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 15 March 2013

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 © 2013 Elsevier B.V. All rights reserved.
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