Desalination and Water Treatment
www.deswater.com

43 (2012) 185–192
April

Taylor & Francis
Taylor & Francis Group

doi: 10.1080/19443994.2012.672169

External analysis-based fuzzy PLS model for prediction and monitoring in MBR

TaeSuk Oh, Hongbin Liu, MingJung Kim, SeungChul Lee, Min-Kyeong Yeo, ChangKyoo Yoo*

Department of Environmental Science and Engineering, Center for Environmental Studies, College of Engineering, Kyung Hee University, Seocheon-dong, Giheung-gu, Yongin-Si, Gyeonggi-Do 446-701, South Korea
Tel. +82 31 201 3824; Fax: +82 31 202 8854; email: ckyoo@khu.ac.kr

Received 25 December 2011; Accepted 10 February 2012

ABSTRACT

In general, the operation conditions of water treatment plants happen to be affected by external environmental variations such as temperature, viscosity, and loading changes. They sometimes result in bad treatment performance due to fouling or sludge decay and some process faults. Therefore, when designing a process model, negative effects of the external variables are needed to be incorporated. The purposes of this study are to propose a new external fuzzy partial least squares method (eFPLS) and apply it to predict the treatment performance of a pilot-scale membrane bioreactor (MBR). The proposed eFPLS model can represent an interpretability of the original FPLS of the inner and outer relationship with the viewpoint of physical meaning as well as keeping the capability of the original FPLS with handling the nonlinear correlation between inputs and outputs, while incorporating operation condition changes by the external analysis. It was used to predict the transmembrane pressure and the removal rates of chemical oxygen demand (COD) and total nitrogen in the MBR as well as to monitor the fouling progress. The prediction performance of the eFPLS model is compared with the other models of linear PLS and original FPLS. The results obtained in this study confirm that the eFPLS model with external analysis could improve not only the prediction efficiency but also the monitoring performance since it can efficiently remove the effects of external variables.

Keywords: Membrane bioreactor; Fuzzy partial least squares; External analysis; External fuzzy partial least squares; Partial least squares

1. Introduction

Partial least squares (PLS) method which captures the linear relationship between independent variables and response variables has proven to be a popular and effective approach to problems in many scientific and engineering fields [1]. The PLS model has some advantages such as collinearity removal, statistical interpretability, and graphical representation ability. But the algorithms of PLS have some limitations such as overfitting, nonlinearity, and so on [2]. Since many practical data are inherently nonlinear, it needs to be built a robust method that can model the nonlinear relation [3].

Recently, studies on the nonlinear PLS (NLPLS) models have been focused on the algorithm develop-

*Corresponding author.

The 4th International Desalination Workshop (IDW4), 16–18 November 2011, Jeju Island, South Korea

ments and applications. Baffi et al. [4] proposed neural networks PLS (NNPLS) that uses neural networks inner models. Centner et al. [5] developed locally weighted regression PLS (LWR-PLS) that uses linear PLS (LPLS) as a regression method to make a locally weighted model for every data. However, the results of several NLPLS models go through overfitting or local minima. To solve these problems, Bang et al. [3] proposed the fuzzy PLS (FPLS) method that combines the PLS method with the Takagi-Sugeno-Kang (TSK) fuzzy model. The FPLS is an effective general nonlinear regression modeling approach, since it can capture the nonlinearity while maintaining the good interpretability of LPLS.

On the other hand, the operation changes of feed flow rate and set-point change of controllers are expected to be external effects to the process and should be distinguished from process faults. To tackle the problem of external variables, several researches have been suggested in the literature [6–8]. Kano et al. [6] proposed a novel method for multimode process monitoring which is based on external analysis. Ge et al. [7] proposed a robust online monitoring approach based on nonlinear external analysis for monitoring multimode processes. Kim [8] applied external analysis to monitor the indoor air quality in the subway station. In this study, to remove external effect, the new prediction strategy of FPLS considering external information was proposed.

Note that the transmembrane pressure (TMP) in membrane bioreactor (MBR) is directly proportional to flux and the flux is strongly influenced by the temperature. The temperature has an effect on nitrification and denitrification efficiency of microorganism in the biological reaction, since microbial enzymes are easily susceptable by the variations of surrounding temperature. Hence, removal efficiencies of chemical oxygen demand (COD) and nitrogen are strongly influenced by surrounding temperature. Therefore, temperature as an external information should be incorporated when developing a prediction model of TMP, COD, and nitrogen removal rate. Also, the monitoring model needs to remove the effect of temperature when designing the monitoring model.

This study consists of two main components: (1) development of a new prediction model by combining FPLS and external analysis; (2) application of eFPLS to a pilot-scale MBR. The LPLS, original FPLS, external LPLS (eLPLS), and external FPLS (eFPLS) have been compared with the performances of the prediction and the monitoring.

The remaining parts of this paper are organized as follows. The first section introduces the basic theories and algorithms of FPLS and external analysis. Then, the eFPLS method is proposed in the materials and methods section. The results and discussion are presented with the MBR pilot plant. Finally, the conclusions are given.

2. External analysis

Changes in operating conditions which need to be distinguished from faults or malfunctions are assumed to be driven from the external effects [6]. Thus, the monitored variables can be classified into three groups: external variables, main variables, and other variables. The main variables are affected by external variables and other unmeasured variables. Hence, changes of external variables and their effects on main variables should be distinguished from faults. The main variables can be subdivided into two groups: the first group is explained by external variables and the second group is explained by the other variables [9]. The concept of external analysis is shown in Fig. 1.

In external analysis, data matrix X can be expressed as combination of main and external variables.

$$X = [G\,H] \tag{1}$$

where G is the matrix of external variables data and H is the matrix of main variables data. The main matrix H can be subdivided into two groups: GC, which is a group affected by external variables; and E, which is a group affected by other variables [6]. The regression coefficient matrix C can be determined as sum of squared errors.

$$C = (G^T G)^{-1} G^T H \tag{2}$$

$$E = H - GC \tag{3}$$

In this study, we focused on the error matrix E, which can be obtained by getting rid of the effect of changes of temperature as external variables in MBR process. Because the temperature influences microbial activity and TMP in MBR, the temperature should be removed for improving the quality of the prediction as well as the monitoring performance.

3. Fuzzy partial least square (FPLS)

Since many practical data are inherently nonlinear in the MBR process, there is a need for a nonlinear PLS modeling method that can attain the robust regression property of the LPLS method as well as represent any nonlinear relationship. Therefore, FPLS method as a nonlinear modeling method is proposed

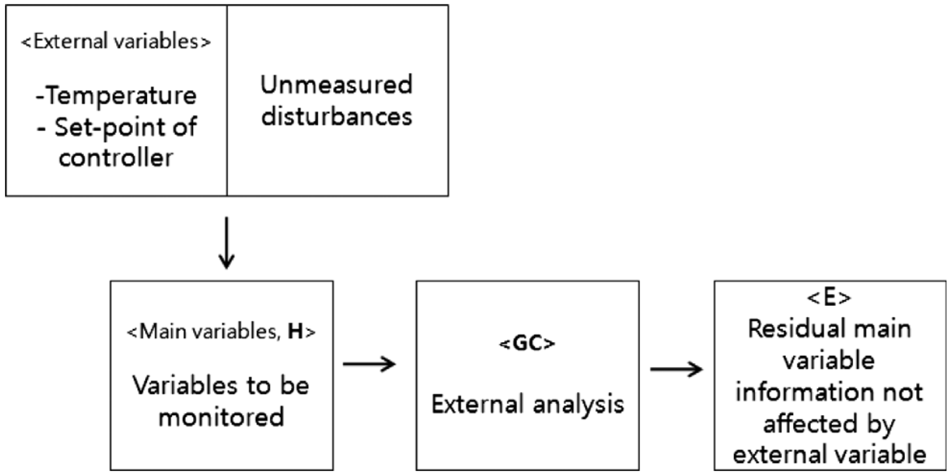

Fig. 1. The concept of external analysis.

in this study. The FPLS method combines the PLS method with the TSK fuzzy model. The PLS outer projection is used to decrease the dimension, and the TSK fuzzy inner model is used to take the nonlinearity in the projected latent space [3]. By using TSK fuzzy model, it also has robust nonlinear regression property and local interpretability. Fig. 2 shows a structure of the basic FPLS method. Score vectors (t_h and u_h) are presented to train the TSK fuzzy inner model $f_h(\cdot)$, which follows Eq. (4).

$$u_\mathrm{h} = f_\mathrm{h}(t_\mathrm{h}) + e_\mathrm{h} \tag{4}$$

where e_h represents the regression error. The components of $f_h(\cdot)$ need to be decided to minimize the regression error without overfitting.

4. Materials and methods

4.1. Proposed method

It is well known that the treatment performance of MBR is dependent on the biological reactions with nonlinear kinetics of microbiology and the temperature. The prediction model of MBR plants should be incorporated with the information of temperature effect on the microbiological reaction kinetics. A new external FPLS (eFPLS) which combines the FPLS method with external analysis is proposed in this paper. It can represent an interpretability of the original FPLS of the inner and outer relationship with the viewpoint of physical meaning and the capability of the original FPLS with handling the nonlinear correlation between inputs and outputs, while incorporating operation condition changes by the external analysis.

The integrated scheme the eFPLS model is shown in Fig. 3. First, the measured data of MBR process are collected and preprocessed. The second step is to detect for outliers missing value imputation. The next step is to select the external variable for external analysis in MBR process. TMP is strongly influenced by temperature as external information in MBR process. Hence, temperature is selected as external variable. After the temperature which is an external variable is removed from the

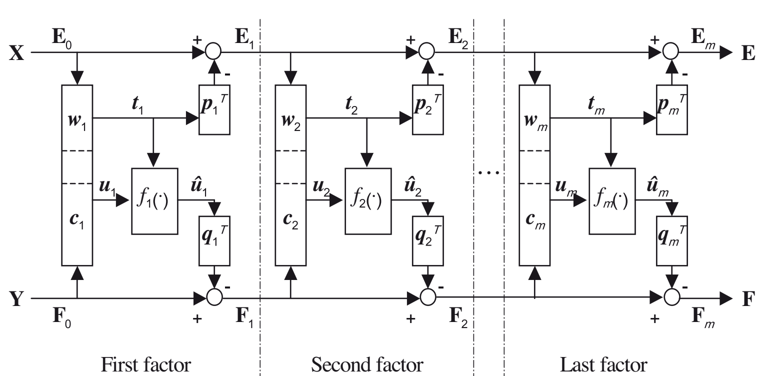

Fig. 2. The structure of FPLS method.

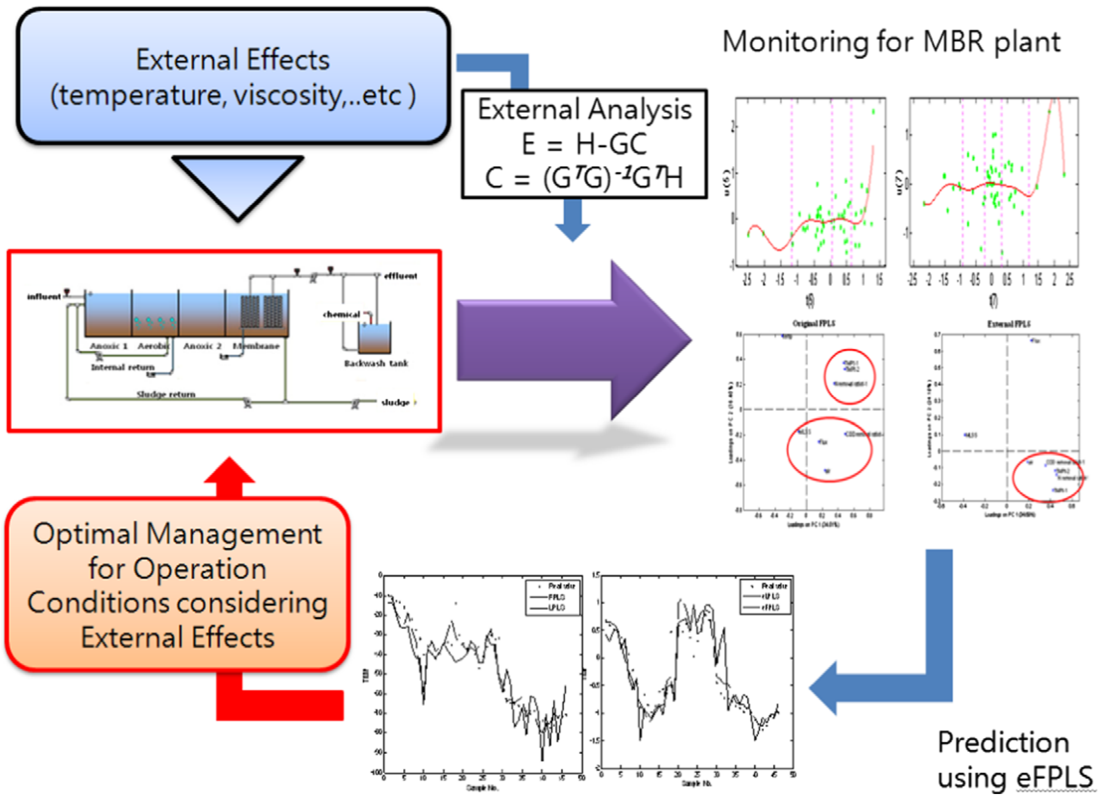

Fig. 3. The integrated scheme of the proposed eFPLS model.

overall process by external analysis, this idea is combined with FPLS and applied to modeling and monitoring as eFPLS in the MBR process. In this study, input variables were consisted of temperature, flux, air flow rate, and mixed liquor suspended solid (MLSS) and response variables were consisted of TMP, COD removal rate, and N removal rate. To raise the efficiency of prediction, input variables of the influent information that are in yesterday and the day before yesterday ($TMP_{t\text{-}1}$, $TMP_{t\text{-}2}$, COD removal rate and N removal rate on yesterday) are added. Then the response variables are predicted using the eFPLS model in the MBR process.

To check the prediction ability of the eFPLS model, eFPLS is compared with linear PLS, external PLS, and conventional FPLS. For the predictive accuracy of these comparisons, a root mean squared error (RMSE) is calculated in Eq. (5).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-1}} \tag{5}$$

where Y_i is the real value, $\hat{Y}_i$ is the predicted value, and n is the number of observations.

4.2. A pilot-scale MBR process

MBR process combines the biological degradation process with a direct solid–liquid separation by membrane filtration [10]. MBR offers major advantages over conventional activated sludge (CAS) system: a smaller footprint, less sludge production, and better effluent quality [11]. Hence, it becomes the important wastewater treatment technology.

MBR plant used in this study locates at Y-city, Korea. It consists of four reactors: anoxic 1, aerobic, anoxic 2, and a membrane bioreactor with capacities of 38, 63.8, 38, and 24.3 m^3, respectively (Fig. 4). This MBR plant has been operated at a constant-flux (or constant-flow rate) mode. The membrane flux of the pilot-scale MBR varied from 17.5 to 25.7 L/(m^2-h), with a suction cycle of nine minutes followed by one-minute relaxation (no suction). The operation temperature was from 15.5 to 25.6°C. Sludge retention time (SRT) was set to more than 9d, and TMP was kept within the period of 18 to 23 kPa. The periodic coarse bubble was supplied to the membrane to minimize the membrane fouling [12,13].

The samples were analyzed for biochemical oxygen demand (BOD), COD, total nitrogen (TN) and

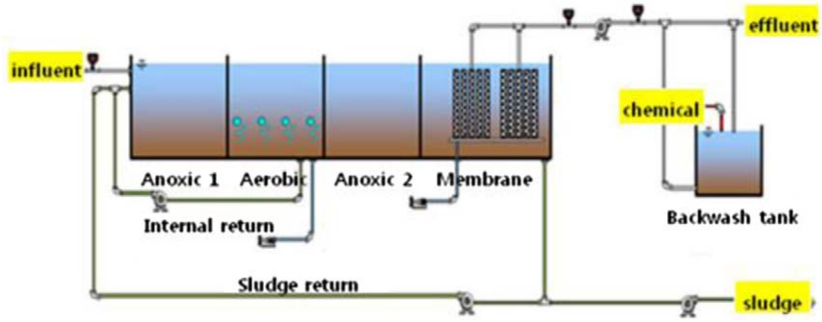

Fig. 4. Layout of a pilot-scale MBR plant [12].

Table 1
Influent conditions in a pilot-scale MBR process [12]

Measurements	Mean concentration
Flow	25 m^3/d
COD	301 g COD/m^3
BOD	166 g O_2/m^3
TN	38 g N/m^3
TP	6.4 g P/m^3

total phosphorus (TP), and total suspended solids (TSS). All influent concentration measurements of BOD, COD, TN, TP, and TSS were conducted gravimetrically in accordance with *Standard Methods* [14]. Table 1 shows the average influent concentrations of BOD, COD, TN, and TP for one year in Y-city.

5. Results and discussion

The total data set from an MBR pilot plant in Y-city, Korea is used from 1 December 2008 to 25 April 2009 with a number of 107 observations. In this study, the data-set is divided as period of cleaning in MBR plant, since TMP is changed by period of the chemical cleaning of the membrane. The first 61 observations are used to develop the prediction model as a training data-set which presents before the cleaning in the MBR plant. The remaining 46 observations after the chemical cleaning are used as test data-set to see the validation capability of the model.

5.1. External analysis

The preliminary study has been done to see the effect of external variable on the treatment performance. The effect of temperature which is an external variable in MBR plant was removed from the overall process. Fig. 5 shows the loading plots obtained for original data and treated data using external analysis and explains the correlation among the variables. Fig. 5(a) presents two clusters in the loading plot. The first cluster contains TMP_{t-1}, TMP_{t-2}, and one previous day's N removal rate. The second cluster is related to MLSS, flux, air flow rate, and one previous day's COD removal rate. On the other hand, there is only one cluster in the loading plot of external data as shown Fig. 5(b). The cluster in Fig. 5(b) contains air flow rate, COD removal rate on yesterday, TMP_{t-1}, TMP_{t-2}, and N removal rate on one day before. It means that the effect of temperature as the external variable was removed from the MBR process by applying external analysis. The effect of air flow rate on TMP, COD, and N removal rate was more significant rather than previous result. After the external analysis, we need to analyze the removed parts for better prediction model of MBR process.

5.2. External FPLS (eFPLS)

Fig. 6 shows the fuzzy PLS inner relation model in the third and fourth latent factors. The scatter points represent nonlinear trends in these plots. The linear PLS cannot cope with this situation which is presented as nonlinear data. However, the FPLS can give a direct and interactive way of treating such nonlinearities [3]. The score plots in third and fourth factor suggest that the inner relation may be caused by a combination of three trends. Hence, three fuzzy rules were chosen to model this relation. And score plots in other factors show the similar nonlinearity (not shown). To know the superiority of FPLS over linear-PLS (LPLS) in terms of nonlinear prediction ability, the predictive results obtained using FPLS and LPLS are compared. In addition, the prediction results obtained using external prediction methods (eFPLS and eLPLS) are compared with those obtained using FPLS and LPLS.

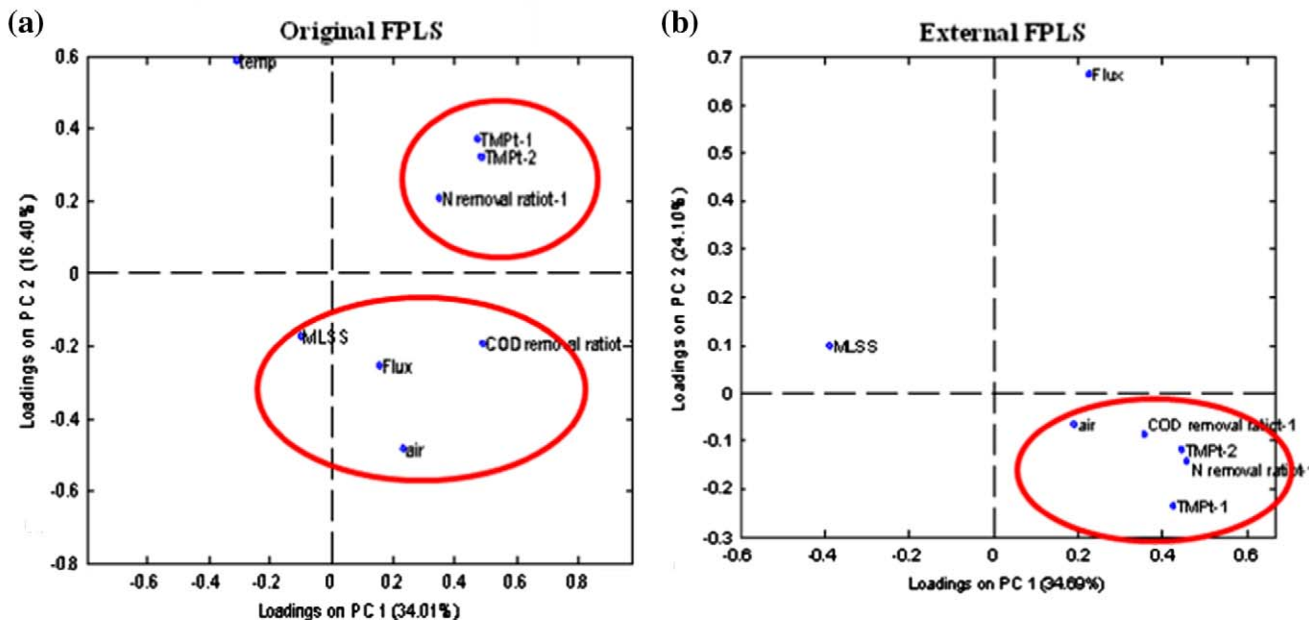

Fig. 5. Loading plots of (a) original data and (b) external data.

Fig. 7 shows the prediction results of TMP, COD, and *N* removal rate using eFPLS and eLPLS models. As shown in Fig. 7(a), the prediction performances of both eFPLS and eLPLS on TMP are quite good since temperature effect was removed from the original data-set by applying external analysis. In Fig. 7(b) and (c), prediction performances of COD and *N* removal rates are worse than those of TMP, since the flux and air flow rate appropriate variables for predicting COD and *N* removal rates among the used X variables are not. On the other hand, other independent variables of flux and air flow rate cannot predict the removal rate of COD and nitrogen. COD and *N* removal rate could not be predicted accurately by eFPLS.

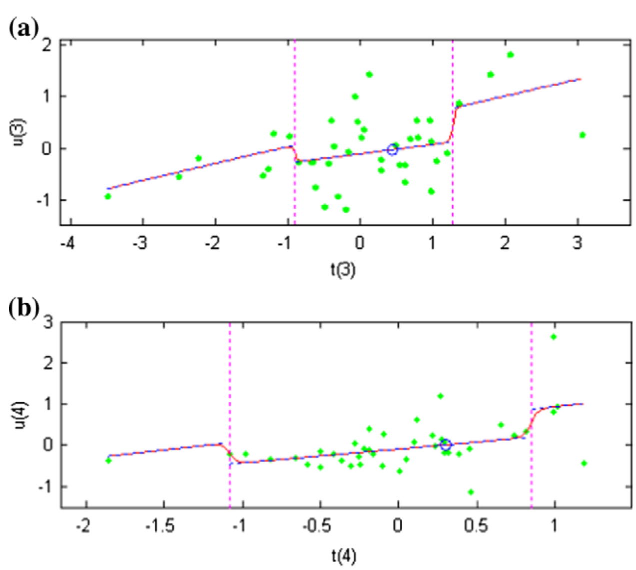

Fig. 6. FPLS scores of (a) third and (b) fourth latent factor.

The RMSE values of TMP, COD, and nitrogen removal rates obtained using LPLS, FPLS, eLPLS and eFPLS are compared in Table 2. It represents that four models have significantly different RMSE values. First, to model TMP, COD, and *N* removal rate, the RMSE values obtained LPLS are 0.46, 0.82, and 0.94, respectively; those obtained using FPLS are 0.32, 0.66, and 0.86, respectively; those obtained using eLPLS are 0.40, 0.73, and 0.90, respectively; those obtained using eFPLS are 0.28, 0.63, and 0.79, respectively. As summarizing these results, it confirms that the FPLS and eFPLS models could predict more accurately than the LPLS and eLPLS, since the non-linear relationship was embedded in most independent variables. Note that the TMP is directly proportional to flux in MBR process and the flux is strongly influenced by external temperature. The COD and *N* removal rates are also strongly influenced by temperature, since the temperature influences on nitrification and denitrification efficiency of microorganism in the biological reaction [13]. Hence, when the effects of external variable were removed by external analysis, it is clear that eLPLS and eFPLS prediction models have for predicting the response variables (especially TMP) superiority over LPLS and FPLS models. On the other hand, it represents that both LPLS and eLPLS models could not predict more

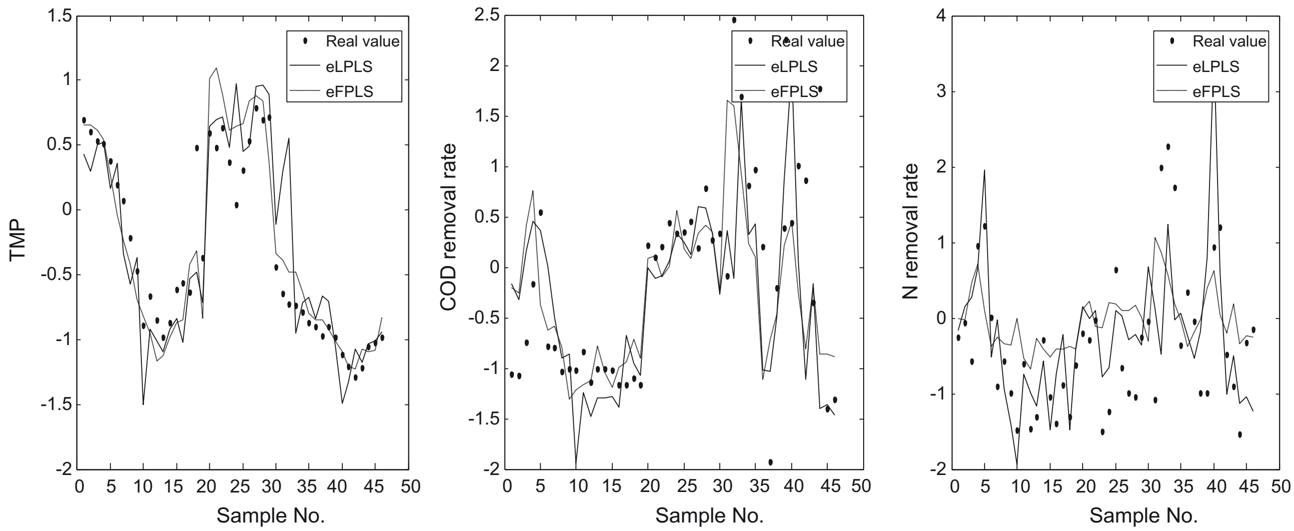

Fig. 7. Prediction performance of eFPLS and eLPLS for (a) TMP, (b) COD removal rate, and (c) *N* removal rate.

Table 2
RMSEs of LPLS, FPLS, eLPLS, and eFPLS

RMSE	LPLS	FPLS	eLPLS	eFPLS
TMP (kPa)	0.46	0.32	0.40	0.28
COD removal rate (g COD/m^3)	0.82	0.66	0.73	0.63
N removal rate (g N/m^3)	0.94	0.86	0.90	0.79

accurately than the original FPLS model, even though external analysis was carried out. Therefore, it confirms that FPLS can predict well for nonlinear data, where MBR data have strong nonlinearity.

6. Conclusions

The eFPLS model was suggested to monitor the effect of temperature on biological process reaction as well as to predict the TMP and the treatment performance in an MBR process. The results in a pilot-scale MBR plant show that the proposed eFPLS model could improve the prediction accuracy as well as the monitoring accuracy of MBR process than the original LPLS and FPLS. It confirms that external variable of wastewater temperature influences the operation change and treatment performance of MBR. Moreover, our ongoing research is focused to develop a new dynamic nonlinear external model, which uses the dynamic external analysis model for external variable modeling purposes and the nonlinear neural network model for monitoring unexpected process faults.

Acknowledgments

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (KRF-2009-0076129) and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2012-0000609).

Nomenclature

BOD	—	biochemical oxygen demand, g O_2/m^3
C	—	regression coefficient matrix
CAS	—	conventional activated sludge system
COD	—	chemical oxygen demand, g COD/m^3
e_h	—	regression error
E	—	error matrix
eFPLS	—	external fuzzy partial least squares
eLPLS	—	external linear partial least squares
$f_h(\cdot)$	—	inner TSK fuzzy model
FPLS	—	fuzzy partial least squares
G	—	matrix of external variables data
GC	—	group affected by external variables
H	—	matrix of main variables data
LPLS	—	linear partial least squares
LWRPLS	—	locally weighted regression partial least squares
MBR	—	membrane bioreactor
MLSS	—	mixed liquor suspended solid, g /m^3
n	—	number of observations
TN	—	total nitrogen, g N/m^3

NLPLS	—	nonlinear partial least squares
NNPLS	—	neural networks partial least squares
PLS	—	partial least squares
RMSE	—	root mean squared error
SRT	—	sludge retention time, d
t_{h}	—	input score vector
TMP	—	transmembrane pressure, kPa
TP	—	total phosphorus, g P/m^3
TSK	—	Takagi-Sugeno-Kang
TSS	—	total suspended solids, g /m^3
u_{h}	—	output score vector
X	—	data matrix
Y_{i}	—	real value
$\hat{Y}_{\mathrm{i}}$	—	predicted value

References

[1] K.P. Bennett, M.J. Embrechts, An optimization perspective on Kernel partial least quares regression, Comput. Syst. Sci. 190 (2003) 227–250.
[2] S. Wold, Nonlinear partial least squares modeling II, Spline inner relation, Chemometr. Intel. Lab. 14 (1992) 71–84.
[3] Y.H. Bang, C.K. Yoo, I.B. Lee, Nonlinear PLS modeling with fuzzy inference system, Chemometr. Intel. Lab. 64 (2003) 137–155.
[4] G. Baffi, E.B. Martin, A.J. Morris, Non-linear projection to latent structures revisited (the neural network PLS algorithm), Comput. Chem. Eng. 23 (1999) 1293–1307.
[5] V. Centner, D.L. Massart, Optimization in locally weighted regression, Anal. Chem. 70 (1998) 4206–4211.
[6] M. Kano, S. Hasebe, I. Hashimoto, H. Ohno, Evolution of multivariate statistical process control: Application of independent component analysis and external analysis, Comput. Chem. Eng. 28 (2004) 1157–1166.
[7] Z. Ge, C. Yang, Z. Song, H. Wang, Robust online monitoring for multimode processes based on nonlinear external analysis, Ind. Eng. Chem. Res. 47 (2008) 4775–4783.
[8] Y.S. Kim, Data analysis and development of prediction model for indoor air quality in subway stations, based on multivariate statistical approach, MSc thesis, Kyung Hee Univ., Yongin, Kyunggi, Korea, 2010.
[9] M.J. Kim, B.S. Rao, O.Y. Kang, J.T. Kim, C.K. Yoo, Monitoring and prediction of indoor air quality (IAQ) in subway or metro systems using season dependent models, Energ. Build. 46 (2012) 48–55.
[10] L.C. Pierre, C. Vicki, A.G.F. Tony, Fouling in membrane bioreactors used in wastewater treatment, J. Membrane Sci. 284 (2006) 17–53.
[11] Z. Wang, Z. Wu, X. Yin, L. Tian, Membrane fouling in a submerged membrane bioreactor (MBR) under sub-critical flux operation: Membrane foulant and gel layer characterization, J. Membrane Sci. 325 (2008) 238–244.
[12] M.J. Kim, O.Y. Kang, B.S. Rao, J.R. Kim, H.J. Hwang, M.H. Kim, C.K. Yoo, Proposing a new fouling index in a membrane bioreactor (MBR) based on mechanistic fouling model, Desalin. Water Treat. 33 (2011) 209–217.
[13] S.Y. Park, S.J. Kim, Y. Rhee, S.S. Yum, T.D. Kwon, T.K. Lee, Genotoxic and neurotoxic potential in marine fishes exposed to sewage effluent from a wastewater treatment plant, Mol. Cell. Toxicol. 5(3) (2009) 265–271.
[14] APHA, AWWA, and WEF, Standard Methods for the Examination of Water and Wastewater, American Public Health Association, Washington, DC, 2005.