Use of support vector machine model to predict membrane permeate flux

Kui Gao[a], Xuejie Xi[a], Zhan Wang[a,*], Yu Ma[b,*], Sha Chen[a], Xiao Ye[c], Yili Li[a]

[a]*Beijing Key Laboratory for Green Catalysis and Separation, Department of Chemistry and Chemical Engineering, Beijing University of Technology, Beijing 100124, P.R. China, Tel. +86 10 67396186; Fax: +86 10 67391983; emails: gaokui5@emails.bjut.edu.cn (K. Gao), xxj@emails.bjut.edu.cn (X. Xi), wangzh@bjut.edu.cn (Z. Wang), chensha@bjut.edu.cn (S. Chen), liyili@bjut.edu.cn (Y. Li)*
[b]*Climatic Center of Xingjiang Uygur Autonomous Region of China, Urumqi 830002, P.R. China, Tel./Fax: +86 991 2656440; email: rainhorse6709@163.com*
[c]*Biochemical Engineering College, Beijing Union University, Beijing 100124, P.R. China, Fax: +86 10 67391983; email: yexiao@163.com*

ABSTRACT

In this paper, the structure optimized support vector machine (SVM) model was applied to predict the membrane permeate flux during dead-end microfiltration of activated sludge suspensions from sequencing batch reactor (SBR) with different experimental samples. The membrane permeate flux was considered as a function of mixed liquor suspended, temperature, dissolved oxygen, hydraulic retention time, transmembrane pressure, and operating time. Excellent agreements between the predicted values of SVM model and the experimental data demonstrated that SVM model has sufficient prediction accuracy. Furthermore, the results showed that the predicted values of SVM model agreed well with experimental data at different experimental samples in comparison with back propagation artificial neural network (BP-ANN) model. From the simulation results, the conclusion can be derived that SVM model outperforms BP-ANN model when the experimental samples sizes are small.

Keywords: SVM; ANN; Membrane permeate flux; Prediction; Comparison

1. Introduction

Microfiltration (MF) technology, integrated with the conventional activated sludge bioreactors, has been widely used in wastewater treatment and reuse [1–3]. However, the occurrence of membrane fouling may cause severe reduction in membrane performance including the loss of permeate flux and selectivity [4]. To solve these problems, membrane cleaning as an essential step has been frequently carried out to restore the membrane permeate flux. Therefore, how to exactly predict the membrane permeate flux is of great significance for the industrial process [5,6].

In industrial operation, once membrane styles and feed systems are assigned, the membrane permeate flux in MBRs is mainly influenced by operating conditions, such as temperature, airflow rate of aeration, cross-flow rate, pH, HRT, sludge retention time, and transmembrane pressure (TMP) [7,8]. Several conventional models have been developed to optimize the operating conditions of MBRs [1,9–11]. For example, Bai et al. [1] proposed a mathematical model to predict the cake thickness and flux decay in a cross-flow microfiltration. Gui et al. [9] have studied the

*Corresponding authors.

influence of operating parameters using orthogonal experimental method and found the best combination of them. In addition, Liu et al. [10] have proposed a model which was used to calculate the cross-flow velocity of the activated sludge over the membrane surface in a submerged membrane bioreactor. However, these models could not give a good prediction accuracy since influence factors limiting its applications is extremely complex [7,8]. So, finding an efficient and accurate method to predict the membrane permeate flux in MBRs is very stringent.

Artificial neutral networks (ANN) is a mathematical model based on the modern neuroscience research, which composes a large number of neurons connected with simple components [12]. ANN, as an efficient method for dealing with complex nonlinear problems, has been applied in business, science and engineering fields [13]. Compared with multiple linear regressions and other traditional methods [14–16], ANN technology shows high accuracy, but it needs a large number of data and some shortcomings will arise from its theoretical statistical basis such as the prediction accuracy is not satisfying when the training set is small. In contrast, support vector machine (SVM), which is based on statistical learning theory, has been proposed to solve the difficult problem for this situation [17,18]. Compared with ANN, SVM has several merits such as efficient utilization of high-dimensional feature space, distinctively solvable optimization problem and theoretically analysis ability using computational learning theory [14].

As a highly effective approach to model with limited training sample sets, SVM has been widely applied in many fields for prediction [18], such as pattern recognition problem [19–23], classification [24,25], regression [26,27], image analysis [28], drug design [29–31], time series analysis [32–34], quality control of food [35,36], protein structure function prediction [37–40], and genomics [41]. Additionally, SVM usually outperformed traditional statistical learning methods [42,43]. Thus, SVM has been arousing more and more interest, and many researches are focused on this subject. For example, to make SVM comprehensively understood by junior learners, support vector classification machines [44] and support vector regression machines [45] were reviewed by Burges et al. and Smola et al., respectively. Specifically, Noble described the definition of SVM and its biological applications [46]. In addition, Xu et al. adopted SVM for classification in chemometrics [47]. Jia et al. showed that SVM was applicable to predict the synthesis characteristics of hydraulic valve in industrial production [19]. A predictor was constructed to predict the true and false splice sites for higher eukaryotes based on SVMs [48].

Liang et al. proposed an effective approach based on biased SVM for content-based sketch retrieval [49]. Gumus et al. presented an evaluation of using various methods (PCA, wavelets and SVM) for face recognition [50]. All these studies make great contributions to the development and application of SVM. Although the prediction performance of SVM is excellent, few papers have reported about the application of SVM in the field of membrane permeate flux prediction.

Therefore, in this paper, SVM was used to validate the application for the predictions of the membrane permeate flux as a function of different operating conditions during dead-end microfiltration fouled with activated sludge suspensions from sequencing batch reactor (SBR). Additionally, the prediction accuracy of SVM/back propagation artificial neural network (BP-ANN) model at different experimental sample sizes was compared.

2. Experimental

2.1. System and methods

The laboratory-scale experimental system (in Fig. 1) described elsewhere in the literature [17] consisted of two parts. The first part is an intermittent mode bioreactor system with an effective volume of 25 L. The
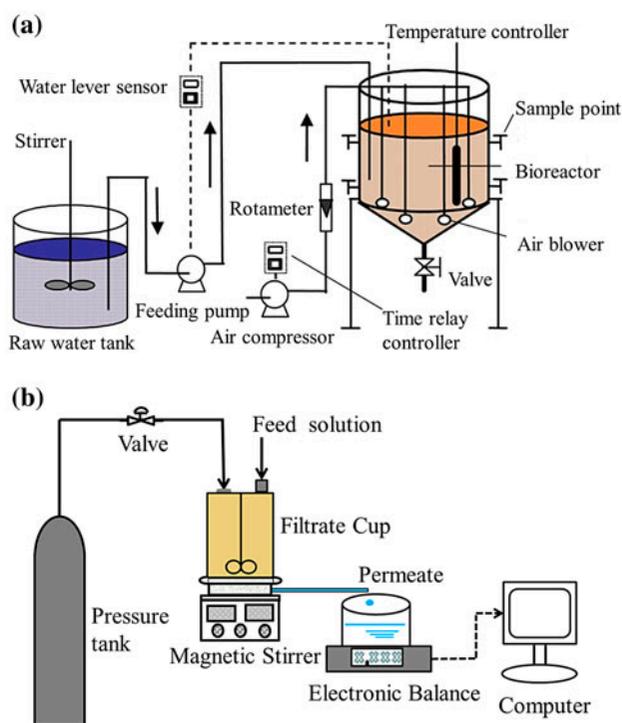
Fig. 1. Schematic of the experimental system: Intermittent bioreactor system (a); Dead-end filtration system (b).

second part is a dead-end microfiltration cell. The feed solution (raw wastewater) was obtained from the storage tank of domestic sewage shown in Table 1. The dead-end microfiltration cell has an effective membrane area of 24.0 cm². Prior to each experiment, the polyacrylonitrile (PAN) hydrophilic membrane with nominal pore size of 0.1 μm was soaked in deionized water for 12 h to remove glycerin (protective agents) at 4℃. The mixed liquor suspended solids (MLSS) concentration was measured by weighing a dried sample and pH was measured with a pHS-3C acidity meter [51].

The different operating parameters for the membrane bioreactor system, such as the MLSS, temperature (*T*), dissolved oxygen (DO), pH, and HRT, are shown in Table 2. No sludge was discharged during the operation or test period.

2.2. Experimental procedure

The experiments were conducted as follows: for each experiment, the permeate weight was measured with an electronic balance connected to a PC for automatic data acquisition per a given time, in which the sampling interval varied from 30 s to 2 min. The temperature of the whole filtration process was controlled by a constant temperature water-bath, and the temperature was set as the bioreactor's temperature. TMP was varied from 0.03 to 0.24 MPa. When the conditions of SBR remained stable, the activated sludge suspension was poured into the filtration cell (shown in Fig. 1(b)) in which a 0.1-μm polyacrylonitrile (PAN) membrane had been loaded, and was filtered through the membrane under a constant TMP using compressed air.

3. The support vector machine

SVM proposed by Vapnik is a classification method which aimed at separating two data-sets based on the maximum distance between them [27]. It is an excellent tool for the classification and regression problems of good generalization ability. This method separates two data-sets into particular classes by looking for an optimal separating hyperplane between

Table 2
Operating conditions of bioreactors

Operating conditions	
MLSS (mg/L)	1,800, 2,600, 3,000, 4,500, 6,000
Temperature (℃)	15, 18, 20, 24, 28
DO (mg/L)	7, 5.2, 6, 4.6, 7.5
HRT (h)	16, 20, 25, 17, 22
pH	7 ± 1

them [52–54]. Bounds between data-sets and OSH are called "support vectors". In this paper, the structure optimized SVM is used as the predicted model.

The training sample set is given as {(x_i, y_i), i = 1, 2, 3, … l}, where x_i is the *i*th input and y_i is the corresponding desired output [52–56]. Supposed that the training data can be linearly separated, the SVM formalism uses the following linear estimation function *f* with the empirical risk minimization as follows:

$$f(x) = (\omega, \varphi(x)) + b \tag{1}$$

where ω ($\omega \in F$) is the weight vector, (,) is the inner product, $\varphi(x)$ denotes a mapping function in the feature space, i.e. it represents the nonlinear mapping from low-dimensional feature spaces to high-dimensional feature space *F*; *b* is designated to be the bias.

The values of ω and *b* in Eq. (1) can be derived by substituting the training data-set (samples) (x_i, y_i) into the following function:

$$R_{\text{reg}}[f] = R_{\text{emp}}[f] + \lambda \|\omega\|^2 = \sum_{i-1}^{s} C(e_i) + \lambda \|\omega\|^2 \tag{2}$$

where $R_{\text{reg}}[f]$ is the sum of empirical risk and experience risk; $R_{\text{emp}}[f]$ is the experience risk; λ is the regularization parameter for controlling the loss of training data-set (samples) and the compromise of model complexity; the minimum distance between any training point and this hyperplane is defined as the margin of the classifier. In order to achieve the minimal of structural risk, larger margins (the optimal hyperplane) should be obtained to get better

Table 1
Quality of raw wastewater used in the experiment

COD (mg L⁻¹)	NH₃-N (mg L⁻¹)	TOC (mg L⁻¹)	pH	Turbidity/NTU
180.6–225.8	45.9–73.6	86.5–115.5	7.5–8.0	20–26

generalization. $\|\omega\|^2$ is the confidence risk and reflects the model complexity in high-dimensional feature space, and the smaller $\|\omega\|^2$ means smaller confidence risk; s is the size of the training set (samples); $C(.)$ is the loss function, $e_i = f(x_i) - y_i = \hat{y}_i - y_i$ represents the difference between the predicted values and the experimental data, and $C(e_i)$ represents the experience error of the model. Based on the principle of structured risk minimization, SVM model sought to minimize the sum of empirical risk and confidence risk.

For a given loss function, the problem of finding function f can be solved as a quadratic programming problem as following:

$$\max J = -\frac{1}{2}\sum_{i,j=1}^{s}(\alpha_i - \alpha_i^*)(\alpha_j^* - \alpha_j)(\varphi(X_i), \varphi(X_j))$$
$$+ \sum_{i=1}^{s}\alpha_i^*(Y_i - \varepsilon) - \sum_{i=1}^{s}\alpha_i^*(Y_i) \tag{3}$$

$$\text{s.t.}\begin{cases}\sum_{i=1}^{s}\alpha_i = \sum_{i=1}^{s}\alpha_i^* \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C\end{cases} \tag{4}$$

By solving the Eqs. (3) and (4), $\omega = \sum_{i-1}^{s}(\alpha_i - \alpha_i^*)\varphi(x_i)$ can be obtained; b can be obtained by substituting any supported vector into the Eqs. (3) and (4). In this way, the function f is transformed into the following representation:

$$f(x) = \sum_{i-1}^{s}(\alpha_i - \alpha_i^*)(\varphi(x_i), \varphi(x)) + b \tag{5}$$

Define the inner product of high-dimensional feature transformation space as the Kernel function of SVM:

$$K(x_i, x_j) = (\varphi(x_i), \varphi(x_j)) \tag{6}$$

The inner product in high-dimensional space can only be obtained by computing the Kernel function in the low-dimensional space. Finally, by introducing Lagrange multipliers and exploiting the optimal constraints, the decision function has the following explicit form:

$$f(x) = \sum_{i-1}^{s}(\alpha_i - \alpha_i^*)K(x_i, x) + b \tag{7}$$

4. Back propagation artificial neural network

Artificial neural network (ANN) is a nonlinear contains many simple computational units and back propagation (BP) is the most typical supervised learning algorithm [57]. BP-ANN commonly consists of input, output, and hidden layers. Neurons interconnected with adjacent layers by a weighting factor. BP-ANN layers by learning to modify the connections between neurons weights, then the final error could be minimal [58]. In this paper, a structure optimized three-layer BP-ANN is made as the predicted model. The number of hidden neurons, learning functions, and learning rate were optimized as follows.

4.1. Selection of the number of hidden neurons

The large number of hidden neurons may lead to the overlong learning time, larger error and poorer fault tolerance. In this paper, the training epochs and average absolute relative deviation were chosen as the reference standard. The relationship between the number of hidden neurons and the predicted error of the model and epochs is depicted in Fig. 2. It can be seen in Fig. 2 that the epochs decreased with increasing the number of hidden neurons. According to Fig. 2, when the number of hidden neurons was 7, the average absolute relative deviations for the membrane permeate flux got the minimum. So, the number of hidden neurons of the BP-ANN model in dead-end microfiltration with the activated sludge suspension was 7.

4.2. Comparison of different training function

Table 3 shows the result of comparison of different training functions (Traingda, Traingdx and Trainlm).
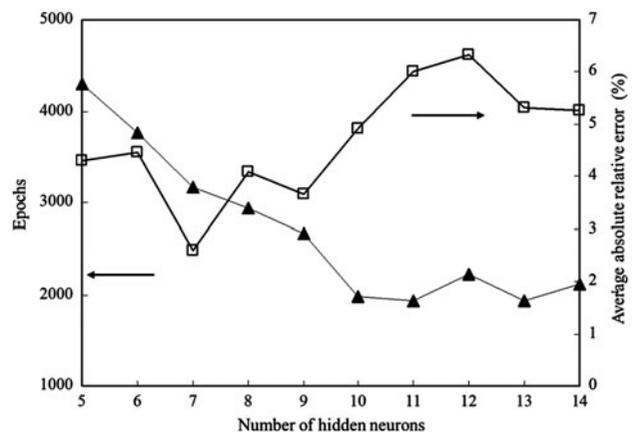
Fig. 2. Average absolute relative error/epoch as function of number of hidden neurons.

Table 3
The influence of training function on performance of BP-ANN network

Training function	Epoch	Average absolute relative error (%)
Traingda	4,735	7.31
Traingdx	3,986	2.62
Trainlm	1,503	5.26

The sequence of average absolute relative error was Traingda (7.31%) > Trainlm (5.26%) > Traingdx (2.62%), which clearly demonstrated that the predicted error using the function of Traingdx was the smallest. Consequently, the Traingdx was chosen as the training function in the further study.

4.3. Comparison of different learning rate

Although the slow learning rate can avoid the problem of local minimum in training the model, it can lead to slow the convergence rate and prolong learning time. Moreover, the high learning rate can induce the instability of the system [17]. Therefore, in order to ensure the convergence rate and learning time, the optimal learning rate should be determined. The influence of learning rate on performance of BP-ANN network is presented in Table 4. It is clearly observed that the predicted error was the smallest (2.62%) when the learning rate was 0.08. In addition, the predicted error was 6.89, 5.11, 4.64%, respectively, when the learning rate was 0.02, 0.05, and 0.1.

5. Structure of the SVM and BP-ANN Model

The prediction structure of SVM and BP-ANN models, which used for modeling the membrane permeate flux of activated sludge suspensions from SBR, is presented in Fig. 3.

In this paper, the optimal structure of SVM model was determined by adjusting the programmer and LIBSVM tool loaded into MATLAB (R2010b). The operating conditions, such as MLSS, temperature (T), DO, HRT, TMP, and operating time (t), were set as

Table 4
The influence of learning rate on performance of BP-ANN network

Learning rate	Epoch	Average absolute relative error (%)
0.02	7,784	6.89
0.05	6,031	5.11
0.08	3,986	2.62
0.1	3,692	4.64

input variables, and the membrane permeate flux was set as output variables. Meantime, the optimized structure of a three-layer BP-ANN with the Traingdx was set as the training function, with the corresponding number of hidden neurons was 7 and the learning rate was 0.08. In further study, the prediction performances of the structure optimized by SVM and BP-ANN models were studied.

6. Results and discussion

6.1. The prediction by SVM and BP-ANN

6.1.1. First group

In this paper, the experiments were divided into three groups with different experimental samples. In this group, a total of 810 sets of experimental data were used, including 510 sets as training samples, 300 sets as predicted sample. The inputs are the operating condition (MLSS, temperature (T), DO, HRT, TMP, and operating time (t)) and the outputs are the membrane permeate flux.

Fig. 4 presents the comparison of the predicted values of SVM/BP-ANN model and experimental data for the membrane permeate flux, one group includes 60 predicted points (point stands for the experimental/filtration time to measure the membrane
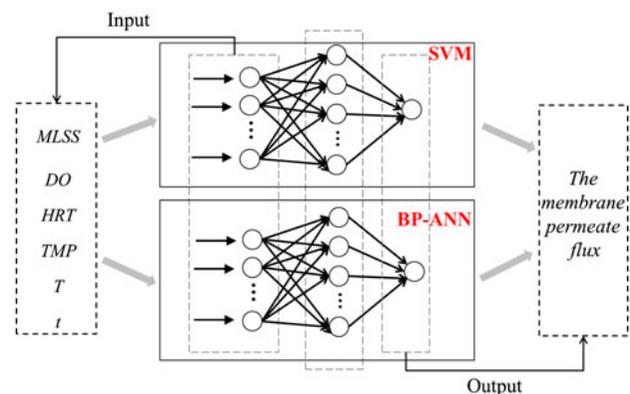
Fig. 3. Predictive structure of the SVM or the BP-ANN used for modeling the membrane permeate flux of activated sludge suspensions from SBR.
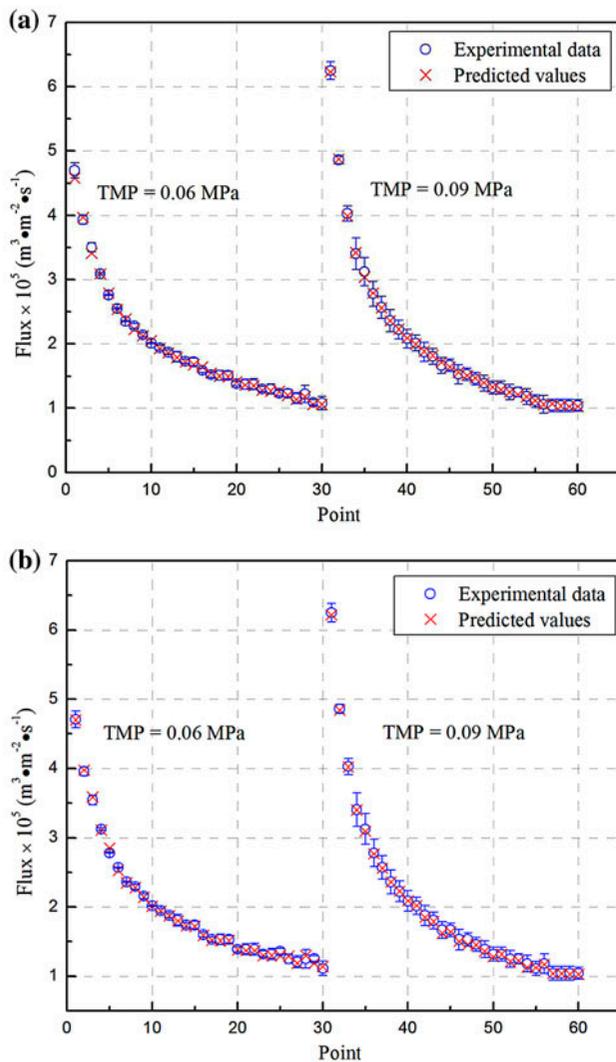
Fig. 4. The comparisons of experimental data for the membrane permeate flux and the predicted values of SVM model (a) and BP-ANN model. (b) Operating conditions: $T = 15°C$, MLSS = 2,600 mg/L, DO = 7.0 mg/L, HRT = 16 h, TMP = 0.06 MPa or 0.09 MPa.

flux during the filtration process) (Operating conditions: $T = 15°C$, MLSS = 2,600 mg/L, DO = 7.0 mg/L, HRT = 16 h, TMP = 0.06 MPa or 0.09 MPa). As seen in Fig. 4(a), the membrane permeate flux at high TMP (0.09 MPa) was clearly higher than that at low TMP (0.06 MPa), as well as the flux decay. The reason for this observation is that with the higher TMPs, the accumulation of molecules near the membrane surface was faster and at higher permeate flux TMP decreased quickly [59]. It is apparent from Fig. 4 that the values of the membrane permeate flux predicted by SVM/BP-ANN model are close to the experimental data, indicating that the SVM/BP-ANN yield a suitable fit at different operating conditions.

Table 5 and Fig. 5 show the relative error between experimental data and predicted values of SVM/BP-ANN model, respectively. All the errors for SVM model were distributed between 7.2 and −6.0%, and that for BP-ANN model were distributed between 5.5 and −5.5%. Simultaneously, the average values of the error of 300 experimental points for SVM and BP-ANN model were 3.43 and 2.62%, respectively. These results meant that the agreement between the experimental data and the predicted data of SVM/BP-ANN model was excellent. Furthermore, Fig. 5 and Table 5 also demonstrated that the prediction accuracy of BP-ANN model was slightly higher than that of SVM model when the large-capacity training set was encountered (810 sets of experimental data).

6.1.2. Second group

In this group, a total of 450 sets of experimental data were used, including 300 sets as training samples, 150 sets as predicted samples.

The comparison between experimental data of the membrane permeate flux and the predicted values by the SVM or ANN model with 30 predicted points is present in Fig. 6. As the number of predicted samples reduced to half in contrast with that of the first group (60 predicted points in Section 6.1.1), the predictions of SVM model also can maintain a high accuracy for the experimental data at different operating conditions (Fig. 6(a)) (Operating conditions: $T = 18°C$, MLSS = 6,000 mg/L, DO = 7.5 mg/L, HRT = 22 h, TMP = 0.13 MPa). It is also observed from Fig. 6(b) that the BP-ANN model agreed well with the
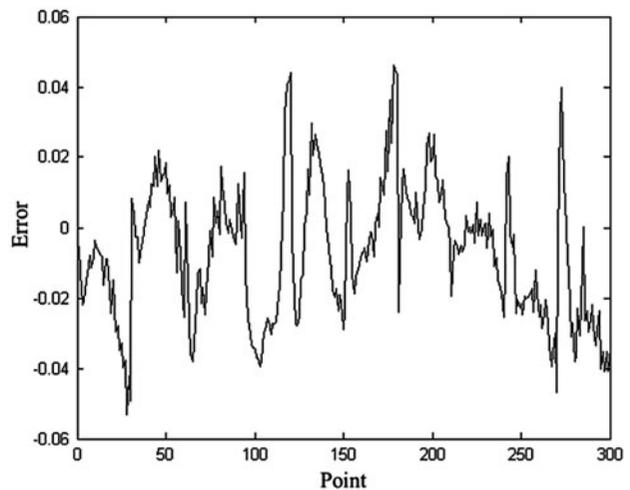
Fig. 5. The relative error between experimental data and predicted values of BP-ANN model.

Fig. 7. The relative error between experimental data and predicted values of BP-ANN model.

±18% as shown in Fig. 7. Moreover, the average values of the error of 150 experimental points of SVM and BP-ANN model were 3.68 and 6.32%, respectively. Table 6 and Fig. 7 also confirmed that the predicted values of SVM and BP-ANN model were in good agreement with the experimental data for simulating the membrane permeates flux of activated sludge suspensions from SBR at the complex operating conditions. Additionally, the value of the relative error between experimental data and predicted values of BP-ANN model was higher than that of SVM model. It is also obviously observed from Fig. 6 that several big relative errors were emerged at a very small part of points using BP-ANN model. These observations implied that the SVM model yielded a much better fit than BP-ANN model as the number of training samples decreased.

6.1.3. Third group

In this group, a total of 225 sets of experimental data were used, including 150 sets as training samples, 75 sets as predicted samples.

The comparison between the predicted values of SVM/BP-ANN model and experimental data for this group (there are 15 predicted points) is depicted in Fig. 8. It can be seen from Fig. 8 that the predicted permeate flux of SVM/BP-ANN model basically maintained a higher accuracy for the experimental data at different operating conditions (Operating conditions: $T = 20°C$, MLSS = 1,800 mg/L, DO = 5.2 mg/L, HRT = 20 h, TMP = 0.09 MPa), although the number of predicted samples reduced by almost a half with the second group (30 predicted samples in Section 6.1.2).

Fig. 6. The comparisons of experimental data for the membrane permeate flux and the predicted values of SVM model (a) and BP-ANN model (b). Operating conditions: $T = 18°C$, MLSS = 6,000 mg/L, DO = 7.5 mg/L, HRT = 22 h, TMP = 0.13 MPa.

experimental data. Comparing with the result for the first group in Section 6.1.1 (Fig. 4(b)), the prediction performance of the BP-ANN model for the second group is not as satisfying as it is. This observation may be explained that a large number of weight factors are typically required for ANN model [53].

Table 6 and Fig. 7 illustrate the relative error between experimental data and predicted values of SVM model and BP-ANN model, respectively. It can be seen from Table 6, all the errors between experimental data and predicted values for SVM model were distributed between 7.4 and −6.8%, in contrast, those for BP-ANN model were distributed between

Table 5
The prediction accuracy of SVM model of the first group

Operating conditions	Max. relative error (%)	Min. relative error (%)	The mean relative error of 300 experimental points (%)
T = 15℃, MLSS = 2,600 mg/L, DO =7.0 mg/L, HRT = 16 h	7.0	−6.0	3.43
T = 18℃, MLSS = 6,000 mg/L, DO = 7.5 mg/L, HRT = 22 h	6.7	−5.3	
T = 20℃, MLSS = 1,800 mg/L, DO = 5.2 mg/L,HRT = 20 h	7.2	−5.0	
T = 24℃, MLSS = 3,000 mg/L, DO = 6.0 mg/L,HRT = 25 h	6.2	−4.0	
T = 28℃, MLSS = 4,500 mg/L, DO = 4.6 mg/L,HRT = 17 h	7.0	−5.2	

Table 6
The prediction accuracy of SVM model of the second group

Experimental Conditions	Max. relative error (%)	Min. relative error (%)	The mean relative error of 150 experimental points (%)
T = 15℃, MLSS = 2,600 mg/L, DO = 7.0 mg/L, HRT = 16 h, TMP = 0.09 MPa	7.4	−6.8	3.68
T = 18℃, MLSS = 6,000 mg/L, DO = 7.5 mg/L, HRT = 22 h, TMP = 0.13 MPa	5.4	−5.6	
T = 20℃, MLSS = 1,800 mg/L, DO = 5.2 mg/L, HRT = 20 h, TMP = 0.09 MPa	6.4	−6.8	
T = 24℃, MLSS = 3,000 mg/L,DO = 6.0 mg/L, HRT = 25 h, TMP = 0.08 MPa	6.2	−5.8	
T = 28℃, MLSS = 4,500 mg/L,DO = 4.6 mg/L, HRT = 17 h, TMP = 0.09 MPa	6.4	−6.2	

Table 7
The prediction accuracy of SVM model of the third group

Experimental Conditions	Max. relative error (%)	Min. relative error (%)	The mean relative error of 75 experimental points (%)
T = 15℃, MLSS = 2,600 mg/L, DO = 7.0 mg/L, HRT = 16 h, TMP = 0.09 MPa	3.0	−3.6	3.63
T = 18℃, MLSS = 6,000 mg/L, DO = 7.5 mg/L, HRT = 22 h, TMP = 0.13 MPa	8.0	−5.0	
T = 20℃, MLSS = 1,800 mg/L, DO = 5.2 mg/L, HRT = 20 h, TMP = 0.09 MPa	4.8	−6.8	
T = 24℃, MLSS = 3,000 mg/L, DO = 6.0 mg/L, HRT = 25 h, TMP = 0.08 MPa	8.0	−4.8	
T = 28℃, MLSS = 4,500 mg/L, DO = 4.6 mg/L, HRT = 17 h, TMP = 0.09 MPa	2.5	−3.4	

As presented in Table 7, all the relative errors between experimental data and predicted values by SVM model were distributed between ±8%, and the average error of 75 experimental points was 3.63%. According to Fig. 8 and Table 7, the phenomenon that SVM model has excellent prediction ability can be
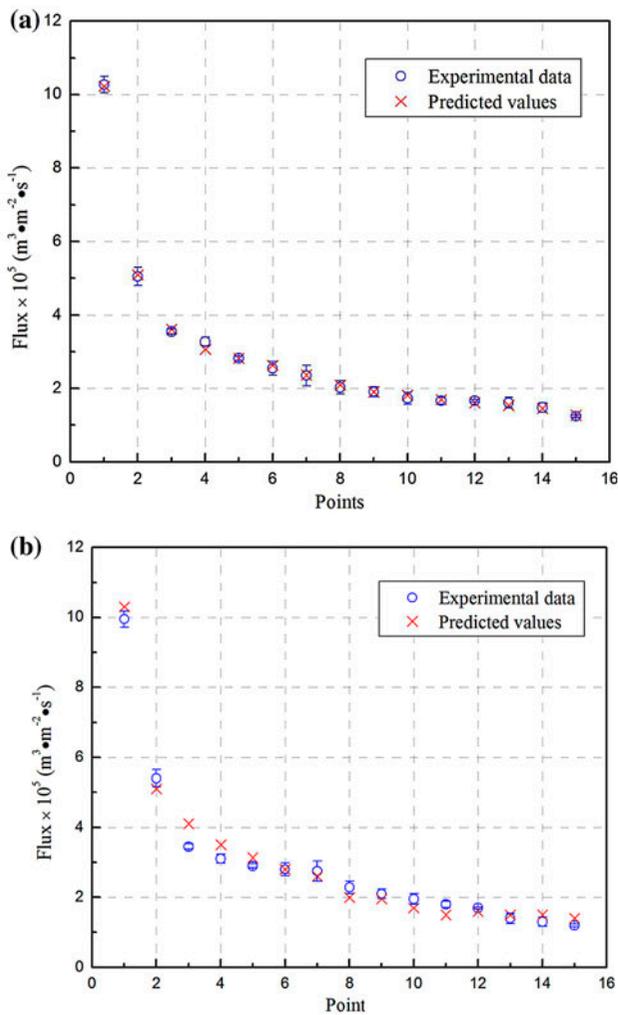
Fig. 8. The comparisons of experimental data for the membrane permeate flux and the predicted values of SVM model (a) and BP-ANN model (b). Operating conditions: $T = 20°C$, MLSS = 1,800 mg/L, DO = 5.2 mg/L, HRT = 20 h, TMP = 0.09 MPa.

Fig. 9. The relative error between experimental data and predicted values of BP-ANN model.

discovered. Furthermore, the relative error between experimental data and predicted values for BP-ANN model is presented in Fig. 9. It is clearly seen that all the errors lie in the range −6 to 5%, and the average error of all 75 experimental points was 9.93% as illustrated in Fig. 9. Compared with the relative errors summarized in Table 7, Fig. 9 illustrates that the prediction performance of BP-ANN model is not ideal with the big error between experimental data and the predicted values, reflecting that the BP-ANN model cannot get good prediction results in this case. Therefore, these observations further validated that SVM model was more suitable than BP-ANN model in the small size of the training set (small training sample or restricted data-sets). This can be explained by the fact
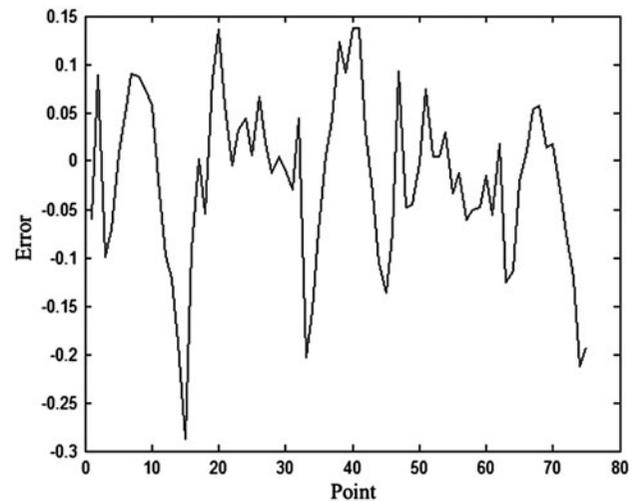
that (i) the SVM method has a rigorous theoretical and mathematical foundation as well as good generalization ability, which is based on statistical theory, while ANN depends on the designer's experience and knowledge; (ii) ANN methods need a large size of training set (samples) to achieve the desired prediction accuracy, while the SVM method can get excellent accuracy only with a small size of training set (samples).

6.2. The average relative errors

Summarizing the above analysis, the average relative errors between the experimental data and predicted values for SVM and BP-ANN model to predict the membrane permeate flux of activated sludge suspensions from SBR in the case of similar sample sizes are illustrated in Fig. 10. It is clearly found in Fig. 10 that the predictions of SVM and BP-ANN models aiming at the large-capacity training set (810 groups and 300 predicted samples in Section 6.1.1) can obtain high accuracy, which indicate that the predictions of these models are satisfactory. And in this case, the average relative error of SVM model (3.43%) was slightly bigger than that of BP-ANN model (2.62%). Aiming at the medium capacity (450 groups and 150 predicted samples in Section 6.1.2) training set, these models also fitted well with the experimental data. However, a small portion of big relative errors for BP-ANN model appeared as depicted in Fig. 6(b). Moreover, the relative error of BP-ANN model was larger than that of SVM model with regard to small volume training set (225 groups and 75 predicted samples in Section 6.1.3) implying that the prediction
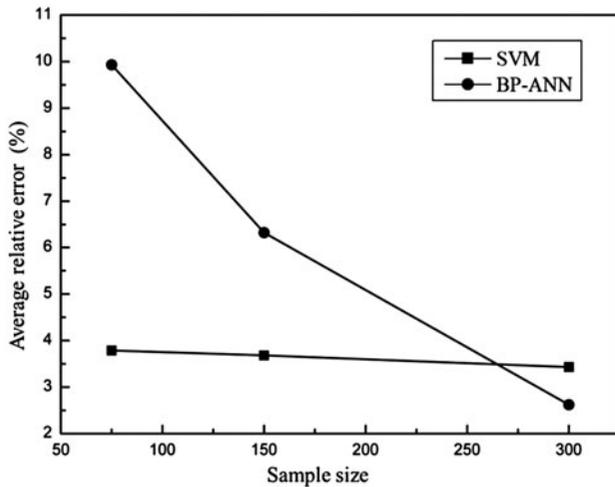
Fig. 10. The average relative errors of membrane permeate flux of bioreactor by SVM models and BP-ANN model in the case of different sample sizes.

performance of BP-ANN model was not ideal. Obviously, the values of relative errors for BP-ANN model increased as the number of sample set decreased. The variation trend of relative errors for SVM model changed inversely in comparison with that of BP-ANN model. Furthermore, all average relative errors of SVM model were below 4%, implying the good performance of SVM model. At the meantime, the average relative errors were significantly lower than that of BP-ANN model with regard to the medium- and small-capacity training set. Thus, the conclusion can be obtained that the predictions of SVM model are in good agreement with the experimental data and the prediction average relative error is stabilized in a lesser extent. The results were in consistent with other studies [60].

7. Conclusion

In this paper, the membrane permeates flux of activated sludge suspensions from SBR was predicted by SVM models with the optimized network structure. During the prediction process, the operating conditions, such as MLSS, temperature (T), DO, HRT, TMP, and operating time (t), were set as input node, and the membrane permeate flux of activated sludge suspensions from SBR was set as output node. The training set (sample) were divided into different groups (810 groups, 450 groups and 225 groups), respectively. Then, the predicted values of SVM or BP-ANN models were compared with the experimental data, respectively.

The results showed that the predicted values of SVM or BP-ANN model are essentially in agreement with the experimental data at different operating condition with different experimental sizes. SVM model is based on statistical theory and it has a rigorous theoretical and mathematical foundation, while BP-ANN model needs to rely on the designer's experience and knowledge. When the sample size is large, the relative errors between the experimental data and predicted values for SVM and BP-ANN models are relatively small, reflecting the high prediction accuracy. When the sample size is small, the prediction of SVM model still shows a better accuracy in comparison with BP-ANN model. In the meantime, the predicted average relative error of SVM model is stabilized in a lesser extent than that of BP-ANN model. So, BP-ANN model needs a large number of experimental data to achieve desired prediction accuracy, while the SVM model can get the higher accuracy only with the small size of the training set. In another word, SVM model is superior to BP-ANN model for modeling the limited number of training set so as to avoid the requirement of large numbers of training samples.

Nomenclature

F	—	a high-dimensional feature space
φ	—	a nonlinear mapping
ω	—	weight vector
$(,)$	—	inner product
$\varphi(x)$	—	a mapping function in the feature space
b	—	a constant (the bias)
$R_{emp}[f]$	—	experience risk
λ	—	regularization parameter
$\|\omega\|^2$	—	confidence risk
S	—	size of the sample
$C(.)$	—	loss function
e_i	—	difference between the predicted values and the experimental data
$C(e_i)$	—	experience loss of the model
J	—	the membrane flux

References

[1] R. Bai, H.F. Leow, Microfiltration of activated sludge wastewater-the effect of system operation parameters, Sep. Purif. Technol. 29 (2002) 189–198.
[2] E.H. Bouhabila, R.B. Aïm, H. Buisson, Microfiltration of activated sludge using submerged membrane with air bubbling (application to wastewater treatment), Desalination 118 (1998) 315–322.

[3] T. Melin, B. Jefferson, D. Bixio, C. Thoeye, W.D. Wilde, J.D. Koning, J. van der Graaf, T. Wintgens, Membrane bioreactor technology for wastewater treatment and reuse, Desalination 187 (2006) 271–282.

[4] W. Lee, S. Kang, H. Shin, Sludge characteristics and their contribution to microfiltration in submerged membrane bioreactors, J. Membr. Sci. 216 (2003) 217–227.

[5] C. Duclos-Orsello, W. Li, C.C. Ho, A three mechanism model to describe fouling of microfiltration membranes, J. Membr. Sci. 280 (2006) 856–866.

[6] K.J. Hwang, C.Y. Liao, K.L. Tung, Analysis of particle fouling during microfiltration by use of blocking models, J. Membr. Sci. 287 (2007) 287–293.

[7] J. Zhang, H.C. Chua, J. Zhou, A.G. Fane, Factors affecting the membrane performance in submerged membrane bioreactors, J. Membr. Sci. 284 (2006) 54–66.

[8] S.P. Hong, T.H. Bae, T.M. Tak, S. Hong, A. Randall, Fouling control in activated sludge submerged hollow fiber membrane bioreactors, Desalination 143 (2002) 219–228.

[9] P. Gui, X. Huang, Y. Chen, Y. Qian, Effect of operational parameters on sludge accumulation on membrane surfaces in a submerged membrane bioreactor, Desalination 151 (2002) 185–194.

[10] R. Liu, X. Huang, Y.F. Sun, Y. Qian, Hydrodynamic effect on sludge accumulation over membrane surfaces in a submerged membrane bioreactor, Process Biochem. 39 (2003) 157–163.

[11] Y. Shimizu, Y.I. Okuno, K. Uryu, S. Ohtsubo, A. Watanabe, Filtration characteristics of hollow fiber microfiltration membranes used in membrane bioreactor for domestic wastewater treatment, Water Res. 30 (1996) 2385–2392.

[12] D. Libotean, J. Giralt, F. Giralt, R. Rallo, T. Wolfe, Y. Cohen, Neural network approach for modeling the performance of reverse osmosis membrane desalting, J. Membr. Sci. 326 (2009) 408–419.

[13] Y.G. Lee, Y.S. Lee, J.J. Jeon, S. Lee, D.R. Yang, I.S. Kim, J.H. Kim, Artificial neural network model for optimizing operation of a seawater reverse osmosis desalination plant, Desalination 247 (2009) 180–189.

[14] Ł. Jedliński, J. Jonak, Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform, Appl. Soft Comput. 30 (2015) 636–641.

[15] K. Boussu, C. Vandecasteele, Relation between membrane characteristics and performance in nanofiltration, J. Membr. Sci. 310 (2008) 51–65.

[16] C. Zhou, Z. Wang, Y. Liang, J. Yao, Study on the control of pore sizes of membranes using chemical methods part II. Optimization factors for preparation of membranes, Desalination 225 (2008) 123–138.

[17] X.J. Xi, Y.J. Cui, Z. Wang, J.H. Qian, J. Wang, L.Y. Yang, S.S. Zhao, Study of dead-end microfiltration features in sequencing batch reactor (SBR) by optimized neural networks, Desalination 272 (2011) 27–35.

[18] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, NY, 1999.

[19] Z. Jia, J. Ma, F. Wang, Hybrid of simulated annealing and SVM for hydraulic valve characteristics prediction, Expert Syst. App. 38 (2011) 8030–8036.

[20] V. Vapnik, Statistical Learning Theory, Wiley, New York, NY, 1998.

[21] C. Cortes, V. Vapnik, Support vector networks, Mach. Learn. 20 (1995) 273–297.

[22] N. Cristianini, J.S. Taylor, An Introduction to Support Vector Machines, Cambridge University Press, Cambridge, 2000, pp. 93–120.

[23] R.J. Vong, T.V. Larson, W.H. Zoller, A multivariate chemical classification of rainwater samples, Chemom. Intell. Lab. Syst. 3 (1988) 99–109.

[24] B. Scholkopf, K. Sung, C. Burges, F. Girosi, P. Niyogi, T. Poggio, V. Vapnik, Comparing support vector machines with Gaussian kernels to radial basis function classifiers, IEEE Transactions on Signal Processing 45 (1997) 2758–2765.

[25] V. Vapnik, A. Lerner, Pattern recognition using generalized portrait method, Autom. Rem. Control 24 (1963) 774–780.

[26] M. Stitson, A. Gammerman, V. Vapnik, V. Vovk, C. Watkins, J. Weston, Advances in Kernel Methods—Support Vector Learning, MIT Press, Cambridge, MA, 1999, pp. 285–292.

[27] V. Vapnik, S. Golowich, A.J. Smola, Support vector method for function approximation, regression estimation and signal processing, Adv. Neural Inf. Process Syst. 9 (1996) 281–287.

[28] J.J. Liu, M.H. Bharati1, K.G. Dunn, J.F. MacGregor, Automatic masking in multivariate image analysis using support vector machines, Chemom. Intell. Lab. Syst. 79 (2005) 42–54.

[29] R. Burbidge, M. Trotter, B. Buxton, S. Holden, Drug design by machine learning: Support vector machines for pharmaceutical data analysis, Comput. Chem. 26 (2001) 5–14.

[30] U. Norinder, Support vector machine models in drug design: Applications to drug transport processes and QSAR using simplex optimisations and variable selection, Neurocomputing 55 (2003) 337–346.

[31] L. Bao, Z.R. Sun, Identifying genes related to drug anticancer mechanisms using support vector machine, FEBS Lett. 521 (2002) 109–114.

[32] J.J. Rodríguez, C.J. Alonso, J.A. Maestro, Support vector machines of interval-based features for time series classification, Knowl-based Syst. 18 (2005) 171–178.

[33] U. Thissen, R. van Brakel, A.P. de Weijer, W.J. Melssen, L.M.C. Buydens, Using support vector machines for time series prediction, Chemom. Intell. Lab. Syst. 69 (2003) 35–49.

[34] K.J. Kim, Financial time series forecasting using support vector machines, Neurocomputing 55 (2003) 307–319.

[35] A. Borin, M.F. Ferrão, C. Mello, D.A. Maretto, R.J. Poppi, Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk, Anal. Chim. Acta 579 (2006) 25–32.

[36] D. Wu, Y. He, S.J. Feng, D.W. Sun, Study on infrared spectroscopy technique for fast measurement of protein content in milk powder based on LS-SVM, J. Food Eng. 84 (2008) 124–131.

[37] S.J. Hua, Z.R. Sun, Support vector machine approach for protein subcellular localization prediction, Bioinformatics 308 (2001) 397–407.

[38] Y.D. Cai, X.J. Liu, X.B. Xu, K.C. Chou, Prediction of protein structural classes by support vector machines, Comput. Chem. 26 (2002) 293–296.

[39] A.C. Lorena, A.C.P.L.F. de Carvalho, Protein cellular localization prediction with support vector machines and decision trees, Comput. Biol. Med. 37 (2007) 115–125.

[40] C.Z. Cai, W.L. Wang, L.Z. Sun, Y.Z. Chen, Protein function classification via support vector machine approach, Math. Biosci. 185 (2003) 111–122.

[41] R. Teramoto, M. Aoki, T. Kimura, M. Kanaoka, Prediction of siRNA functionality using generalized string kernel and support vector machine, FEBS Lett. 579 (2005) 2878–2882.

[42] S.R. Amendolia, G. Cossu, M.L. Ganadu, B. Golosio, G.L. Masala, G.M. Mura, A comparative study of K-nearest neighbour, support vector machine and multi-layer perceptron for thalassemia screening, Chemom. Intell. Lab. Syst. 69 (2003) 13–20.

[43] H.X. Liu, X.J. Yao, R.S. Zhang, M.C. Liu, Z.D.E. Hu, B.T. Fan, The accurate QSPR models to predict the bioconcentration factors of nonionic organic compounds based on the heuristic method and support vector machine, Chemosphere 63 (2006) 722–733.

[44] C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discovery 2 (1998) 1–43.

[45] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (2004) 199–222.

[46] W.S. Noble, What is a support vector machine? Nat. Biotechnol. 24 (2006) 1565–1567.

[47] Y. Xu, S. Zomer, R.G. Brereton, Support vector machines: A recent method for classification in chemometrics, Crit. Rev. Anal. Chem. 36 (2006) 177–188.

[48] J. Huang, T. Li, K. Chen, J. Wu, An approach of encoding for prediction of splice sites using SVM, Biochimie 88 (2006) 923–929.

[49] S. Liang, Z. Sun, Sketch retrieval and relevance feedback with biased SVM classification, Pattern Recogn. Lett. 29 (2008) 1733–1741.

[50] E. Gumus, N. Kilic, A. Sertbas, O.N. Ucan, Evaluation of face recognition techniques using PCA, wavelets and SVM, Expert Syst. App. 37 (2010) 6404–6408.

[51] C. SEPA, Water and Wastewater Monitoring Methods, third ed., Chinese Environmental Science Publishing House, Beijing, China, 1997.

[52] S.Y. Mei, SVM ensemble based transfer learning for large-scale membrane proteins discrimination, J. Theor. Biol. 340 (2014) 105–110.

[53] H. Xiao, B. Biggio, B. Nelson, H. Xiao, C. Eckert, F. Roli, Support vector machines under adversarial label contamination, Neurocomputing 160 (2015) 53–62.

[54] H. Su, X.B. Wu, X.H. Yan, A. Kidwell, Estimation of subsurface temperature anomaly in the Indian Ocean during recent global surface warming hiatus from satellite measurements: A support vector machine approach, Remote Sens. Environ. 160 (2015) 63–71.

[55] Z.W. Guo, J.G. Yao, Y.C. Jiang, X.Q. Zhu, Z.W. Tan, W. Wen, A novel distributed unit transient protection algorithm using support vector machines, Electr. Power Syst. Res. 123 (2015) 13–20.

[56] S. Kim, Z. Yu, R.M. Kil, M. Lee, Deep learning of support vector machines with class probability output networks, Neural Networks 64 (2015) 19–28.

[57] Y. Li, C.K. Li, J. Tao, L.D. Wang, Study on spatial distribution of soil heavy metals in Huizhou City based on BP–ANN modeling and GIS, Procedia Environ. Sci. 10 (2011) 1953–1960.

[58] R. Prasad, A. Pandey, K.P. Singh, V.P. Singh, R.K. Mishra, D. Singh, Retrieval of spinach crop parameters by microwave remote sensing with back propagation artificial neural networks: A comparison of different transfer functions, Adv. Space Res. 50 (2012) 363–370.

[59] F. Li, Q. Tian, B. Yang, L. Wu, C.H. Deng, Effect of polyvinyl alcohol addition to model extracellular polymeric substances (EPS) on membrane filtration performance, Desalination 286 (2012) 34–40.

[60] X.J Xi, Z. Wang, J. Zhang, Y.N Zhou, N. Chen, L.Y Shi, W.Y Dong, L.N Cheng, W.T Yang, Prediction of impacts of fabrication conditions on the filtration performance of homemade VC-co-VAc-OH microfiltration membrane by support vector machine (SVM), Desalin. Water Treat. 51 (2013) 3970-3978.