Desalination and Water Treatment
www.deswater.com

doi: 10.1080/19443994.2016.1140080

57 (2016) 24132–24145
November

Taylor & Francis
Taylor & Francis Group

Modeling of membrane fouling in a submerged membrane reactor using support vector regression

Serhan Aydin Aya[a,*], Turkan Ormanci Acar[b], Nese Tufekci[b]

[a]*Faculty of Mechanical Engineering, Department of Mechanical Engineering, Istanbul Technical University, Gumussuyu 34437, Istanbul, Turkey, Tel. +90 212 293 13 00, ext. 2489; Fax: +90 212 245 07 95; email: ayas@itu.edu.tr*
[b]*Engineering Faculty, Departmant of Environmental Engineering, Istanbul University, Istanbul, Turkey, Tel. +90 212 473 7070; emails: turkanormanci@gmail.com (T. Ormanci Acar), nese@istanbul.edu.tr (N. Tufekci)*

Received 27 July 2015; Accepted 30 December 2015

ABSTRACT

Removal rate of Fe^{2+} and Mn^{2+} using submerged membrane reactor for drinking water in the presence of fulvic acid and iron hydroxide is studied using the data from the experiments obtained from various concentrations of Fe^{2+}, Mn^{2+}, fulvic acid, and iron hydroxide. The relationship between these contaminants and membrane fouling is investigated. In the experiments, flux is kept as constant, and the pressure change with time is observed. To model the relationship, a regression analysis using the support vector regression (SVR) model is presented. Hyperparameter optimization for SVR is important, that is, wrong selection may cause underfitting/overfitting phenomena. In order to find optimal values, grid search method is performed with various parameters such as different kernel functions (radial basis functions, polynomial, linear), cost parameter (C), and scale parameters γ and ε. The results obtained by SVR show that proposed method is feasible.

Keywords: Submerged membrane; Fulvic acid; Iron hydroxide; Membrane fouling; Support vector regression

1. Introduction

Iron, manganese, and fulvic acid are commonly found together in natural water sources. Although they are non-hazardous materials for human health, they may cause bad taste and esthetic problems such as the staining of clothes and the deterioration of plumbing fixtures. In industry, ferrous iron and manganese cause severe economic losses due to the discoloration of products; the specks in finished paper, textile, food, and beverage products; and the reduction in the capacity of pipelines. Natural organic matter (NOM) includes high fractions of fulvic acid that can affect many chemical and biological processes in drinking water treatment plants. If iron and manganese form complexes with NOM and present in dissolved form, oxidation can be prevented. When they form complex structures with NOMs, conventional filtration methods are not enough to remove these complexes. For the removal of iron and manganese from drinking water, low-pressure hollow-fiber membranes are widely used as a replacement for the conventional filtration technologies.

Several statistical learning methods such as multiple linear regression (MLR) and artificial neural networks (ANN) are intensively used for different

*Corresponding author.

studies concerning various fields of environmental engineering [1,2]. Recent developments in statistical learning methods resulted in finding a novel theory called support vector machines (SVMs) in the early 1990s as a non-linear solution for classification and regression tasks [3,4]. The SVM method relies on the statistical learning theory to find a prediction rule for output values that works well for new, so far unseen input values. ANN and SVM can be used for same task; however, ANNs can suffer from multiple local minima, while the solution to a SVM is global and unique, which is a significant advantage of SVMs. SVMs have other advantages that they have a simple geometric interpretation and give a sparse solution. Unlike ANNs, computational complexity of SVMs does not depend on the dimensionality of the input space. Optimization processes also differ between these methods: ANNs use empirical risk minimization, while SVMs use structural risk minimization.

For the investigation of membrane fouling, Gao et al. established a SVM network structure [5]. The results were compared to the results of artificial neural network model. It was concluded that SVM regression models are superior to the traditional ANN models, because of the better generalization ability and classification accuracy for sewage treatment using Membrane Bioreactor. Bouamar and Ladjal presented a method using SVM to determine classification of water quality [6]. The proposed method was applied to control of risks in the factories that produce and distribute water.

Using SVM method, Aryafar et al. predicted the heavy metals included in the acid mine drainage (AMD) [7]. To assess the model accuracy, the results of SVM with those of the general regression neural network (GRNN) were compared. It has been found that SVM makes the running time considerably faster with a higher accuracy.

Gao et al. [8] applied structure optimized SVM model to predict the membrane permeate flux during dead-end microfiltration of activated sludge suspensions from sequencing batch reactor (SBR) with different experimental samples. Good agreement between the experimental data and predicted values proved that the SVM model has sufficient prediction accuracy.

SVMs achieve a global solution in the search for optimal parameter values and there is no need for trial and error procedures to determine the final machine architecture, which is directly obtained through structural risk minimization principle. This method effectively solves the over-fitting phenomena, assures good generalization ability, and better accuracy. In this study, SVM regression method (SVR) is proposed for the mathematical modeling of the fouling effects of Fe^{2+}, Mn^{2+}, fulvic acid, and iron hydroxide on membrane for the various concentrations of Fe^{2+}, Mn^{2+}, fulvic acid, and iron hydroxide. Parameter selection for SVR is significantly important factor for choosing an appropriate and high-performance data-driven model, because wrong selection of these parameters may cause under-fitting/over-fitting phenomena. Cross-validation method is employed to determine the optimum values of the model parameters, namely kernel functions, cost parameter, ε value. All the SVM regression models constructed here have performed well for given data.

2. Method and materials

2.1. Materials

One grams of Fe^{2+} and 1 g/L Mn^{2+} stock solutions are prepared using $FeSO_4{\cdot}7H_2O$ and $MnCl_2{\cdot}2H_2O$. Fulvic acid is supplied by International Humic Substances Society, University of Minnesota. Iron hydroxide is prepared as 50 mg/L by aerating Fe^{2+} stock solution for 3 or 4 h at 8×10^{-3} eq/L alkalinity with addition of $NaHCO_3$ before the experiment. Zee Weed-1 (ZW-1) module membranes are supplied by GE Water and Process Technologies. The properties of ZW-1 membrane are given in Table 1.

2.2. Experimental procedure

Synthetic solutions are fed into 10 cm × 20 cm × 45 cm-sized Plexiglas reactor by a peristaltic pump. During the experiment, air is fed from the bottom of the reactor by using fine bubble diffusers. Backwash is maintained through feedback of filtered water with a backwash pump. Pressure changes are continuously monitored by a pressure gauge. The experimental setup is shown schematically in Fig. 1.

Experiments are conducted with synthetic solutions. Every experiment lasted for three days. The concentrations of iron, manganese, and fulvic acid are chosen as 1–5 mg/L, 1–2 mg/L, and 1–7 mg/L, respectively. Iron hydroxide is used to increase iron and manganese removal, and it is obtained from aeration of solution containing Fe^{2+}.

3. Support vector regression analysis

SVMs are formulated from the principles of statistical learning theory that aims to minimize an upper bound of generalization error based on the structural risk minimization. SVM algorithms are used for classification and regression. In SVMs for regression (SVR), $\{(x_i, y_i)\}_{i=1}^{l}$ is considered as a training set where each

Table 1
The properties of ZW-1 membrane

Membrane type	Capillary, hydrophilic
Surface area (m^2)	0.047
Membrane material	PVDF
Nominal membrane pore size (nm)	40
Capillary outer diameter (mm)	2
Module length (cm)	17.5
Module width (cm)	5.8
Module inner diameter volume (ml)	10
Transmembrane pressure	max. 62 kPa at 40°C
Flux (L/m^2 sa)	15–35

$x_i \subset R^n$ represents the input space of the sample and has a corresponding scalar measure output value $y_i \subset R$ for $i = 1 \ldots l$, *as* l denotes the size of the training data. The aim of this method is to find a function that predicts the corresponding response value, in best possible way. The formulation of SVRs generally results in a function estimation equation analogous to the following form:

$$f(\boldsymbol{x}) = \langle \boldsymbol{w}, x \rangle + b \tag{1}$$

where $w \subset R^n$ and $b \subset R$ denote the n-dimensional weight vector and the offset of the linear regression function, respectively. The main goal is to find a function having at most ε deviation from the actual target vectors for all given training data. The ε-insensitive loss function is the most widely used cost function. This function is in the form:

$$L(y, f(\boldsymbol{x})) = |y - f(\boldsymbol{x})|_{\varepsilon} = \begin{cases} 0, & |y - f(\boldsymbol{x})| \leq \varepsilon \\ |y - f(\boldsymbol{x})| - \varepsilon, & \text{otherwise} \end{cases} \tag{2}$$

where f is a real value function on the field $\boldsymbol{x}$. Loss function describes ε as the fitting precision and if the difference between predicted value and the actual value is less than ε, the loss is equal to 0. Considering fitting error, solving regression function (i.e. Eq. (1)) can be expressed as a constrained optimization problem:

$$\text{minimize } R(\boldsymbol{w}) = C \sum_{i=1}^{l} (\xi_i + \xi_i^*) + \frac{1}{2} \|\boldsymbol{w}\|^2 \tag{3}$$

$$\text{subject to} \begin{cases} f(\boldsymbol{x_i}) - y_i \leq \xi_i^* + \varepsilon \\ y_i - f(\boldsymbol{x_i}) \leq \xi_i^* + \varepsilon \\ \xi_i, \xi_i^* \geq 0 \quad (i = 1, 2, \ldots, l) \end{cases}$$

where the constant $C > 0$ represents a regularization parameter that allows tuning the trade-off between the

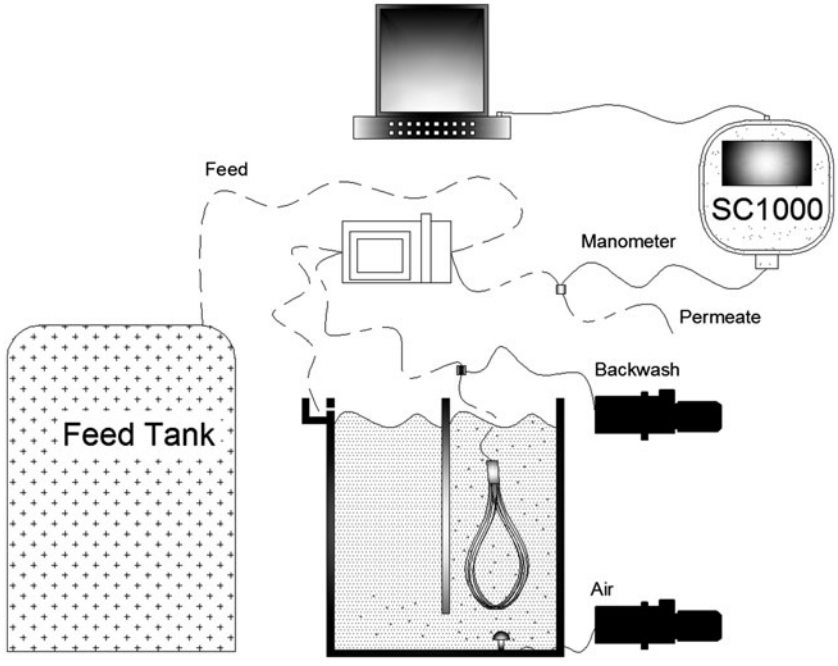

Fig. 1. UF membrane experimental setup.

smoothness of the function f and the value of that allowed error is larger than ε. The cost parameter is the main tool for adjusting the complexity of the model. Larger values result flexible models since the effect of errors is amplified. However, when the cost is small, the model will stiffen and become less likely to overfit and more likely to underfit because the contribution of the squared parameters is proportionally large in the modified error function. ξ_i, ξ_i^* are named as slack variables and they account for samples which do not lie in the ε-deviation tube. Using the slack variables allows some error to deal with noise in the training data.

The optimization problem given in Eq. (3) can be reformulated through a Lagrange function from the objective function with Lagrange multipliers. The Lagrange multipliers α and α^* can be found by a dual optimization leading to quadratic programming (QP) solution:

$$\text{minimize } \frac{1}{2}\sum_{i,j}^{l}(\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)(\boldsymbol{x_i} \cdot \boldsymbol{x_j}) + \varepsilon\sum_{i=1}^{l}(\alpha_i^* + \alpha_i) - \varepsilon\sum_{i=1}^{l}\boldsymbol{y_i}(\alpha_i^* - \alpha_i) \tag{4}$$

$$\text{subject to } \begin{cases} \sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \\ 0 \le \alpha_i^*, \alpha_i \le C, \quad i = 1, 2, \ldots, l \end{cases}$$

The vector $\boldsymbol{w}$ can be written in terms of training data points as:

$$\boldsymbol{w} = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)\boldsymbol{x_i} \tag{5}$$

By substituting Eq. (5) into Eq. (1), SVR function $f(\boldsymbol{x})$can be rewritten as:

$$f(\boldsymbol{x}) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)(\boldsymbol{x_i} \cdot \boldsymbol{x_i}) + b = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)k(\boldsymbol{x_i}, \boldsymbol{x}) + b \tag{6}$$

In Eq. (6), the dot product can be replaced with function $k(\boldsymbol{x_i}, \boldsymbol{x})$, known as the kernel function. The basic idea behind SVMs is mapping the input space into feature space utilizing kernels. Thus, SVR models are linear models obtained in a new feature space, which is the result of this transformation. This procedure is known as "kernel trick" and enables the SVM to work with nonlinear mapping in the feature spaces having very high dimensions. The flexibility of the SVR is provided by the use of kernel functions since a linear solution in the higher dimensional feature space corresponds to a nonlinear solution in the original, lower dimensional input space. Linear, polynomial, or Gaussian radial basis function can be selected as the kernel function K.

In Table 2, d stands for degree and σ is a constant parameter of the kernel and can control the amplitude of the radial basis function and the generalization ability of SVR. Similar with other multivariate statistical models, the performances of SVM for regression depend on the optimal selection of parameters. Also, preprocessing of data (standardization, detection of outliers, missing values, etc.) greatly affects the results. The following is a summary of preprocessing and parameter selection steps for this study.

3.1. Standardization

For a given vector, subtracting a measure of location and dividing by a measure of scale is often referred as standardization. The first step of data processing is to standardize all sample values by subtracting the mean from the observed value and dividing the difference by the standard deviation of all samples for a given variable. Standardizing either input or target variables tend to make the training process better behaved by improving the numerical condition of the optimization problem and ensuring that various default values involved in initialization and termination are appropriate. In case of SVR, the raw data are standardized to an interval by transformation. Many elements used in the objective function of a learning algorithm (such as the RBF kernel of SVMs or the $L1$ and $L2$ regularizers of linear models) assume that all features are centered at zero and have variance in the same order. If a feature has a variance that is orders of magnitude larger than others, it might dominate the objective function and make the estimator unable to learn from other features correctly as expected.

Table 2
Kernel functions for SVMs

Kernel type	Kernel function
Linear kernel	$K(x_i, x) = \langle \boldsymbol{x}, \boldsymbol{x_i} \rangle$
Polynomial kernel	$K(x_i, x) = (\langle \boldsymbol{x}, \boldsymbol{x_i} \rangle + 1)^{\mathrm{d}}$
Radial basis function kernel	$K(x_i, x) = \exp\left(-\lvert \boldsymbol{x} - \boldsymbol{x_i} \rvert^2 / 2\sigma^2\right)$

Since the predictors enter into the model as the sum of cross products, differences in the predictor scales can affect the model. Therefore, centering and scaling the predictors prior to building a SVM model is important. In this study, all the exploratory variables and target values are transformed to the same ground-uniform distributions on −1, +1.

3.2. Parameter selection

To avoid over-fitting, it is a common practice when performing a supervised machine learning experiment to hold out part of the available data as a test set. In this study, for each model, a training data-set is used to establish the SVR models, and the remaining data are used to evaluate the performance of the models. In the models, we developed, 3 of 5 of the data-set is used to train the model and the remaining 2 of 5 is used to test it.

Selecting which kernel function to be used in the analysis depends on the problem. The radial basis function has been shown to be very effective. However, when the regression line is truly linear, the linear kernel function will be a better choice. Note that some of the kernel functions have extra parameters. For example, the polynomial degree in the polynomial kernel must be specified. Similarly, the radial basis function has a parameter (γ) that controls the scale. These parameters, along with the cost value, constitute the tuning parameters for the model. Therefore, the application of SVR involves the optimization of the regularization cost parameter (C), type of kernel, and the kernel-specific parameter γ.

3.2.1. Cross-validation

Determining appropriate values of regularization cost parameter (C), type of kernel, and the kernel-specific parameter γ are often achieved by trial and error. It is important to conduct cross-validation tests by applying the tree built from one set of observations (train data) to another completely independent set of observations (test data). This approach involves randomly dividing the set of observations into k groups, or folds, of approximately equal size. The first fold is treated as a validation set, and the method is fit on the remaining $k - 1$ folds. The mean squared error is then computed on the observations in the held-out fold. This procedure is repeated k times; each time, a different group of observations is treated as a validation set. If most of the splits in the development sample are driven by noise, then the prediction on the validation sample would be poor. Ten-fold cross-validation is used to examine the best performing regression. The entire sample is randomly divided into 10 mutually exclusive subsets of roughly the same size. Each of the 10 subsets is reserved as the validation sample and the model estimated using the remaining 9 subsets.

3.2.2. Grid search

Grid search is simply an exhaustive searching through a manually specified subset of the hyperparameter space of a learning algorithm. A grid search algorithm must be guided by some performance metric, typically measured by cross-validation on the training set or evaluation on a held-out validation set [9]. In this study, the following parameters are investigated for optimal value selection: Radial basis function, polynomial and linear kernels for kernel function; 10^{-1}, 10^{-2}, …, 10^{-5} for ε and γ values (hyperparameter $\gamma = 1/2\sigma^2$); 1, 2, 3, and 4 for degree used in polynomial kernels and 1, 10, 100, and 1,000 for cost value (C). Grid search trains an SVR with each parameter and evaluates their performance on a held-out validation set. Finally, the grid search algorithm outputs the settings that achieved the highest R^2 score in the validation procedure. As a consequence, the regression result has optimal properties.

4. Results

Iron, manganese, fulvic acid, and iron hydroxide concentrations for each run and the mean value of the resulting pressure differences are given in Table 3. Different concentrations and resulting pressure values give information about the effects of the ingredients. In view of this table, the following is the detailed explanation of the relationship between iron, manganese, iron hydroxide, and fulvic acid for clarification of experimental results.

Iron, manganese, and fulvic acid are commonly found together in natural water sources. Fouling is the main problem while operating membrane systems. In this study, fouling effects of Fe^{2+}, Mn^{2+}, fulvic acid, and iron hydroxides and removal of Fe^{2+} and Mn^{2+} were investigated. Drinking water treatment plants are generally operated at pH 8.5 in the aeration basin. It is well known that the oxidation rate of Mn^{2+} is quite slow under the pH 9.5. Thus, the fouling effect of Mn^{2+} is expected to be quite high, but removal efficiency is expected to be lower. Also, fulvic acid is notably a foulant to membrane systems due to its ability to solute at every pH, and it effects the removal of Fe^{2+} and Mn^{2+}. Since these contaminants increase the membrane fouling, adding iron hydroxide decreased

Table 3
Combination of iron, manganese, fulvic acid, and iron hydroxide for oxidation and membrane filtration experiments

Run	Fe^{2+} (mg/L)	Mn^{2+} (mg/L)	Fulvic acid (mg/L)	Ferric hydroxide (mg/L)	Mean pressure (mbar)
1	1	–	–	–	39.76
2	5	–	–	–	38.50
3	–	1	–	–	122.69
4	–	2	–	–	133.11
5	5	1	–	–	64.85
6	–	–	1	–	50.14
7	5	1	1	–	66.25
8	5	1	1	50	49.39
9	–	–	7	–	109.37
10	5	1	7	–	48.52
11	5	1	7	50	46.91

the membrane fouling originated from Mn^{2+} and fulvic acid.

In the experiments, $Fe(OH)_3$ and MnO_2 flocs are occurred with the oxidation of Fe^{2+} and Mn^{2+}. At pH 8.5, flocculation of Fe^{2+} is extremely rapid. Oxidation of Fe^{2+} and Mn^{2+} is both catalyzed by $Fe(OH)_3$, which accelerates the oxidation process. In the beginning of the reaction of Fe^{2+} and Mn^{2+} with fulvic acid, the oxidation process is accelerated. However, with the increase in fulvic acid concentration (7 mg/L), the oxidation process slows down because of the complexation between Fe^{2+} and fulvic acid. On the other hand, catalytic effect of $Fe(OH)_3$ overrides that decrease factor. Comparing the removal efficiency of Mn^{2+} between (Fe^{2+}, Mn^{2+}, and FA) and (Fe^{2+}, Mn^{2+}, and Fe $(OH)_3$) combinations, higher removal efficiency is observed in latter group which contains $Fe(OH)_3$.

Experimental part of this study is focused on removal of Fe^{2+} and Mn^{2+}. Improvement of removal efficiency is originated from iron hydroxide flocs. After membrane filtration, Mn^{2+} removal efficiency is significantly increased. In the working pH, iron hydroxide has positive charge because of point of zero charge. Thus, fulvic acid (negative charge) can be blocked by iron hydroxide and flocculation. Thus, both removal efficiency and membrane fouling improve by contribution of iron hydroxide.

In the next step, a unified SVR approach is presented for analysis and prediction of the membrane performance under various combinations of input parameters. Cleaning up the data is the first step of analysis process. First, by visually inspecting the data, some of the experimental errors can easily be detected. For example, we know that the pressure value should not be below zero (Fig. 2); therefore, these data should be cleaned before we apply the robust modern algorithms for regression analysis. Another important concept is outliers. Outliers in training or test subset may cause higher prediction errors, therefore affect the regression modeling results. Since the outliers are physically located far away from the normal data, they will be rarely predicted correctly. The outliers may be detected if the regression process is implemented several times with changed training and test sets and record the predicted cases for every single run. In this study, median absolute deviation (MAD) is used to detect outliers. In statistics, MAD is a measure of statistical dispersion that represents a measure of the variability of a univariate sample of quantitative data [10].

The clean datasets along with their sample size and contents are summarized in Table 4. The parameters for the support vector regression (SVR) analysis found by grid search method are given in Table 5. The model performance criteria parameters for the test sets that represent the quality of predictions are also included in this table. The metrics that are applied to the estimators are the coefficient of determination R^2 and root-mean-square error RMSE.

$$R^2 = 1 - \frac{\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2}{\sum_{i=0}^{n-1}(y_i - \bar{y}_i)^2} \tag{7}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2}{n}} \tag{8}$$

In Eqs. (7) and (8), y_i stands for the measured value, $\bar{y}_i$ is the mean of the measured values, $\hat{y}_i$ denotes the predicted value, and n is the number of samples.

The data are obtained from the experiments with various concentrations of Fe^{2+}, Mn^{2+}, fulvic acid, and iron hydroxide. The data are separated into two subsets, namely train and test sets with a ratio of 0.6 and 0.4, respectively.

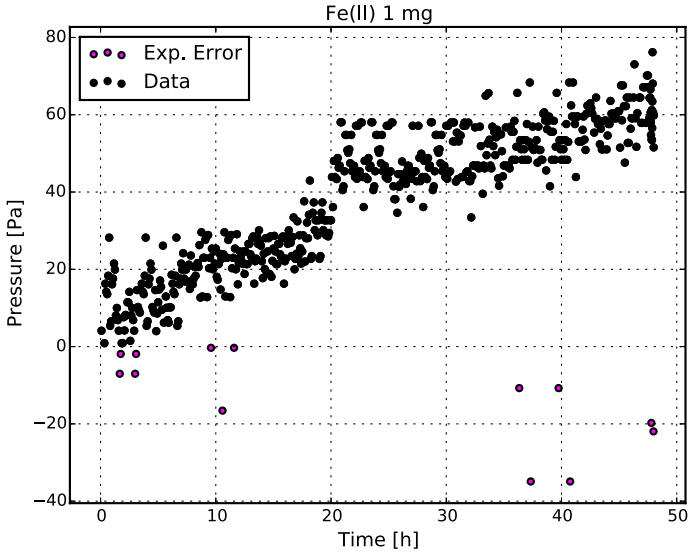

Fig. 2. Detecting experimental errors.

Table 4
Summary of the observation data

Data	Size	Contents
Dataset 1	573	Fe^{2+} 1 mg
Dataset 2	596	Fe^{2+} 5 mg/L
Dataset 3	593	Mn^{2+} 1 mg/L
Dataset 4	593	Mn^{2+} 2 mg/L
Dataset 5	579	Fe^{2+} 5 mg/L; Mn^{2+} 1 mg/L
Dataset 6	596	Fulvic Acid 1 mg/L
Dataset 7	592	Fe^{2+} 5 mg/L; Mn^{2+} 1 mg/L; Fulvic Acid 1 mg/L
Dataset 8	595	Fe^{2+} 5 mg/L; Mn^{2+} 1 mg/L; Fulvic Acid 1 mg/L; FeO_x 50 mg/L
Dataset 9	586	Fulvic Acid 7 mg/L
Dataset 10	573	Fe^{2+} 5 mg/L; Mn^{2+} 1 mg/L; Fulvic Acid 7 mg/L
Dataset 11	530	Fe^{2+} 5 mg/L; Mn^{2+} 1 mg/L; Fulvic Acid 7 mg/L; FeO_x 50 mg/L

Table 5
Parameters and performance metrics for SVR

Data	Kernel	C	ε	γ	R^2	RMSE
Dataset 1	rbf	1,000	0.1	0.1	0.8502	6.4193
Dataset 2	rbf	1,000	0.01	0.1	0.8891	6.6864
Dataset 3	rbf	100	0.1	0.1	0.8987	12.0301
Dataset 4	rbf	1,000	0.01	0.1	0.9335	11.4695
Dataset 5	rbf	1,000	0.1	0.1	0.5302	7.3795
Dataset 6	rbf	1,000	0.01	0.1	0.5859	4.9690
Dataset 7	rbf	1,000	0.1	0.1	0.8339	4.9530
Dataset 8	rbf	10	0.1	0.1	0.7935	4.5126
Dataset 9	rbf	1,000	0.01	0.1	0.9475	7.4144
Dataset 10	rbf	10	0.1	0.1	0.4299	4.6841
Dataset 11	rbf	100	0.1	0.1	0.8760	3.2522

In SVR, pressure is the dependent variable, whereas the time is the independent variable. A detailed grid search method with a 10-fold cross-validation is used to derive the optimal SVR model parameters. It is important to note that in this study, all the exploratory variables and target values are transformed to the same ground-uniform distributions on −1, +1. The model is tuned over four cost values between 1 and 1,000, three kernel functions radial basis function, polynomial and linear kernel functions, five different values 10^{-1}, 10^{-2}, …, 10^{-5} for ε and γ values. For the polynomial model, we tuned over the polynomial degree, the cost, and scale factor ε. The polynomial degree is changed from 1 to 4. Tuning the radial basis function kernel parameter is easier than tuning the polynomial model, which has three tuning parameters. In general, quadratic models

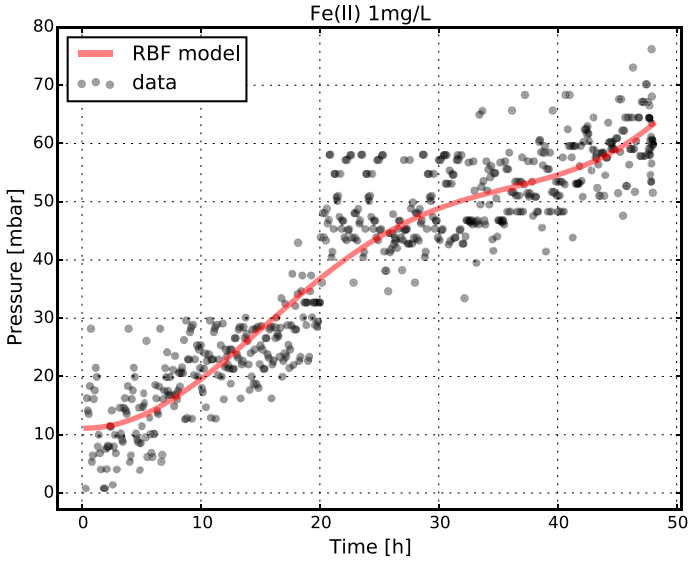

Fig. 3. Effect of 1 mg/L Fe^{2+} on membrane fouling.

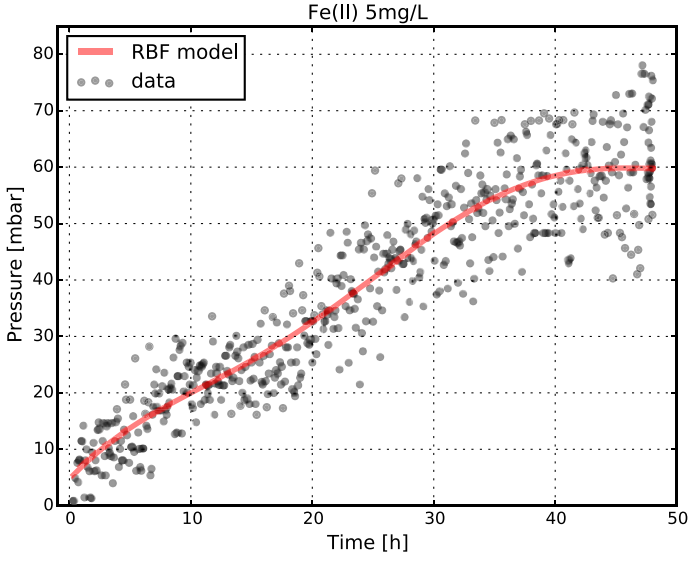

Fig. 4. Effect of 5 mg/L Fe^{2+} on membrane fouling.

tend to have smaller error rates than the linear models, and the models associated with larger scale factors have better performance [11].

Among the linear, polynomial, and RBF kernel functions, the latter is finally selected to be used SVR models as it yielded the highest R^2 and corresponding parameter is estimated to be $\gamma = 0.1$ for all datasets. After the SVR parameters are found, the data scaled back to original range. We give the corresponding model performance criteria parameters (R^2 and RMSE) for each analysis, obtained from test sets. Here, R^2 is the value corresponds to grid search best model, which is calculated with the standardized test set; and RMSE is the other performance criteria obtained from the predicted and the real values of the test set in original data scale. To sum up; kernel, cost, scale parameters ε and γ are calculated using the standardized train sets, and RMSE is determined from the predicted values and the real values from test set in original scale. For most of the datasets, when the cost

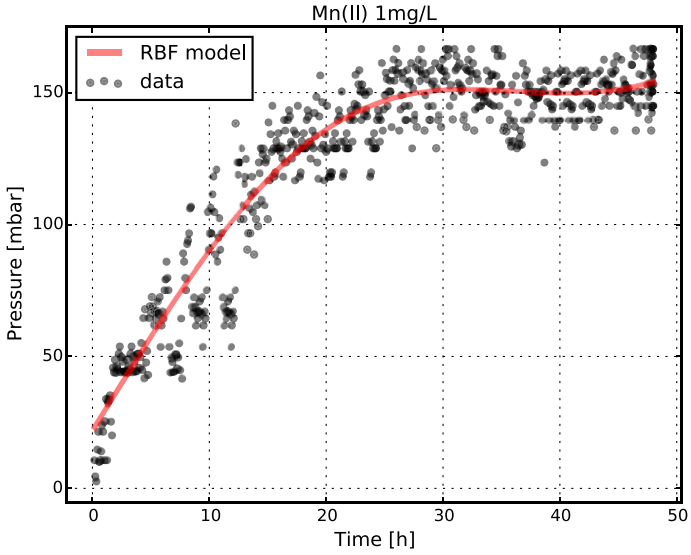

Fig. 5. Effect of 1 mg/L Mn^{2+} on membrane fouling.

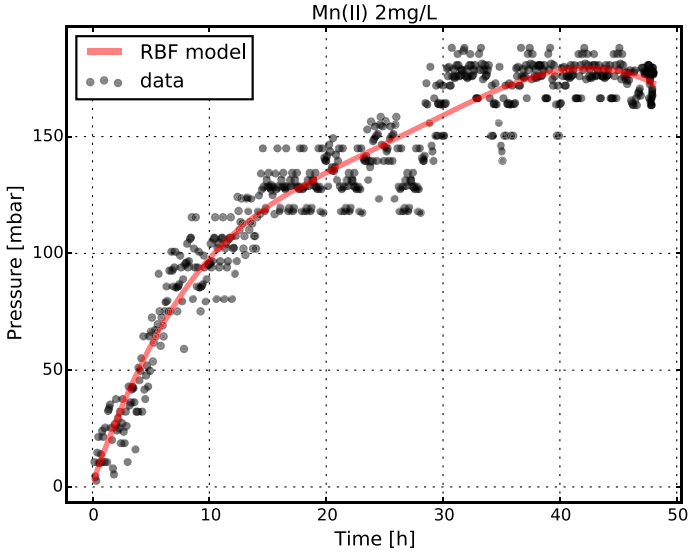

Fig. 6. Effect of 2 mg/L Mn^{2+} on membrane fouling.

values are small, the model underfits the data, but, as the error starts to increase when the cost approaches 1,000, overfitting begins (Table 5).

For the first dataset, the optimal model is created with radial basis function and the cost value associated with the largest R^2 is 1,000. The value of epsilon determines the level of accuracy of the approximated function. It relies entirely on the target values in the training set. If epsilon is larger than the range of the target values, we cannot expect a good result. If epsilon is zero, we should expect overfitting. After the grid search method is performed largest score is obtained with $\varepsilon = 0.1$, which reflects the data in best possible way. The red line in Fig. 3 represents an SVR model with a radial basis kernel function with these parameters. This line better describes the overall structure of the data. As a comparison, both the optimal radial basis and the polynomial SVM models use a similar number of support vectors, 291 and 280, respectively (of 343 training samples).

For the second dataset, a similar uptrend is seen with previous set for pressure change with time (Fig. 4). Also, for both experiments, the pressure stops increasing at around 60 mbar, which proves that

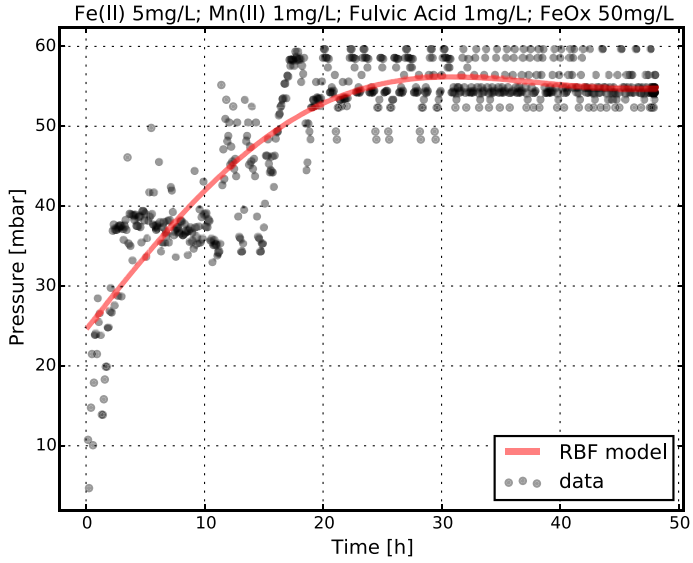

Fig. 7. Effects of 5 mg/L Fe^{2+}, 1 mg/L Mn^{2+} 1 mg/L fulvic acid, and 50 mg/L FeO_x on membrane fouling.

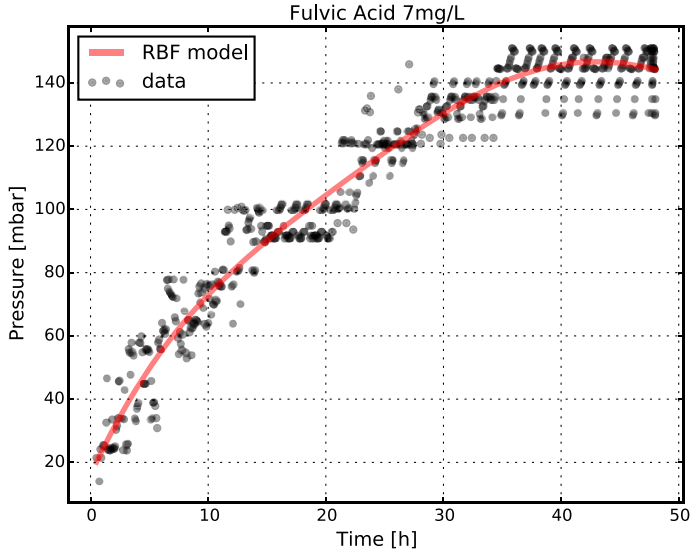

Fig. 8. Effect of 7 mg/L fulvic acid on membrane fouling
.

increasing concentration of Fe^{2+} does not affect the pressure change, as so membrane fouling. Fouling slightly decreases to the end because of the catalytic effect of iron hydroxide on oxidation when the Fe^{2+} concentration is 5 mg/L.

As seen from Fig. 4, the cost value associated with the largest R^2 is 1,000 in second dataset. Generally, choosing ε to a certain accuracy only guarantee that accuracy on the training set. In order to achieve better accuracy overall, we need to choose a slightly smaller ε. Keeping that in mind, we selected other parameters as $\varepsilon = 0.01$ and $\gamma = 0.1$, according to largest R^2. The optimal model is created with radial basis function, which is the same as the model created for the first dataset. Third, fourth, eighth, and ninth datasets have similar trends for pressure change with time (Figs. 5–8). Figs. 5 and 6 show the effect of Mn^{2+} on membrane fouling. Since higher pressure differences are observed, it is

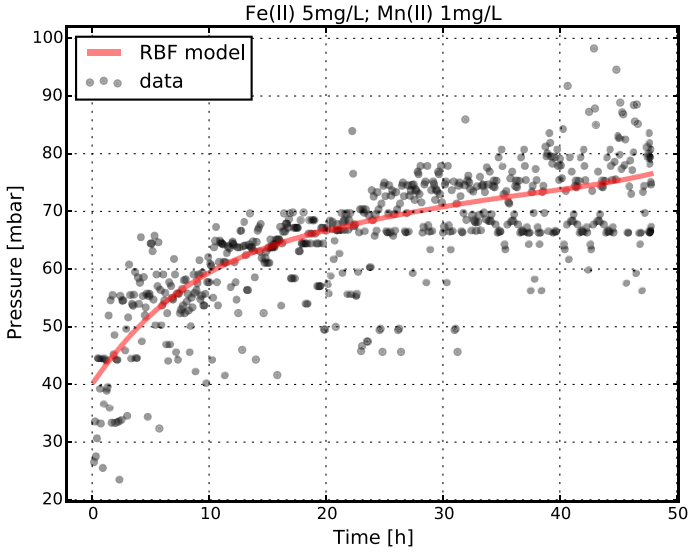

Fig. 9. Effects of 5 mg/L Fe^{2+} and 1 mg/L Mn^{2+} on membrane fouling.

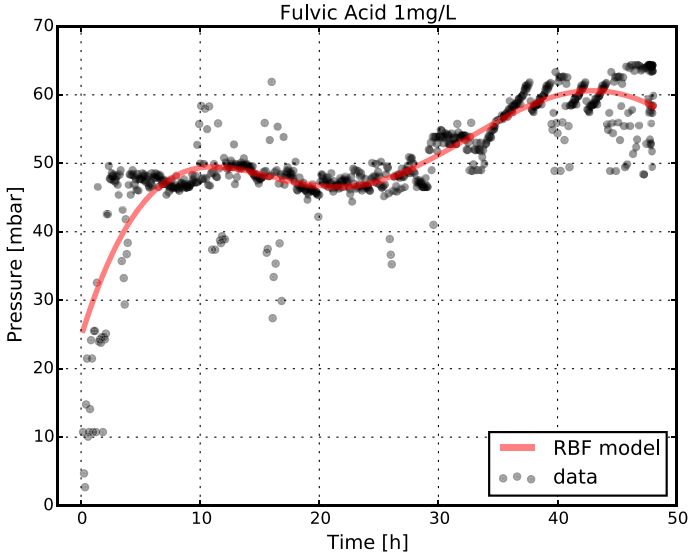

Fig. 10. Effect of 1 mg/L fulvic acid on membrane fouling.

concluded that the Mn^{2+} is notably contaminant compared to Fe^{2+}. However, it is also observed that membrane fouling is decreased in the presence of 5 mg/L Fe^{2+} (Fig. 9). This is originated from the auto-catalytic effect of oxidized Fe^{2+} and its adsorption capacity. Figs. 10 and 8 show single effects of 1 mg/L fulvic acid (FA) and 7 mg/L FA, respectively, and it is clear from that the increment of FA have adverse effect on membrane fouling. However, from Figs. 11 and 12, in which FA is present together with Fe^{2+} and Mn^{2+}, it is observed that the membrane fouling decreases with the increase in FA concentration. It can be evaluated at the second membrane formation effect, which is formed by complexation of fulvic acid and Fe^{2+} and possibly Mn^{2+}. This second membrane cake avoids fouling of membrane pores and decreases the pressure changes. The similar effect can also be observed from iron hydroxide as shown in Figs. 7 and 13.

Although the SVR model clearly catches overall trend as seen from Figs. 9 and 10, the corresponding metrics $R^2 = 0.5302$ for dataset 5 and $R^2 = 0.5859$ for dataset 6 shows the model's weak performance. The reason for the performance decline is the noisy data. Several sources on SVM suggested that the optimal ε

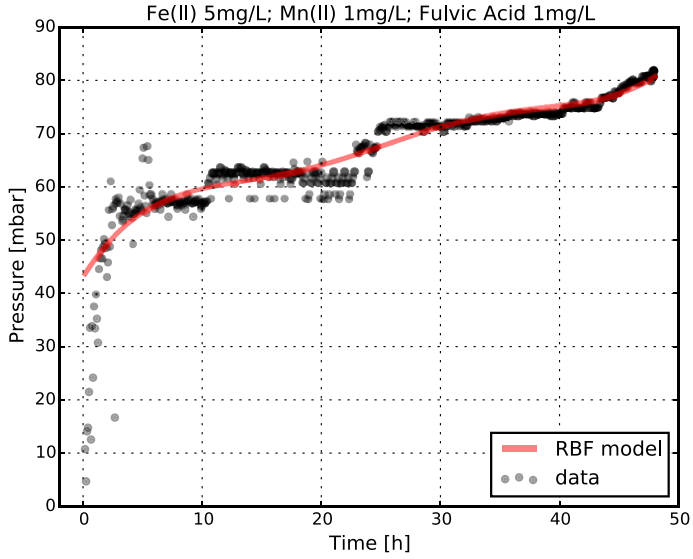

Fig. 11. Effects of 5 mg/L Fe^{2+}, 1 mg/L Mn^{2+}, 1 mg/L fulvic acid on membrane fouling.

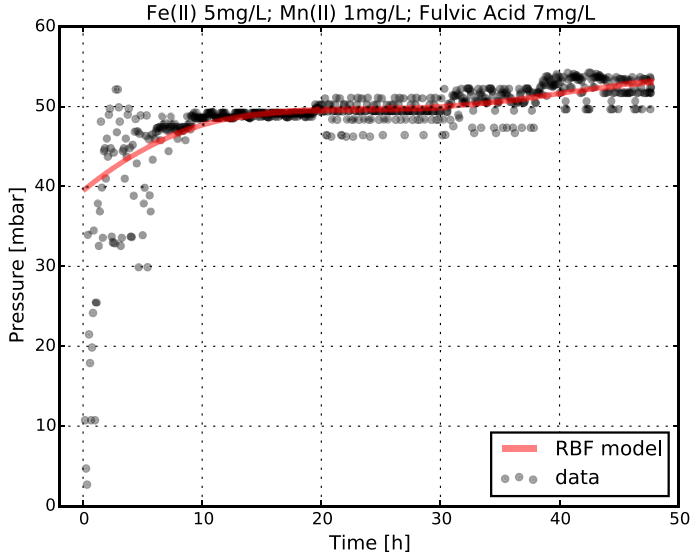

Fig. 12. Effects of 5 mg/L Fe^{2+}, 1 mg/L Mn^{2+}, 7 mg/L fulvic acid on membrane fouling.

values are proportional to noise variance [12,13]. However, the effect of the sample size should also be considered. In the fifth dataset, we have $C = 1{,}000$ and $\varepsilon = 0.01$ as the best model parameters. To see the effect of cost value, we fixed epsilon to 0.01 and changed cost to lower values. This procedure is repeated for the sixth dataset and it is observed that with optimal choice of ε, the value of regularization parameter C has negligible effect on the generalization performance. But one should remember that the cost value should be larger than a certain level as it is observed that the error increases when the parameter is selected far from the optimal value. For datasets 6, 7, 10, and 11, the pressure values increase linearly with time at the first couple of hours. In order to achieve better regression model performance on given data, we separated the linear part, which approximately corresponds to the first 20 values for each dataset, and trained the SVR model with the 3 of 5 of the remaining nonlinear data. In the meantime, for the linear part, linear regression method is applied and corresponding regression coefficients are obtained.

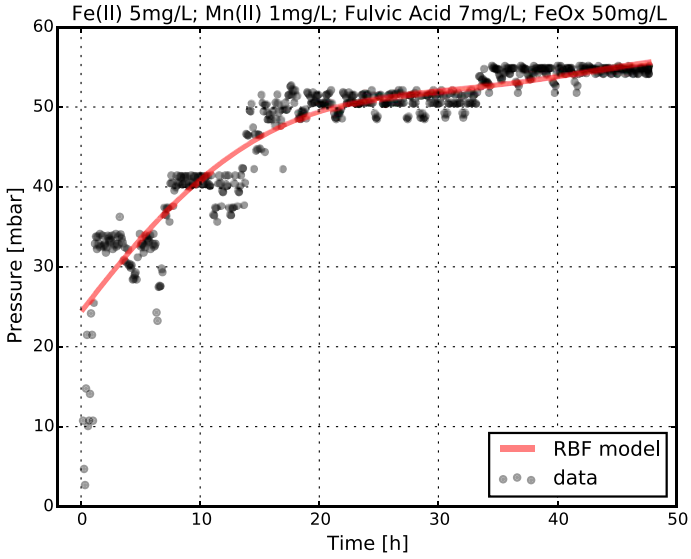

Fig. 13. Effects of 5 mg/L Fe^{2+}, 1 mg/L Mn^{2+} 7 mg/L fulvic acid, and 50 mg/L FeO_x on membrane fouling.

This procedure increased the overall performance of the SVR by ≈10 compared to given metrics in Table 5, since the linear part does not reflect the nonlinear structure of the remaining data.

5. Conclusion

In this study, the fouling effects of Fe^{2+} and Mn^{2+} on membrane in a submerged membrane system for various concentrations of Fe^{2+}, Mn^{2+}, fulvic acid, and iron hydroxide are modeled using SVM regression. Since the complex microphenomena occur during membrane filtration, the modeling of the procedure is difficult and the conventional theoretical models have only been able to predict the filtration procedure under limited conditions and mostly with various assumptions. SVR approach provided relatively robust model less prone to overfitting due to minimized error and regularization terms. A unified SVR approach is presented for analysis and prediction of the membrane performance under various combinations of input parameters. The input parameters were Fe^{2+}, Mn^{2+}, fulvic acid, and iron hydroxide, and the observed output parameter is the membrane pressure change. The contribution of iron hydroxide and fulvic acid on reduction of membrane fouling is observed clearly. These contaminants have significant effect on membrane fouling and removal of Fe^{2+} and Mn^{2+}. The iron hydroxide increases removal efficiency of Fe^{2+} and Mn^{2+} via adsorption/surface oxidation. Also, the size of resulting iron hydroxide flocks exceeds membrane pore sizes which in turn increases membrane effluent efficiency.

In the modeling process, first, the effect of data preprocessing on the model performance is studied in detail. In order to select optimal SVR parameters, grid search with 10-fold cross-validation is performed with different parameters such as different kernel functions, regularization parameter, and scale parameters using the training sets. Based on the results using different SVR parameters, various model selection methods are discussed. The results showed that the nonlinear behavior of the membrane fouling during submerged membrane filtration can be predicted by SVR method with high accuracy. It is therefore not necessary to carry out entire full-scale tests to collect and verify filtration data. The SVR method can be used with a range of sparse data points, which in turn helps optimizing the filtration monitoring procedure through reduction of the time and costs required.

Acknowledgments

This work is supported by the Research Fund of the University of Istanbul (Project number BAP-UDP-15035) and Istanbul Metropolitan Municipality within the scope of My Project Istanbul.

References

[1] M. Khandelwal, T.N. Singh, Prediction of mine water quality by physical parameters, J. Sci. Ind. Res. 64 (2005) 564–570.
[2] D. Hanbay, I. Turkoglu, Y. Demir, Prediction of wastewater treatment plant performance based on wavelet packet decomposition and neural networks, Expert Syst. Appl. 34(2) (2008) 1038–1043.

[3] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20(3) (1995) 273–297.
[4] C.J.C. Burges, B. Schölkopf, A.J. Smola, Advances in Kernel Methods: Support Vector Machines, MIT Press, Cambridge, MA, 1998.
[5] M.J. Gao, J.W. Tian, J. Li, The study of membrane fouling modeling method based on support vector machine for sewage treatment membrane bioreactor, Ind. Electron. Appl. 1–4 (2007) 1393–1398.
[6] M. Bouamar, M. Ladjal, Multisensor system using support vector machines for water quality classification, 9th Int. Symp. Signal Process. Appl. 1–3 (2007) 756–759.
[7] A. Aryafar, R. Gholami, R. Rooki, F.D. Ardejani, Heavy metal pollution assessment using support vector machine in the Shur River, Sarcheshmeh copper mine, Iran, Environ. Earth Sci. 67(4) (2012) 1191–1199.
[8] K. Gao, X. Xi, Z. Wang, Y. Ma, S. Chen, X. Ye, Y. Li, Use of support vector machine model to predict membrane permeate flux, Desalin. Water Treat. (in press), doi: 10.1080/19443994.2015.1086691.
[9] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, J. Mach. Learn. Res. 13 (2012) 281–305.
[10] C. Leys, C. Ley, O. Klein, P. Bernard, L. Licata, Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median, J. Exp. Soc. Psychol. 49(4) (2013) 764–766.
[11] M. Kuhn, K. Johnson, Applied Predictive Modeling, Springer, New York, NY, 2013.
[12] V. Cherkassky, X. Shao, F. Mulier, V. Vapnik, Model complexity control for regression using VC generalization bounds, IEEE Trans. Neural Networks 10(5) (1999) 1075–1089.
[13] V. Vapnik, The Nature of Statistical Learning Theory, Springer, Berlin, 1999.