



## Fluorescence spectrum-based biofouling prediction method for RO membrane systems

Sun-Nyoung Hwang<sup>a</sup>, Wooyeol Choi<sup>b</sup>, Hyuk Lim<sup>a,b,c,\*</sup>, Jinhee Choi<sup>d</sup>, Hyunjung Kim<sup>d</sup>, In Seop Chang<sup>d</sup>

<sup>a</sup>Department of Medical System Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 500-172, Republic of Korea

Tel. +82 62 715 2229; Fax: +82 62 715 2204; email: hlim@gist.ac.kr

<sup>b</sup>School of Information and Communications, Gwangju Institute of Science and Technology (GIST), Gwangju 500-172, Republic of Korea

<sup>c</sup>Department of Nanobio Materials and Electronics, Gwangju Institute of Science and Technology (GIST), Gwangju 500-172, Republic of Korea

<sup>d</sup>School of Environmental Science and Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju 500-172, Republic of Korea

Received 25 December 2011; Accepted 10 February 2012

---

### ABSTRACT

Monitoring reverse osmosis (RO) membrane conditions is an important task because it helps reduce the operation and maintenance cost in the RO membrane desalination systems by achieving long membrane lifetime and energy saving. As biological interactions between the membrane itself and microorganism cause the rapid degradation of membrane performance, it is crucial to identify and quantify potential biofoulants that are sensitive to each specific RO membrane. This study proposed a biofouling prediction method that indirectly quantifies the degree of biofouling by comparing the fluorescence excitation-emission matrix (EEM) of foulants sampled on the fully fouled RO membrane and those of brine samples from currently operating RO system. The experiment showed that the similarity distance measured from the comparison between the two fluorescence EEMs tends to increase when brine samples were secured from relatively clean RO membranes.

*Keywords:* Similarity measuring; Biofouling; Reverse osmosis membrane; Fluorescence excitation-emission matrix; Principal component analysis

---

### 1. Introduction

Reverse osmosis (RO) desalination plants have nowadays played a significant role in water supply not only to compensate the water shortage caused by environmental contamination, but also to meet the fresh-water demand arising from growing population.

Even though RO membrane system is recognized as one of the leading desalination technologies, it has some economical issues to be resolved such as low energy efficiency and membrane replacement costs.

Membrane biofouling, which occasionally occurs as time elapses due to the developed biofilms, has been considered as a significant problem that affects the overall RO system performance because it causes a drop of permeate flux, which requires more energy to

---

\*Corresponding author.

generate more pressure for keeping the same flux of potable water, and shortens the membrane lifetime leading to an increase in the replacement cost of RO membranes. Timely membrane treatments such as chemical cleaning can prevent an abrupt performance degradation of RO membranes and effectively defer membrane replacement due to fouling, scaling, and degradation, resulting in operation and maintenance cost reduction of RO systems [1]. To perform cost-effective membrane treatments, the development of a fouling index that represents the degree of membrane fouling is crucial. Conventionally, membrane autopsy was performed for investigation of the characteristics of foulants on membrane; however, it is an expensive method because damage of the membrane is inevitable for the examination [2]. Instead, several indirect fouling prediction methods such as silt density index (SDI), modified fouling index (MFI), ultrasonic time-domain reflectometry (UTDR), and membrane fouling simulator (MFS) were developed as an alternative to the direct and destructive method. SDI, which is the most commonly used method of the indirect and non-destructive measurements of membrane fouling, is used as an indicator representing the amount of particles related to the membrane fouling. It is based on the ratio of the time it takes for 500 ml feed water to flow through a membrane filter with a pore size of 0.45  $\mu\text{m}$  and 47 mm in diameter at a certain pressure to the time it takes for the same amount of feed water to flow through the same filter after 15 min. However, it has been reported that SDI is not suitable for the fouling indicator of RO membranes with a pore size smaller than 0.45  $\mu\text{m}$  because it does not consider the particulate matter less than 0.45  $\mu\text{m}$  [3]. MFI has been thought to be an upgraded version of SDI. However, because it uses the same membrane filter as that of SDI, it has the same problem in representing the degree of membrane fouling [4]. UTDR monitors the development and growth of fouling layer in real-time by exploiting the amplitudes and arrival time differences of the ultrasonic waves reflected from the membrane surface and the layer of foulants on the membrane [5]. MFS is a miniaturized testing device made of the same material as spiral-wound RO membranes, which allows detecting the pressure drop, observing the membrane surface through the window, and analyzing the coupons of the MFS membrane [6]. Since these previous approaches try to evaluate the fouling potential capacity of the feed water before passing the RO membrane, it is hard to estimate the fouling state of RO membrane itself.

Membrane foulants can be classified as inorganic compounds, colloidal or particulate matter, dissolved organics, chemical reactants, and microorganisms. Bio-fouling caused by microorganism is thought to be the

major problem because all other foulants can usually be removed by pre-treatment steps, and it only requires a few colonies to be present to develop into a biofilm. Furthermore, microorganisms are ubiquitous in most water systems and tend to adhere to surfaces and multiply on any surface in contact with the water treatment system. Once attached to the membrane surface, microorganisms can then grow and rapidly increase the amount of extracellular polymeric substances (EPS) (polysaccharides together with proteins and compounds such as DNA derivatives) in order to survive and form a mature biofilm [7]. These EPS compounds are high-molecular weight complexes, and include carbohydrates, proteins, nucleic acids, lipids, and other polymeric compounds that can be secreted by microorganisms into their aggregates [8]. It has also been reported that the proteins and humic-like substances of these compounds generate fluorescence signals because they contain a large aromatic structure with functional groups [9]. This fluorescence, emission spectra when molecules re-emit absorbed light at a different wavelength, can be measured via fluorescence spectroscopy. In previous studies, fluorescence spectroscopy has been shown to be very sensitive and hence has been widely used to identify the structure and type of dissolved organic matter (DOM) in fields such as marine and fresh water environments [10]. Indeed, a great deal of research has been conducted in attempts to classify and characterize the DOM type using a fluorescence excitation emission matrix (EEM), in which regions respond to different types of DOM functional groups. As time passed, the RO membrane is covered with foulants that might detach from or attach to the RO membrane; therefore, the fluorescence EEM changes in the brine are closely related to foulants directly affecting the RO membrane. For this reason, we propose an online and non-intrusive method to predict the condition of RO membranes by comparing the fluorescence EEM of brine and that of foulants sampled on fouled RO membranes. This method takes the fluorescence EEM of sampled foulants as a reference since fluorescence EEM of foulants on RO membrane directly shows the affected materials on the membrane, and quantifies the similarity change between the reference and the fluorescence EEM of brine sampled during RO operation.

## 2. Materials

### 2.1. Samples collection and treatment

The water samples were collected from a local plant using an RO membrane system plant located close to the Yellow Sea in South Korea. In this system,

raw seawater was treated using a three-stage filtration process: (1) a multimedia filter, (2) a microfiltration (MF) membrane filter, and (3) an RO membrane module. Samples were collected at the following stages: raw seawater (R) in the storage tank, feed water of the RO membrane (F) that was treated by multilayer and MF membrane filtration, permeate (P) that passes through the RO membrane, and finally the brine (B), which is concentrated; a schematic of the system is shown in Fig. 1. The water samples were stored in sampling bottles at 4°C and measured within 3 days after sampling; the pH of the water samples was not adjusted.

The specifications of RO membrane were a polyamide thin-film composite, 8.0 inch (203 mm) in size. The RO membrane was replaced due to a decline in membrane performance. At that time, the feed side of the fouled RO membrane was collected because it was seriously fouled. To obtain small fractions, the fouled RO membrane was cut into pieces; 10 fragments (1 × 1 cm) were extracted at a clean bench and then immersed in 20 mL deionized water. These fragments were treated with ultrasound pulses for 1-min periods, which were repeated 10 times to avoid generating heat. To detach the remaining foulants from the membrane fragments, the treated fragments were rinsed with 30 mL deionized water.

## 2.2. Fluorescence analysis

The water samples and treated fouled RO membrane were analyzed by a fluorescence spectrophotometer (F-2500 FL spectrophotometer, Hitachi High-Technologies Corporation, Japan) at room temperature (~22°C). The fluorescence EEM data of the samples were then modeled at an excitation wavelength ranging from 220 to 490 nm and emission wavelength from 220 to 490 nm, with 10 nm sampling intervals. The excitation and emission slits were maintained at 5 nm and the scanning speed was set at

3,000 nm/min. The voltage of the photomultiplier tube was 700 V. During the fluorescence analyses, the Raman scattering peak intensity recorded for deionized water at Ex/Em = 348/397 nm was examined in order to confirm there were no significant fluctuations in the performance of the spectrophotometer lamp or other hardware; deionized water was used as the blank sample.

## 2.3. Fluorescence data pretreatment and determination of membrane fouling indicator

Obtained fluorescence EEM was pretreated for investigation of Rayleigh and Raman light scattering regions using MATLAB 7.9.0 with EEMcut [11]. First, the fluorescence EEM of feed water is subtracted from that of brine in order to exclude the feed-specific effects. In other words, we can eliminate the feed-specific characteristics through the relative differences between the fluorescence EEM of feed water and brine.

## 3. Proposed algorithm

We propose a principal component analysis (PCA)-based method that quantitatively shows how much the fluorescence EEM of brine is correlated to that of fouled membrane in order to indirectly predict the degree of fouling developed in the RO membranes. The PCA is a very powerful technique to extract a set of characteristic features from a large data by dimension reduction [12]. Our method aims to quantify the similarity change of the fluorescence EEM of brine samples from the reference fluorescence EEM of fouled membrane over time. We generate a set of multiple synthetic data from the reference fluorescence EEM and use it for the training process of PCA technique. The synthetic data ( $C_1$  and  $C_2$  data in Fig. 2) are generated by selecting peak points in the fluorescence EEM data and applying the Gaussian filter to the original data (R in Fig. 2) at each selected

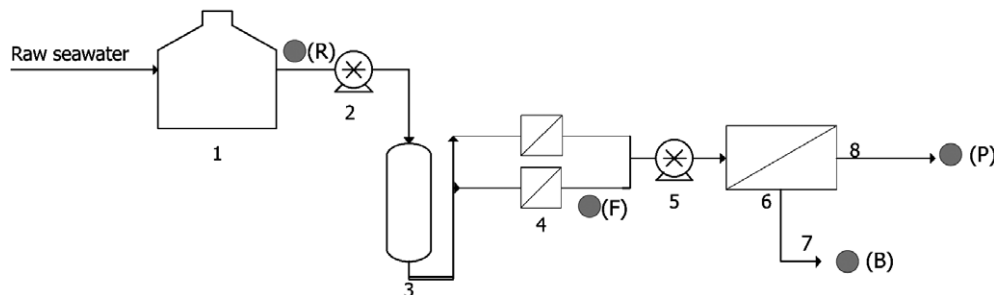


Fig. 1. Schematic of desalination system using an RO membrane (1: raw seawater storage tank; 2: feed pump; 3: multimedia filter; 4: 1st and 2nd micro filter; 5: high pressure pump; 6: RO membrane module; 7: brine; 8: permeate). The circles indicate the sampling point of water samples: R-raw seawater; F-feed water for RO membrane; P-permeate, and B-brine.

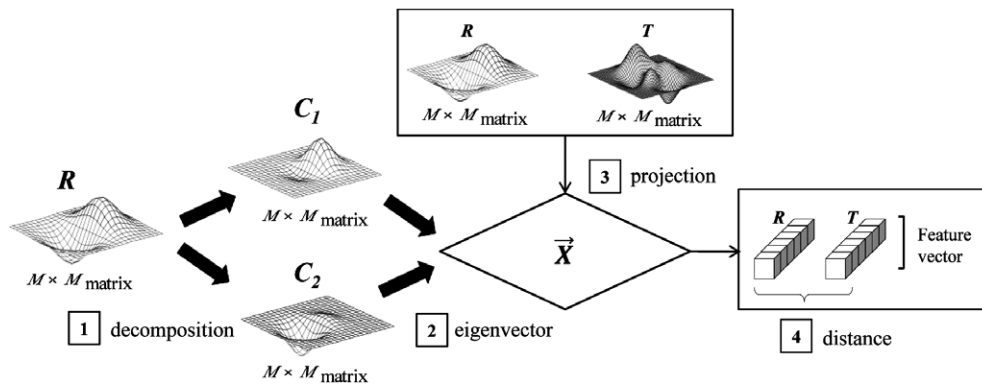


Fig. 2. Procedure of the proposed algorithm ( $R$ : fluorescence EEM of fouled membrane,  $C_1$  and  $C_2$ : synthetic fluorescence EEM generated from  $R$ ,  $T$ : fluorescence EEM of brine sample to be evaluated,  $\bar{X}$ : eigenvectors extracted from synthetic fluorescence EEMs).

peak point. Note that peaks of fluorescence EEM in a wavelength range indicate the existence of particular substances that can cause the biofouling of RO membrane [13]. Therefore, each synthetic data can be considered as an emphasized variation of the reference fluorescence EEM data for the corresponding fouling substances. The eigenvectors can be obtained in the same manner as done in conventional approaches. Then, the feature vector for a fluorescence EEM data is obtained by projecting the data onto the eigenvectors. The similarity between the fluorescence EEM of the reference and a brine sample is finally computed by calculating the Euclidean distance between two corresponding feature vectors.

---

#### Algorithm 1. Generating the multiple synthetic data

---

**Require:** A two-dimensional  $M \times M$  matrix  $R$  of intensity value.  $\theta_d$  indicates the minimum distance.  $N$  is the number of peak points to be selected.  $\kappa$  means covariance  $\begin{bmatrix} \kappa & 0 \\ 0 & \kappa \end{bmatrix}$ . Both  $x$  and  $y$  are an  $N \times 1$  column vector for representation of the  $x$ - and  $y$ -coordinate of the selected peak point, respectively.  $G_i$  is the  $M \times M$  Gaussian distribution matrix with the mean of  $[x_i, y_i]$  and the covariance  $\kappa$ .  $C_i$  is the  $M \times M$  matrix generated by element-by-element multiplication of  $G_i$  and  $R$ .

- 1:  $[x, y] \leftarrow \text{FindPeak}(R, \theta_d, N)$ ;
  - 2: **for**  $i=1$  to  $N$  **do**
  - 3:  $G_i \leftarrow \text{GaussianDistribution}([x_i, y_i], \kappa)$ ;
  - 4:  $C_i = G_i \cdot R$ ;
  - 5: **end for**
- 

#### Algorithm 2. Computing the eigenvectors

---

**Require:**  $M \times M$  synthetic data  $C_i$  for  $i = 1, \dots, N$ . SVD ( $\cdot$ ) is the singular value decomposition function.  $k$  is the number of the chosen column vectors of  $V$ .

- 1: **for**  $i=1$  to  $N$  **do**
  - 2:  $s_i \leftarrow \text{vectorized } C_i$ ;
  - 3: **end for**
  - 4:  $\bar{s} \leftarrow \sum_{i=1}^N s_i / N$ ;
  - 5: **for**  $i=1$  to  $N$  **do**
  - 6:  $\phi_i \leftarrow s_i - \bar{s}$ ;
  - 7: **end for**
  - 8:  $\Psi \leftarrow \sum_{i=1}^N \phi_i * \phi_i^T / N$ ;
  - 9:  $[U, S, V] \leftarrow \text{SVD}(\Psi)$ ;
  - 10: Store the first  $k$  columns of  $V$ .
- 

#### Algorithm 3. Computing the Euclidean distance

---

**Require:**  $k$  columns of  $V$ .  $M^2 \times 1$  vectorized reference  $r$  and test data  $t$ .  $v_i$  denotes the  $i$ -th column vector of  $V$ .  $\text{ED}(\cdot, \cdot)$  symbolizes the function for measuring the Euclidean distance.

- 1: **for**  $i=1$  to  $k$  **do**
  - 2:  $u_i \leftarrow (r - \bar{s}) \cdot v_i$ ;
  - 3:  $w_i \leftarrow (t - \bar{s}) \cdot v_i$ ;
  - 4: **end for**
  - 5: distance  $\leftarrow \text{ED}(u, w)$ ;
- 

#### 3.1. Generating the multiple synthetic data

Let  $R$  denote the fluorescence EEM of the foulants on a fully fouled RO membrane, which is a two-dimensional  $M \times M$  matrix. From the reference  $R$ , we generate  $N$  synthetic data (i.e.  $C_i$  for  $i=1, \dots, N$ ). In order to make each synthetic data  $C_i$  represent excitation-emission pair of the  $i$ -th dominant foulants, we first select the  $i$ -th largest peak points and apply the Gaussian filter, which can filter out all the measurement data except the points near the selected peak point. During the peak selection procedure, it is very important to mitigate the effect of local minima

due to measurement noises of fluorescence EEMs. The selected peak points need to be away from each other by more than a minimum distance  $\theta_d$ . Starting from the largest peak, we select  $N$  peak points. If the  $i$ -th peak point candidate is within  $\theta_d$  from the  $j$ -th peaks with  $j < i$ , we ignore the peak point candidate and select the next largest one as the candidate point. It can prevent insignificant peaks that are close to each other from being selected. Then,  $C_i$  is obtained by applying multivariate Gaussian distribution centered at the  $i$ -th peak point to  $R$ . Each multivariate Gaussian distribution has the mean of the corresponding peak point and the covariance of  $\begin{bmatrix} \kappa & 0 \\ 0 & \kappa \end{bmatrix}$ . The detailed procedures are shown in Algorithm 1.

### 3.2. Extracting the eigenvectors from the training set

In order to obtain the principal components of  $R$ , we apply PCA to the synthetic data-set. First, we convert  $C_i$  into  $s_i$  in an  $M^2 \times 1$  vector format, which serves as a training set of PCA. Then,  $s_i$  is centered by extracting the average of the training set in order to remove a positional bias of the training set as follows:  $\phi_i = s_i - \bar{s}$  where  $\bar{s} = \sum_{i=1}^N s_i / N$ . The covariance matrix  $\Psi = \sum_{i=1}^N \phi_i * \phi_i^T / N$  is obtained to capture the variation among the training set. By using SVD, the eigenvectors of the training set are obtained. The eigenvectors are the principal components that can efficiently represent the training data-set. In order to exclude a trivial variation of the training set, we select the most significant  $k$  eigenvectors, which are denoted by  $v_i$  for  $i=1, \dots, k$ . A symbolic description is provided in Algorithm 2.

### 3.3. Calculating the euclidean distance using feature vectors

Given the eigenvectors of the training data-set, we obtain the feature vector of fluorescence EEM data-set by projecting the data-set onto the eigenvectors. The vectorization of fluorescence EEMs of the reference  $R$  and the test data  $T$  is centered and projected onto  $r$  and  $t$ , respectively, as follows:

$$u = [u_1, u_2, \dots, u_k]$$

$$w = [w_1, w_2, \dots, w_k]$$

where

$$u_i \leftarrow (r - \bar{s}) \cdot v_i$$

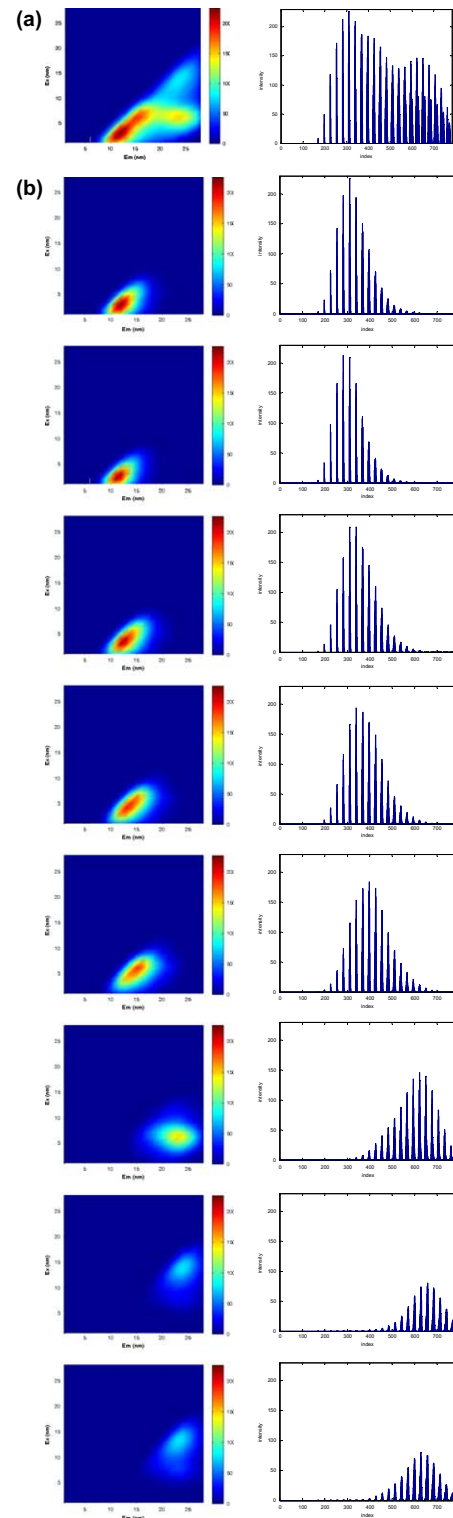


Fig. 3. (a) Fluorescence EEM of the foulants on the fouled RO membrane and its vectorization. (b) Eight synthetic fluorescence EEMs generated from (a) and their vectorizations.

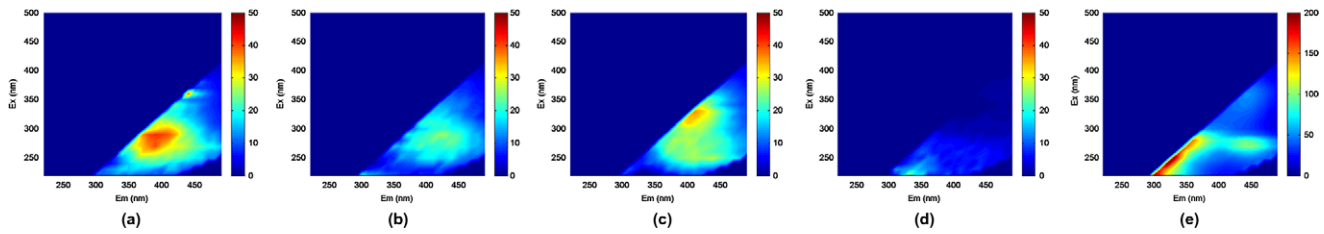


Fig. 4. Fluorescence EEM spectra of brine sampled at (a) fully fouled RO membrane, (b) clean state after membrane replacement, (c) intermediate state after a two-month operation, (d) refreshed state after chemical cleaning, and (e) fluorescence EEM of foulants sampled from (a).

Table 1  
The first five features of each fluorescence EEM spectra

	(a)	(b)	(c)	(d)	(e)
PC1	326.39	356.93	373.11	304.84	−45.34
PC2	35.20	−35.33	−1.22	−106.75	182.98
PC3	29.38	13.14	38.39	22.01	−164.24
PC4	−96.47	−93.88	−74.78	−113.75	25.35
PC5	32.43	61.06	29.19	129.29	−111.71

$$w_i \leftarrow (t - \bar{s}) \cdot v_i,$$

where  $r$  and  $t$  are the vectorization of  $R$  and  $T$ , respectively.

The similarity distance between  $u$  and  $w$  is obtained by calculating the Euclidean distance of each feature vector as shown in Algorithm 3.

$$ED(u, w) = \sqrt{\sum_{i=1}^k (u_i - w_i)^2}.$$

#### 4. Results

In order to evaluate our proposed algorithm, we prepared a reference fluorescence EEM data of foulants

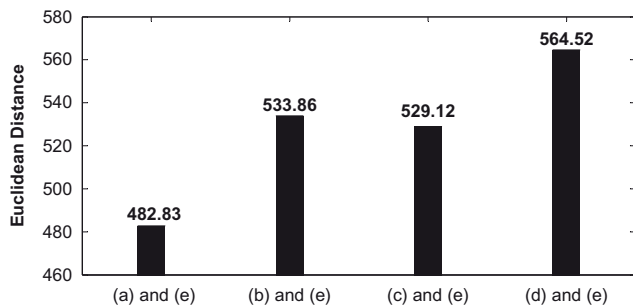


Fig. 5. Euclidean distance between the feature vector of foulants on RO membrane and that of brine samples obtained from different fouling conditions in Table 1.

on fully fouled RO membrane and performed the pre-processing, which is explained in detail in the previous section 2. The reference fluorescence EEM has the vector with a dimension of 784 as shown in Fig. 3(a). We selected eight peak points from the reference data and generated the corresponding synthetic reference data for the training set. Those generated data  $C_i$  and the corresponding vector  $s_i$  are depicted in Fig. 3(b). After performing a PCA on the training set, we chose the eigenvectors corresponding to the first five largest eigenvalues that account for almost 100% of the total variance of 784 eigenvalues. Fig. 4 shows the fluorescence EEM data of brine sampled at a different time from RO membrane, which correspond to different fouling conditions, i.e. fully fouled state, clean state after membrane replacement, after a two-month operation, and after chemical treatment, as well as the fluorescence EEM of the foulants on fully fouled RO membrane. Once those data were projected onto the selected five eigenvectors, each data can be represented as a set of five features. Table 1 shows the features of the corresponding fluorescence EEM spectra in Fig. 4. The Euclidean distance between the feature vector of brine sample and that of foulants is shown in Fig. 5. We observed that the fluorescence EEM of brine sampled from (a) had the smallest value when comparing with (e). It implies that they are closely correlated with each other. The distance for (b) and (e) apparently increased due to the membrane replacement. After the time elapse, the difference between the fluorescence EEM of (e) and (c) decreased as the membrane was being contaminated. When the chemical treatment was performed, the distance between (e) and (d) increased again. This indicates that the membrane at (d) became considerably clearer than (c).

##### 4.1. Brief description of the application

The software implementing the proposed algorithm was developed with Java programming language of Java development kit (JDK) 1.7.0. Fig. 6 shows a snapshot of the graphic user interface

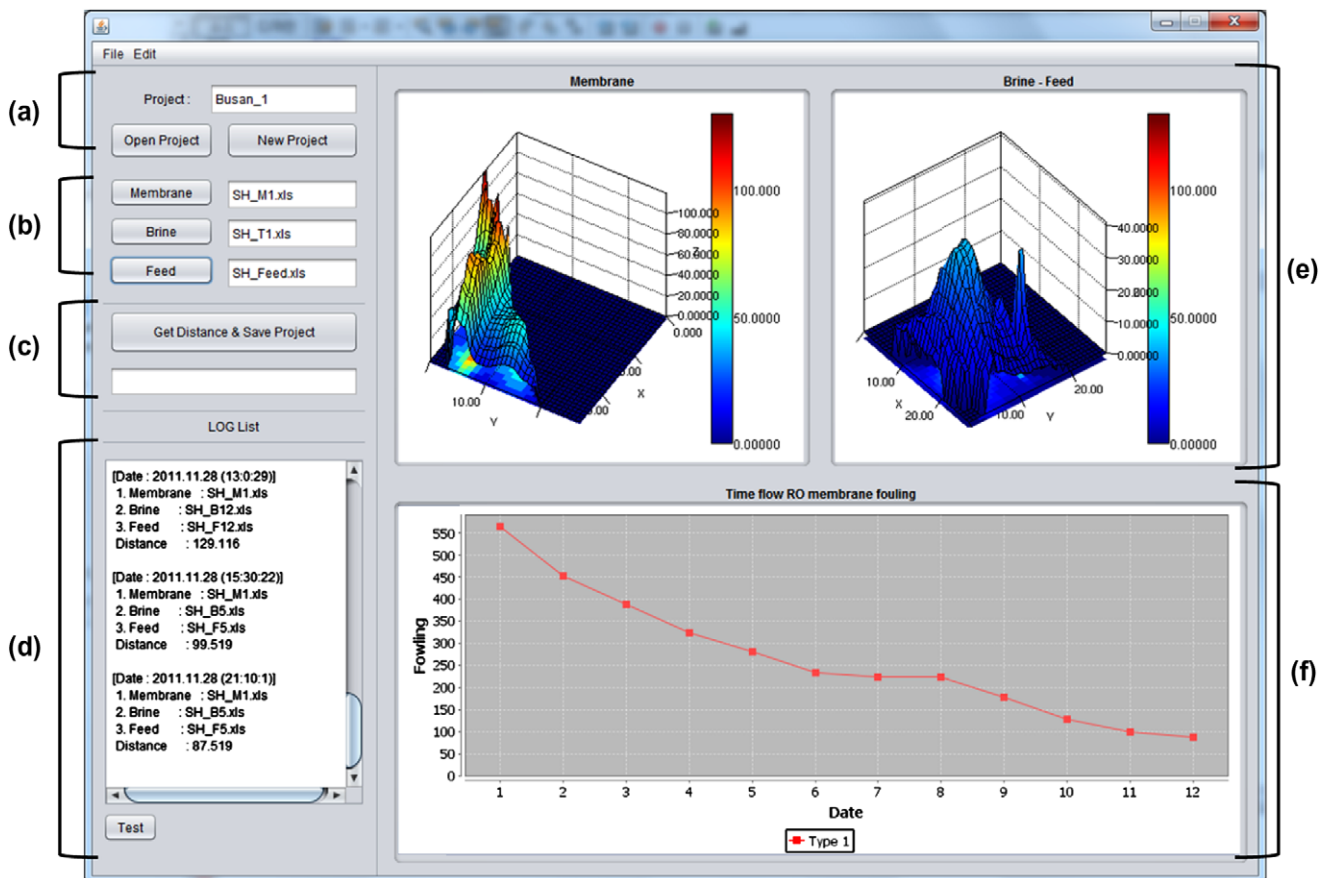


Fig. 6. Software for monitoring the degree of membrane fouling developed in RO membrane using JDK 1.7.0.

(GUI). We provide a brief explanation for the GUI of the developed software as follows: Part (a) includes the controls for creating a new project or opening an existing project of a certain membrane module. In part (b), there are three buttons for loading the fluorescence EEMs of foulants taken from the RO membrane, brine, and feed water, respectively. As the data-sets are stored in an MS Excel file format, we implemented the data I/O functions with JExcelApi or a JAVA library. Part (c) enables us to calculate the distance by applying the proposed algorithm to the loaded data, and then to save the similarity distance results in an XML format. Note that XML is a well-designed and standardized textual format for structured documents to be readable over the Internet, and has been used in a wide variety of applications thanks to its simplicity, generality, and usability. Part (d) shows the overall information about the current project. Parts (e) and (f) illustrate the 3-dimensional plots for fluorescence EEM spectra of the reference and brine samples and a trend of variation in distance, respectively.

## 5. Conclusion

In this study, a novel method to evaluate the degree of biofouling of currently operating RO membrane was developed. Unlike several previous indirect and nondestructive methods that measure fouling potential based on feed water, this proposed method can predict the membrane condition based not only on feed water and brine samples, but also on foulants on fouled RO membranes. Experimental results support our hypothesis that the far distance appears in the case of comparing with the brine sampled from the relatively clean RO membrane. This approach can be used for assessment of the cleaning performance and extended to other applications related to the membrane biofouling as well.

## Acknowledgments

This research was supported by a grant (07SeaHeroB01-03) from Plant Technology Advancement Program funded by the Ministry of Construction & Transportation of Korea and by a grant (R31-10026)

from the World Class University program provided by the MEST of Korea.

## References

- [1] C. Fritzmann, J. Lowenberg, T. Wintgensand, T. Melin, State-of-the-art of reverse osmosis desalination, *Desalination* 216 (2007) 1–76.
- [2] A. Al-Amoudi, R.W. Lovitt, Fouling strategies and the cleaning system of NF membranes and factors affecting cleaning efficiency, *J. Membr. Sci.* 303 (2007) 4–28.
- [3] S.G. Yiantsios, A.J. Karabelas, An assessment of the silt density index based on RO membrane colloidal fouling experiments with iron oxide particles, *Desalination* 151 (2003) 229–238.
- [4] S. Khirani, R.B. Aim, M. Manero, Improving the measurement of the modified fouling index using nanofiltration membranes (NF-MFI), *Desalination* 191 (2006) 1–7.
- [5] J. Li, R.D. Sanderson, In situ measurement of particle deposition and its removal in microfiltration by ultrasonic time-domain reflectometry, *Desalination* 146 (2002) 169–175.
- [6] J.S. Vrouwenvelder, S.M. Bakker, L.P. Wessels, J.A.M. van Passen, The membrane fouling simulator as a new tool for biofouling control of spiral-wound membranes, *Desalination* 204 (2007) 170–174.
- [7] C.M. Pang, P. Hong, H. Guo, W.T. Liu, Biofilm formation characteristics of bacterial isolates retrieved from a reverse osmosis membrane, *Environ. Sci. Technol.* 39 (2005) 7541–7550.
- [8] H.C. Flemming, J. Wingender, Relevance of microbial extracellular polymeric substances (EPSs)—Part I: Structural and ecological aspects, *Water Sci. Technol.* 43 (2001) 1–8.
- [9] G.P. Sheng, H.Q. Yu, Characterization of extracellular polymeric substances of aerobic and anaerobic sludge using three-dimensional excitation and emission matrix fluorescence spectroscopy, *Water Res.* 40 (2006) 1233–1239.
- [10] A. Baker, Fluorescence excitation–emission matrix characterization of some sewage-impacted rivers, *Environ. Sci. Technol.* 35 (2001) 948–953.
- [11] C.A. Stedmon, R. Bro, Characterizing dissolved organic matter fluorescence with parallel factor analysis: A tutorial, *Limnol. Oceanogr.: Methods* 6 (2008) 572–579.
- [12] M. Turk, A. Pentland, Eigenfaces for recognition, *J. Cognit. Neurosci.* 3 (1991) 71–86.
- [13] Z. Wang, Z. Wu, S. Tang, Characterization of dissolved organic matter in a submerged membrane bioreactor by using three-dimensional excitation and emission matrix fluorescence spectroscopy, *Water Res.* 43 (2009) 1533–1540.