBFNN is generated. The number of neurons in the input layer is the same as the number of input variables. In addition, the number of neurons in the hidden layer is determined by the experimental results. Figs. 5 and 6 show the predicting performance of the proposed PLS-RBFNN-based soft-sensor in Case 1 and Case 2, respectively.

In order to evaluate the performances in Cases 1 and 2, the experimental results are assessed and compared in terms of computational time, the RMSE as defined in Eq. (8) and the predicting accuracy ( $p$ ) as defined in Eq. (9). The details are presented in Table 4. The results in Table 4 show that the predicting performance of the proposed PLS-RBFNN-based soft-sensor in Case 2 is better than that in Case 1.

Moreover, the proposed PLS-RBFNN-based soft-sensor is compared with the other two methods: the compact portable flow analysis system using ultra-violet photo-reactor and the spectrophotometry method [8] and a soft-sensor method combining multi-layer neural network and multi-way PCA [18]. To make the results comparable, all experiments have been performed 20 times using the same data. The details are shown in Table 5 in terms of the computational time and RMSE. For the first method, the results show that the computational time of the compact portable flow analysis system is bigger than that of the PLS-RBFNN-based soft-sensor method. In addition, the predicting accuracy the compact portable flow analysis system is lower than that of the proposed method.

For the second method, the results illustrate that the performance of the soft-sensor method based on MPCA and multi-layer neural network is better than that of the compact portable flow analysis system. However, the computational time and the RMSE are worse than those of the proposed PLS-RBFNN-based soft-sensor method. Moreover, the PLS-RBFNN-based soft-sensor method can save almost 15 s to obtain a reportable result in the continuous measurement mode. This means that it can obtain over 210 measurement results per hour. Further, the results demonstrate that the ETP trends in WWTP can be predicted with acceptable accuracy using  $T$ , TSS, pH, and DO<sub>1</sub> as the input variables.

## 4. Results Analysis

### 4.1. Results analysis

Figs. 5 and 6 show that the proposed PLS-RBFNN-based soft-sensor method can predict the values of ETP online. Further, the performance

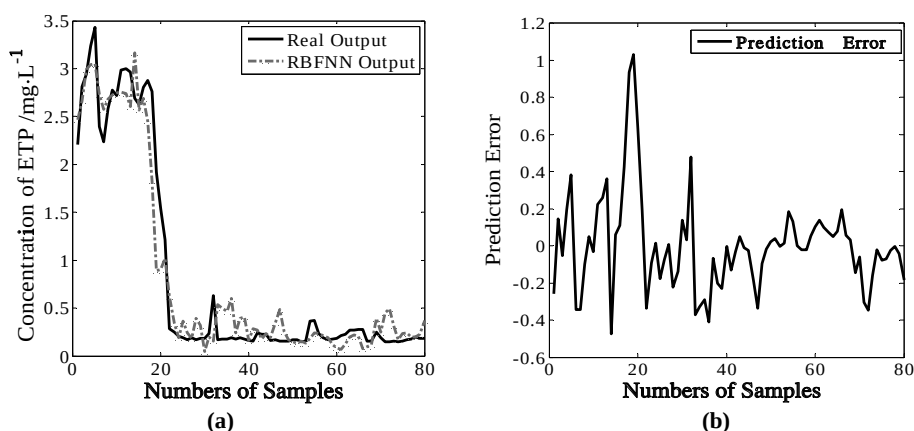


Fig. 5. The predicting performance in Case 1: (a) concentration of ETP, (b) predicting error.

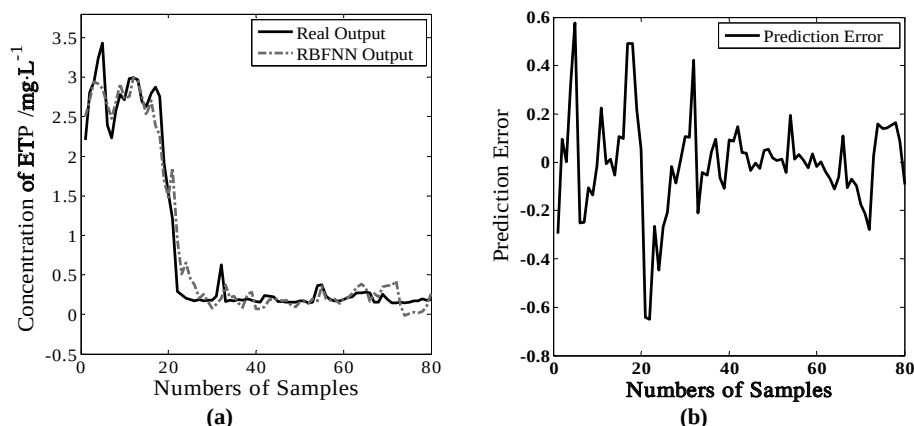


Fig. 6. The predicting performance in Case 2: (a) concentration of ETP, (b) predicting error.

Table 4

Comparison of the performance with different inputs (All results were averaged on 20 independent runs)

| Input numbers | Computational time/s | RMSE         | Predicting accuracy |
|---------------|----------------------|--------------|---------------------|
| 9             | 43.6                 | 0.237        | 75%                 |
| <b>4</b>      | <b>16.8</b>          | <b>0.104</b> | <b>83%</b>          |

Table 5

Comparison of the performance with different methods (All results were averaged on 20 independent runs)

| Methods                                   | Computational time/s | RMSE         |
|-------------------------------------------|----------------------|--------------|
| <b>PLS-RBFNN-based soft-sensor method</b> | <b>16.8</b>          | <b>0.104</b> |
| Flow analysis system [8]                  | 31.3 <sup>①</sup>    | –            |
| MPCA-MLNN-based soft-sensor method [16]   | 30.1                 | 0.530        |

<sup>①</sup> The results are listed in the original papers.

of the proposed PLS-RBFNN-based soft-sensor method in Case 2 is better than in Case 1, which means that the model with selected secondary variables has smaller predicting error than the initial process variables. The results in Table 3 demonstrate that the proposed soft-sensor method with the selected secondary variables can obtain better predicting accuracy with less computational time and smaller RMSE, since it only takes 16 s to get a reportable result with an average accuracy rate of 83%.

Moreover, compared to the other methods, the results in Table 5 show that the proposed PLS-RBFNN-based soft-sensor method can achieve better predicting accuracy and the computational time. From a practical point of view, the predicting values with large errors undoubtedly have a serious impact on the performance of further quality control loops and deteriorate the performance of WWTP. Since the predicting values of ETP will act as the feedback signal for the control loops. In addition, another practical issue of a soft-sensor is its online computational cost for a real application. Once the online instruments collect a group of new samples, they are immediately fed to the soft-sensor as inputs. However, it usually takes some time for computers to calculate the online inference of quality indices. If the soft-sensor method is too complicated, the computation time may be considerably long. The results in Table 5 demonstrate that the proposed PLS-RBFNN-based soft-sensor requires much less computational time than the other methods.

## 5. Conclusions

The aim of this paper has been to develop a new data-derived soft-sensor for predicting the values of ETP online in a real A<sup>2</sup>/O WWTP.

The PLS technique and RBFNN algorithm were designed to develop the soft-sensor. The learning algorithm of PLS-RBFNN-based soft-sensor could handle the characteristics of the nonlinear system. Further, compared with the other methods, the proposed PLS-RBFNN-based soft-sensor method can predict the values of ETP online with suitable accuracy. Meanwhile, the monitoring system has been successfully used to detect the values of ETP online in a real WWTP with short computation time. Based on the predicting values of ETP by the proposed monitoring system, an efficient model predictive controller (MPC) will be designed for improving the treatment of WWTP.

## References

- [1] M. Lüring, G. Waajen, F. van Oosterhout, Humic substances interfere with phosphate removal by lanthanum modified clay in controlling eutrophication, *Water Res.* 54 (1) (2014) 78–88.
- [2] P. Roccaro, M. Yan, G.V. Korshin, Use of log-transformed absorbance spectra for online monitoring of the reactivity of natural organic matter, *Water Res.* 84 (1) (2015) 136–143.
- [3] H. Hauduc, I. Takacs, S. Smith, A. Szabo, S. Murthy, G.T. Daigger, M. Sperandio, A dynamic physicochemical model for chemical phosphorus removal, *Water Res.* 73 (1) (2015) 157–170.
- [4] W.L. Huang, W. Cai, H. Huang, Z.F. Lei, Z.Y. Zhang, J.H. Tay, D.J. Lee, Identification of inorganic and organic species of phosphorus and its bio-availability in nitrifying aerobic granular sludge, *Water Res.* 68 (1) (2015) 423–431.
- [5] C.Y. Ki, K.H. Kwon, S.W. Kim, K.S. Min, T.U. Lee, D.J. Park, Effects of injection of acetic acid and propionic acid for total phosphorus removal at high temperature in enhanced biological phosphorus removal process, *Water Sci. Technol.* 69 (10) (2014) 2023–2028.
- [6] M.Q. Hui, J.G. Jin, H. Chao, Determination of phosphite, phosphate, glyphosate and aminomethylphosphonic acid by two-dimensional ion chromatography system coupled with capillary ion chromatography, *Chin. J. Anal. Chem.* 41 (12) (2013) 1910–1914.
- [7] Y. Zhang, P. Thepsithar, X. Jiang, J.H. Tay, Direct determination of phosphate in raw *Jatropha curcas* oil by ion chromatography, *Ind. Crop. Prod.* 44 (1) (2013) 459–464.
- [8] B.S. Gentle, P.S. Ellis, P.A. Faber, M.R. Grace, I.D. McKelvie, A compact portable flow analysis system for the rapid determination of total phosphorus in estuarine and marine waters, *Anal. Chim. Acta* 674 (2) (2010) 117–122.
- [9] E.M. Wilson, J. Turekova, P. Rotwein, Micro cobalt electrodes for detection of total phosphorus in water, *Micro Nano Lett.* 71 (12) (2012) 1176–1179.
- [10] C. Valls-Cantenys, M. Iglesias, J.L. Todoli, Speciation of phosphorus oxoacids in natural and waste water samples, *J. Chromatogr. A* 1231 (2012) 16–21.
- [11] H.C. Kim, High-rate MIEX filtration for simultaneous removal of phosphorus and membrane foulants from secondary effluent, *Water Res.* 69 (1) (2015) 40–50.
- [12] M.P. Ginige, A.S. Kayaalp, K.Y. Cheng, J. Wylie, A.H. Kaksonen, Biological phosphorus and nitrogen removal in sequencing batch reactors: effects of cycle length, dissolved oxygen concentration and influent particulate matter, *Water Sci. Technol.* 68 (5) (2013) 982–990.
- [13] H. Haimi, M. Mulas, F. Corona, R. Vahala, Data-derived soft-sensors for biological wastewater treatment plants: an overview, *Environ. Model. Softw.* 47 (1) (2013) 88–107.
- [14] Y.Q. Liu, J.D. Chen, Z.H. Sun, Y. Li, D.P. Huang, A probabilistic self-validating soft-sensor with application to wastewater treatment, *Comput. Chem. Eng.* 71 (1) (2014) 263–280.
- [15] G. Leonhardt, S. Fach, C. Engelhard, H. Kinzel, W. Rauch, A software-based sensor for combined sewer overflows, *Water Sci. Technol.* 66 (7) (2012) 1475–1482.
- [16] M.H. Kim, Y.S. Kim, A.A. Prabhu, C.K. Yoo, A systematic approach to data-driven modeling and soft sensing in a full-scale plant, *Water Sci. Technol.* 60 (2) (2009) 363–370.

- [17] J.Q. Wan, M.Z. Huang, Y.W. Ma, Prediction of effluent quality of a paper mill wastewater treatment using an adaptive network-based fuzzy inference system, *Appl. Soft Comput.* 11 (3) (2011) 3238–3246.
- [18] X.S. Qin, F.R. Gao, G.H. Chen, Wastewater quality monitoring system using sensor fusion and machine learning techniques, *Water Res.* 46 (4) (2012) 1133–1144.
- [19] M.W. Lee, S.H. Hong, H. Choi, Real-time remote monitoring of small-scaled biological wastewater treatment plants by a multivariate statistical process control and neural network-based software sensors, *Process Biochem.* 43 (10) (2008) 1107–1113.
- [20] Ö. Çinar, H. Hasar, C. Kinaci, Modeling of submerged membrane bioreactor treating cheese whey wastewater by artificial neural network, *J. Biotechnol.* 123 (2) (2006) 204–209.
- [21] K. Hu, J.Q. Wan, Y.W. Ma, A fuzzy neural network model for monitoring A2/O process using on-line monitoring parameters, *J. Environ. Sci. Health Part A* 47 (5) (2012) 744–754.
- [22] H.G. Han, Y. Li, Y.N. Guo, J.F. Qiao, A recurrent self-organizing neural network for predicting sludge volume index, *Applied Soft Computing* 38 (1) (2016) 477–486.
- [23] H.G. Han, J.F. Qiao, Prediction of activated sludge bulking based on a self-organizing RBF neural network, *J. Process Control* 22 (6) (2012) 1103–1112.
- [24] P.F. Cao, X.G. Luo, Soft sensor model derived from wiener model structure: modeling and identification, *Chin. J. Chem. Eng.* 22 (5) (2014) 538–548.
- [25] K. Hiromasa, F. Kimoto, Estimation of predictive accuracy of soft sensor models based on data density, *Chemom. Intell. Lab. Syst.* 128 (2013) 111–117.
- [26] D. Cecil, M. Kozłowska, Software sensors are a real alternative to true sensors, *Environ. Model. Softw.* 25 (5) (2010) 622–625.
- [27] W.M. Shao, X.M. Tian, Adaptive soft sensor for quality prediction of chemical processes based on selective ensemble of local partial least squares models, *Chem. Eng. Res. Des.* 95 (1) (2015) 113–132.
- [28] P. Kadlec, B. Gabrys, S. Strandt, Data-driven soft sensors in the process industry, *Comput. Chem. Eng.* 33 (4) (2009) 795–814.
- [29] I.T. Jolliffe, *Principal Component Analysis*, Springer, Heidelberg, 2002.
- [30] P. Geladi, B.R. Kowalski, Partial least-square regression: a tutorial, *Anal. Chim. Acta* 185 (1) (1986) 1–17.
- [31] J.L. Liu, Developing a soft sensor based on sparse partial least squares with variable selection, *J. Process Control* 24 (7) (2014) 1046–1056.
- [32] P. Shan, S. Peng, L. Tang, C.X. Yang, Y.H. Zhao, Q. Xie, C.W. Li, A nonlinear partial least squares with slice transform based piecewise linear inner relation, *Chemom. Intell. Lab. Syst.* 143 (1) (2015) 97–110.
- [33] Y.W. Lei, L.X. Ding, W.S. Zhang, Generalization performance of radial basis function networks, *IEEE Trans. Neural Netw. Learn. Syst.* 26 (3) (2015) 551–564.
- [34] H.G. Han, Q.L. Chen, J.F. Qiao, An efficient self-organizing RBF neural network for water quality prediction, *Neural Netw.* 24 (7) (2011) 717–725.
- [35] H. Yu, T. Xie, S. Paszczynski, B.M. Wilamowski, Advantages of radial basis function networks for dynamic system design, *IEEE Trans. Ind. Electron.* 58 (12) (2011) 5438–5450.
- [36] J. Ghosh, A. Nag, *Radial Basis Function Networks 2*, Springer, Heidelberg, 2001.
- [37] H.G. Han, J.F. Qiao, Adaptive computation algorithm for RBF neural network, *IEEE Trans. Neural Netw. Learn. Syst.* 23 (2) (2012) 342–347.