Korean J. Chem. Eng., 32(8), 1486-1497 (2015)
DOI: 10.1007/s11814-014-0356-0

pISSN: 0256-1115
eISSN: 1975-7220

Evaluation of multivariate statistical analyses for monitoring and prediction of processes in an seawater reverse osmosis desalination plant

Srinivas Sahan Kolluri, Iman Janghorban Esfahani, Prithvi Sai Nadh Garikiparthy, and ChangKyooYoo†

Department of Environmental Science and Engineering, Center for Environmental Studies, College of Engineering, Kyung Hee University, Seocheon-dong 1, Giheung-gu, Yongin-si, Gyeonggi-do 446-701, Korea

(*Received 18 July 2014 • accepted 1 December 2014*)

Abstract−Our aim was to analyze, monitor, and predict the outcomes of processes in a full-scale seawater reverse osmosis (SWRO) desalination plant using multivariate statistical techniques. Multivariate analysis of variance (MANOVA) was used to investigate the performance and efficiencies of two SWRO processes, namely, pore controllable fiber filter-reverse osmosis (PCF-SWRO) and sand filtration-ultra filtration-reverse osmosis (SF-UF-SWRO). Principal component analysis (PCA) was applied to monitor the two SWRO processes. PCA monitoring revealed that the SF-UF-SWRO process could be analyzed reliably with a low number of outliers and disturbances. Partial least squares (PLS) analysis was then conducted to predict which of the seven input parameters of feed flow rate, PCF/SF-UF filtrate flow rate, temperature of feed water, turbidity feed, pH, reverse osmosis (RO)flow rate, and pressure had a significant effect on the outcome variables of permeate flow rate and concentration. Root mean squared errors (RMSEs) of the PLS models for permeate flow rates were 31.5 and 28.6 for the PCF-SWRO process and SF-UF-SWRO process, respectively, while RMSEs of permeate concentrations were 350.44 and 289.4, respectively. These results indicate that the SF-UF-SWRO process can be modeled more accurately than the PCF-SWRO process, because the RMSE values of permeate flowrate and concentration obtained using a PLS regression model of the SF-UF-SWRO process were lower than those obtained for the PCF-SWRO process.

Keywords: MANOVA, PCA, PLS, Desalination Plant

INTRODUCTION

Water is available in substantial quantities on earth, but only a limited amount is adequately low enough in salinity for drinking and irrigation [1]. Over the past few decades, desalination techniques have emerged as an effective tool to produce potable water from seawater and brackish [2]. In this regard, reverse osmosis (RO) and thermal desalination, which includes multi-effect distillation (MED) and multi-stage flash (MSF) [3] are two most widely used desalination techniques. Further, a total of 9,000 reverse osmosis (RO) desalination plants have been installed over the past two decades [4]. In desalination plants, process controls systems are used for plant automation and these systems generate a large amount of information. However, storing a large amount of information is very difficult when enormous numbers of process variables need to be accounted for. Because SWRO processes are very sensitive to variations in operating conditions, reliable monitoring and prediction of SWRO plant processes is needed to maintain system performance as close as possible to optimized conditions. Multivariate statistical techniques can be used to analyze data obtained from various SWRO processes and address the issues described above. Furthermore, the use of multivariate statistical techniques can allow the state of an SWRO plant to be monitored from a remote centralized operation center [5].

Several multivariate statistical methods to analyze processes have been developed, and used successfully for monitoring and fault detection [6]. Multivariate statistical approaches to monitor processes, detect faults, and diagnose problems have advanced over the past 15 years [7]. Principal component analysis (PCA) is one of the most popular multivariate statistical-based monitoring methods [8]. Multivariate statistical projection methods like PCA and PLS can be used in SWRO desalination plants to make data more comprehensible and to extract relevant information [9].

Several studies recently investigated monitoring of reverse osmosis desalination systems. Bourouni [10] analyzed and compared the performance of graphical methods such as reliability block diagram and fault tree analysis methods for availability assessment in RO plants. Quintanilla et al. [11] developed a statistical approach for membrane rejection and organic chemicals, both of which reduce physical-chemical compound properties. Alvarez et al. [5] investigated adaptation of unfolded PCA (UPCA)-based monitoring and fault detection for desalination plants and reported that this resulted in a decrease in the number of false alarms and improved fault detection. Mcfall et al. [12] used a dynamic model for a high recovery RO desalination plant to detect and isolate faults in the case of system failure.

Recent studies have focused on fault detection and control of SWRO plant performance. Few studies have performed multivariate statistical monitoring and prediction of processes in full-scale SWRO desalination plants. We performed this study to analyze,

†To whom correspondence should be addressed.
E-mail: ckyoo@khu.ac.kr, ChangKyoo.Yoo@biomath.ugent.be

monitor, and predict processes in a full-scale seawater reverse osmosis desalination plant using multivariate statistics techniques.

This paper consists of three sections. In the first section, multivariate analysis of variance (MANOVA) was used to evaluate the performance and efficiencies of the pretreatment processes of an SWRO desalination plant based on evaluation of the silt density index (SDI) and turbidity parameters. In the second part, PCA was applied to reduce the dimensions of the collected data and to determine the number of principal components. Process input data of the two SWRO processes were easy to analyze using PCA monitoring: outliers were detected and unwanted measurements were identified based on Hotelling's T^2 statistic and SPE charts. In the third section, PLS regression was used to develop a prediction performance model for the output responses of two SWRO desalination processes. RMSE values of permeate flow rate and permeate concentration obtained using PLS regression model were compared to determine which of these two SWRO pretreatment processes was superior to the other.

MATERIALS AND METHODS

1. Plant and Data Description

An overall block diagram of the SWRO desalination plant located in the shipyard on Geojae Island, Korea, is shown in Fig. 1 [13]. Two SWRO processes with different pretreatment units are used in this plant. The first process is pore-controllable fiber filter-reverse osmosis, and the second process is sand filtration-ultra filtration-reverse osmosis, which involves ultra-filtration (UF) with a sand filter. For the first process, two PCF filtration processes with four PCF filters are used a teach PCF filtration stage, with a maximum capacity of 5,000 m^3/day.

The process consists of ultra-filtration with a maximum capacity of 75 m^3/day using sand filters (SF). In both processes, the SWRO unit comprises two membrane modules and these modules are placed within a single vessel. We collected data from PCF-SWRO and SF-UF-SWRO processes performed between March 2011 and May 2011.

2. Multivariate Analysis of Variance (MANOVA)

MANOVA was used to investigate whether the population mean-vectors were the same or not; the latter case would mean the components differed significantly [14-16].

F-test is a statistical hypothesis test that has an *F*-distribution as the null hypothesis. It is most frequently used to compare the fit of statistical models for a data set. MANOVA is the most common and useful application of the *F*-test. Suppose the number of elements in a vector are greater than or equal to one (q≥1) and there are two populations (h=2). The *F*-test statistic value of MANOVA can then be calculated using Eq. (4) [14-16]:

$$F=\left(\frac{\sum_{j=1}^{h}n_j-q-2}{q}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\Lambda^*}\right) \tag{1}$$

where n_j is the number of vectors in the j^{th} population and Λ^* is Wilky's lambda that is calculated by expression

$$\Lambda^*=\frac{|\mathbf{W}|}{|\mathbf{B}+\mathbf{W}|} \tag{2}$$

where $|\mathbf{W}|$ and $|\mathbf{B}+\mathbf{W}|$ are the determinant values of $\mathbf{W}$ and $\mathbf{B}+\mathbf{W}$. The expressions for calculating matrices $|\mathbf{W}|$ and $|\mathbf{B}+\mathbf{W}|$ are provided in Table 1.

MANOVA was conducted using the equations presented in Table 1. The *F*-test rejects the null hypothesis H_0: $\tau_1=\tau_2=\ldots\tau_h=0$ (where τ_1 is the mean-vector of the j^{th} population) at level α if

$$F=\left(\frac{\sum_{j=1}^{h}n_j-q-2}{q}\right)\left(\frac{1-\sqrt{\Lambda^*}}{\Lambda^*}\right)>F_{2q,\,2\left(\sum_{k=1}^{h}n_k-q-2\right)}(\alpha) \tag{3}$$

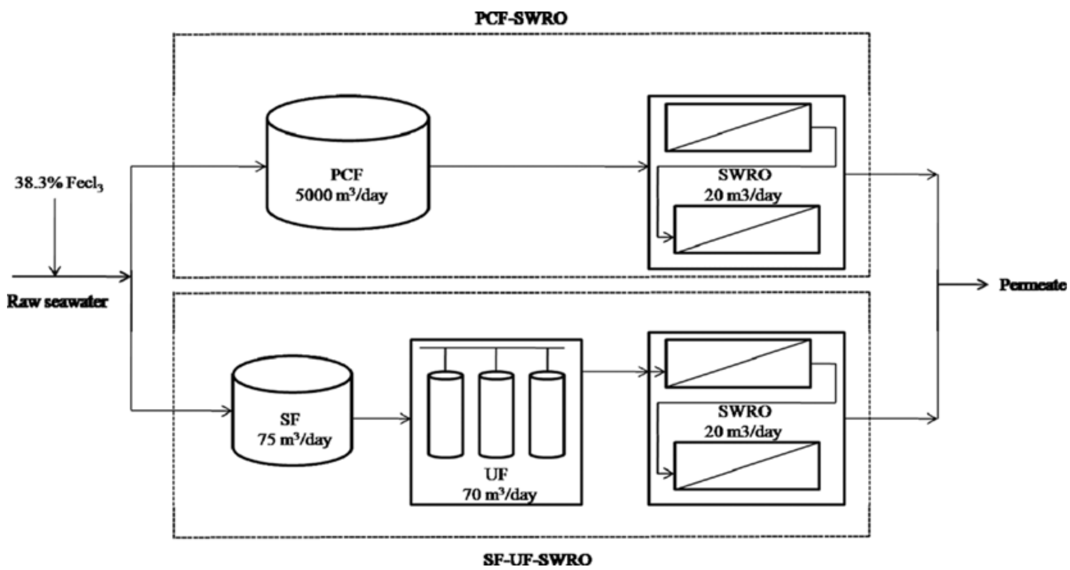

Fig. 1. Block diagram of a full-scale SWRO desalination plant on Geojae Island, Korea.

Table 1. MANOVA table comparing population mean vectors [15-17]

Source of variation	Matrix of sum of square and cross products (SSP)	Degrees of freedom (d.f.)
Treatment	$B=\sum_{l=1}^{g} n_1(\bar{x}_1-\bar{x})(\bar{x}_1-\bar{x})^T$	$h-1$
Residual (error)	$W=\sum_{j=1}^{h}\sum_{i=1}^{m}(x_{ji}-\bar{x}_j)(x_{ji}-\bar{x}_j)^T$	$\sum_{j=1}^{h} m_j-h$
Total (corrected for the mean)	$B+W=\sum_{j=1}^{h}\sum_{i=1}^{m}(x_{ji}-\bar{x}_j)(x_{ji}-\bar{x}_j)^T$	$\sum_{j=1}^{h} m_j-1$

where $F_{2q,\, 2\left(\sum_{k=1}^{h} n_k-q-2\right)}(\alpha)$ is the upper $(100\alpha)^{th}$ percentile of the *F*-distribution with 2q and $2\left(\sum_{j=1}^{h} n_j-q-2\right)$ is the degrees of freedom [14-16].

In this study, MANOVA was applied to two SWRO processes, pore controllable fiber filter-reverse osmosis and sand filtration-ultra filtration-reverse osmosis, to statistically evaluate the performance and efficiencies of the pretreatment processes based on evaluation of SDI and turbidity parameters. SDI and turbidity are proximately related to each; in general, if feed water turbidity increases, then SDI increases, though the converse is not always true.

3. Multivariate Statistical Analysis

Multivariate statistical methods are used to find the dependency among different variables from a dataset. Commonly used multivariate statistical analysis methods are principal component analysis (PCA) and partial least squares (PLS).

3-1. Principal Component Analysis (PCA)

PCA projects a vector of high-dimensional measurement space onto a space with significantly fewer dimensions. To determine the relevant number of principal components, a scree plot is used, in which eigenvalues are plotted against principal components. The matrix X can be written as the sum of the outer product of the vector t_i, which is a score vector that contains information on the relationship between different samples and m_i, which is a loading vector that contains information on the relation between different variables plus the residual matrix G as shown in Eq. (7) [17-19]:

$$X=TM^T+G=\sum_{i=1}^{n} t_i m_i^T+G \tag{4}$$

where 'n' is the number of independent variables.

In this study, PCA was performed to determine the number of PCs for all input parameters: feed flow rate of the plant, PCF/SF-UF filtrate flow rate, temperature of the feed water, turbidity feed, pH, reverse osmosis (RO) flow rate, and pressure.

3-2. Multivariate Statistical Process Monitoring

Mutivariate monitoring methods are used to detect disturbances in measured data. Major multivariate control charts used for monitoring are Hotelling's T^2 statistic and the square prediction error (Q) statistic, which is also called SPE.

The confidence limits for T^2 are given by Eq. (7) using the *F*-distribution [20-22]:

$$T^2\le T_\alpha^2=\frac{(n^2-1)a}{n(n-a)}F_\alpha(a,\, n-a) \tag{5}$$

where $F_a(a,\, n-a)$ is the critical value of the *F*-distribution with n degrees of freedom and α is the level of significance. Typical values of alpha for warning and action limits are 0.05 and 0.01, respectively.

The Q statistic is defined as the sum of squares that takes A principal components, and the approximate distribution of Q is given by Eq. (9) [23-26] as

$$SPE_\alpha=\Theta_1\left[\frac{c_\alpha\sqrt{2\Theta_2 h_0^2}}{\Theta_1}+1+\frac{\Theta_2 h_0(h_0-1)}{\Theta_1^2}\right]^{1/h_0} \tag{6}$$

where c_α is the standard normal deviation corresponding to the upper $(1-\alpha)$ percentile, and λ_j is associated with the j^{th} loading vector.

In this study, Hotelling's T^2 statistic, SPE charts, and contribution plots were applied to the two SWRO processes to identify and define outliers.

3-3. Partial Least Squares (PLS)

Since data in process industries are highly correlated, we used a PLS model to correlate and determine the relationship between the set of independent variables or process inputs (X) and the set of dependent or output variables (or responses) (Y). PLS model builds up a linear model that relates the matrices X and Y. These matrices can be decomposed into bilinear terms using the following equations [27-30]:

$$X=TM^T+G=\sum_{i=1}^{p} t_i m_i^T+G \tag{7}$$

$$Y=UN^T+H=\sum_{i=1}^{p} u_i n_i^T+H \tag{8}$$

where m and n are loading vectors that contain information on the relationship between different process and response variables, respectively; p is a vector of the number of latent variables, T and U are score matrices, and G and H are residuals.

In this study, PLS regression was used to develop a prediction performance model for the output responses of the two SWRO desalination processes: permeate flow rate and permeate concentration.

3-4. Analysis, Monitoring, and Prediction of SWRO Desalination

Data used in this study were collected over a period of three months (March 2011 to May 2011). SWRO desalination plant data were divided into two datasets: one for the PCF-SWRO process and the other for the SF-UF-SWRO process; each dataset comprised 92 samples.

Multivariate analysis of variance (MANOVA) was used to analyze the two SWRO processes. Pretreatment units of the SWRO desalination plant were divided into two populations: the PCF-SWRO process and other SF-UF-SWRO process. We used MANOVA to evaluate the performance and efficiencies of the two pretreatment processes by using the input variables of silt density index (SDI) and turbidity.

It is important to operate an SWRO plant under optimal conditions to enhance process performance. Sophisticated monitoring and prediction of the performance of desalination plant processes should decrease operational costs and increase process performance. We therefore used multivariate statistical methods to monitor and predict the performance of SWRO desalination plant processes. PCA was first used to determine the number of PCs using a scree plot. In addition, PCA was used for monitoring based on the score plot, T^2 statistic and SPE charts to analyze and define outliers. We defined the responses (or output) variables (**Y**) as permeate flowrate and permeate concentration, and the input variables (**X**) as feed flow rate of the plant, PCF/SF-UF flow rate, filtrate flow rate, temperature of feed water, pH, RO flow rate, turbidity feed, and pressure. Partial least squares (PLS) regression was conducted to predict which of the seven input variables had a significant effect on the output variables. PLS regression was subsequently used to develop a prediction performance model for the output responses of the two SWRO desalination processes. RMSE values of the permeate flow rate and permeate concentration were obtained using the PLS regression model and compared.

RESULTS AND DISCUSSION

1. Evaluation of SWRO Plant Performance by Multivariate Analysis

To supply high quality feed water to a reverse osmosis system, an optimized and sufficient pretreatment system is required. This pretreatment can improve long-term operation, maintain high efficiency, and improve fouling sensitivity, leading to an increment in RO membrane lifespan. The most commonly used parameters to evaluate the performance and efficiencies of pretreatment processes are SDI and turbidity. SDI is typically used to characterize the fouling potential of the influent. Often membrane suppliers suggests an SDI value <3; moreover, in practice, it is commonly accepted that the SDI of the RO feed water should be lower than 3 [31] to minimize fouling potential and pressure loss in the RO modules. Turbidity can be caused by solid deposits and chemical precipitates. It is important to measure the turbidity of feed water in an SWRO desalination plant because high turbidity blocks pretreatment filters and decreases pretreatment efficiency.

As shown in the Fig. 2(a) and (b), the average SDI values for the PCF-SWRO and SF-UF-SWRO processes were 3.17 and 2.29, respectively. On the other hand, as illustrated in Fig. 2(a) and Fig. 3(a), the average turbidity value of feed water for PCF-SWRO and SF-UF-SWRO processes was 2.6 and 2.5 NTU, respectively. Therefore, the average turbidity of the SF-UF-SWRO process was lower than that of the PCF-SWRO process; this indicated that the SF-

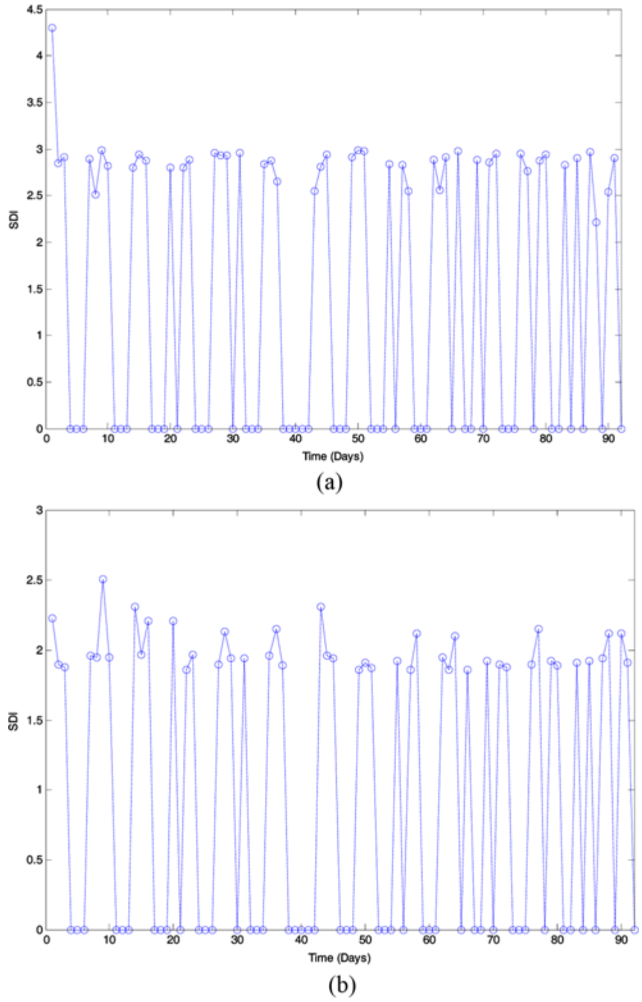

Fig. 2. Time evolutionary plot for SDI (a) PCF-SWRO process, and (b) SF-UF-SWRO process.

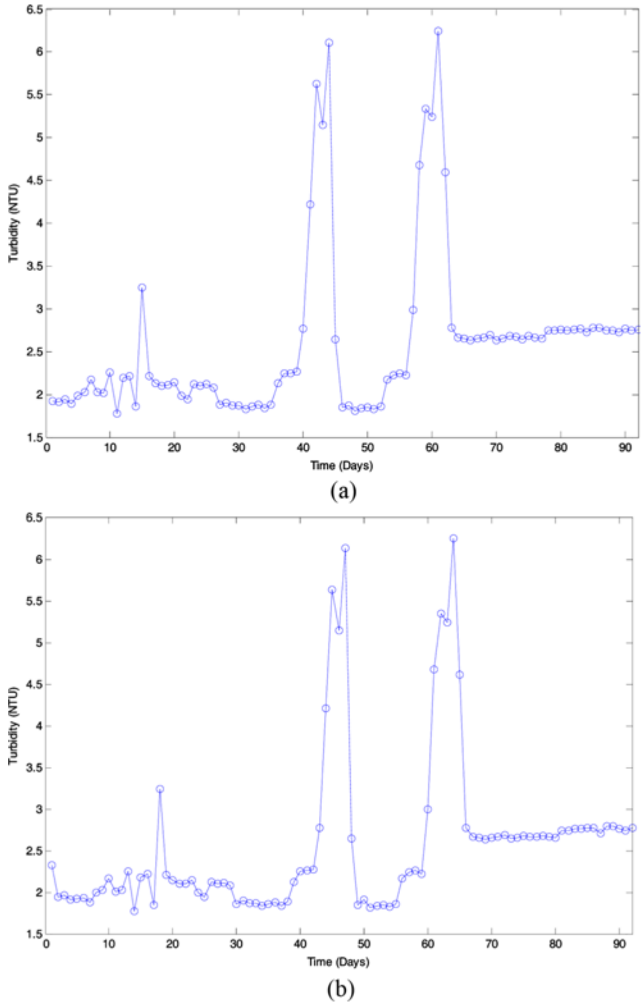

Fig. 3. Time evolutionary plot for Turbidity (a) PCF-SWRO process, and (b) SF-UFSWRO process.

UF-SWRO removed turbidity contributors more effectively than the PCF-SWRO process. PCF-SWRO process exhibited a maximum turbidity of 6.24 NTU during the end of April, because of an increase in influent flow rate. Periodic variations in turbidity values were observed for the PCF-SWRO process, perhaps due to fouling or solid deposits on the pores of filter.

We concluded that the overall pretreatment efficiency and performance of the PCF-SWRO process were lower than those of the SF-UF-SWRO process, potentially due to membrane fouling and low effluent water quality. To maintain and operate an efficient SWRO desalination plant, an optimized and monitored pretreatment process is necessary to maintain effluent standards.

To statistically verify the performance and efficiency of the pretreatment processes, we performed a MANOVA by comparing the population mean vectors. To statistically verify the performance and efficiency of the pretreatment processes, we performed a MANOVA by comparing the population mean vectors. MANOVA test will give the quantitative measure of the differences among the data sets of two pretreatment processes. For MANOVA test, the original data obtained from SWRO desalination plant is divided into two populations: (1) PCF-SWRO process, and (2) SF-UF-SWRO process, which has the effect of SDI and turbidity parameters. Null hypothesis for MANOVA is that there is no significant difference between the two pretreatment processes mean-vector (H_0: $\tau_{PCF\text{-}SWRO}=\tau_{SF\text{-}UF\text{-}SWRO}=0$), if Eq. (4) does not satisfy. As the result of MANOVA test, the value of F is 3.2649×10^4, while F_{limted} is 3.2849 (where F_{limted} is the upper 0.95th percentile of the F-distribution). The null hypothesis was rejected because the F value was much higher than $F_{limited}$ at the 95% confidence level, revealing that there were significant variations in the measurements of SDI and turbidity parameters between the two pretreatment processes. An inefficient pretreatment process requires periodic membrane cleaning and an increase in operating pressure, resulting in accumulation of solid deposits and membrane fouling, which decreases membrane life. Therefore, the PCF pretreatment unit should be scrutinized during the design and operation of a desalination plant.

2. PCA Monitoring of the PCF-SWRO Process

Since MANOVA investigates is not able to detect and analyze the outliers for two pretreatment processes including PCF-SWRO and SF-UF-SWRO, therefore, multivariate statistical monitoring and prediction methods like PCA, PLS have the capability for analyzing and interpretation of data.

PCA was applied to the seven input variables of feed flow rate of the plant, PCF/SF-UF filtrate flow rate, temperature of the feed water, turbidity of the feed, pH, reverse osmosis flow rate, and pressure. Significant parameters were obtained through PCA, which was applied to all 92 observations (from March 2011 to May 2011). In addition, a scree plot was used to determine the optimal number of principle components.

Cumulative percentage variance (CPV) is the percentage captured by first c PCs. A measure of CPV(c)≥90% captured by first c principal components and is given as [5,32]

$$\mathrm{CPV}(c)=\left[\frac{\sum_{j=1}^{c}\lambda_j}{\sum_{j=1}^{m}\lambda_j}\right]\times 100 \tag{9}$$

where c is the first principal component and m is the number of input variables.

The number of principal components, calculated using the CPV approach with 92% variance level, is five. Moreover, Fig. 4 shows

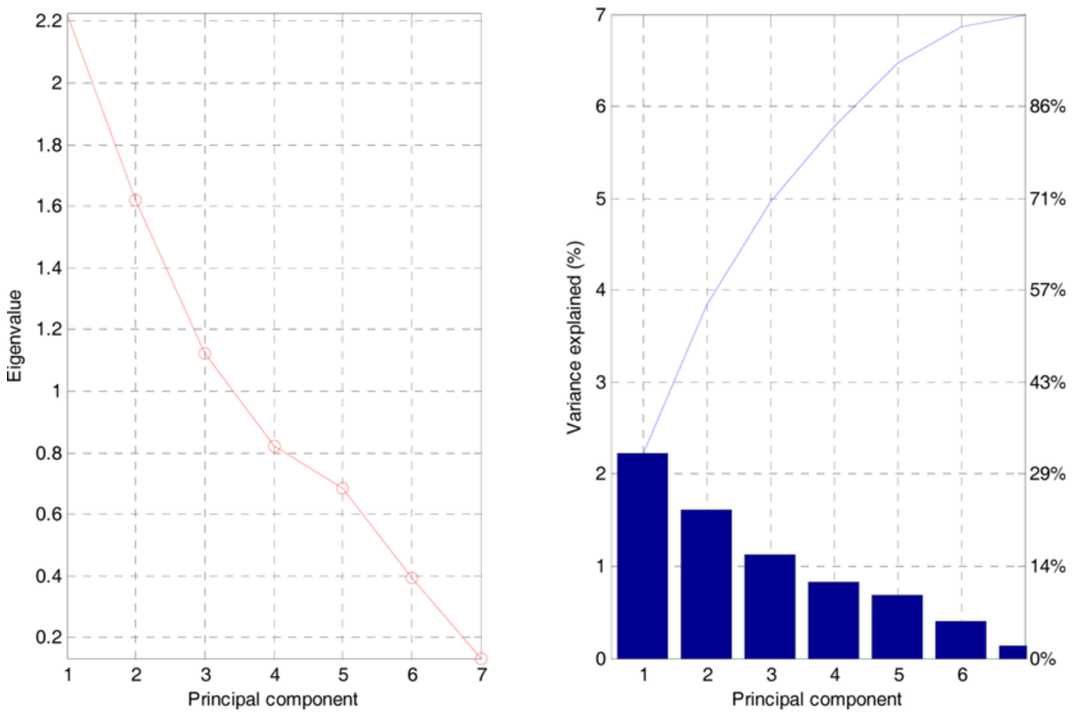

Fig. 4. Scree plot for principal component selection for the PCF-SWRO process.

the scree plot results for PCF-SWRO process. As shown in Fig. 4, five variables capture 92% variance and information from the dataset. Therefore, seven input variables could be replaced with five key variables which are linear combinations of the original variables.

Monitoring charts for the PCF-SWRO process obtained using PCA are shown in Fig. 5; score plot, T2 plot, and SPE chart results are shown. Results of score plots on the PC_1-PC_2 plane for 92 samples are shown in Fig. 5(a). The circle repersents the 95% confidence level. Samples within the circle were statistically controlled for, whereas those outside the circle were considered outliers or contaminated

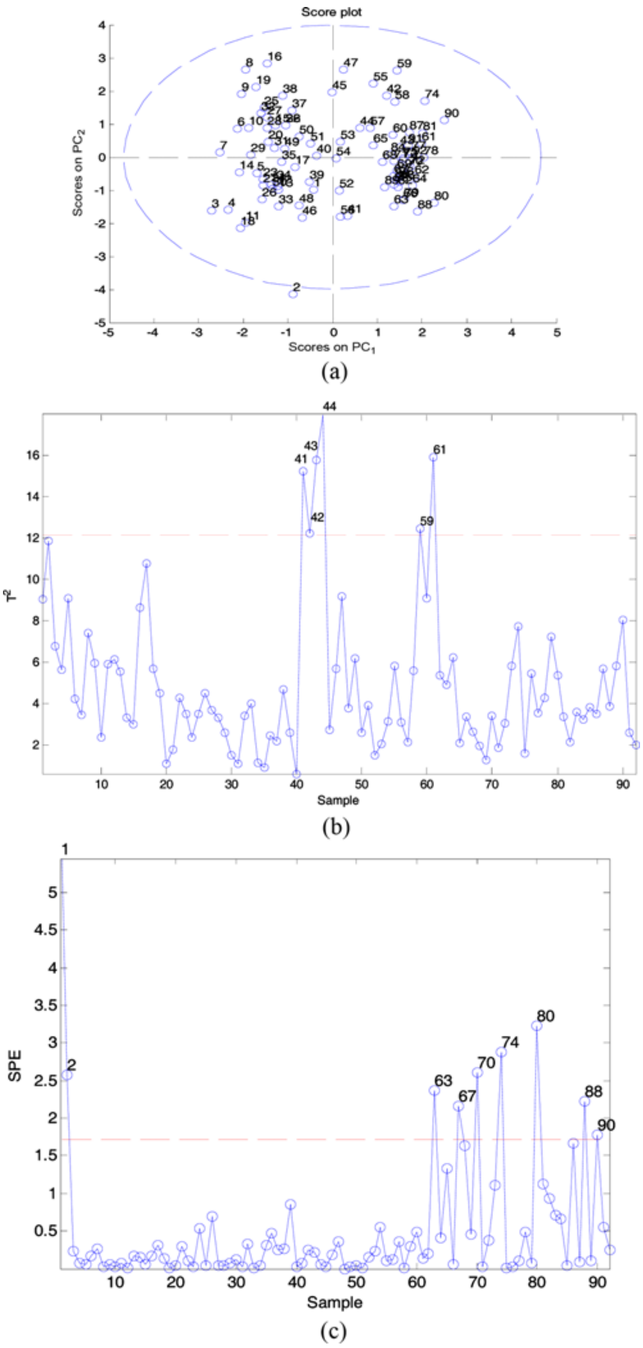

Fig. 5. PCA monitoring of the PCF-SWRO process: (a) score plot, (b) T^2 plot, and (c) SPE chart.

samples. As shown in Fig. 5(a), only sample number 2 (March 2, 2011) fell outside the circle. For this sample, the flow rates of the feed and filtrate were lower than average.

Fig. 5(b) and (c) shows the results of Hotelling's T^2 statistic and SPE charts based on the PCA model. T^2 and SPE statistics include a summary of all input variables, and can be used to detect strong deviations in the data. As shown in Fig. 5(b), samples 41 to 44, 59, and 61 in the T^2 plot are believed to be outliers. To determine the most likely cause of the outliers, the contribution of the input for these particular samples (41-44, 59, and 61) was plotted. Fig. 6 shows the results of the contributions of samples 41 to 44, 59, and 61. Note that turbidity feed flowrate had the highest value among all input parameters. At sample 41, the process variables, RO flow and turbidity, show higher contributions. The turbidities of PCF-SWRO process were low in early stage of opertation, but after the fluctuations in inffulent water, they increased to higher level. In can be concluded that in higher turbidity periods, the turbidity of PCF-SWRO process increases. It is significant that this poor pretreatment process could result in low quality water from the RO system. On the other hand, process variables feed flow rate, filtrate flow rate, temperature and pressure show low contibution plot. Hence, these variables seem to maintain the average value without any fluctuations. The contribution plot for samples in between 42 and 44 were considered and illustrated in Fig. 6. In this period, the SWRO desalination plant received influent with a high turbidity and pH and a small flow rates, i.e., a highly solid deposited water. Later, contribution plot for samples 59 and 61 were considered. In this period, turbidity feed and RO feed flow rate showed largest contibutions. For sample 59, it is apparent that the of process variables, feed flow rate and filtrate, also show somewhat higher contribution than other samples. On the other hand, 61 as shown in Fig. 6, it can be clearly seen that the input variables, temperature and pressure, were lower than usual. It can be clearly seen that PCA systematically identifies the abnormal data. Therefore, by detecting and interpretation of process outliers plays an important role in ensuring process monitoring and process efficiency. As shown in Fig. 5(c) in the SPE chart, many data samples were located outside the confidence level far out of the expected range of the process. Raw data revealed an apparent disturbance. To construct a reliable model, outliers were excluded.

3. PCA Monitoring of the SF-UF-SWRO Process

The SF-UF-SWRO process managed to capture 93% variance level are five. In addition, Fig. 7 shows the scree plot for SF-UF-SWRO process. From Fig. we observe that the five PCs capture 93% variance.

Score plot, T^2 plot, and SPE chart for the SF-UF-SWRO process obtained using PCA are shown in Fig. 8. Fig. 8(a) shows the results of score plots on the PC_1-PC_2 plane for 92 samples; all samples fell within the 95% confidence levels.

Fig. 8(b) shows the Hotellings T^2 results; samples 43, 44, 59, and 61 of the T^2 plot exceeded the confidence level. The likely cause for this deviation is shown in the contribution plot in Fig. 9. For these samples, the turbidity feed flowrate was very high because the SF-UF-SWRO process received a turbidity feed almost twice the average flowrate. The SPE chart presented in Fig. 8(c) shows that samples 1, 42, 43, 83, and 86 exceeded the confidence limits. This is

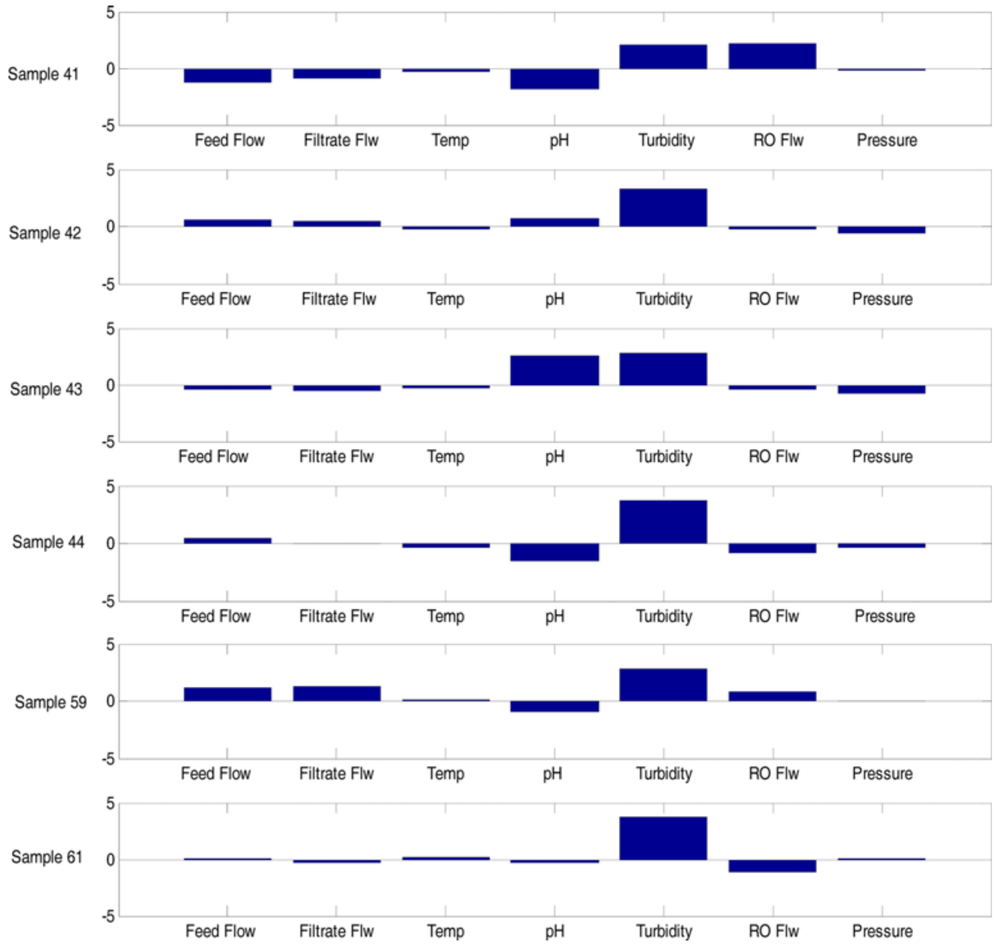

Fig. 6. Contribution plot of T^2 for sample numbers 41, 42, 43, 44, 59, and 61 of the PCF-SWRO process.

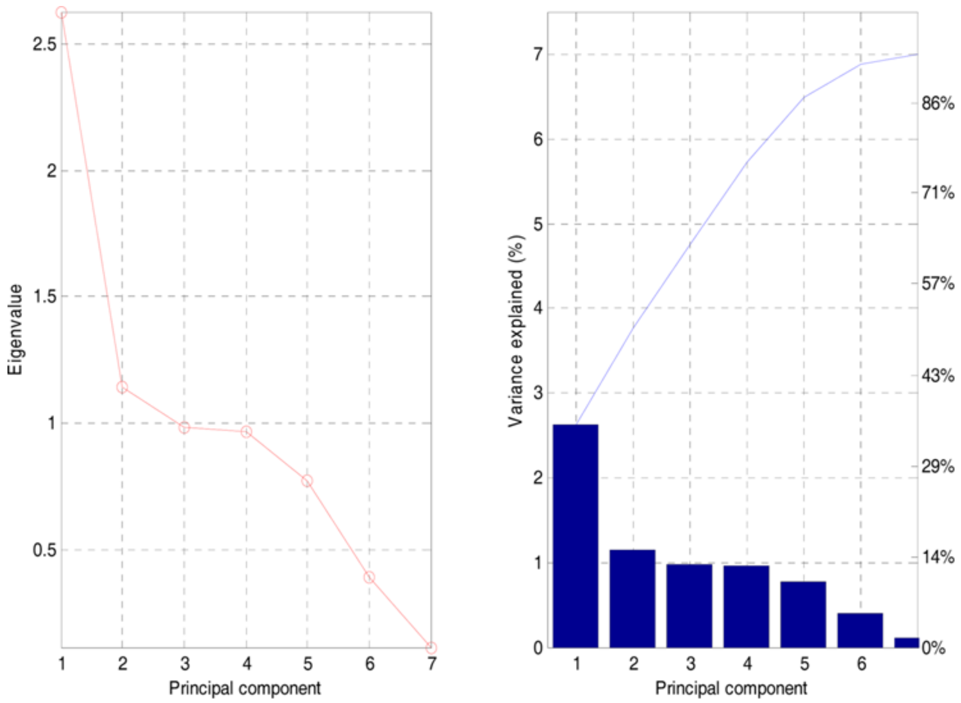

Fig. 7. Scree plot for principal component selection for the SF-UF-SWRO process.

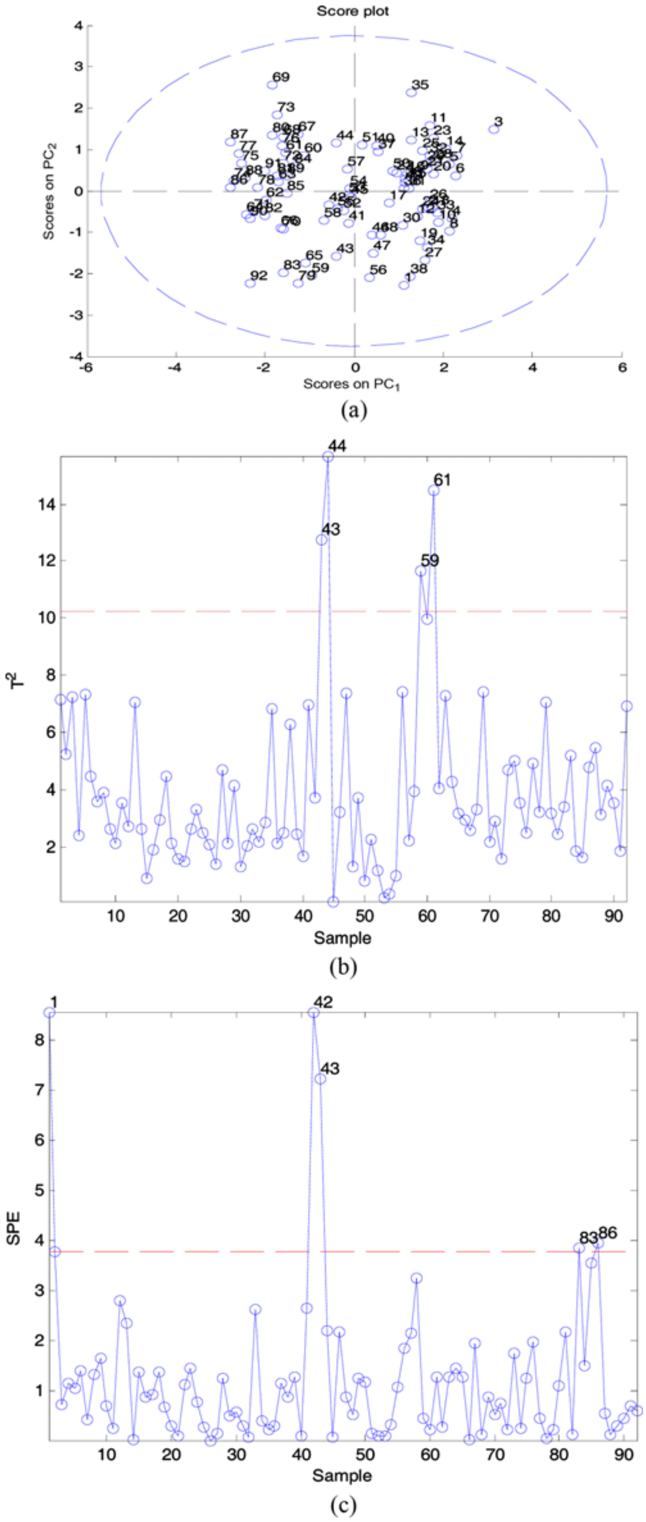

Fig. 8. PCA monitoring charts for the SF-UF-SWRO process: (a) Score plot, (b) T^2, and(c) SPE chart.

because the SF-UF pretreatment unit acts as control process, and samples from the SF-UF-SWRO process might attain a new steady state when a disturbance occurs. Moreover, during this period, the turbidity feed flowrate was almost twice the mean flowrate, which would likely have disturbed the system.

4. PLS Prediction of the SWRO Process

PLS as a multivariate linear regression algorithm was used for prediction analysis of the PCF-SWRO process. A loading plot of PC1 and PC2 is shown in Fig. 10. Nine variables were included in the loading plot: seven input variables and two output variables. Output variables (Y) were permeate flowrate and permeate concentration, while the input variables (X) were feed flow rate of the plant, intake PCF/SF-UF flow rate, temperature of the feed water, pH, RO flow rate, turbidity feed, and pressure.

A loading plot of the PCF-SWRO process with three clusters is shown in Fig. 10(a). The first cluster consisted of temperature and permeate flowrate, the second cluster included feed flowrate and permeate concentration, and the last cluster was comprised of PCF flowrate, pH, pressure, RO flowrate, and turbidity feed. Clusters were divided according to the correlation among the variables. In the first cluster, the proximity of temperature and permeate flowrate reveal that these input variables were highly correleated and that their effect on permeate flowrate was considerable. In the second cluster, feed water flowrate was strongly correlated with permeate concentration. The third cluster consisting of PCF flowrate, pH, pressure, RO flowrate, and turbidity feed indicated that these variables were correlated with one other.-Moreover, the dataset of the input variables in the cluster are related to each other. Therefore, if the value of one input variable changed, so did the values of the other input variables. This third cluster also contained input variables that influenced each other. Fig. 10(b), which shows the loading plot for the SF-UF-SWRO process, was also characterized by three clusters. Note that the dependancy variables in the clusters in Fig. 10(b) were similar to the clusters shown in Fig. 10(a). However, the input variables for the SF-UF-SWRO process had different values than for the PCF-SWRO process. These results indicated that a change in the value of the influent feedwater flowrate resulted in a change in the input variables for both SWRO processes.

Fig. 11 shows the results of a VIP plot based on PLS loading weight. The variable with the largest value among all parameters in VIP plots has the most significant influence on process performance. As shown in Fig. 11(a), for the PCF-SWRO process, the input variables of feed flow rate, temperature of feed water, pressure, RO flow rate, pH, turbidity feed, and filtrate flow rate had the highest to lowest Q_p and C_p values in rank order. Fig. 11(b) shows the VIP plot for the SF-UF-SWRO process. Input variables for the VIP plot shown in Fig. 11(b) are different from those in Fig. 11(a). Feed flow rate, temperature of feed water, RO flow rate, pH, turbidity feed, pressure, and filtrate flow rate had the highest to lowest Q_p and C_p values in rank order. The loading and VIP plot results indicated that feed flow rate and temperature had a highly correlated effect on permeate flow rate and permeate concentration.

Seawater desalination is generally used to produce and supply freshwater for many industrial and domestic applications. Also, it is important to focus whether product water is able to meet requirement in quality and quality. In this regard, an efficient prediction model is essential for the output variables including the permeate flow rate and concentration [5,13] with respect to various operating variables including the feed flow rate, PCF/SF-UF filtrate flow rate, temperature of feed water, turbidity feed, pH, reverse osmo-

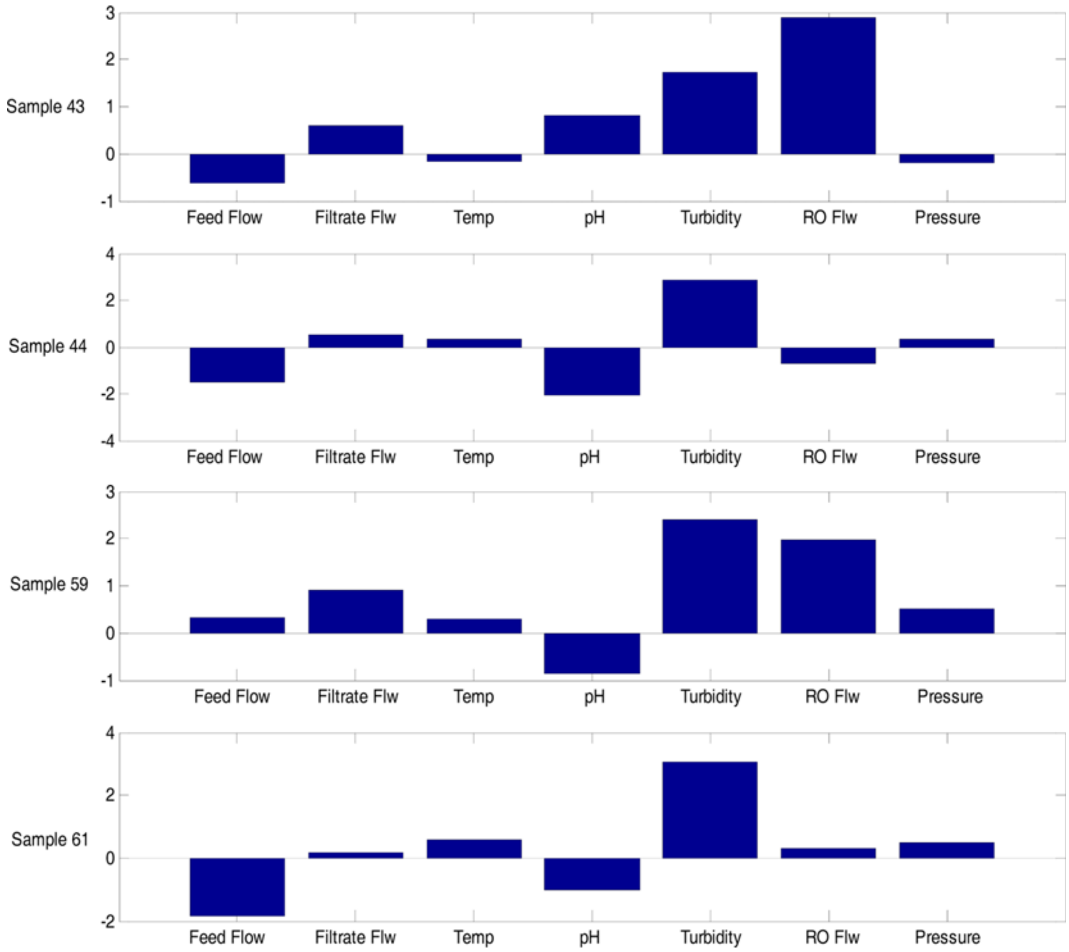

Fig. 9. Contribution plot of T^2 for sample numbers 43, 44, 59, and 61 of the SF-UF-SWRO process.

sis (RO) flow rate, and pressure. Therefore, PLS is a multivariate regression method used to propose the prediction models for both output variables including permeate flow rate and permeate concentration with respect to both PCF-SWRO and SF-UF-SWRO processes. The values in the dataset of PCF-SWRO process is mean centered and auto-scaled to unit variance. Table 2 represents the variables of PCF-SWRO process along with their means and standard deviations (SDs). The feed flow rate shows a particularly large variation with respect to its mean. On the other hand, Table 3 represents the variables of SF-UF-SWRO process with their mean and standard deviations (SDs). Of the variables considered, feed flow rate of SF-UF-SWRO shows a large variation with respect to its mean.

To know the superiority of the SF-UF-SWRO process over the PCF-SWRO process, root mean square error (RMSE) values of flow rate and concentration prediction curves (permeate flow rate and permeate concentration) are compared. RMSE was used to determine the modeling error between the modelled and measuredvalues of responses. RMSE is defined as follows [30]:

$$\mathrm{RMSE}=\sqrt{\frac{\sum_{i=1}^{n}(y_{i,\,observed}-\hat{y}_{i,\,model})^2}{n}} \tag{10}$$

where $y_{i,observed}$ are the actual observed values, $\hat{y}_{i,\,model}$ are the predicted values, and n is the number of experiments.

Fig. 12(a) and (b) shows the predicted permeate flow rate and permeate concetration using the PCF-SWRO process. In this model, PLS regression for the PCF-SWRO process is constructed with three latent variables (LVs), capturing about 73% of the original data. It means, the seven input variables (feed flow rate, PCF/SF-UF filtrate flow rate, temperature of feed water, turbidity feed, pH, reverse osmosis (RO) flow rate, and pressure) of X are reduced to three LVs, and represent a strong linear correlation with the Y variables (permeate flowrate and permeate concentration). The PLS regression model for PCF-SWRO process is expressed as:

$$\text{Permeate flow rate}_{\,PCF\text{-}SWRO}=0.6882\ \mathrm{X}\ (\mathrm{LV}_1)_{\,permeate\ flow\ rate} +0.2768\ \mathrm{X}\ (\mathrm{LV}_2)_{\,permeate\ flow\ rate}+0.1711\ \mathrm{X}\ (\mathrm{LV}_3)+\mathrm{H}_{\,permeate\ flowrate}$$

$$\text{Permeate concentration}_{\,PCF\text{-}SWRO}=0.6882\ \mathrm{X}\ (\mathrm{LV}_1)_{\,permeate\ concentration} +0.2768\ \mathrm{X}\ (\mathrm{LV}_2)_{\,permeate\ concentration}+0.1711\ \mathrm{X}\ (\mathrm{LV}_3)+\mathrm{H}_{\,permeate\ concentration}$$

where permeate flow rate is the vector of permeate flow rate for all samples in PCF-SWRO process. $(\mathrm{LV}_1)_{\,permeate\ flow\ rate}$ is a vector of the values of first latent variable, which corresponds to the permeate flow rate. $(\mathrm{LV}_2)_{\,permeate\ flow\ rate}$ is a vector of the values of second latent variable, which corresponds to the permeate flow rate. $(\mathrm{LV}_3)_{\,permeate\ flow\ rate}$ is a vector of the values of third latent variable, which corresponds

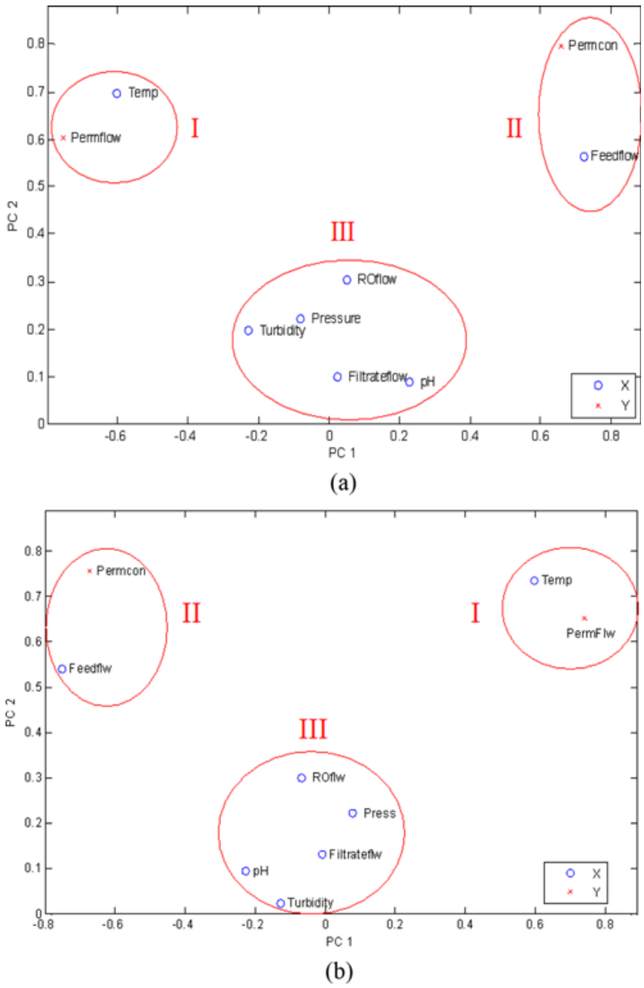

Fig. 10. Loading plot for a SWRO desalination plant using PLS: (a) PCF-SWRO process and (b) SF-UF-SWRO process.

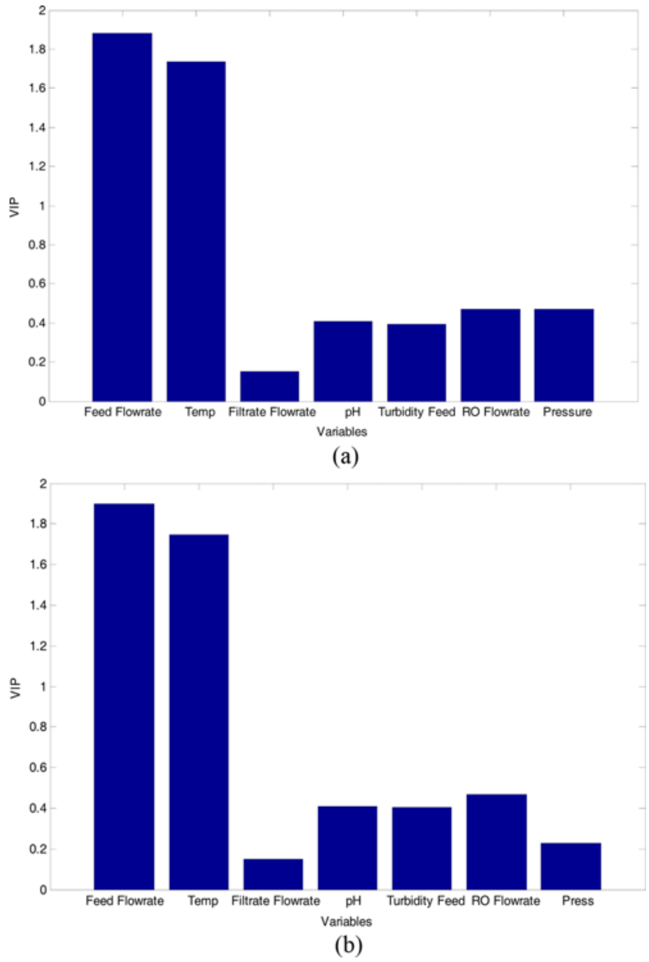

Fig. 11. VIP for SWRO desalination plant using PLS: (a) PCF-SWRO process and (b) SF-UF-SWRO process.

to the permeate flow rate. $H_{permeate\ flowrate}$ is the vector of the residuals, which corresponds to the permeate flow rate. On the other hand, permeate concentration is the vector of concentrations of permeate flow rate for all samples. $(LV_1)_{permeate\ concentration}$ is a vector of the values of first latent variable, which corresponds to the permeate concentration. $(LV_2)_{permeate\ concentration}$ is a vector of the values of second latent variable, which corresponds to the permeate concentration. $(LV_3)_{permeate\ concentration}$ is a vector of the values of third latent variable, which corresponds to the permeate concentration. $H_{permeate\ concentration}$ is the vector of the residuals, which corresponds to the permeate concentration.

Fig. 13(a) and (b) show the prediction results of permeate flowrate and permeate concentration using the SF-UF-SWRO regression model. The model for SF-UF-SWRO process was developed in a similar manner to the procedure mentioned in PCF-SWRO

Table 2. Process variables in a PCF-SWRO process taken from March 2011 to May 2011

Variable	Unit	Mean	Standard deviation (SD)
Feed flow rate	m^3/day	5458.8	19.91
Filtrate flowrate	m^3/day	5159.8	17.103
Temperature of feed	°C	13.56	4.14
pH		7.42	0.055
Turbidity feed	NTU	2.60	0.98
RO feed flow rate	m^3/day	75.48	0.075
Pressure	kgf/cm^2	52.17	0.946

Table 3. Process variables in a SF-UF-SWRO process taken from March 2011 to May 2011

Variable	Unit	Mean	Standard deviance (SD)
Feed flow rate	m^3/day	84.02	1.455
Filtrate flowrate	m^3/day	72.71	0.823
Temperature of feed	°C	13.56	4.14
pH		7.423	0.058
Turbidity feed	NTU	2.59	0.97
RO feed flow rate	m^3/day	72.48	0.081
Pressure	kgf/cm^2	52.08	0.943

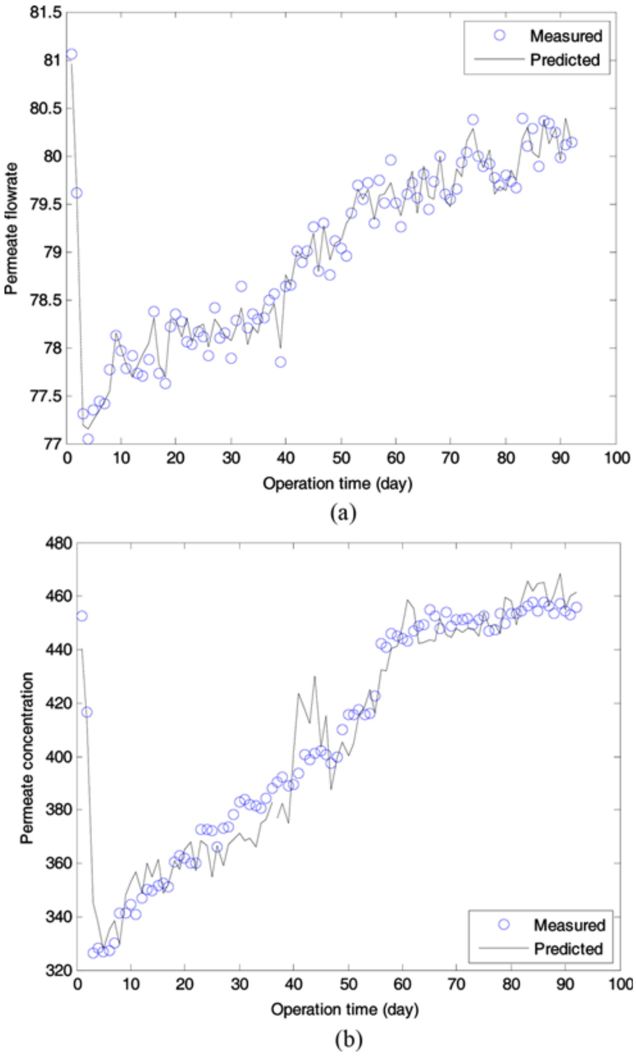

Fig. 12. PLS prediction results for the PCF-SWRO process: (a) permeate concentration and (b) permeate flow rate.

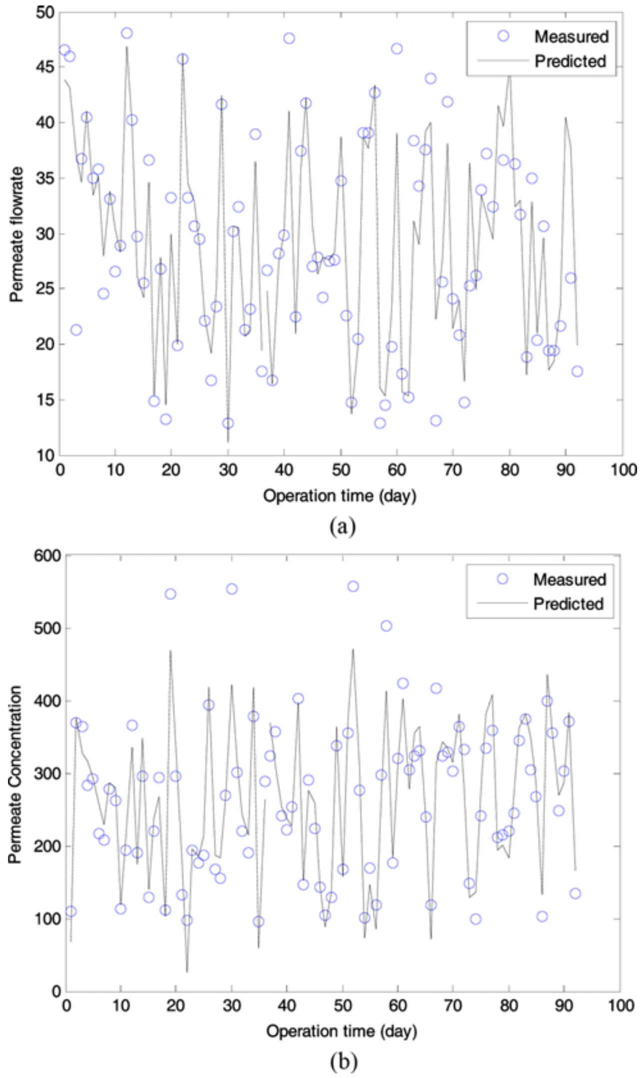

Fig. 13. PLS prediction results for SF-UF-SWRO process: (a) permeate flow rate and (b) permeate concentration.

model. This model captures 78% of the original data. The PLS regression model for PCF-SWRO process is expressed as:

$$\text{Permeate flow rate}_{\text{SF-UF-SWRO}}=0.6941\text{ X }(LV_1)_{\text{permeate flow rate}}+0.2703\text{ X }(LV_2)_{\text{permeate flow rate}}+0.1711\text{ X }(LV_3)+H_{\text{permeate flowrate}}$$

$$\text{Permeate concentration}_{\text{SF-UF-SWRO}}=0.6941\text{ X }(LV_1)_{\text{permeate concentration}}+0.2703\text{ X }(LV_2)_{\text{permeate concentration}}+0.1711\text{ X }(LV_3)+H_{\text{permeate flowrate}}$$

where permeate flow rate is the vector of permeate flow rate for all samples in SF-UF-SWRO process. On the other hand, permeate concentration is the vector of concentrations of permeate flow rate for all samples in SF-UF-SWRO process.

To know the superiority of developed PCF-SWRO and SF-UF-SWRO processes, RMSE values of flow rate and concentration prediction curves of permeate flow rate and concentration obatined using PLS regression models are compared. Table 4 summarizes the RMSE values of PLS prediction results for two pretreatment processes. As shown in Table 4, RMSE values of permeate flowrate, and permeate concentration are 28.6 and 26.27, and 289.4 and 280.02 for training and testting data, respectively, when considering SF-UF-RO process. On the other hand, the RMSE values of permeate flow rate, and permeate concentration are 31.5 and 28.25, and 350.44 and 303.08 for training and testing data respectively, when considering PCF-SWRO process. The comparision shows that the RMSE values of output variables, namely permeate flow rate and permeate concentration, were lower for the SF-UF-SWRO process than the PCF-SWRO process as shown in Table 4. This indicates that the SF-UF-SWRO model results in more accurate prediction than the PCF-SWRO model.

Table 4. Comparison of RMSE values between the PCF-SWRO process and SF-UF-SWRO process

Process	Permeate flow rate (Y1)		Permeate concentration (Y2)	
	Train data	Test data	Train data	Test data
PCF-SWRO	31.5	28.25	350.44	303.0865
SF-UF-SWRO	28.6	26.27	289.4	280.02

CONCLUSIONS

After analyzing, monitoring, and predicting the performance of two SWRO desalination pretreatment processes using multivariate statistical techniques, we were able to draw the following conclusions:

1. MANOVA analysis revealed significant differences in the performance and efficiencies of the two SWRO processes based on measurements of SDI and turbidity. Overall performance and efficiency of PCF was lower than that of SF-UF.

2. Among the two SWRO desalination processes, the SF-UF-SWRO process was modeled more reliably based on T^2 and a SPE chart with a considerably lower number of outliers and disturbances than the PCF-SWRO process.

3. PLS results showed that the RMSE values of permeate flowrate and permeate concentration for the SF-UF-SWRO process were lower than those for the PCF-SWRO process, which indicated that the SF-UF-SWRO prediction model was more accurate than the PCF-SWRO prediction model.

Based on these conclusions, the overall cost of the PCF-SWRO process is likely to be higher than that of the SF-UF-SWRO process due to the lower preatment efficiency of the former process, which is likely to decrease membrane life.

ACKNOWLEDGEMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2015R1A2A2A11001120).

REFERENCES

1. I. Janghorban Esfahani and C. Yoo, *Desalination*, **332**, 18 (2014).
2. I. Janghorban Esfahani, J. Kim and C. Yoo, *Ind. Eng. Chem. Res.*, **52**, 11099 (2013).
3. I. Janghorban Esfahani and C. Yoo, *Energy*, **59**, 340 (2013).
4. W. Lawler, Z. B. Hartke, J. M. Cran, M. Duke, G. Lesil, P. D. Ladewig and P. Le-Clech, *Desalination*, **299**, 103 (2012).
5. D. Garcia-Alvarez and M. J. Funente, *Desalin. Water Treat.*, **52**, 1272 (2014).
6. D. Garcıa-Alvarez, Fault detection using principal component analysis (PCA) in a wastewater treatment plant (WWTP), *Proceedings of the International Student's Scientific Conference* (2009).
7. C. Yoo, S. Choi and I. Lee, *Ind. Eng. Chem. Res.*, **41**, 4303 (2002).
8. C. Rosen and J. A. Lennox, *Water Res.*, **35**, 3402 (2001).
9. D. Aguado and C. Rosen, *Eng. Appl. Artif. Intell.*, **21**, 1080 (2008).
10. S. I. Al-mutaz and A. B. Al-sultan, *Desalination*, **120**, 153 (1998).
11. V. Yangali-Quintanilla, T.-U. Kim, M. Kennedy and G. Amy, *Drink. Water Eng. Sci.*, **1**, 7 (2008).
12. W. C. McFall, A. Bartman, D. P. Christofides and Y. Cohen, *Ind. Eng. Chem. Res.*, **47**, 17 (2008).
13. I. Janghorban Esfahani, M. Kim, C. Yun and C. Yoo, *J. Membr. Sci.*, **442**, 83 (2013).
14. J. Jackson, *A user guide to principal component analysis*, Wile (1991).
15. R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis*, 3rd Ed., Prentice Hall (1992).
16. M. Kim, B. S. Rao, O. Kang, J. Kim and C. Yoo, *Energy Build.*, **46**, 48 (2012).
17. C. Yoo, J. Lee, P. Vanrolleghem and I.-B. Lee, *Chemometrics Intell. Lab. Syst.*, **71**, 151 (2004).
18. H. Liu, M. Kim, O. Kang, B. S. Rao, C. Kim, J. Kim and C. Yoo, *Indoor Built Environ.*, **21**, 205 (2012).
19. C. Yoo and I. Lee, *Bioprocess. Biosyst. Eng.*, **29**, 213 (2006).
20. W. Lin, Y. Qian and X. Li, *Comput. Chem. Eng.*, **24**, 423 (2000).
21. E. L. Russel, L. H. Chiang and R. D. Braatz, *Chemometrics Intell. Lab. Syst.*, **51**, 81 (2000).
22. B. M. Wise and N. B. Gallagher, *J. Process. Contr.*, **6**, 329 (1996).
23. A. Raich and A. Cinar, *AIChE J.*, **42**, 995 (1996).
24. M. J. Kim, H. Liu, J. T. Kim and C. K. Yoo, *Energy Build.*, **66**, 384 (2013).
25. C. Yoo, J. Lee, I. Lee and P. Vanrolleghem, *Automation in Water Quality Monitoring II*, **50**, 163 (2004).
26. Y. Kim, J. Kim, I. Kim, J. Kim and C. Yoo, *Environ. Eng. Sci.*, **27**, 721 (2010).
27. B. S. Dayal and J. F. MacGregor, *J. Process. Contr.*, **7**, 169 (1997).
28. T. Yamamoto, A. Shimameguri, M. Ogawa, M. Kano and I. Hashimoto, *IFAC Symposium on Advanced Control of Chemical Processes (ADCHEM)* (2004).
29. Y. Kim, M. Kim, J. Lim, J. Kim and C. Yoo, *J. Hazard. Mater.*, **183**, 448 (2010).
30. H. Liu, M. Huang, J. Kim and C. Yoo, *Korean J. Chem. Eng.*, **22**, 94 (2013).
31. A. Alhadidi, A. J. B. Kemperman, B. Blankert, J. C. Schippers, M. Wessling and W. G. J. van der Meer, *Desalination*, **273**, 48 (2011).
32. S. Valle, W. Li and S. J. Qin, *Ind. Eng. Chem. Res.*, **38**, 11 (1999).