Desalination 316 (2013) 154–161

Contents lists available at SciVerse ScienceDirect

Desalination

journal homepage: www.elsevier.com/locate/desal

Data-driven models of steady state and transient operations of spiral-wound RO plant

Xavier Pascual [a], Han Gu [b], Alex R. Bartman [b], Aihua Zhu [b], Anditya Rahardianto [b], Jaume Giralt [a], Robert Rallo [c], Panagiotis D. Christofides [b], Yoram Cohen [b,*]

[a] SCITA, Departament d'Enginyeria Química, Universitat Rovira i Virgili, Av. Països Catalans 26, 43007 Tarragona, Catalunya, Spain
[b] Department of Chemical and Biomolecular Engineering, Water Technology Research Center, University of California, 5531 Boelter Hall, Los Angeles, CA 90095-1592, USA
[c] BioCENIT, Departament d'Enginyeria Informàtica i Matemàtiques, Universitat Rovira i Virgili, Av. Països Catalans 26, 43007 Tarragona, Catalunya, Spain

HIGHLIGHTS

- RO plant performance data-driven models are developed via support vector regression.
- Models of steady state and transient RO plant operation were constructed.
- Permeate and retentate flows and conductivities were accurately described.
- Transient short-time permeate and retentate conductivity forecasting was feasible.

ARTICLE INFO

Article history:
Received 20 November 2012
Received in revised form 31 January 2013
Accepted 2 February 2013
Available online xxxx

Keywords:
Desalination
Data-driven models
Support vector regression
Spiral-wound RO plant
Process models

ABSTRACT

The development of data-driven RO plant performance models was demonstrated using the support vector regression model building approach. Models of both steady state and unsteady state plant operation were developed based on a wide range of operational data obtained from a fully automated small spiral-wound RO pilot. Single output variable steady state plant models for flow rates and conductivities of the permeate and retentate streams were of high accuracy, with average absolute relative errors (AARE) of 0.70%–2.46%. Performance of a composite support vector regression (SVR) based model (for both streams) for flow rates and conductivities was of comparable accuracy to the single output variable models (AARE of 0.71%–2.54%). The temporal change in conductivity, as a result of transient system operation (induced by perturbation of either system pressure or flow rate), was described by SVR model, which utilizes a time forecasting approach, with performance level of less than 1% AARE for forecasting periods of 2 s to 3.5 min. The high level of performance obtained with the present modeling approach suggests that short-term performance forecasting models that are based on plant data, could be useful for advanced RO plant control algorithms, fault tolerant control and process optimization.

1. Introduction

Water desalination by reverse osmosis (RO) membrane technology has been increasingly deployed for potable water production from seawater and water reuse applications including municipal wastewater and agricultural drainage (AD) water. Most RO plants are designed to operate at relatively steady state conditions with traditional control strategies to attain the target permeate productivity and quality. Given the complexity of RO plants, plant process models, which consider specific plant characteristics and equipment, are needed to describe both steady state and dynamic plant operations in order to optimize water production and design robust process control strategies [1–4].

The development of first principle deterministic models of RO plant behavior requires fundamental knowledge of the complex physical phenomena that govern plant operational dynamics including, but not limited to, behavior of sensors and actuators, concentration polarization [5], membrane fouling [6] and mineral scaling [7]. For example, membrane fouling can lead to permeate flux decline when operating under constant transmembrane pressure or increased net driving pressure under constant flux operation [7]. Deterministic plant models that are a priori predictive of fouling and mineral scaling would clearly be of practical value; however, given the challenge of accounting for the complex interplay among various fouling [8] and scaling mechanisms [9], such models are lacking for industrial size plants [8]. Admittedly, commercial RO system design software (e.g., Winflows [10], CAROL [11] and ROSA [12]), which are built on the basis of deterministic and semi-empirical models, can be used to simulate steady state operation of RO plants. Mechanistic computational (CFD) models of RO desalination have also been advanced over the last few decades [13] focusing on either simple membrane channel geometries or modeling of single membrane modules. Efforts to incorporate the impact of fouling on the operation of spiral-wound membranes in theoretical and CFD models have also been recently proposed [13–15] and hold promise for adoption in full-scale plant models. The use of both computational

* Corresponding author. Tel.: +1 310 825 8766; fax: +1 310 206 4107.
E-mail address: yoram@ucla.edu (Y. Cohen).

0011-9164/$ – see front matter
http://dx.doi.org/10.1016/j.desal.2013.02.006

CFD models and software design packages for accurate simulation of real-time plant performance is limited since such models typically do not account for complex plant hydraulics, evolution of fouling and mineral scaling throughout the plant, and plant equipment performance over time (e.g., pumps, valves, sensors, etc.).

Data-driven algorithms (i.e., models based on plant data) represent another class of models capable of providing an effective way of describing plant behavior making use of historical plant data without having to rely on predetermined process parameters that are needed by deterministic models [16]. Data-driven models are well suited for accurately describing complex and non-linear systems. Such models can be trained to recognize and learn the characteristics of the plant that affect overall process performance. One advantage of data-driven models is that they can self-adapt (through incremental learning) to changes in operating conditions [17]. Data-driven models (e.g., based on process operational data) can be integrated within control systems [18] by facilitating the development of virtual sensors capable of inferring the properties of manufactured products [19], that provide the basis to improve plant control strategies [20], in addition to data-driven models of membrane based separation processes [21].

Over the past two decades, there has been a growing interest in developing data-driven models, based on machine learning methods, to describe membrane performances (e.g., transmembrane flux and rejection) and fouling in membrane separation processes that include microfiltration (MF), ultrafiltration (UF), nanofiltration (NF) and reverse osmosis (RO). Artificial neural networks (ANN) based models for NF membrane salt rejection [22] were reported with average absolute deviation not greater than 5%. ANN based models of fouling of hollow fiber membranes were reported for a bench-scale system [8] enabling a single composite model for transmembrane pressure covering the various stages of fouling, while piecewise fitting was required when using deterministic fouling models (cake formation, surface blocking and pore blocking models). ANN based models [23] were also developed for resistance of UF membranes with a reported performance of average absolute error of 10%.

Data-driven models of membrane desalination (NF and RO) have been proposed to describe various aspects of steady state process performance with respect to salt rejection [16,22,24,25], permeate flux [24,26], as well as modeling of membrane fouling [8,27,28]. For example, back-propagation ANN models were used [25] to model the rejection of NaCl and $MgCl_2$ salts (based on laboratory scale steady state NF plant data for salt feed concentration of 5000–25,000 mg/L) demonstrating average absolute deviation of 5%. In the above work, models of different ANN architectures and training algorithms were assessed for two different sets of input variables (feed pressure or permeate flux and feed salt concentration) with salt rejection as the output variable. More recently, an interesting approach to modeling steady state RO plant performance was proposed in which the use of the product of salt rejection and permeate flux was introduced as an index of plant performance [24]. Using data for spiral-wound RO desalting of aqueous sodium chloride solutions, an ANN model was developed (input parameters included salt concentration, feed temperature, feed flow rate and feed pressure) for the plant performance index which demonstrated higher performance for salt concentrations of 6000 mg/L and 30,000 mg/L.

The majority of efforts on the development of data-driven RO process models have focused on the use of ANN algorithms given their ability to describe complex non-linear behavior [34]. However, such models require optimal ANN architecture while avoiding over-fitting and convergence to local minima [29]. Support vector machine (SVM) algorithm is an alternative method for developing data-driven models since it is based on the Structural Risk Minimization principle and thus avoids the convergence to local minima, while avoiding over fitting through control of the number of support vectors [30]. The use of SVM is especially useful for developing non-linear controllers as has been demonstrated in recent studies involving membrane based and other industrial processes [31–33]. For example, SVM based non-linear predictive functional control design was applied to a coking furnace, improving the regulatory capacity for both reference input tracking and load disturbance rejection compared with traditional PFC and PID control strategies [31]. A Least squares (LS)-SVM model was shown to be effective for developing a non-linear temperature controller for a proton exchange membrane fuel cell plant [32,34]. SVM, in addition to radial basis function (RBF)-based ANN, was also reported effective in developing a data-driven model [33] of fouling of a membrane bioreactor (quantified via flux decline) making use of eight input parameters (e.g., membrane aperture, aeration gas quantity, initial membrane flux, operating pressure, water temperature, pumping time, sludge concentration and sludge granule). SVM as well as back-propagation ANN algorithms were also applied to developing forecasting models of brackish water RO plant performance [35], with respect to permeate flow rate and salt passage, where variability of up to 25% and 10% was experienced with respect to the normalized permeate flux and salt rejection, respectively. It was shown, that time-series ANN based models enabled forecasting of salt passage and permeate flow rate up to 24 h with similar prediction errors for the SVM and ANN models (average absolute relative errors of 1.2% and 6.6%, respectively). The above models while suitable for long-term plant response (order of hours and above), do not capture short-time scale dynamic responses (order of seconds to minutes) of the system (e.g., due to sudden changes in input pressure or feed flow rate) which would be necessary, for example, for real-time plant control and fault detection.

Data-driven models could be particularly useful for use in plant controllers, identification of deviation of plant behavior from the expected norm, for sensor fault detection and even for smoothing of fluctuations in sensor data. However, such models must be able to accurately describe plant performance not only under steady state conditions, but more importantly under unsteady state operation and over time scales that capture short-time transients. Accordingly, the present work presents an approach for the development and integration of both steady and unsteady state data-driven plant models of RO desalting based on support vector regression (SVR). It is shown that SVR models can accurately describe RO plant performance (e.g., permeate and retentate flow rates and their respective salinities) based on basic operational plant parameters (i.e., feed flow rate, feed pressure and feed conductivity). Moreover, data-driven models for transient plant operation can provide accurate performance forecasting that is suitable for fault-tolerant control of RO plants.

2. Experimental procedure

2.1. Feed solution and materials

Aqueous salt feed solutions were prepared using analytical grade sodium chloride (Fisher Scientific, ACS grade, Pittsburgh, Pennsylvania) in deionized (DI) water. Solutions of two different salt concentrations were utilized (7500 and 5000 mg/L) with the feed solutions maintained at pH ~ 7. Spiral-wound RO membranes that were utilized in pilot RO system (Dow Filmtec XLE-2540, The Dow Chemical Company, Midland, Michigan) were 2.5 inch (outer diameter) 40 inch long elements (0.0635 m and 1.02 m, respectively) with an average surface area of 2.6 m^2. A single membrane element had a manufacturer reported permeate flow rate of 3.2 m^3/day, and a salt rejection of 99%, as determined at a pressure of 6.9 bar for a 500 mg/L NaCl solution. Each membrane element was contained in a separate pressure vessel with six membranes connected in series.

2.2. Description of experimental equipment

Data for model development were generated using the UCLA spiral-wound Mini-Mobile-Modular (M3) pilot RO desalination system shown schematically in Fig. 1 [3,36,37]. The M3 system was designed for

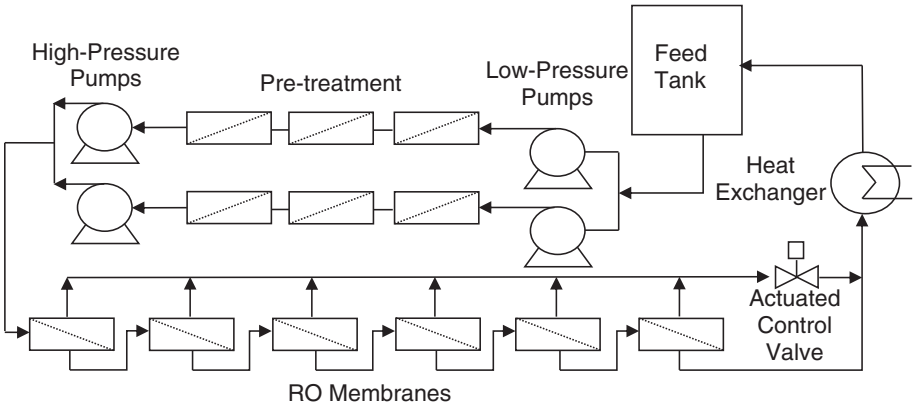

Fig. 1. Configuration of the Spiral-wound RO pilot plant.

permeate water production capacity up to 1.2 m^3/h (7560 gallons/day) for brackish water (5000 mg/L TDS) operating at 75% recovery and up to 0.64 m^3/h (4058 gallons/day) for seawater desalination (at recovery of 40%) using up to 18 spiral-wound elements in various configurations. In the present study, a configuration of six elements in series was utilized with the system operating in a total recycle mode with the permeate and concentrate streams returned to the feed tank. Briefly, the M3 RO plant consisted of a 450 L feed tank with two low-pressure feed pumps (Model JM3460-SRM, Sea Recovery, Carson, CA) pumping the RO feed through a series of cartridge microfilters (5 μm, 0.45 μm and 0.2 μm; 08P GIANT, pleated 177 polypropylene filter cartridges, Keystone Filter, Hatfield, PA). The M3 was operated such that the filtered feed was fed to the RO membranes via two-high pressure pumps (Danfoss Model CM 3559, 3HP, 3450RPM, Baldor Reliance Motor, Sea Recovery Corp. Carson, California) operating in parallel and controlled by variable frequency drives (VFDs) (Model FM50, TECO Fluxmaster, Round Rock, Texas). The retentate flow rate and pressure in the RO unit were set by a model-based controller [36] that adjusts both electrically actuated needle valve (valve V-1) (model VA8V-7-0-10, ETI Systems, Carlsbad, California) on the retentate stream of the M3 RO system and the pump VFD. In order to maintain the temperature of the RO feed, a heat exchanger (Model BP 410-030 Refrigerant heat exchanger, ITT Industries) was installed on the retentate side of the RO system. Permeate and retentate streams of the M3 were monitored in-line via conductivity sensors, conductivity/resistivity sensor electronics (Signet 2839 to 2842 and Signet 2850, George Fischer Signet, Inc. El Monte, California) and pH sensor (DryLoc pH electrodes 2775, George Fischer Signet, Inc. El Monte, California). The conductivity meters were calibrated over the expected concentration range for the study. The M3 plant was equipped with a centralized data acquisition system which receives all sensor outputs (0–5 V, 0–10 V, and 4–20 mA) which are then converted to process variable values. The data were logged into a local computer as well as onto a network database. Data could be logged at a frequency range as high as 1 kHz, although for the present study a frequency of a 1 Hz was employed.

2.3. Experimental procedure

RO desalting experiments covered the range of operating conditions permitted by the operability limit of the M3 system for the specific feed salinities. The M3 control system was programmed to autonomously step through a range of feed flow rates and transmembrane pressures. Feed flow rate and pressure were varied by changing the speed of the high-pressure pumps and the actuated valve settings (Fig. 1). Feed pressure was varied by changing the actuated valve position while maintaining a constant feed flow rate (constant VFD), allowing the plant to operate until the attainment of steady state. The above experiments covered feed pressure and feed flow rate ranges of 6.9–26 bar (or 100–375 psi), and 0.23–0.68 m^3/h (or 1–3 gpm), respectively.

3. Model development

3.1. Data pre-processing

Operational M3 pilot data were acquired for both steady state operations, as well as for transient periods (except plant startup and shutdown), for which there were pressure and flow rate step changes of up to 15% and 25%, respectively. Operational parameters were recorded at a frequency of 1 Hz for each steady and during transitions between states. The recorded steady state data (Table 1) included natural plant fluctuations (i.e., due to operation of pumps and valves) as well as noise from normal operation of sensors, actuators and system pumps. However, in developing the data-driven models data smoothing was not employed given the small variance of the noise, and the objective of developing a simulator for the actual plant operation.

Data-driven models of the state-the-plant (i.e., steady state) for permeate and retentate flow rates and conductivities were first developed from recorded data selected at 0.1 Hz sampling frequency in order to accelerate model training. The steady state period was established as that for which the measured process variables did not vary with time by more than 3% with respect to the time-averaged values. In all cases only the last five minute trace from the steady state period was utilized. Models for transient operation describing the evolution of permeate and retentate conductivities were developed based on data acquired from experiments in which the operating conditions were perturbed

Table 1
Relative standard deviation for the range of steady state plant operating parameters covered in the study[a].

Variable	Minimum	Maximum	Av. STDV (%)
Feed flow rate (m^3/h)	0.26	0.67	0.85
Feed conductivity (μS)	9842	15,828	0.44
Feed pressure (bar)	8.13	24.62	0.43
Retentate flow rate (m^3/h)	0.14	0.53	0.80
Retentate conductivity (μS)	11,539	24,907	0.29
Retentate pressure (bar)	3.59	11.29	0.99
Permeate flow rate (m^3/h)	0.03	0.28	1.66
Permeate conductivity (μS)	628.58	4024	0.44

[a] The Av. STDV $= \left(\sum_{i=1}^{N} STDV_i/V_{i,\mathrm{ave}}\right)/N$ where STDV is the standard deviation for the given variable within a steady state trace i, $V_{i,\mathrm{ave}}$ is the average variable value for the given trace, and N is the number of the experimental steady state traces (or experiments). Note: The salinity conversion factors for the permeate and retentate streams were $S = 4.337 \cdot 10^{-4} \cdot C^{1.0201}$ and $S = 3.833 \cdot 10^{-4} \cdot C^{1.0391}$ respectively, where S is salinity (mg/L total dissolved solids) and C is conductivity (μS).

from steady state. Steady state was generally reached within a period of about 10 min after the perturbation. Higher frequency captured data (0.5 Hz) were utilized for modeling transient operation. Although data were logged by the M3 at an even higher frequency, use of higher frequency data would increase data redundancy and correspondingly also increase the computational time for model training.

Data for model developments were normalized in the range of [0,1] using min–max normalization:

$$y_i^{(n)} = \frac{y_i - \min(y)}{\max(y) - \min(y)} \quad (1)$$

where $y_i^{(n)}$ is the normalized value of the experimental data (y_i) and $\min(y)$ and $\max(y)$ are the minimum and maximum variable values in the data set. Subsequently, the data were divided into two sets, one for model training and the other for model testing. Data selected for training were used to adjust the model while test data (i.e., data that have not been used for model development) were used to evaluate model performance. In order to obtain state-of-the-plant models with good generalization capability, the training dataset was selected to cover the entire plant operational domain (i.e., applicability domain; [40]). Data for model testing were also selected within the applicability domain to avoid extrapolation during predictions. In all cases, complete steady state and transient sequences were selected for the training and test sets. For the transient models, self-organizing-maps (SOM) analysis [38] of the transient data was first performed whereby operational data of similar patterns were clustered in SOM cells. Training and test data for the transient models were then selected from the different cells in order to ensure that adequate data representation is achieved for the whole operational domain.

3.2. Support vector regression (SVR) models

Since the relationships among process variables in the RO process are highly non-linear, both the steady state and unsteady state plant performance models (Sections 3.3 and 3.4) were developed using the support vector regression (SVR) algorithm [39,40]. Briefly, given a vector x of RO process variables (e.g., feed pressure, feed flow rate, and feed conductivity), the goal of SVR is to find a function $f(x)$ that has at most ε deviation with respect to the actual values of a given target RO process variable y (e.g., permeate conductivity) and at the same time is as flat as possible (Fig. 2). SVR can be formulated as a convex optimization problem where a set of coefficients w for the regression model are computed in such a way that the flatness and the accuracy of $f(x)$ are maximized. Since it is not always possible to keep the error within the margin ε for all the available data points, a pair of slack variables ξ and ξ^* must be introduced within the SVR formulation to cope with otherwise infeasible constraints in the optimization (Fig. 2). In most cases, the optimization problem can be solved more readily by projecting x (i.e., the vector of the input RO process variables) onto a higher-dimensional space where linear regression models can be developed for the target RO process variable of interest. The functions used to perform the above linear to non-linear mapping are known as kernel functions. Finally, non-linear models relating input and target variables can be obtained by mapping the data back to its original (i.e., non-linear) space of RO process variables. In the SVR formulation only a subset of the training data points, representing the overall data behavior (i.e. support vectors), are used to generate the regression model (Fig. 2).

In the current work, the kernel function used for model development was the radial basis function (RBF) which is suitable for highly non-linear behaviors [40]. The parameter characterizing this kernel is the width of the Gaussian, σ, which determines the area of influence of the support vectors over the data space and here its optimal value was determined via grid search. The SVR based models were developed in MATLAB using the LS-SVMlab1.7 package [41,42]. This SVR implementation utilized a regularization parameter γ, which controls the tradeoff between the flatness (or smoothness) of the models and their accuracy, whose optimal value was also determined via a grid search. Table 2 summarizes the optimal values of σ and γ obtained via grid optimization for each model.

The performance of the different models (Sections 3.3 and 3.4) was quantified using the linear r^2 correlation coefficient (between the predicted, y_i^*, and experimental, y_i, variables) and by the percent average absolute (AAE) and average absolute relative (AARE) errors defined as:

$$AAE = \frac{1}{n}\sum_{i=1}^{n} |y_i^* - y_i| \quad (2)$$

$$AARE = \frac{1}{n}\sum_{i=1}^{n} \frac{|y_i^* - y_i|}{y_i} \times 100 \quad (3)$$

where n is the total number of data points.

3.3. State-of-the-plant models

State-of-the-plant (STP) models were developed for steady state operation using feed flow rate, feed conductivity and feed pressure as input parameters. The data for the steady state segments (for the range of operating conditions listed in Table 1) were divided into training (60% of data, 2394 samples) and test (40% of data, 1596 samples) sets. Two different modeling approaches were implemented. First, individual models (Eq. (4)) were constructed for each of the four target output variables (permeate flow rate, permeate conductivity, retentate

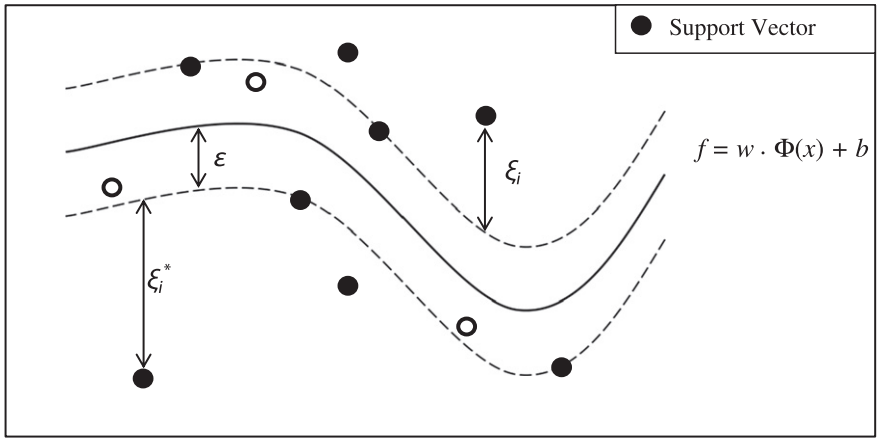

Fig. 2. Support vector regression structural parameters. Regression function f supported on the most representative vectors of information (support vectors), w the normal vector to the hyperplane generated, the kernel function ϕ, the vector of biases b, slack variables ξ and radius of the insensitive tube ε.

Table 2
Performance of SVR based models for predicting steady state RO plant performance based on individual single output parameter models and a composite model. [a]

Model	Predicted variable	γ	σ	AAE	AARE (%)	r^2
Individual	Permeate flow rate (m^3/h)	4200	8.8	0.013	2.46	0.994
Individual	Permeate conductivity (μS)	4200	8.8	21.72	1.24	0.999
Individual	Retentate flow rate (m^3/h)	4200	12	0.011	0.70	0.998
Individual	Retentate conductivity (μS)	4200	8.8	110.16	0.75	0.998
Composite	Permeate flow rate (m^3/h)	1000	7	0.011	2.54	0.993
	Permeate conductivity (μS)			22.74	1.32	0.999
	Retentate flow rate (m^3/h)			0.011	0.71	0.998
	Retentate conductivity (μS)			117.01	0.80	0.997

[a] γ — regularization parameter; σ — width of the Gaussian; AAE and AARE are the average absolute and average absolute relative errors, respectively, and r^2 is the linear correlation coefficient.

flow rate and retentate conductivity), followed by a composite output parameters (OP) model to simultaneously predict the four output variables with the same three input parameters (Eq. (5)),

$$F_i = f\left(C_f, Q_f, P_f\right) \tag{4}$$

$$\left[C_p, Q_p, C_r, Q_r\right] = f\left(C_f, Q_f, P_f\right) \tag{5}$$

where F_i is one of (permeate flow rate, permeate conductivity, retentate flow rate, or retentate conductivity), C, Q and P are conductivity, flow rate and pressure, respectively, and the subscripts f, p and r refer to feed, permeate and retentate streams, respectively. For the individual OP models the optimal value of γ was 4200 for all the target variables. The optimal value of σ was 12 for the retentate flow rate model and 8.8 for the permeate flow rate, permeate conductivity and retentate conductivity models. The optimal γ and σ parameters for the composite OP model were 1000 and 7, respectively.

3.4. Unsteady state models

Data-driven unsteady state plant models for the conductivity of the retentate and permeate streams were developed using a modeling structure that considers a time marching approach (Fig. 3). In this approach, the data-driven transient model is constructed to enable a marching forecasting prediction of stream conductivities at a prescribed period of time, Δt, forward of elapsed time τ from the change in operating conditions at t_0 (i.e., $\tau = t - t_o$). Accordingly, the transient model is expressed as

$$[C(\tau + \Delta t)] = f(C(\tau), C_{\text{initial}}, C_{\text{final}}, \tau, \Delta t) \tag{6a}$$

in which steady state data for the modeled stream conductivities, prior to and post the perturbation of the steady state operating condition, are input variables that are in turn predicted from the steady state plant model (Section 3.3, Eq. (5))

$$[C_{\text{initial}}, C_{\text{final}}] = f\left(C_f, Q_f, P_f\right) \tag{6b}$$

in which C_{initial} and C_{final} are the stream conductivities at time t_o and t_f, at the given initial steady state operation and the new steady state, respectively. The optimal γ and σ parameters for building the SVR composite OP model were 10,000 and 18, respectively. Finally, the transient models were developed based on a dataset of 122 different unsteady state runs (60 were utilized for model testing), covering the range of operating conditions given in Table 1, in which either pressure and flow rates were perturbed by up to 15% and 25%, respectively.

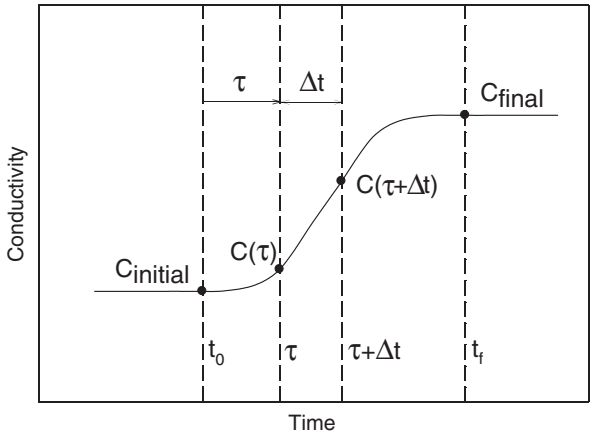

Fig. 3. Illustration of a transient conductivity trace with input and output parameters used in Eq. (6a) for predicting the evolution of conductivity between the initial and final steady states (C_{initial} and C_{final}) at times t_0 and t_f, respectively. The parameters τ and $\tau + \Delta t$ refer to the elapsed time relative to t_o and forecasting period, respectively.

4. Results and discussion

4.1. State-of-the-plant model

State-of-the-plant (STP) model (Section 3.3) was first developed in order to establish the base model for steady state operation of the RO plant (Section 2.3). Performances of the individual and composite OP models are provided in Table 2. Individual OP models predicted permeate and retentate flow rates with average absolute errors (AAEs) of 0.013 m^3/h (2.46% AARE) and 0.011 m^3/h (0.70% AARE), respectively, for permeate and retentate flow rate ranges of 0.023–0.295 m^3/h (0.1–1.3 gpm) and 0.136–0.522 m^3/h (0.6–2.3 gpm), respectively. Similarly, permeate and retentate conductivities were predicted with AAE levels of 21.7 μS (1.24% AARE) and 110.2 μS (0.75% AARE) (Table 2), respectively, for corresponding conductivity ranges of 630–4000 μS and 11,500–25,000 μS. The correlation coefficient (r^2) for the linear correlation (predicted versus measurements) was in all cases ≥0.994.

The composite OP model provided simultaneous prediction of all four output parameters at a similar level of accuracy (Table 2; Fig. 4). Permeate and retentate flow rates (Fig. 4a and b, respectively) were predicted with average absolute errors of 0.011 m^3/h (corresponding to 2.54% and 0.71% AARE for the above two streams). Conductivities of the permeate and retentate stream (Fig. 4c and d respectively) were predicted with AAE values of 22.7 μS (1.32% AARE) and 117.0 μS (0.80% AARE), respectively (Table 2). The r^2 linear correlation coefficient was ≥0.993 for the composite OP model. For both the individual and composite OP models, AARE values for the predictions were greater for the permeate flow rate and conductivity relative to predictions of these parameters for the retentate stream. This trend is not surprising given the higher absolute values of the above parameters for the retentate stream.

The difference in performance of the composite and individual OP models was on average less than 0.05% (Table 2). Overall, the AARE for the steady state individual and composite OP models were for the permeate flow rate (2.46% and 2.54%, respectively). The above performance level suggests that there is little advantage to using the individual OP models relative to the composite OP model, particularly given that the latter is more convenient. Compared with the present SVR models, predictions obtained using the commercial ROSA software [12] (suitable for the present membrane elements, Section 2.1) are of lower accuracy, e.g., with AARE of 23% and 27% for the salinity of the permeate and retentate streams, respectively. The retentate and permeate flow rates were underpredicted and overpredicted, respectively, by ROSA (AARE of 19% and 68%, respectively). The above behavior is not surprising since ROSA predictions were based on the manufacturer membrane permeability and without consideration of the efficiency of

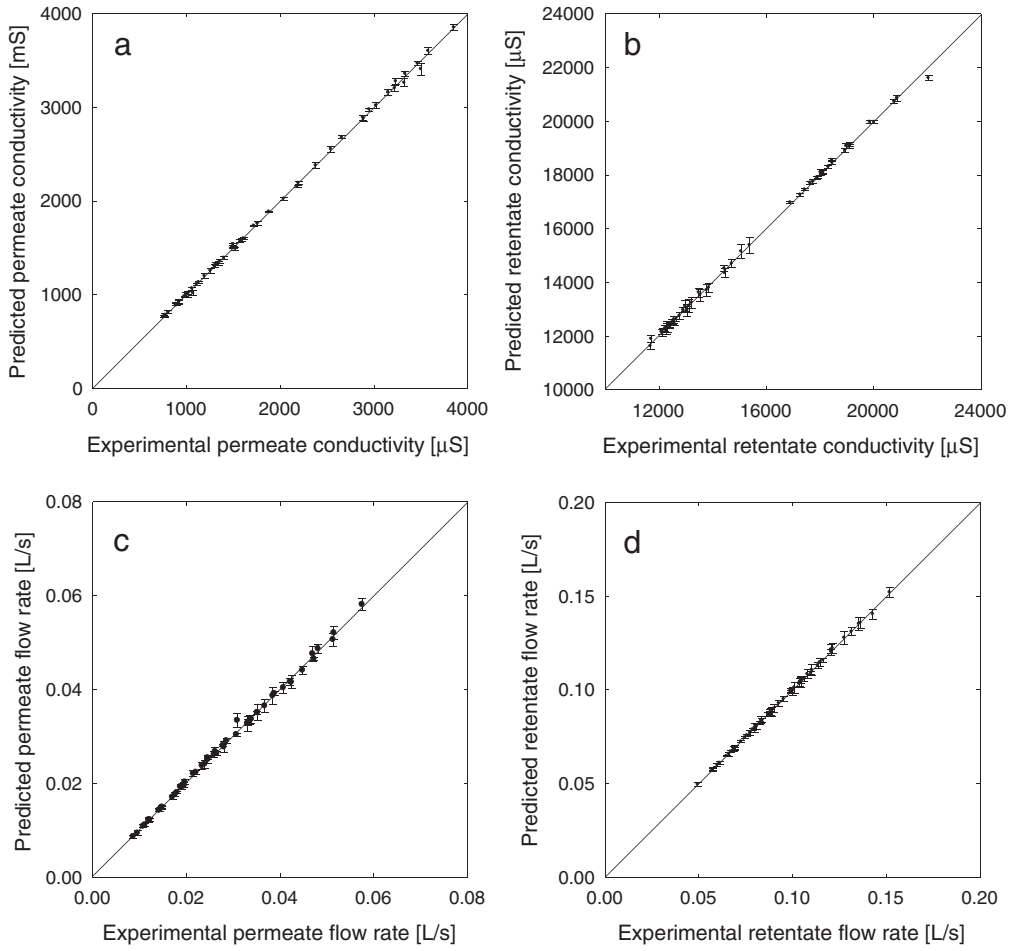

Fig. 4. Comparison of experimental and composite model predictions for: (a) permeate, and (b) retentate conductivities; and (c) permeate and (d) retentate flow rates. Note: In order to maintain clarity of presentation only the predicted averages are plotted (along with a bar depicting the standard deviation) for each steady state data trace.

various plant components. In contrast, the SVR models were based on the actual experimental performance data and thus such data-driven models can serve as more reliable plant-specific operational models.

4.2. Transient operation models

Data-driven models of unsteady state (or transient) plant operation can be useful in refining plant operational control strategies by enabling prediction of plant dynamic response post perturbation from steady state operation. In this regard, it is noted that plant hydraulic response is generally fast. In the present RO plant flow rate transients were typically of the order of a few seconds relative to much longer system response (order of minutes) with respect to transients of retentate and permeate conductivities. Accordingly, from the perspective of plant control, state-of-the plant models may be sufficiently effective for predicting the retentate and permeate flow rates upon pressure or feed flow rate transitions. In contrast, the dynamics of the conductivities of these streams (e.g., due to feed flow rate, feed salinity and pressure perturbations) are influenced by salt dispersion through the system and stabilization of salt flux through the membrane. For example, for the present RO pilot, transient periods of up to 10 min were required to reach steady state, with respect to salinity of the permeate and retentate streams, as a result of feed pressure perturbations.

Several data-driven models were developed to predict the transient response of permeate and retentate conductivities at different forecasting periods (i.e., Δt, Section 3.4, Eq. (6a)) in the range of 2 to 210 s (Table 3). Longer forecasting periods (i.e., >210 s) were not considered since model accuracy decreased significantly due to reduction of available training data with increasing Δt (see Fig. 3). Illustrations of model tracking of the measured permeate conductivities, for a set of different transient trajectories (induced by 5–15% perturbation of the applied pressure), are provided in Fig. 5, for model forecasting period (i.e., Δt) of 30 s. Forecasting at $\Delta t = 30$ s (Fig. 5) was highly accurate, for a range of different dynamics, with average AAE and AARE for the above transients being 6.16 µS and 0.47%, respectively. The average AAE corresponding to the retentate conductivity (Table 3) at the same conditions was slightly higher with a value of 38.12 µS (0.27% AARE). At the longer forecasting period of 120 s, permeate conductivity AAE and AARE were 9.54 µS and 0.66% respectively. Errors in retentate conductivity predictions were somewhat higher with AAE in the range of 50.85 µS (0.36% AARE).

Table 3
Performance of the transient operation models for prediction of permeate and retentate conductivities for different forecasting periods.

Permeate				Retentate		
Δt (s)[a]	AAE (µS)	AARE (%)	r^2	AAE (µS)	AARE (%)	r^2
2	1.55	0.12	1	8.98	0.07	1
30	6.16	0.47	0.999	38.12	0.27	0.999
60	7.91	0.58	0.999	44.86	0.32	0.999
80	8.49	0.60	0.999	47.30	0.34	0.999
90	8.72	0.62	0.999	48.62	0.34	0.999
120	9.54	0.66	0.999	50.85	0.36	0.999
180	10.91	0.73	0.999	51.87	0.37	0.999
210	11.58	0.76	0.999	54.21	0.38	0.999

(a) Forecasting period Δt, ($t = \tau + \Delta t$; Eq. (6a)).

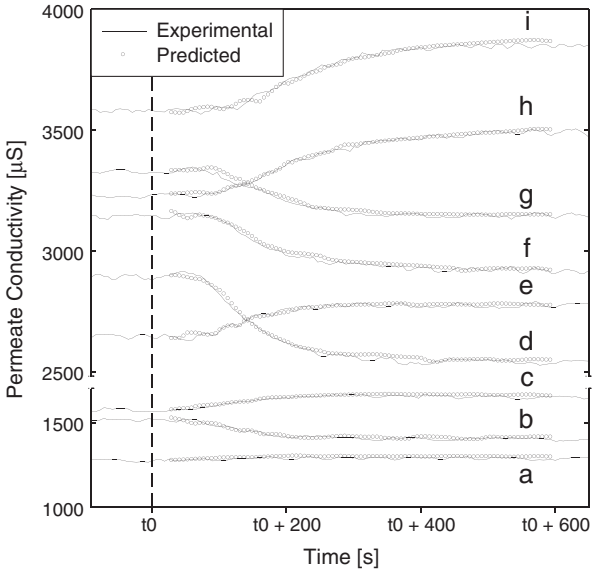

Fig. 5. Comparison of transient model predictions and experimental permeate conductivity measurements for forecasting period of 30 s (Note: t_0 denotes the starting time for the change in operating conditions). Unsteady state operation was induced by the following changes in either applied pressure or feed flow rate: a) 10.2–9.6 bar, 0.27 m^3/h, b) 9.6–9.2 bar, 0.31 m^3/h, c) 10.1–10.5 bar, 0.31 m^3/h, d) 11.8–12.2, 0.40 m^3/h, e) 12.0–11.6 bar, 0.35 m^3/h, f) 13.7–14.4 bar, 0.49 m^3/h, g) 11.0–10.5 bar, 0.36 m^3/h, h) 9.8–10.9 bar, 0.35–0.41 m^3/h, and i) 8.8–8.5 bar, 0.28 m^3/h.

As expected, forecasting errors for conductivities increased (Table 3), for both the permeate and retentate streams, as the amount data available for model training decreased with increasing forecasting period (Fig. 3). The above trend is depicted in Fig. 6 for both the permeate and retentate conductivities. For the forecasting period range of $\Delta t = 2$–210 s, errors in predictions of the permeate and retentate streams were in the range of AAE = 1.55–11.58 µS (AARE 0.12–0.76%) and AAE = 8.98–54.21 µS (AARE 0.07–0.38%), respectively (Table 3). Forecasting at very short times (e.g., $\Delta t = 2$ s) is likely to be of limited value for plant control since differences in the transient data are very small over such short time intervals. However, predictions at longer forecasting periods (i.e., larger Δt) of up to 210 s (Fig. 6, Table 3) were excellent with an AARE of 0.44% and 0.23% for the conductivity of the permeate and retentate streams, respectively. The somewhat lower AARE for the retentate conductivity predictions can be attributed to the greater accuracy of the state-of-the-plant models for predicting $C_{initial}$ and C_{final} (Fig. 3)

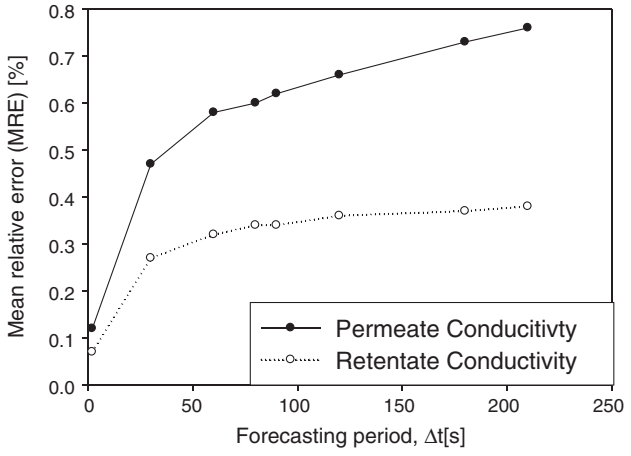

Fig. 6. Performance of the unsteady state model for forecasting of permeate and retentate conductivities as a function of the forecasting period (Δt) averaged for each forecasting period over the transient test traces for the range of operating conditions given in Table 1.

for the retentate (0.8% AARE) relative to the permeate (1.32% AARE) streams. However transient predictions for both streams were clearly of reasonable comparable accuracy.

5. Conclusions

The development of data-driven RO plant performance models was demonstrated using the support vector regression model building approach. Models of both steady state and unsteady state plant operation were developed based on a wide range of operational data obtained from a fully automated small spiral-wound RO pilot. Single output variable steady state plant models for flow rates and conductivities of the permeate and retentate streams were of high accuracy. Performance of a composite SVR based model (for both streams) for flow rates and conductivities was of comparable accuracy to the single output variable models. Predictions of stream conductivities for transient operation were achieved, making use of both predictions from state-of-the-plant model along with a performance time forecasting approach, with less than 1% AARE for forecasting periods in the range of 2–210 s. This higher level of accuracy suggests that short-term performance forecasting, based on plant performance data, can be particularly useful for the development of advanced RO plant control and for process optimization. For example, plant controllers can utilize data-driven models to aid in smoothing operational fluctuations, supplementing missing sensor data (e.g., due to sensor faults or short-term interruption in communication between sensors and data acquisition systems and/or controllers), and for fault tolerant control. Admittedly, data-driven models are plant specific and must be developed and applied for the desired applicability domain. In this regard, online model training can be implemented to improve model performance and its applicability domain as new plant data are acquired.

Nomenclature

AAE	average absolute error
AARE	average absolute relative error
b	bias vector
C_f	feed conductivity
C_{final}	conductivity at the final steady state
$C_{initial}$	conductivity at the initial steady state
C_p	permeate conductivity
C_r	retentate conductivity
$C(\tau)$	conductivity at $t_0+\tau$
$C(\tau+\Delta t)$	predicted stream conductivity
$f(x)$	function obtained by SVR
n	number of experimental values
P_f	feed pressure
Q_f	feed flow rate
Q_p	permeate flow rate
Q_r	retentate flow rate
r^2	correlation coefficient
STDV	standard deviation
t_0	time at the change in the operating conditions
t_f	time at the new steady state
V_i	average variable value for the given steady state
w	normal vector to hyperplane
x	vector of input RO parameters
y_i	experimental value for a given variable
y_i^*	predicted value for a given variable
$y_i^{(n)}$	normalized value for a given variable
Δt	forecasting period
ε	tolerance of the SVR models
γ	regularization parameter
ξ, ξ^*	upper and lower slack variables
σ	width of the Gaussian
τ	elapsed time relative to t_0
Φ	kernel function

Acknowledgements

This work was supported in part by funding from the Catalan Government (2009 SGR 1529), the Spanish Ministry of Education (CTQ2009-14627), the California Department of Water Resources, the United States Environmental Protection Agency, the National Water Research Institute, the Office of Naval Research, and the UCLA Water Technology Research Center.

References

[1] K. Jamal, M.A. Khan, M. Kamil, Mathematical modeling of reverse osmosis systems, Desalination 160 (2004) 29–42.
[2] A. Abbas, Model predictive control of a reverse osmosis desalination unit, Desalination 194 (2006) 268–280.
[3] A.R. Bartman, P.D. Christofides, Y. Cohen, Nonlinear model-based control of an experimental reverse-osmosis water desalination system, Ind. Eng. Chem. Res. 48 (2009) 6126–6136.
[4] A.R. Bartman, C.W. McFall, P.D. Christofides, Y. Cohen, Model-predictive control of feed flow reversal in a reverse osmosis desalination process, J. Process. Control. 19 (2009) 433–442.
[5] S.S. Sablani, M.F.A. Goosen, R. Al-Belushi, M. Wilf, Concentration polarization in ultrafiltration and reverse osmosis: a critical review, Desalination 141 (2001) 269–289.
[6] G. Belfort, R.H. Davis, A.L. Zydney, The behavior of suspensions and macromolecular solutions in crossflow microfiltration, J. Membr. Sci. 96 (1994) 1–58.
[7] S. Shirazi, C.-J. Lin, D. Chen, Inorganic fouling of pressure-driven membrane processes — a critical review, Desalination 250 (2010) 236–248.
[8] Q. Liu, S. Kim, Evaluation of membrane fouling models based on bench-scale experiments: a comparison between constant flowrate blocking laws and artificial neural network (ANNs) model, J. Membr. Sci. 310 (2008) 393–401.
[9] S. Gray, R. Semiat, M. Duke, A. Rahardianto, Y. Cohen, Seawater Use and Desalination Technology, in: W. Peter (Ed.), Treatise on Water Science, Elsevier, Oxford, 2011, pp. 73–109.
[10] Windflows, GE power & water, in, 2010.
[11] CAROL, Toray Membrane Europe AG, in, Münchenstein, Switzerland, 2010.
[12] ROSA, Reverse Osmosis System Analysis, The Dow Chemical Company, 2010.
[13] J.S. Vrouwenvelder, C. Picioreanu, J.C. Kruithof, M.C.M. van Loosdrecht, Biofouling in spiral wound membrane systems: three-dimensional CFD model based evaluation of experimental data, J. Membr. Sci. 346 (2010) 71–85.
[14] K.L. Chen, L. Song, S.L. Ong, W.J. Ng, The development of membrane fouling in full-scale RO processes, J. Membr. Sci. 232 (2004) 63–72.
[15] E. Alhseinat, R. Sheikholeslami, A completely theoretical approach for assessing fouling propensity along a full-scale reverse osmosis process, Desalination 301 (2012) 1–9.
[16] W.R. Bowen, M.G. Jones, J.S. Welfoot, H.N.S. Yousef, Predicting salt rejections at nanofiltration membranes using artificial neural networks, Desalination 129 (2000) 147–162.
[17] J.L. Dirion, M. Cabassud, M.V. Le Lann, G. Casamatta, Development of adaptive neural networks for flexible control of batch processes, Chem. Eng. J. Biochem. Eng. J. 63 (1996) 65–77.
[18] W.T. Miller, R.S. Sutton, P.J. Werbos, Neural Networks for Control, The MIT Press, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1995.
[19] R. Rallo, J. Ferre-Giné, A. Arenas, F. Giralt, Neural virtual sensor for the inferential prediction of product quality from process variables, Comput. Chem. Eng. 26 (2002) 1735–1754.
[20] M. Azlan Hussain, Review of the applications of neural networks in chemical process control — simulation and online implementation, Artif. Intell. Eng. 13 (1999) 55–68.
[21] A. Shahsavand, M.P. Chenar, Neural networks modeling of hollow fiber membrane processes, J. Membr. Sci. 297 (2007) 59–73.
[22] H. Al-Zoubi, N. Hilal, N.A. Darwish, A.W. Mohammad, Rejection and modelling of sulphate and potassium salts by nanofiltration membranes: neural network and Spiegler–Kedem model, Desalination 206 (2007) 42–60.
[23] N. Delgrange-Vincent, C. Cabassud, M. Cabassud, L. Durand-Bourlier, J.M. Laîné, Neural networks for long term prediction of fouling and backwash efficiency in ultrafiltration for drinking water production, Desalination 131 (2000) 353–362.
[24] M. Khayet, C. Cojocaru, M. Essalhi, Artificial neural network modeling and response surface methodology of desalination by reverse osmosis, J. Membr. Sci. 368 (2011) 202–214.
[25] N.A. Darwish, N. Hilal, H. Al-Zoubi, A.W. Mohammad, Neural networks simulation of the filtration of sodium chloride and magnesium chloride solutions using nanofiltration membranes, Chem. Eng. Res. Des. 85 (2007) 417–430.
[26] A. Abbas, N. Al-Bastaki, Modeling of an RO water desalination unit using neural networks, Chem. Eng. J. 114 (2005) 139–143.
[27] N. Delgrange, C. Cabassud, M. Cabassud, L. Durand-Bourlier, J.M. Lainé, Modelling of ultrafiltration fouling by neural network, Desalination 118 (1998) 213–227.
[28] G.B. Sahoo, C. Ray, Predicting flux decline in crossflow membranes using artificial neural networks and genetic algorithms, J. Membr. Sci. 283 (2006) 147–157.
[29] P. Bhagat, An introduction to neural nets, Chem. Eng. Prog. 86 (1990) 55–60.
[30] N. Cristianini, J. Shawe-Taylor, An introduction to support vector machines and other kernel-based learning methods, Cambridge University Press, 2000.
[31] R. Zhang, S. Wang, Support vector machine based predictive functional control design for output temperature of coking furnace, J. Process. Control. 18 (2008) 439–448.
[32] X. Li, G.-Y. Cao, X.-J. Zhu, Modeling and control of PEMFC based on least squares support vector machines, Energy Convers. Manag. 47 (2006) 1032–1050.
[33] G. Meijuan, T. Jingwen, L. Jin, The Study of Membrane Fouling Modeling Method Based on Support Vector Machine for Sewage Treatment Membrane Bioreactor, Industrial electronics and applications, ICIEA 2007. 2nd IEEE Conference on, 2007, 2007, pp. 1393–1398.
[34] R. Yuan, C. Guang-Yi, Z. Xin-Jian, Predictive control of proton exchange membrane fuel cell (PEMFC) based on support vector regression machine, Machine Learning and Cybernetics, Proceedings of 2005 International Conference on, 2005, vol. 4027, 2005, pp. 4028–4031.
[35] 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.
[36] A.R. Bartman, A. Zhu, P.D. Christofides, Y. Cohen, Minimizing energy consumption in reverse osmosis membrane desalination using optimization-based control, J. Process. Control. 20 (2010) 1261–1269.
[37] H. Gu, A.R. Bartman, M. Uchymiak, P.D. Christofides, Y. Cohen, Self-adaptive feed flow reversal operation of reverse osmosis desalination, Desalination 308 (2013) 63–72.
[38] T. Kohonen, The self-organizing map, Proc. IEEE 78 (1990) 1464–1480.
[39] M. Vogt, V. Kecman, Active-set methods for support vector machines, in: L. Wang (Ed.), Support vector machines: theory and applications, Springer, Berlin, 2005, pp. 133–158.
[40] C.M. Bishop, Pattern Recognition and Machine Learning, Springer Science + Business Media, LLC, New York, NY, 2006.
[41] V.N. Vapnik, The Nature of Statistical Learning Theory, Springer-Verlag New York, Inc., 1995
[42] K. De Brabanter, P. Karsmakers, F. Ojeda, C. Alzate, J. De Brabanter, K. Pelckmans, B. De Moor, J. Vandewalle, A.K. Suykens, LS-SVMlab Toolbox user's guide, K.U. Leuven Leuven, Belgium, 2011.