Prediction of sludge bulking using the knowledge-leverage-based fuzzy neural network

Honggui Han, Zheng Liu, Luming Ge and Junfei Qiao

ABSTRACT

One of the most important steps and the main bottleneck of the activated sludge wastewater treatment process (WWTP) is the secondary clarification, where sludge bulking is still a widespread problem. In this paper, an intelligent method, based on a knowledge-leverage-based fuzzy neural network (KL-FNN), is developed to predict sludge bulking online. This proposed KL-FNN can make full use of the data and the existing knowledge from the operation of WWTP. Meanwhile, a transfer learning mechanism is applied to adjust the parameters of the proposed method to improve the predicting accuracy. Finally, this proposed method is applied to a real wastewater treatment plant for predicting the sludge bulking risk, and then for predicting the sludge bulking. The experimental results indicate that the proposed prediction method can be used as a tool to achieve better performance and adaptability than the existing methods in terms of predicting accuracy for sludge bulking.

Key words | clarification, fuzzy neural network, knowledge leverage, sludge bulking, wastewater treatment process

Honggui Han (corresponding author)
Zheng Liu
Luming Ge
Junfei Qiao
Beijing Key Laboratory of Computational Intelligence and Intelligence System, College of Automation, Faculty of Information Technology, Beijing University of Technology, Chaoyang District, Beijing 100124, China
E-mail: *rechardhan@bjut.edu.cn*

INTRODUCTION

Activated sludge process (ASP) is the most commonly used technology in the wastewater treatment process (WWTP) (Wagner *et al.* 2015; Roohian & Mehranbod 2017). However, sludge bulking, a term used to describe the excessive growth of filamentous bacteria, is a common operational problem in ASP. Sludge bulking can cause the decrease of the sludge settling ability and effluent quality deterioration, and may destroy WWTP seriously (Lou & Zhao 2012). The process operation conditions of bulking sludge, however, are usually only marginally documented (Van 1992; Martins *et al.* 2004). Therefore, the prediction of sludge bulking is still an open problem.

To find a general explanation for sludge bulking, several theories have been reported, such as the diffusion-based selection (Flores-Alsina *et al.* 2009), the kinetic selection theory (Lou 2012), the storage selection theory (Nielsen *et al.* 2002) and the nitric oxide hypothesis (Han & Qiao 2012), etc. It is true that these theories describe the hindered and settling compression, but they assume no variation in the sludge's physical properties. Thus, the influences of sludge bulking have not been considered on the overall performance (Ramin *et al.* 2014). Moreover, even though the activated sludge models (ASMs), developed by International Water Association (IWA), are implemented and widely used for the benchmarking, diagnosis and design of WWTP, the problems of sludge bulking have not been covered in these models (Fenu *et al.* 2010; Keskitalo *et al.* 2010; Nogaj *et al.* 2015). Recently, many other advanced models have been proposed. However, the growth of filamentous bacteria has only been sparingly treated in activated sludge modelling studies (Hug *et al.* 2006; De Clercq *et al.* 2008). The current kinetic properties cannot be considered satisfactory, making a general model of sludge bulking more difficult.

To overcome the lack of a suitable mechanistic understanding of the growth of filamentous organisms, sludge volume index (SVI), one important parameter of sludge settling performance, is used to measure the settleability of activated sludge to quantify sludge bulking (Smets *et al.* 2006; Adonadaga 2015). But the values of SVI are hard to measure online. Recently, the data-driven methods, based on the neural networks, have been developed as an efficient alternative way for measuring sludge bulking, in which the necessary process information can be extracted directly from the process data. The neural networks, based on a

doi: 10.2166/wst.2017.574

biological prototype of the human brain, have the ability to model nonlinear systems with the input–output data. For example, Bagheri *et al.* developed a hybrid artificial neural network-genetic algorithm model for predicting the SVI values in (Bagheri *et al.* 2015). The results indicated that the intelligent model is more accurate than the traditional methods. A self-organizing radial basis function (RBF) neural network was introduced to predict the SVI values online in Han & Qiao 2012. In this method, a growing and pruning algorithm is designed to improve the generalization performance of the model to have the accurate values. Some other methods based on the artificial neural networks were widely reported for process monitoring in WWTPs (Mjalli *et al.* 2007; Haimi *et al.* 2013). Although these methods have the advantages of highly adaptive ability and strong learning ability, they cannot produce accurate results with the incomplete datasets (Gao *et al.* 2015). In order to solve this problem, some knowledge based methods have been proposed to predict sludge bulking (Wan *et al.* 2011), such as decision trees, expert systems (Wu 2009; Kim & Ahn 2011; Saif *et al.* 2013). However, the knowledge acquisition of the systems is a complex task (Hurdle *et al.* 2009). The effective integration of knowledge representation and the adaptation of knowledge models are the key challenges in applications (Breuker 2013; Gao *et al.* 2015). Moreover, it is difficult to combine the different information, including data and knowledge (Deng *et al.* 2013, 2016; Lu *et al.* 2015; Shao *et al.* 2015).

Based on the above analysis, an intelligent method, based on a knowledge-leverage-based fuzzy neural network (KL-FNN), is developed to predict sludge bulking online in this paper. This proposed KL-FNN can make full use of the data from the current scene, as well as the existing knowledge in the reference scene (the current scene means that the scene with insufficient data for proper modeling while a modeling task is required to be effectively implemented, and the reference scene is the scene related to the current scene, with similar data distribution and learning task. The reference scene can provide some useful information for the modeling task of the current scene) to deal with situations where the data are scarce, missing or protected. The main contributions of this study include the following aspects.

Firstly, a KL-FNN, combining the learning ability of neural networks and the interpretability of fuzzy systems, is developed for combining the data and knowledge to overcome the situations, where the data are available from the current scene while some useful knowledge exists in the reference scene.

Secondly, a knowledge-leverage-based transfer learning mechanism is applied to adjust the parameters of the prediction method to improve the generalization capability for predicting sludge bulking online with suitable accuracy.

Thirdly, this proposed intelligent method is applied to a real wastewater treatment plant for predicting sludge bulking online. Comprehensive experimental results demonstrate remarkable performance and superiority of the KL-FNN method.

The rest of this paper is organized as follows. The next section briefly introduces hardware setup. Then, the proposed KL-FNN is introduced and the knowledge-leverage-based transfer learning mechanism is discussed. In addition, the intelligent method is developed for predicting sludge bulking in detail. Furthermore, this proposed intelligent method is evaluated with the experiments, along with the results and discussion. Finally, the conclusion of the paper is given.

HARDWARE SETUP

WWTP, including physical, chemical and biological processes, has been used to reduce or eliminate suspended solids, organic matter, nitrogen, and phosphate. WWTP is a dynamic system due to its nonlinear feature, large uncertainty, multiple time scales in the internal process reactions and multivariable structure. In most WWTPs, it requires the removal of the microbial biomass produced in its biological reactors by sedimentation. To design the prediction system for predicting sludge bulking, the experimental setup, including a municipal WWTP and a KL-FNN-based intelligent method, is schematically shown in Figure 1.

In WWTP, there are primary settling tank, biochemical reaction tank and secondary settling tank. The biochemical reaction tank is composed of anaerobic tank and aerobic tank. Meanwhile, the separation and thickening of sludge is implemented in the secondary settling tank. The sludge from the bottom of secondary settling tank is recirculated into the biological reactor to guarantee the continuous reaction of ASP, and the treated wastewater from the top of secondary settling tank is discharged into the rivers. In this work, practical observations are performed in this A^2O WWTP as the preparation for the experiment. There is a conflict in the requirement for the removal of phosphorus and nitrogen. Since the removal of phosphorus is mainly through the discharge of sludge, it is better to shorten sludge retention time (SRT). The removal of ammonia should take the effect of the metabolism time of the nitrifying bacterium into account,

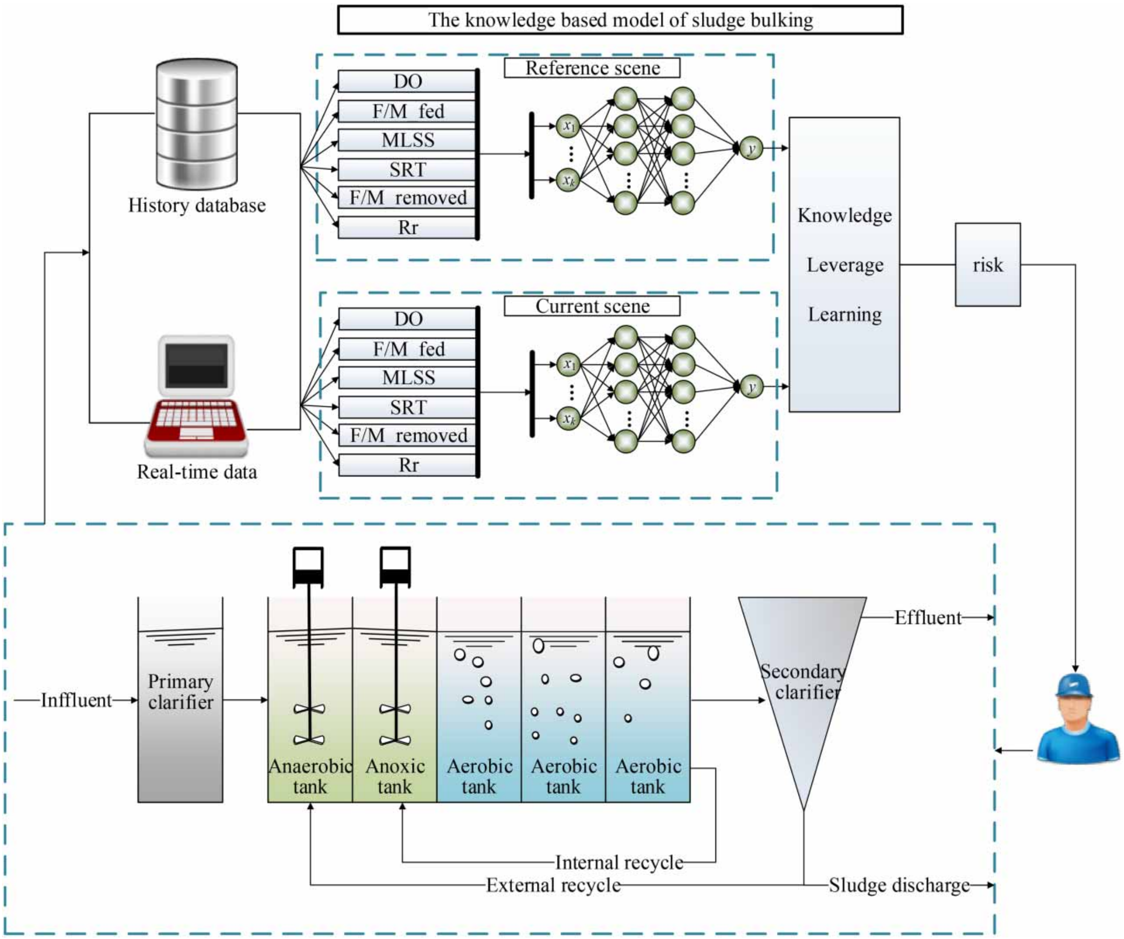

Figure 1 | The framework of the intelligent prediction system for sludge bulking.

which requires a long SRT. Therefore, SRT was maintained 8–25 d to meet different requirements. As shown in Figure 1, in this KL-FNN model, two kinds of information – the data of the current scene and the knowledge of the reference scenes – are used for predicting sludge bulking. Thus, the knowledge-leverage-based transfer learning mechanism is developed to adjust the parameters of the model to improve the predicting accuracy.

In addition, the online sensors are schematically distributed in the hardware setup. The online sensors are: the influent flow rate (Q_{in}) mater, the external circulating flow rate (Q_{et}) mater, the effluent flow rate (Q_w) mater, the biological oxygen demand (BOD) instrument, the dissolved oxygen (DO) concentration probe, and the mixed liquor suspended solids (MLSS) instrument. These main apparatus and instruments used in this setup are explained in Table 1. Moreover, the values of influent food-to-microorganism ratio (F/M_fed), effluent food-to-microorganism ratio (F/M_removed), sludge reflux ratio (Rr), SRT are calculated based on the above parameters.

Table 1 | The information of the online measured variables

Variables	Instrument type
Q_{in} (L/h)	WTW-FP
Q_{et} (L/h)	WTW-FP
Q_w (L/h)	WTW-FP
BOD (mg/L)	BSB/BOD 7400
DO (mg/L)	TriOxmatic700IQ
MLSS (mg/L)	ViSolid700IQ

KNOWLEDGE-LEVERAGE-BASED FUZZY-NEURAL-NETWORK

In order to develop a KL-FNN model, a multi-input and single-output (MISO) FNN is introduced. The mathematical description of FNN is given as follows.

(1) Input layer: there are k neurons in this layer, which represent the input variables of FNN. The output values of input layer can be expressed as:

$$u_i(t) = x_i(t) \tag{1}$$

where $u_i(t)$ is ith output value at time t, $i = 1, 2, \ldots, k$, $\mathbf{x}(t) = [x_1(t), x_2(t), \ldots, x_k(t)]$ is the input vector.

(2) RBF layer: there are P neurons in this layer, and each neuron is a RBF form which consists k fuzzy rules. The output values of RBF neurons are

$$\varphi_j(t) = \prod_{i=1}^{k} e^{-(x_i(t)-c_{ij}(t))^2/2\sigma_{ij}(t)^2} = e^{-\sum_{i=1}^{k}(x_i(t)-c_{ij}(t))^2/2\sigma_{ij}(t)^2} \tag{2}$$

where $\mathbf{c}_j(t) = [c_{1j}(t), c_{2j}(t), \ldots, c_{kj}(t)]$ and $\boldsymbol{\sigma}_j(t) = [\sigma_{1j}(t), \sigma_{2j}(t), \ldots, \sigma_{kj}(t)]$ are the vectors of centers and widths of the jth RBF neuron, respectively, $\varphi_j(t)$ is the output value of the jth neuron.

(3) Normalized layer: there are P neurons in this layer (the number is same as the RBF layer).

$$v_j(t) = \frac{\varphi_j(t)}{\sum_{j=1}^{P}\varphi_j(t)} = \frac{e^{-\sum_{i=1}^{k}(x_i(t)-c_{ij}(t))^2/2\sigma_{ij}(t)^2}}{\sum_{j=1}^{P} e^{-\sum_{i=1}^{k}(x_i(t)-c_{ij}(t))^2/2\sigma_{ij}(t)^2}}, \quad j = 1, 2, \cdots, P \tag{3}$$

where $v_j(t)$ is the jth output, and $\mathbf{v}(t) = [v_1(t), v_2(t), \ldots, v_P(t)]^{\mathrm{T}}$.

(4) Output layer: the output is clarified using the gravity method

$$y(t) = \mathbf{w}(t)\mathbf{v}(t) = \frac{\sum_{j=1}^{P} w_j(t) e^{-\sum_{i=1}^{k}(x_i(t)-c_{ij}(t))^2/2\sigma_{ij}(t)^2}}{\sum_{j=1}^{P} e^{-\sum_{i=1}^{k}(x_i(t)-c_{ij}(t))^2/2\sigma_{ij}(t)^2}} \tag{4}$$

where

$$\mathbf{w}(t) = [w_1(t), w_2(t), \cdots, w_j(t)]^{\mathrm{T}} \tag{5}$$

where $y(t)$ is the output of FNN, $\mathbf{w}(t)$ are the weights between the output neuron and normalized layer.

Without loss of generality, the goal of FNN is to describe the nonlinear systems by building a suitable model. The learning algorithm for FNN is to minimize the least square criterion function:

$$E_1(t) = \frac{1}{2}\sum_{t=1}^{m}(y(t) - y_d(t))^2 \tag{6}$$

where m is the number of the training dataset, $y_d(t)$ is the real output of the nonlinear systems.

To develop an efficient learning algorithm to make up for the deficiency of data in the current scene, the KL-FNN is proposed by combining the least square criterion and L2-norm penalty based learning method. To obtain the advantage of knowledge leverage for FNN, the proposed KL-FNN is developed by integrating the knowledge from the reference scenes. The new optimization criterion is defined as:

$$E_2(t) = \frac{1}{2}\sum_{t=1}^{m}(y(t) - y_d(t))^2 + \frac{1}{2}\sum_{t=1}^{m}\gamma(t)\sum_{j=1}^{p}(w_j(t) - w'_j(t))^2 \tag{7}$$

where $\mathbf{w}'(t) = [w_1'(t), \ldots, w_j'(t)]$ is the weights between the neurons in the output layer and the normalized layer in the reference scene, $\gamma(t)$ is the parameter for balancing the influence of the two terms.

In this paper, the updating formulas of the vectors of centers $\mathbf{c}_j(t)$, the vectors of widths $\boldsymbol{\sigma}_j(t)$, the weight parameters $\mathbf{w}(t)$ and the balance parameter $\gamma(t)$ are given as:

$$c_{ij}(t+1) = c_{ij}(t) - \eta\frac{\partial E_2(t)}{\partial c_{ij}(t)} \tag{8}$$

$$\sigma_{ij}(t+1) = \sigma_{ij}(t) - \eta\frac{\partial E_2(t)}{\partial \sigma_{ij}(t)} \tag{9}$$

$$w_j(t+1) = w_j(t) - \eta\frac{\partial E_2(t)}{\partial w_j(t)} \tag{10}$$

$$\gamma(t+1) = \gamma(t) - \eta\frac{\partial E_2(t)}{\partial \gamma(t)} \tag{11}$$

where η is the learning rate, and

$$\frac{\partial E_2(t)}{\partial \gamma(t)} = \frac{1}{2}\sum_{t=1}^{m}\sum_{j=1}^{P}(w_j(t) - w'_j(t))^2 \tag{12}$$

$$\frac{\partial E_2(t)}{\partial c_{ij}(t)} = \sum_{j=1}^{P}\left(\frac{\partial E_2(t)}{\partial y(t)}\frac{\partial y(t)}{\partial v_j(t)}\frac{\partial v_j(t)}{\partial \varphi_j(t)}\frac{\partial \varphi_j(t)}{\partial c_{ij}(t)}\right) \tag{13}$$

$$\frac{\partial E_2(t)}{\partial \sigma_{ij}(t)} = \sum_{j=1}^{P}\left(\frac{\partial E_2(t)}{\partial y(t)}\frac{\partial y(t)}{\partial v_j(t)}\frac{\partial v_j(t)}{\partial \varphi_j(t)}\frac{\partial \varphi_j(t)}{\partial \sigma_{ij}(t)}\right) \tag{14}$$

$$\frac{\partial E_2(t)}{\partial w_j(t)} = \sum_{j=1}^{P}\left(\frac{\partial E_2(t)}{\partial y(t)}\frac{\partial y(t)}{\partial w_j(t)} + \gamma(t)(w_j(t) - w'_j(t))\right). \tag{15}$$

Based on the above analysis, the learning process of KL-FNN is summarized in Figure 2.

EXPERIMENTAL STUDIES

In this study, the datasets of water quality are selected from a real WWTP (Beijing, China). The input variables of the datasets include the DO(t), F/M_fed(t), F/M_removed(t), Rr(t), SRT(t), MLSS(t) (Flores-Alsina *et al.* 2009). The output is the sludge bulking risk, which is used as an index to quantify sludge bulking: 0 suggests no risk and 1 implies the highest likelihood of risk. The data contain two parts: there are 2,500 samples for the reference scene over the year 2014. Meanwhile, 800 samples from 1/1/2015 to 1/6/2015, including the missing and noise data, are used for the current scene.

In order to evaluate the performance of the proposed intelligent prediction system, the root mean square error (*RMSE*), the mean absolute error (*MAE*), and non-dimensional error (*NDE*), defined in (Han & Qiao 2012; Deng *et al.* 2013; Seim *et al.* 2017), are used in this study. To show the dataset clearly, the notations of the datasets and their definitions are listed in Table 2. The dataset used for the reference scene is denoted by D1. Meanwhile, the datasets used for the current scene are denoted by D2 (training set) and D2_test (testing set).

The proposed intelligent prediction method is tested in four cases: (1) comparing with other FNNs without disturbance, (2) comparing with other FNNs with disturbance, (3) comparing with other methods with missing data, and (4) comparing with other methods with missing data, as well as disturbance. Cases 1–2 discuss the performance of the proposed KL-FNN without missing data, regardless of the disturbance. To evaluate the performance of KL-FNN for missing data, cases 3–4 are discussed in detail. The comprehensive experiments are used to demonstrate that the proposed KL-FNN can learn from not only the data of the current scene but also the knowledge of the reference. Moreover, the promising results illustrate that the KL-FNN has better adaptive abilities and better prediction effect than other methods. All the experiments were run on a PC with a clock speed of 2.6 GHz and 8 GB RAM, under Microsoft Windows 7.0 environment. For a fair comparison, each method was run 50 times, and the presented performance is the average value of 50 trials.

Initialize the FNN parameters ($\mathbf{w}'(0)$, $\mathbf{c}_j'(0)$, $\boldsymbol{\sigma}_j'(0)$) and the parameters $\mathbf{c}_j(0)=\mathbf{c}_j'(t)$, $\boldsymbol{\sigma}_j(0)=\boldsymbol{\sigma}_j'(t)$, $\mathbf{w}(0)=\gamma(0)\,\mathbf{w}'(t)$. Set the balance factor $\gamma(0)$ and the learning rate η.
For each sample in the reference scene $\mathbf{x}'(t)$
  for i=1:k
    for j=1:P
      Calculate the outputs of RBF layer $\phi_j'(t)$, the output of normalized layer $v_j'(t)$, and the output of FNN $y'(t)$; %Eqs.(2)-(4)
    end
    Calculate the criterion function $E_1(t)$; %Eq.(6)
  end
end
For each sample in the current scene $\mathbf{x}(t)$
  for i=1:k
    for j=1:P
      Calculate the outputs of RBF layer $\phi_j(t)$, the output of normalized layer $v_j(t)$ and the output of FNN $y(t)$; %Eqs.(2)-(4)
    end
    Calculate the criterion function $E_2(t)$; %Eq.(7)
    Update the center parameter $c_{ij}(t)$, the width parameter $\sigma_{ij}(t)$, the weight parameter $w_j(t)$, and the balance factor $\gamma(t)$; %Eqs.(8)-(11)
  end
end

Figure 2 | The learning process of the KL-FNN.

Table 2 | The notations of the datasets

		Data for the current scene	
	Data for the reference scene (D1)	Training set (D2)	Testing set (D2_test)
The amount of dataset	2,500	500	300

Case 1: Comparing with other FNNs without disturbance

In this case, four different FNNs are introduced to show the performance: (1) the FNN with the traditional least square (LS-FNN) criterion based on the data in the reference scene, i.e. LS-FNN (D1), (2) the LS-FNN based on the data in the current scene, i.e. LS-FNN (D2), (3) the LS-FNN based on the data in both the current scene and the reference scene, i.e. LS-FNN (D1 + D2), and (4) the proposed KL-FNN, i.e. KL-FNN (D2 + Knowledge). The testing data of the current scene were used for evaluating the generalization performance.

The predicting results of different FNNs are shown in Figure 3(a) and 3(b). Figure 3(a) shows the predicting values of sludge bulking risk. Figure 3(b) gives the predicting errors of sludge bulking risk. It can be seen from Figure 3(a) and 3(b) that the results of LS-FNN (D1) and LS-FNN (D2) are worse than those of LS-FNN (D1 + D2) and KL-FNN (D2 + Knowledge). Moreover, the results indicate that the proposed KL-FNN (D2 + Knowledge) is better than LS-FNN (D1 + D2) for predicting the sludge bulking risk. Moreover, Table 3 shows a detailed comparison. The results show that the performance of KL-FNN (D2 + Knowledge) is better than the other FNNs. And the performance can be listed as: (1) the proposed KL-FNN can exploit useful knowledge from the reference scene in the training procedure; (2) the proposed KL-FNN (D2 + Knowledge) owns the best *RMSE*, *MAE*, and *NDE* in this case.

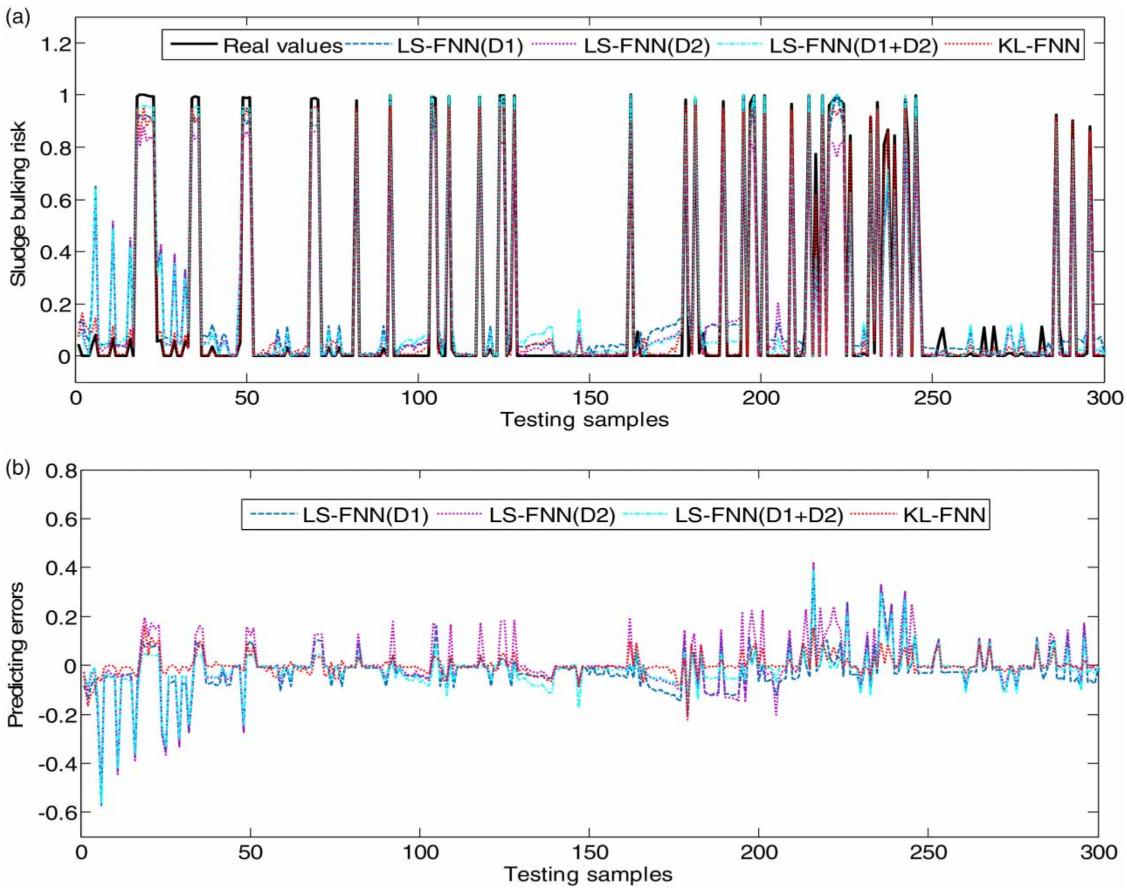

Figure 3 | The predicting results without disturbance. (a) The predicting values of sludge bulking risk without disturbance. (b) The predicting errors of sludge bulking risk without disturbance.

Table 3 | Performance of different FNNs without disturbance

Method	*RMSE*	*MAE*	*NDE*
LS-FNN(D1)	0.0913	0.0651	0.2792
LS-FNN(D2)	0.1002	0.0631	0.3063
LS-FNN(D1 + D2)	0.0753	0.0447	0.2302
KL-FNN	**0.0411**	**0.0254**	**0.1258**

Case 2: Comparing with other FNNs with disturbance

In case 2, the testing data with disturbance in the current scene were used to estimate the generalization performance. And the proposed KL-FNN is also compared with LS-FNN (D1), LS-FNN (D2) and LS-FNN (D1 + D2).

The results are shown in Figure 4(a) and 4(b). Figure 4(a) and 4(b) show the predicting values and the predicting errors, respectively. Based on the results, it is clearly that the predicting errors of KL-FNN is smaller than the other FNNs and less than 0.2. From the results in Table 4, it can be seen that the generalization performance of KL-FNN is better than others. Based on the above analysis, in this case, it is concluded that: (1) the proposed KL-FNN is valuable for WWTP with disturbance; (2) the proposed KL-FNN can adopt useful knowledge from the reference scene in this case.

Case 3: Comparing with other methods with missing data

As mentioned in the above section, when the data are insufficient, the existing data-driven methods will be no longer effective and the generalization performance will be deterioration. In this case, the generalization performance is tested with missing data. And the proposed KL-FNN was compared with the other methods: (1) TS-fuzzy system-based support vector regression (TSFS-SVR) (Juang *et al.* 2007), (2) fuzzy system learned through fuzzy clustering and support vector machine (FS-FCSVM)

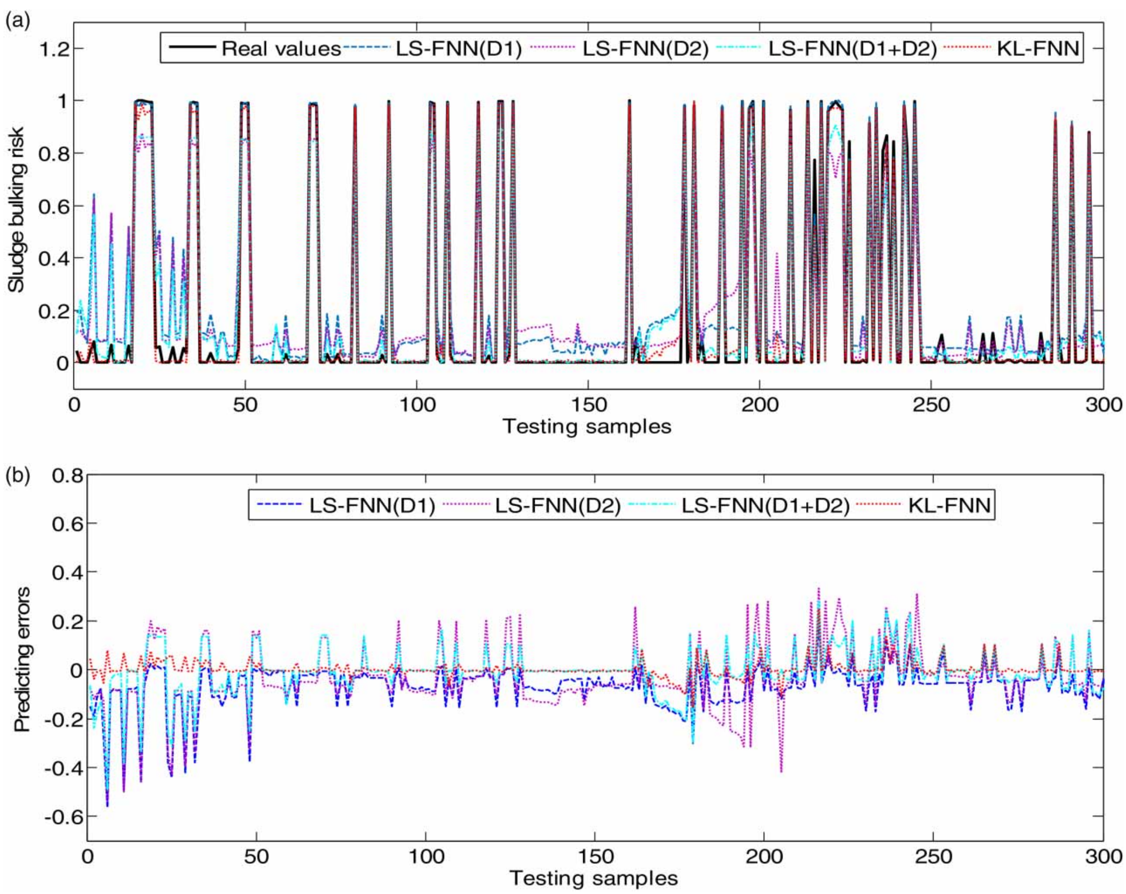

Figure 4 | The predicting results of different FNNs with disturbance. (a) The predicting values of different FNNs with disturbance. (b) The predicting errors of different FNNs with disturbance.

Table 4 | Performance of different FNNs with disturbance

Method	*RMSE*	*MAE*	*NDE*
LS-FNN(D1)	0.1081	0.0823	0.3303
LS-FNN(D2)	0.1206	0.0921	0.3685
LS-FNN(D1 + D2)	0.0955	0.0698	0.2918
KL-FNN	**0.0454**	**0.0240**	**0.1389**

(Juang & Hsieh 2009), (3) Bayesian task-level transfer learning for non-linear regression method (HiRBF) (Yang *et al.* 2008), and (4) knowledge-leverage-based fuzzy system (KL-FS) (Deng *et al.* 2013).

Figure 5(a) and 5(b) show the predicting results and errors of different methods. It is noted that the predicting values of KL-FNN are better than those of the other methods. Moreover, the results in Table 5 demonstrate that the proposed KL-FNN obtains the lowest values of *RMSE*, *MAE* and *NDE*. Then, the proposed KL-FNN owns better generalization ability than other methods. It is also demonstrated that the efficiency of the knowledge leverage mechanism and the proposed method can effectively exploit useful knowledge from the reference scene with missing data.

Case 4: Comparing with other methods with missing data as well as disturbance

In this case, the generalization performance is evaluated using the testing data of the current scene with missing data as well as disturbance. The proposed KL-FNN is also compared with other four schemes: (1) TSFS-SVR (Juang *et al.* 2007), (2) FS-FCSVM (Juang & Hsieh 2009), (3) HiRBF (Yang *et al.* 2008), and (4) KL-FS (Deng *et al.* 2013).

The results of TSFS-SVR, FCSVM, HiRBF, KL-FS and KL-FNN are illustrated in Figure 6(a) and 6(b). Figure 6(a) shows the predicting results and Figure 6(b) gives the predicting errors. The results demonstrate that the

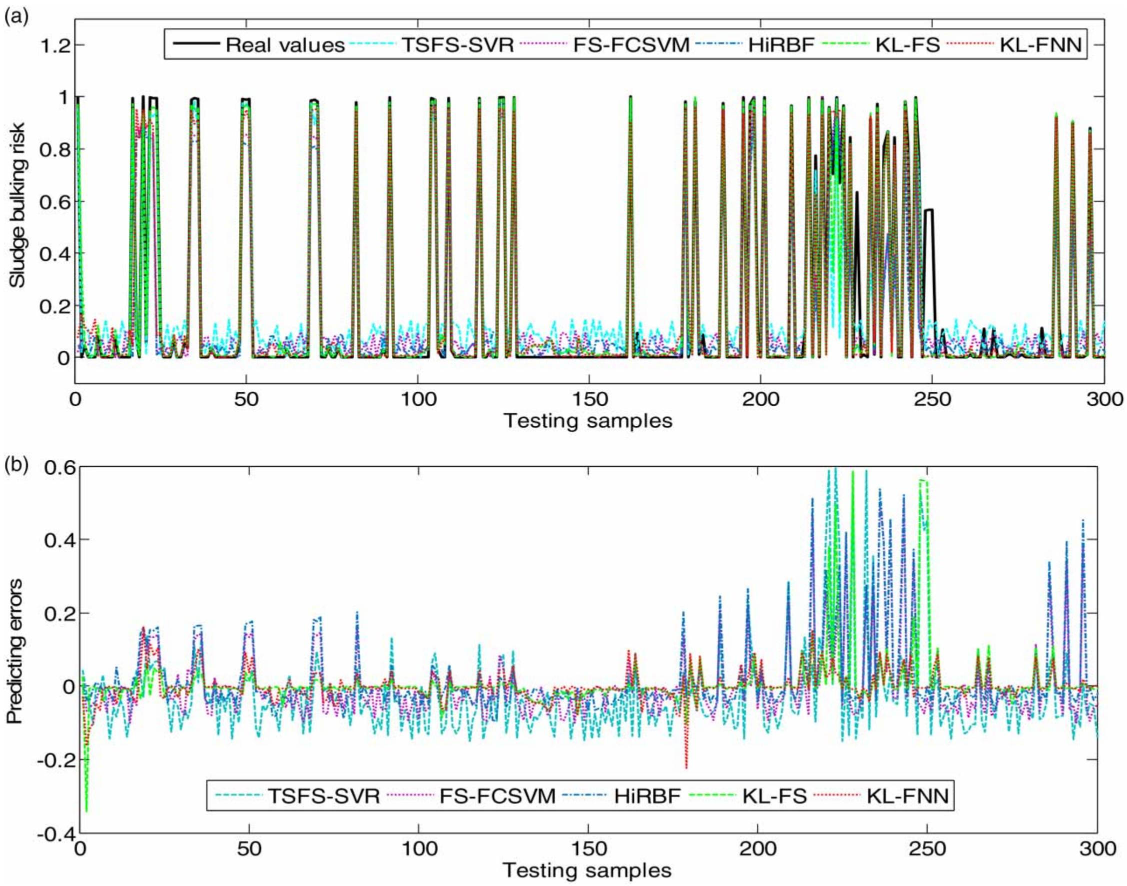

Figure 5 | The predicting results with missing data. (a) The predicting values of sludge bulking risk with missing data. (b) The predicting errors of sludge bulking risk with missing data.

Table 5 | Performance of KL-FNN and other methods with missing data

Method	*RMSE*	*MAE*	*NDE*
TSFS-SVR (Juang *et al.* 2007)	0.1091	0.0778	0.3331
FS-FCSVM (Juang & Hsieh 2009)	0.1044	0.0686	0.3190
HiRBF (Yang *et al.* 2008)	0.1236	0.0648	0.3777
KL-FS (Deng *et al.* 2013)	0.0734	0.0285	0.2239
KL-FNN	**0.0411**	**0.0254**	**0.1258**

predicting values of KL-FNN and KL-FS are more close to the real values. The predicting errors of KL-FNN are in a tight range and the generalization ability of KL-FNN is given in Table 6. For all methods, the performance of testing *RMSE*, *MAE* and *NDE* is degraded because of the disturbance. However, based on the results in Table 6, the proposed KL-FNN still owns better generalization performance than that of the other methods.

DISCUSSION

In the above experiments, cases 1–2 are about the results of KL-FNN and other FNNs without missing data. Based on the details in Tables 3 and 4, the predicting results of KL-FNN are better than that of the other FNNs. The proposed KL-FNN can effectively exploit not only the data of the current scene but also the useful knowledge of the reference scenes to obtain good adaptive abilities and generation performance. Moreover, in order to evaluate the performance of KL-FNN for the situations, where the data are available from the current scene while some useful knowledge exists in the reference scene, cases 3–4 are about the results of KL-FNN and other methods, including TSFS-SVR, FS-FCSVM, HiRBF and KL-FS with missing date. The results demonstrate that the proposed KL-FNN owns better prediction performance than the other three methods. It can be seen from Figures 3(a)–6(b) that the proposed KL-FNN is able to obtain better generalization performances

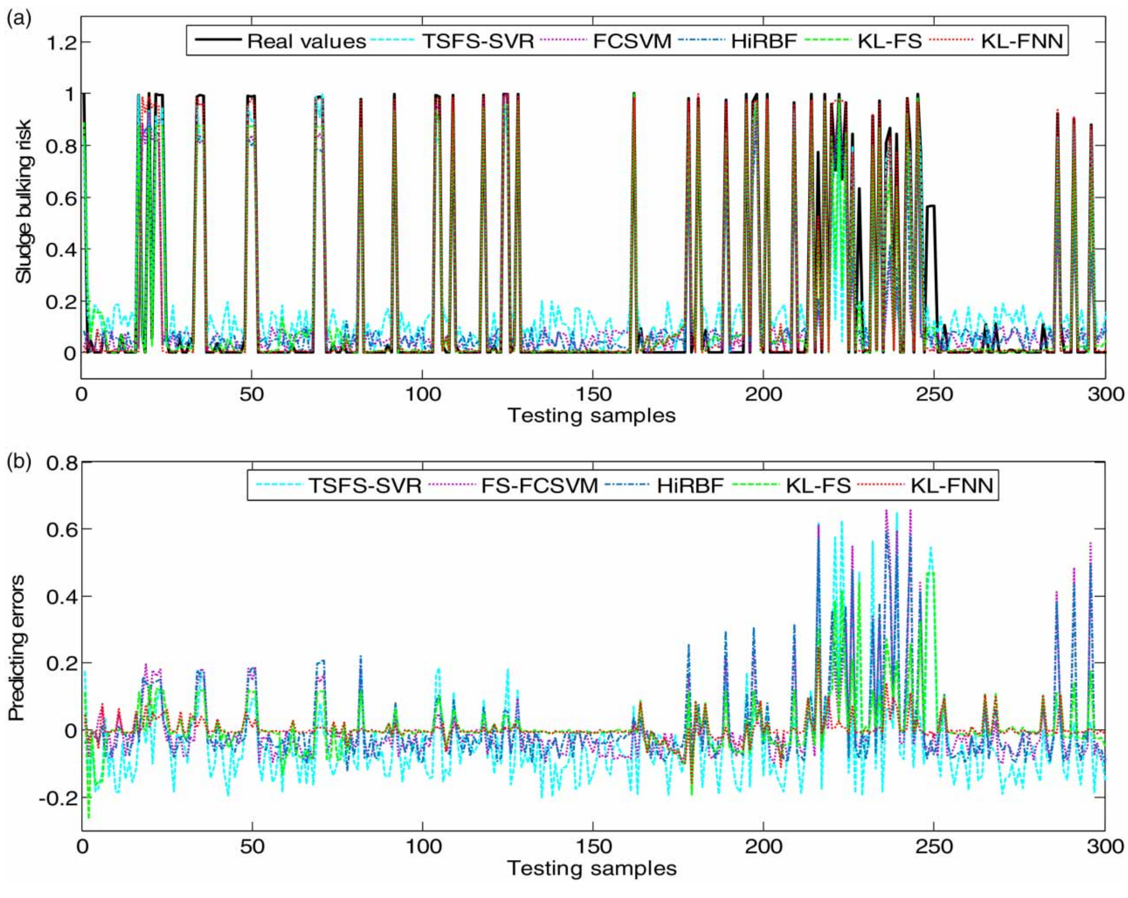

Figure 6 | The predicting results with missing data as well as disturbance for. (a) The predicting values of sludge bulking risk with missing data as well as disturbance. (b) The predicting errors of sludge bulking risk with missing data as well as disturbance.

Table 6 | Performance of KL-FNN and other methods with missing data as well as disturbance

Methods	*RMSE*	*MAE*	*NDE*
TSFS-SVR (Juang *et al.* 2007)	0.1302	0.0994	0.3973
FS-FCSVM (Juang & Hsieh 2009)	0.1400	0.0814	0.4280
HiRBF (Yang *et al.* 2008)	0.1404	0.0803	0.4291
KL-FS (Deng *et al.* 2013)	0.0814	0.0469	0.2484
KL-FNN	**0.0454**	**0.0240**	**0.1389**

than the other four methods. The predicting values of KL-FNN are close to the real values in Figures 3(a), 4(a), 5(a) and 6(a). The predicting errors are between −0.2 and 0.2 as shown in Figures 3(b), 4(b), 5(b) and 6(b). In Figures 4 and 6, the results show that the generalization capability of KL-FNN will not be degraded significantly with disturbance. Moreover, the details in Tables 5 and 6 show that, even if the data in the training data of the current scene are missing, the generalization capability of KL-FNN is still better than the other methods. This remarkable feature is valuable for WWTP as data insufficiency is a common problem due to the poor sensor sensitivity and disturbance environment.

CONCLUSION

In this paper, a KL-FNN-based intelligent method is developed to predict sludge bulking for WWTP. The goal was to remedy the deficiency of the data insufficiency in the current scene by leveraging the useful knowledge available from the reference scene. Meanwhile, a transfer learning mechanism, which can preserve data confidentiality in the reference scene, is applied to adjust the parameters of KL-FNN to improve the predicting accuracy of the intelligent method. In order to illustrate the performance of KL-FNN, this proposed intelligent method is applied to a real wastewater treatment plant. The experimental results demonstrate the effectiveness of the proposed method. Based on the above analysis, the proposed prediction method owns three advantages: (1) this proposed KL-FNN can make full use of the data and the existing knowledge of WWTP; (2) the method has a better anti-interference ability than other methods through the experimental analysis of disturbance test; (3) the proposed KL-FNN-based prediction method is a useful method for the other applications with poor sensor sensitivity and disturbance environment.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation of China under Grants 61622301 and 61533002, Beijing Natural Science Foundation under Grant 4172005, and Major National Science and Technology Project under Grant 2017ZX07104.

REFERENCES

Adonadaga, M. G. 2015 Effect of dissolved oxygen concentration on morphology and settleability of activated sludge flocs. *Journal of Applied and Environmental Microbiology* **3** (2), 31–37.

Bagheri, M., Mirbagheri, S. A., Bagheri, Z. & Kamarkhani, A. M. 2015 Modeling and optimization of activated sludge bulking for a real wastewater treatment plant using hybrid artificial neural networks-genetic algorithm approach. *Process Safety and Environmental Protection* **95** (1), 12–25.

Breuker, J. 2013 A cognitive science perspective on knowledge acquisition. *International Journal of Human-Computer Studies* **71** (2), 177–183.

De Clercq, J., Nopens, I., Defrancq, J. & Vanrolleghem, P. A. 2008 Extending and calibrating a mechanistic hindered and compression settling model for activated sludge using in-depth batch experiments. *Water Research* **42** (3), 781–791.

Deng, Z., Jiang, Y., Choi, K. S., Chung, F. L. & Wang, S. 2013 Knowledge-leverage-based TSK fuzzy system modeling. *IEEE Transactions on Neural Networks and Learning Systems* **24** (8), 1200–1212.

Deng, Z., Jiang, Y., Ishibuchi, H., Choi, K. S. & Wang, S. 2016 Enhanced knowledge-leverage-based TSK fuzzy system modeling for inductive transfer learning. *Acm Transactions on Intelligent Systems & Technology* **8** (1), 11–21.

Fenu, A., Guglielmi, G., Jimenez, J., Sperandio, M., Saroj, D., Lesjean, B., Brepols, C., Thoeye, C. & Nopens, I. 2010 Activated sludge model (ASM) based modelling of membrane bioreactor (MBR) processes: a critical review with special regard to MBR specificities. *Water Research* **44** (15), 4272–4294.

Flores-Alsina, X., Comas, J., Rodriguez-Roda, I., Gernaey, K. V. & Rosen, C. 2009 Including the effects of filamentous bulking sludge during the simulation of wastewater treatment plants using a risk assessment model. *Water Research* **43** (18), 4527–4538.

Gao, Z. W., Cecati, C. & Ding, S. X. 2015 A survey of fault diagnosis and fault-tolerant techniques-part II: fault diagnosis with knowledge-based and hybrid/active approaches. *IEEE Transactions on Industrial Electronics* **62** (6), 3768–3774.

Haimi, H., Mulas, M., Corona, F. & Vahala, R. 2013 Data-derived soft-sensors for biological wastewater treatment plants: an overview. *Environmental Modelling and Software* **47** (1), 88–107.

Han, H. G. & Qiao, J. F. 2012 Prediction of activated sludge bulking based on a self-organizing RBF neural network. *Journal of Process Control* **22** (6), 1103–1112.

Hug, T., Gujer, W. & Siegrist, H. 2006 Modelling seasonal dynamics of *Microthrix parvicella*. *Water Science and Technology* **54** (1), 189–198.

Hurdle, E. E., Bartlett, L. M. & Andrews, J. D. 2009 Fault diagnostics of dynamic system operation using a fault tree based method. *Reliability Engineering and System Safety* **94** (9), 1371–1380.

Juang, C. F. & Hsieh, C. D. 2009 TS-fuzzy system-based support vector regression. *Fuzzy Sets and Systems* **160** (17), 2486–2504.

Juang, C. F., Chiu, S. H. & Shiu, S. J. 2007 Fuzzy system learned through fuzzy clustering and support vector machine for human skin color segmentation. *IEEE Transactions on Systems Man and Cybernetics Part A Systems and Humans* **37** (6), 1077–1087.

Keskitalo, J., Jansen, J. & Leiviska, K. 2010 Calibration and validation of a modified ASM1 using long-term simulation of a full-scale pulp mill wastewater treatment plant. *Environmental Technology* **31** (5), 555–566.

Kim, K. J. & Ahn, H. 2011 Simultaneous optimization of artificial neural networks for financial forecasting. *Applied Intelligence* **36** (4), 887–898.

Lou, I. 2012 Combination of respirometry and molecular approach for re-evaluating microbial kinetic selection of filamentous bulking in wastewater treatment system. *Advanced Science Letters* **9** (1), 540–544.

Lou, I. & Zhao, Y. 2012 Sludge bulking prediction using principle component regression and artificial neural network. *Mathematical Problems in Engineering* **583** (3), 295–308.

Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S. & Zhang, G. 2015 Transfer learning using computational intelligence: a survey. *Knowledge-Based Systems* **80** (C), 14–23.

Martins, A. M., Pagilla, K., Heijnen, J. J. & van Loosdrecht, M. C. 2004 Filamentous bulking sludge – a critical review. *Water Research* **38** (4), 793–817.

Mjalli, F. S., Al-Asheh, S. & Alfadala, H. E. 2007 Use of artificial neural network black-box modeling for the prediction of wastewater treatment plants performance. *Journal of Environmental Management* **83** (3), 329–338.

Nielsen, P. H., Roslev, P., Dueholm, T. E. & Nielsen, J. L. 2002 *Microthrix parvicella*, a specialized lipid consumer in anaerobic-aerobic activated sludge plants. *Water Science and Technology* **46** (1–2), 73–80.

Nogaj, T., Randall, A., Jimenez, J., Takacs, I., Bott, C., Miller, M., Murthy, S. & Wett, B. 2015 Modeling of organic substrate transformation in the high-rate activated sludge process. *Water Science and Technology* **71** (7), 971–979.

Ramin, E., Wagner, D. S., Yde, L., Binning, P. J., Rasmussen, M. R., Mikkelsen, P. S. & Plosz, B. G. 2014 A new settling velocity model to describe secondary sedimentation. *Water Research* **66** (1), 447–458.

Roohian, H. & Mehranbod, N. 2017 Investigation of bio-augmentation of overloaded activated sludge plant operation by computer simulation. *Computers and Chemical Engineering* **104** (1), 11–24.

Saif, S. M., Sarikhani, M. & Ebrahimi, F. 2013 An expert system with neural network and decision tree for predicting audit opinions. *IAES International Journal of Artificial Intelligence* **2** (4), 151–158.

Seim, F., Gravdahl, A. R. & Adaramola, M. S. 2017 Validation of kinematic wind turbine wake models in complex terrain using actual windfarm production data. *Energy* **123** (1), 742–753.

Shao, L., Zhu, F. & Li, X. 2015 Transfer learning for visual categorization: a survey. *IEEE Transactions on Neural Networks & Learning Systems* **26** (5), 1019–1034.

Smets, I. Y., Banadda, E. N., Deurinck, J., Renders, N., Jenné, R. & Van Impe, J. F. 2006 Dynamic modeling of filamentous bulking in lab-scale activated sludge processes. *Journal of Process Control* **16** (3), 313–319.

Van, L. 1992 A review of the potential application of non-specific activated sludge bulking control. *Water S A* **18** (2), 101–106.

Wagner, D. S., Ramin, E., Szabo, P., Dechesne, A. & Plosz, B. G. 2015 *Microthrix parvicella* abundance associates with activated sludge settling velocity and rheology – quantifying and modelling filamentous bulking. *Water Research* **78** (1), 121–132.

Wan, J., Huang, M., Ma, Y., Guo, W., Wang, Y., Zhang, H., Li, W. & Sun, X. 2011 Prediction of effluent quality of a paper mill wastewater treatment using an adaptive network-based fuzzy inference system. *Applied Soft Computing* **11** (3), 3238–3246.

Wu, D. 2009 Supplier selection: a hybrid model using DEA, decision tree and neural network. *Expert Systems with Applications* **36** (5), 9105–9112.

Yang, P., Tan, Q. & Ding, Y. 2008 Bayesian task-level transfer learning for non-linear regression. In: *International Conference on Computer Science and Software Engineering*, IEEE, Wuhan, China, pp. 62–65.

First received 6 June 2017; accepted in revised form 30 October 2017. Available online 14 November 2017