Journal of Hazardous Materials 161 (2009) 538–544

Contents lists available at ScienceDirect

Journal of Hazardous Materials

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

On-line estimation of key process variables based on kernel partial least squares in an industrial cokes wastewater treatment plant

Seung Han Woo [a], Che Ok Jeon [b], Yeoung-Sang Yun [c], Hyeoksun Choi [d], Chang-Soo Lee [e], Dae Sung Lee [f,*]

[a] Department of Chemical Engineering, Hanbat National University, Daejeon 305-719, Republic of Korea
[b] Department of Life Science, Chung-Ang University, Seoul 156-756, Republic of Korea
[c] Division of Environmental and Chemical Engineering and Research Institute of Industrial Technology, Chonbuk National University, Chonbuk 561-756, Republic of Korea
[d] School of Environmental Science and Engineering, Pohang University of Science and Technology, Gyeongbuk 790-784, Republic of Korea
[e] Division of Architecture, Uiduk University, Gyeongbuk 790-784, Republic of Korea
[f] Department of Environmental Engineering, Kyungpook National University, Daegu 702-701, Republic of Korea

ARTICLE INFO

Article history:
Received 8 February 2008
Received in revised form 31 March 2008
Accepted 1 April 2008
Available online 6 April 2008

Keywords:
Kernel-based algorithm
Industrial wastewater treatment plant
Partial least squares
Nonlinearity measure
On-line estimation

ABSTRACT

A kernel-based algorithm is potentially very efficient for predicting key quality variables of nonlinear chemical and biological processes by mapping an original input space into a high-dimensional feature space. Nonlinear data structure in the original space is most likely to be linear at the high-dimensional feature space. In this work, kernel partial least squares (PLS) was applied to predict inferentially key process variables in an industrial cokes wastewater treatment plant. The primary motive was to give operators and process engineers a reliable and accurate estimation of key process variables such as chemical oxygen demand, total nitrogen, and cyanides concentrations in real time. This would allow them to arrive at the optimum operational strategy in an early stage and minimize damage to the operating units as shock loadings of toxic compounds in the influent often cause process instability. The proposed kernel-based algorithm could effectively capture the nonlinear relationship in the process variables and show far better performance in prediction of the quality variables compared to the conventional linear PLS and other nonlinear PLS method.

1. Introduction

Cokes wastewater is generated at coal coking, coal gas purification, and by-product recovery operations in an integrated steel-making plant [1,2]. The wastewater consists of ammonia, cyanide, thiocyanate, sulfides, and a wide variety of complex hydrocarbons such as phenolics, polynuclear aromatic hydrocarbons, and various heterocyclic compounds [3,4]. Most of the chemical oxygen demand (COD) originates from phenol, which is a toxic inhibitory substrate but is also a useful carbon source for acclimatized microorganisms [5]. Cyanide is a highly toxic organic compound even at low concentrations and its presence in aqueous media is severely restricted by regulations [6,7].

Conventional treatment of high-strength cokes wastewater comprises expensive caustic treatment and steam stripping to reduce the pollutant loads, followed by biological treatment [8,9]. Among the proven biological treatment methods for cokes wastewater, activated sludge processes have been widely used [10–12]. However, these treatment systems exhibit varying degrees of performance efficiency. In particular, a single-sludge process with recycle of nitrified effluent, *i.e.* pre-denitrification process, has been preferred in Korea, because of its simplicity and economic benefits [3,8,13]. The pre-denitrification process consists of two distinct microbial reactions under anoxic followed by aerobic conditions. In the anoxic condition, heterotrophic denitrifying bacteria convert nitrate into nitrogen gas using phenols as a carbon source, thus most of phenols are removed in this step. Besides, very toxic free cyanide can be removed in some degree by anaerobes. In spite of its acute toxicity, it has been reported that various anaerobes, which are acclimated indigenous microorganisms in biological wastewater treatment plants, can degrade cyanide compounds [14]. In the aerobic condition, autotrophic nitrifying bacteria convert ammonia into nitrate, while autotrophic thiocyanate-degrading bacteria convert thiocyanate into ammonia, sulfate, and bicarbonate [15]. These successive microbial reactions can remove most of toxic compounds within the cokes wastewater. However, a full-scale pre-denitrification process has been occasionally unstable due to inhibitory effects of toxic compounds on nitrification which is commonly the rate-limiting step of the overall nitrogen removal. In particular, the increased loading of toxic compounds such as phenol

* Corresponding author. Tel.: +82 53 950 7286; fax: +82 53 950 6579.
E-mail address: daesung@knu.ac.kr (D.S. Lee).

0304-3894/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jhazmat.2008.04.004

and free cyanide often caused a failure in nitrogen removal. Furthermore, the recovery of nitrification after an inhibitory event takes very long time (weeks), leaving the treatment system vulnerable to permit violations and the downstream environment vulnerable to ecological damage.

With increasingly stringent regulations of effluent quality, process monitoring and control have become important to enhance process performance by detecting disturbances leading to abnormal process operation in an early stage and to cope with influent variations that are typical of the cokes wastewater treatment plant. However, the lack of suitable on-line sensors for monitoring key process variables such as COD, total nitrogen (TN), and cyanides concentrations limits the effective control of effluent quality. Although these key process variables can be measured by laboratory analyses, a significant time delay in the range of a few hours is usually unavoidable. It is normally too late to achieve well-timed adaptive process control accommodating influent fluctuation and other disturbances. To overcome these problems, inferential sensors, as software sensors, can be developed to estimate hard-to-measure process variables from other easily measurable process variables and historical operation data. In general, a structured process model developed from the information of process behaviors has been considered to be the most effective way of simulating and predicting processes [16,17]. But establishing a structured model is a formidable task in biological wastewater treatment plants as a multitude of microbial reactions coupled with environmental interactions is normally nonlinear, time-variable, and still uncertain.

Partial least squares (PLS) is a projection method for analyzing a historical reference distribution of the measurement trajectories from past successful operations in a reduced latent vector space and comparing the behaviors of new operations to this reference distribution. However, when PLS is applied to chemical and biological processes, there have been some difficulties in its practical applications. Although it is an intrinsically linear method in the basic form, most real problems are inherently nonlinear. The minor latent variables from linear PLS models cannot always be discarded as they may not only describe noise or negligible variance-covariance structures in the data, but they may also encapsulate significant information about the nonlinear nature of the problem [18]. A number of methods have been proposed to integrate nonlinear features within the linear PLS framework. A quadratic PLS method was proposed to fit the functional relationship between each pair of latent scores by quadratic regression [19]. Neural networks were also incorporated into linear PLS to identify the relationship between the input and the output scores, while retaining the outer mapping framework of the linear PLS algorithm [20,21]. In recent years, nonlinear kernel-based algorithms as kernel partial least squares (KPLS) have been proposed [22,23]. The basic idea of KPLS is first to map each point in an original data space into a feature space via nonlinear mapping and then to develop a linear PLS model in the mapped space. According to Cover's theorem, nonlinear data structure in the original space is most likely to be linear after high-dimensional nonlinear mapping [24]. Therefore, KPLS can efficiently compute latent variables in the feature space by means of integral operators and nonlinear kernel functions. Compared to other nonlinear methods, the main advantage of the kernel-based algorithm is that it does not involve nonlinear optimization. It essentially requires only linear algebra, making it as simple as the conventional linear PLS. In addition, because of its ability to use different kernel functions, KPLS can handle a wide range of nonlinearities.

In the application investigated here, different PLS modeling approaches are employed to develop on-line estimation of the key process variables in minimal time and with minimal cost. The only information required to infer the key process variables is the historical data collected from the past successful operations and easily measurable on-line sensor values such as volumetric flowrate, dissolved oxygen, pH, etc. The primary motive is to give operators and process engineers a guideline that would allow them to arrive at the optimum operational strategy and minimize damage to the operating unit as shock loadings of toxic compounds such as phenol and cyanide in the influent often cause process instability, which can lead to the death of the effective microorganisms and leakage of organic carbon. The effectiveness of the kernel-based method is demonstrated by modeling capabilities based on its performance characteristics and prediction accuracy compared with the conventional linear PLS and other nonlinear PLS methods.

2. Materials and methods

2.1. Industrial cokes wastewater treatment plant

The cokes wastewater treatment plant (CWTP) at the steelmaking company in Korea is a conventional activated sludge unit as shown in Fig. 1. It was designed for the removal of toxic organic pollutants from the cokes-making plant. Since a high concentration of nitrogen compounds was found inhibitory to biodegradation, pretreatment steps such as ammonia stripping were employed to render the wastewater more amendable to biodegradation. To alleviate the impact of high concentrations of deleterious substances on the biological treatment, an equalization tank was installed after the preliminary treatment stage and before the aeration tanks of the activated sludge process. The hydraulic retention time of the CWTP was approximately 2.7 d. Oxygen was supplied by submerged mechanical aerators. In the aerobic zone, the nitrification was carried out, and the nitrate produced in this zone was recycled with the mixed liquor to the first anoxic tank. Concentrated sludge from the bottom of the settler was split into two streams: the first was recycled to the beginning of the first anoxic tank and the other was treated in view of incineration of the waste sludge. The effluent from the settler was passed through chemical treatment units to remove hazardous heavy metal ions and to reduce the level of suspended solids and organic matter. Operational data of almost 5 months were collected at 8-h intervals (k). All samples were analyzed for mixed liquor suspended solids (MLSS), COD, TN, total cyanides, and phenol according to the Standard Methods [25]. DO, pH, oxidation reduction potential (ORP), and volumetric flow rates were also measured at each sampling time. This measuring campaign resulted in 420 operational data sets in total. All the measurement variables are automatically stored at a database by a data acquisition system. All variables used in model identification are shown in Table 1. The predictor matrix **X** consists of 27 measurement variables either at time $k-1$ or at time k. The response matrix **Y** consists of the key process variables as COD, TN, and total cyanides in the effluent at time k. These three variables were selected as they are key indicators for performance of the CWTP. The first 200 sets of data were used for training and the remaining 220 data sets for the validation of the developed models.

2.2. Partial least squares

PLS is a linear multivariate method for relating the process variables **X** with responses **Y**. PLS can analyze data with strongly collinear, noisy, and numerous variables in both **X** and **Y** [26]. PLS reduces the dimension of the predictor variables by extracting factors or latent variables that are correlated with **Y** while capturing a large amount of the variations in **X**. This means that PLS maximizes the covariance between matrices **X** and **Y**.

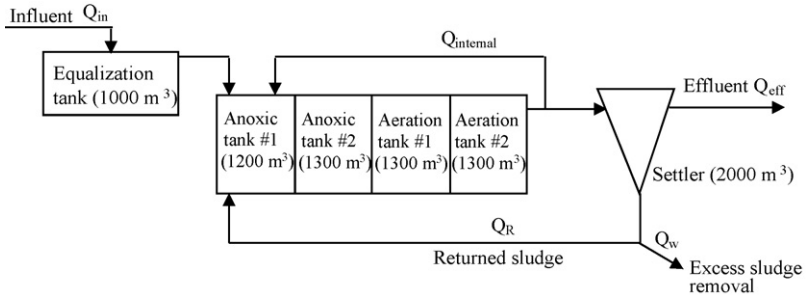

Fig. 1. Schematic diagram of the cokes wastewater treatment plant.

In PLS, the scaled matrices **X** and **Y** are decomposed into score vectors (**t** and **u**), loading vectors (**p** and **q**), and residual error matrices (**E** and **F**):

$$\mathbf{X} = \sum_{i=1}^{a} \mathbf{t}_i \mathbf{p}_i^{\mathrm{T}} + \mathbf{E}$$
$$\mathbf{Y} = \sum_{i=1}^{a} \mathbf{u}_i \mathbf{q}_i^{\mathrm{T}} + \mathbf{F} \quad (1)$$

where a is the number of latent variables. In an inner relation, the score vector **t** is linearly regressed against the score vector **u**.

$$\mathbf{u}_i = b_i \mathbf{t}_i + h_i \quad (2)$$

where b is a regression coefficient that is determined by minimizing the residual h. It is crucial to determine the optimal number of latent variables and cross-validation is a practical and reliable way to test the predictive significance of each PLS component. There are several algorithms to calculate the PLS model parameters. In this work, the NIPALS algorithm was used with the exchange of scores [27].

2.3. Neural network partial least squares

To capture nonlinear structures between the predictor block and the responses, the PLS model can be extended to nonlinear PLS models [18]. Neural network PLS (NNPLS) is an integration of neural networks with PLS to model nonlinear processes with input collinearity [20]. The input and output variables are projected onto the latent space to remove collinearity, and then each latent variable pair is mapped with a single-input single-output (SISO) neural network as follows:

$$\mathbf{u}_i = F(\mathbf{t}_i) + \boldsymbol{\nu}_i \quad (3)$$

where $F(\cdot)$ stands for the inner relation represented by a neural network and $\boldsymbol{\nu}$ is the residuals. The neural network is trained to capture the nonlinearity in the projected latent space. The major advantage of NNPLS method is that it decomposes a multivariate regression problem into a number of univariate regressors so that it can circumvent the over-parameterization problem. In this application, a feed-forward back-propagation neural network (FBNN) with sigmoidal functions was used to identify the nonlinear inner regression model.

Table 1
Process variables in the cokes wastewater treatment plant

Variable	Symbol	Unit	Description
Predictor (**X**)	Q_{in}	m^3/h	Volumetric flowrate in the influent at time $k-1$
	CN_{in}	mg/L	Total cyanides concentration in the influent at time $k-1$
	SCN_{in}	mg/L	Thiocyanate in the influent at time $k-1$
	COD_{in}	mg/L	COD in the influent at time $k-1$
	$Phenol_{in}$	mg/L	Phenol concentration in the influent at time $k-1$
	$NH_4{}^+{}_{in}$	mg/L	Ammonium concentration in the influent at time $k-1$
	TN_{in}	mg/L	Total nitrogen in the influent at time $k-1$
	pH_{in}	–	pH in the influent at time k
	$Temp_{in}$	°C	Temperature in the influent at time k
	pH_{anoxic}	–	pH in the anoxic tank at time k
	ORP_{anoxic}	mV	ORP in the anoxic tank at time k
	$Temp_{anoxic}$	°C	Temperature in the anoxic tank at time k
	$pH_{aerobic}$	–	pH in the aerobic tank at time k
	$ORP_{aerobic}$	mV	ORP in the aerobic tank at time k
	$Temp_{aerobic}$	°C	Temperature in the aerobic tank at time k
	$DO_{aerobic}$	mg/L	DO in the aerobic tank at time k
	$MLSS_{aerobic}$	mg/L	MLSS concentration at time $k-1$
	Q_R	m^3/h	Returned sludge flowrate at time k
	Q_w	m^3/h	Waste sludge volumetric flowrate at time k
	$Q_{internal}$	m^3/h	Internal recycle flowrate at time k
	pH_{eff}	–	pH in the effluent at time k
	ORP_{eff}	mV	ORP in the effluent at time k
	$Temp_{eff}$	°C	Temperature in the effluent at time k
	SVI	mL/g	Sludge volume index at time $k-1$
	COD_s	mg/L	COD in the settler at time $k-1$
	TN_s	mg/L	TN in the settler at time $k-1$
	CN_s	mg/L	Total cyanides concentration in the settler at time $k-1$
Response (**Y**)	COD_{eff}	mg/L	COD in the effluent at time k
	TN_{eff}	mg/L	TN in the effluent at time k
	CN_{eff}	mg/L	Total cyanides concentration in the effluent at time k

2.4. Kernel partial least squares

The KPLS method is based on mapping of the original input data into a high-dimensional feature space $\Im$ where a linear PLS model is created. By nonlinear mapping $\Phi: \mathbf{x} \in \Re^n \rightarrow \Phi(\mathbf{x}) \in \Im$, a KPLS algorithm can be derived from a sequence of NIPALS steps and has the following formulation [23].

1. Initialize score vector $\mathbf{w}$ as equal to any column of $\mathbf{Y}$.
2. Calculate scores $\mathbf{u} = \mathbf{\Phi}\mathbf{\Phi}^T\mathbf{w}$ and normalize $\mathbf{u}$ to $||\mathbf{u}|| = 1$, where $\mathbf{\Phi}$ is a matrix of regressors.
3. Regress columns of $\mathbf{Y}$ on $\mathbf{u}$: $\mathbf{c} = \mathbf{Y}^T\mathbf{u}$, where $\mathbf{c}$ is a weight vector.
4. Calculate a new score vector $\mathbf{w}$ for $\mathbf{Y}$: $\mathbf{w} = \mathbf{Yc}$ and then normalize $\mathbf{w}$ to $||\mathbf{w}|| = 1$.
5. Repeat steps 2–4 until convergence of $\mathbf{w}$.
6. Deflate $\mathbf{\Phi}\mathbf{\Phi}^T$ and $\mathbf{Y}$ matrices:

$$\mathbf{\Phi}\mathbf{\Phi}^T = (\mathbf{\Phi} - \mathbf{u}\mathbf{u}^T\mathbf{\Phi})(\mathbf{\Phi} - \mathbf{u}\mathbf{u}^T\mathbf{\Phi})^T \quad (4)$$

$$\mathbf{Y} = \mathbf{Y} - \mathbf{u}\mathbf{u}^T\mathbf{Y} \quad (5)$$

7. Go to step 1 to calculate the next latent variable.

Without explicitly mapping into the high-dimensional feature space, a kernel function can be used to compute the dot products as follows:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T\Phi(\mathbf{x}_j) \quad (6)$$

$\mathbf{\Phi}\mathbf{\Phi}^T$ represents the $(n \times n)$ kernel Gram matrix $\mathbf{K}$ of the cross dot products between all mapped input data points $\Phi(x_i)$, $i = 1, \ldots, n$. The deflation of the $\mathbf{\Phi}\mathbf{\Phi}^T = \mathbf{K}$ matrix after extraction of the $\mathbf{u}$ components is then given by:

$$\mathbf{K} = (\mathbf{I} - \mathbf{u}\mathbf{u}^T)\mathbf{K}(\mathbf{I} - \mathbf{u}\mathbf{u}^T) \quad (7)$$

where $\mathbf{I}$ is an m-dimensional identity matrix. Taking into account normalized scores $\mathbf{u}$, the prediction of the KPLS model on training data $\hat{\mathbf{Y}}$ is defined as:

$$\hat{\mathbf{Y}} = \mathbf{KW}(\mathbf{U}^T\mathbf{KW})^{-1}\mathbf{U}^T\mathbf{Y} = \mathbf{U}\mathbf{U}^T\mathbf{Y} \quad (8)$$

For predictions on new observation data $\hat{\mathbf{Y}}_t$, the regression can be written as:

$$\hat{\mathbf{Y}}_t = \mathbf{K}_t\mathbf{W}(\mathbf{U}^T\mathbf{KW})^{-1}\mathbf{U}^T\mathbf{Y} \quad (9)$$

where $\mathbf{K}_t$ is the test matrix whose elements are $K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j)$ where $\mathbf{x}_i$ and $\mathbf{x}_j$ present the validation and training data points, respectively.

3. Results and discussion

PLS modeling approaches were employed to develop an inferential prediction model in minimal time and with minimal cost. Three different PLS modeling strategies were applied to the CWTP. The capabilities of the PLS modeling approaches were assessed through their prediction accuracy and performance characteristics. The performance of each model was evaluated in terms of the root-mean-square-error (RMSE) criterion. The RMSE performance index was defined as:

$$\text{RMSE} = \sqrt{\sum (\hat{y} - y)^2/m} \quad (10)$$

where y presents the measured values, $\hat{y}$ presents the corresponding predicted values and m is the number of observations.

Table 2
Comparison of the models' performance

Model	$\text{RMSE}_{\text{Training}}$	$\text{RMSE}_{\text{Validation}}$	BIC
Linear PLS	8.427	16.815	−6577
NNPLS	6.825	15.750	−5823
KPLS	6.646	13.259	−5579

3.1. Linear partial least squares model

First, a linear PLS model was built between the predictor variables $\mathbf{X}$ and the response variables $\mathbf{Y}$. The objective was to determine how well the linear model works and to compare the results to those of the nonlinear models later. On the basis of the cross-validation results, five latent variables were included in the model. This explained 39.94% of the variance of matrix $\mathbf{X}$ and 69.96% of matrix $\mathbf{Y}$. The RMSE values for the training and validation data sets were 8.427 and 16.815, respectively (Table 2). The simulation results of the linear PLS model are given in Fig. 2. This model predicted the dynamics of the wastewater treatment process with a relatively good accuracy for the calibration data set, but there was a significant mismatch between the model prediction and actual plant data in the later part of COD, and almost all TN and total cyanides profiles in the validation data set. In particular, it showed a bias in the prediction of the validation data set of total cyanides. This exemplified the weakness of the linear multivariate regression model. It indicates that the linear PLS model could not adequately describe the CWTP that is inherently nonlinear and exposed to various disturbances such as influent composition variations, temperature changes, and equipment defects.

A nonlinearity measure for PLS model was calculated to assess the nonlinearity in data [28]. The operation data sets were divided into four disjunct regions of 105 samples each. The process variables were normalized with respect to the mean and variance of the regions for which the accuracy bounds were computed. Then linear PLS models were developed for each of these disjunct regions where the accuracy bounds for the sum of the discarded eigenvalues were

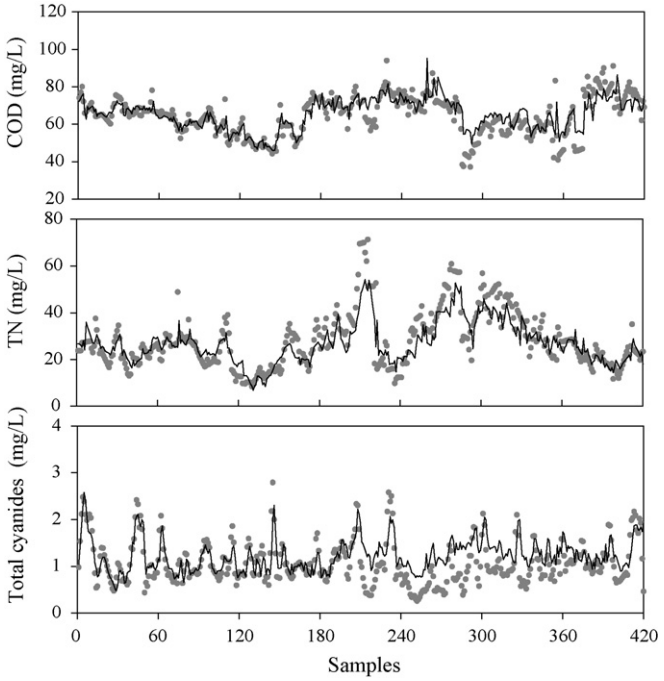

Fig. 2. Prediction results of the linear PLS (grey dot: measured values, solid line: predicted values).

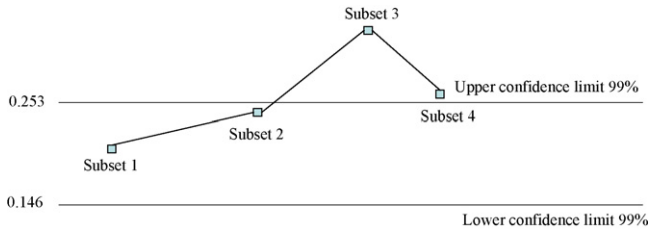

Fig. 3. Graphical representation of nonlinearity measure for accuracy bounds for first of four disjunct regions.

obtained. Fig. 3 is a graphical representation of the case where the accuracy bounds were obtained for the first disjunct region. While the sum of the discarded eigenvalues for the first disjunct region was inside the accuracy bounds, those for other regions fell outside these bounds. This implied that the error variance of the linear PLS model residuals was larger than could be explained by the uncertainty in determining the correlation matrix and hence a nonlinear model is required for the CWTP.

3.2. Neural network partial least squares model

In the application of the NNPLS to the CWPT, a FBNN with sigmoidal functions was used to identify the nonlinear inner regression. The simplified cross-validation was used to determine the optimal number of factors [29]. Each factor was modeled using a SISO network. The neural network was trained to capture the nonlinearity in the projected latent space using a conjugate gradient optimization. The number of hidden units was also determined automatically by simplified cross-validation. The same data was also used to compare the linear PLS method with the NNPLS model. Five latent variables were included into the NNPLS model, which then explained 52.74% of the variance of matrix **X** and 74.03% of matrix **Y**. The RMSE values for the training and validation data sets were 6.825 and 15.750, respectively (Table 2). It shows that the NNPLS gives better prediction results than the linear PLS model. The simulation results of the NNPLS model are given in Fig. 4. Fig. 5 shows the first principal inner relation between the **X** and **Y** blocks from both the linear PLS and NNPLS models. As shown in Fig. 5, the linear PLS regression gives a best linear least-squares model, but the NNPLS model captures the system's nonlinearity and thus outperforms the linear PLS model.

3.3. Kernel partial least squares model

A KPLS model was then developed with a radial basis kernel function. In this application, the radial basis function was determined as the best one from the RMSE value of the validation data set. Both the width parameter σ and the number of principal components in the feature space should be determined to optimize the KPLS model. For this purpose, a three-dimensional response surface, based on the RMSE value, was generated to study the interaction between the two parameters as shown in Fig. 6. The surface plot shows a clear minimum point when the number of principal components and the width of the kernel function were 9 and 13, respectively. These values were used as the optimal parameters in the KPLS model. The developed KPLS model explained 86.52% of the variance of matrix **Y**. The captured variance of matrix **X** could not be calculated because it is impossible to find an inverse mapping function from the feature space to the original space. The RMSE values for the training and validation data sets were 6.646 and 13.259, respectively (Table 2). Based on the RMSE values, the KPLS gave the best prediction performance compared to the linear and nonlinear PLS methods. This clearly indicates that KPLS benefitting from the

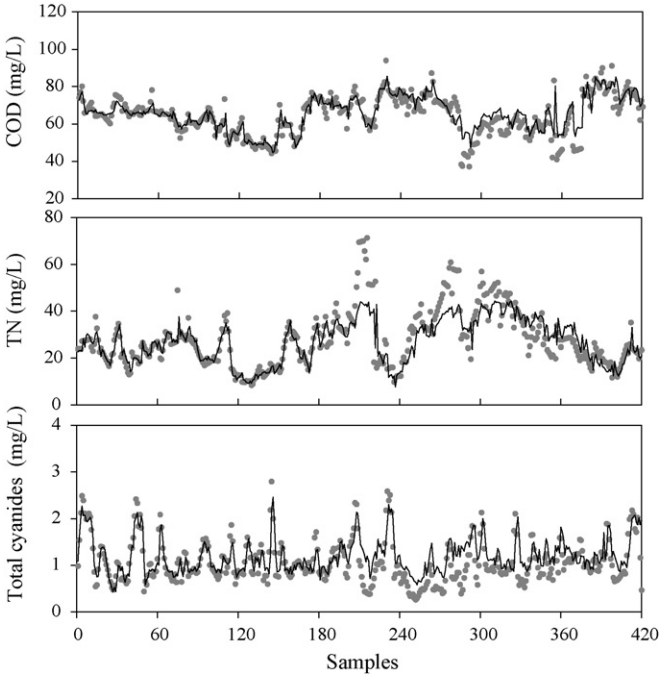

Fig. 4. Prediction results of the NNPLS (grey dot: measured values, solid line: predicted values).

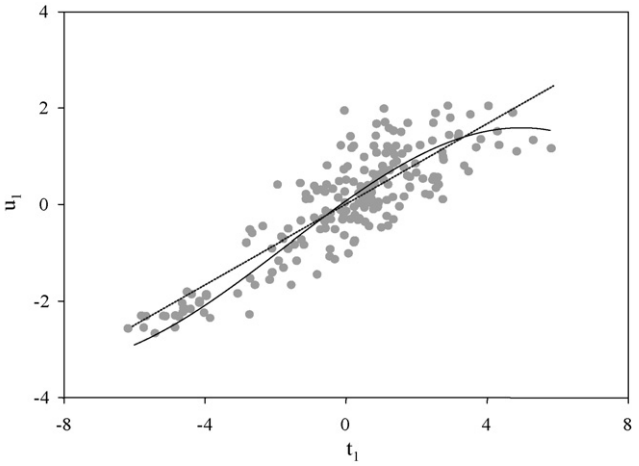

Fig. 5. Score plot of the first latent factor (grey dot: data points, dot line: inner model of the linear PLS, solid line: inner model of the NNPLS).

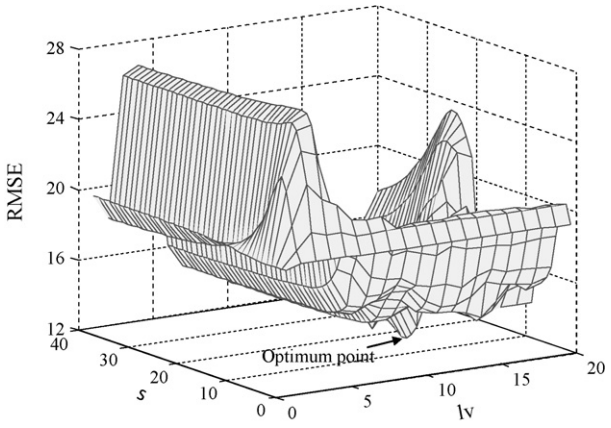

Fig. 6. Three-dimensional surface plot of the KPLS.

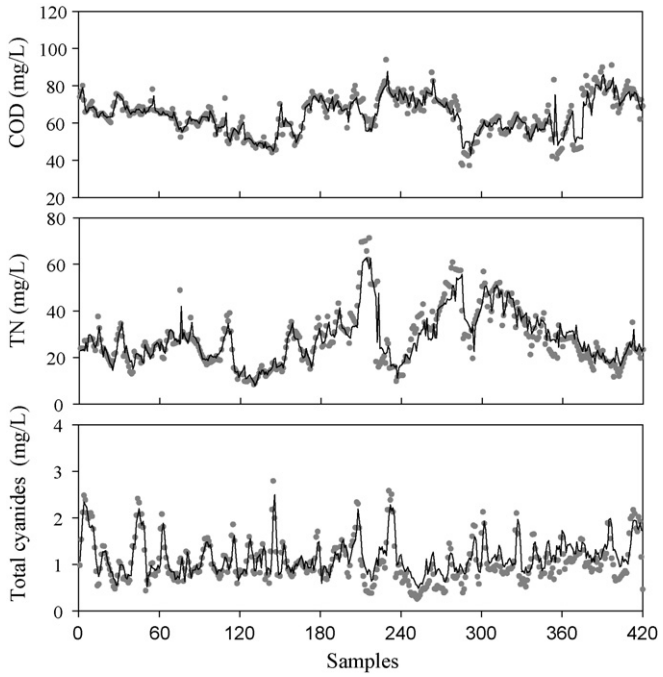

Fig. 7. Prediction results of the KPLS (grey dot: measured values, solid line: predicted values).

linear data structure in the feature space could capture the nonlinearities in the original data space better than the NNPLS method. Fig. 7 shows the simulation results of the KPLS model. This figure, when it is compared with Figs. 2 and 4, also clearly shows that KPLS gives much better prediction results than the linear PLS model. In addition, the KPLS model gave very little bias for the validation data set of total cyanides concentrations.

However, the goodness of fit for the different PLS modeling approaches with different numbers of degrees of freedom cannot be assessed only by the RMSE values. More complex models with larger numbers of parameters will improve the model fit to the data because it reduces the RMSE of the residuals between the model predictions and the corresponding measured values. It is, therefore, necessary to have quantitative measures of model adequacy in order to decide between competing model structures. For large data sets, the appropriate criterion to use is the Bayesian information criterion (BIC) [30]:

$$\mathrm{BIC} = L(\hat{\vartheta}|y_{ij}) - n_\mathrm{p}/2\ \ln(m/2\pi) \tag{11}$$

where L is the likelihood function of the model, $\hat{\vartheta}$ denotes the maximum likelihood estimates of the vector of unknown parameters, n_p is the number of parameters and m is the number of measurements. For the likelihood function L, after taking the natural logarithm and maximizing with respect to the unknown parameters, the maximized logarithmic likelihood can be obtained [31]:

$$L(\hat{\vartheta}|y_{ij}) = -(m/2)\ \ln\left\{\sum [y_{ij} - f(\hat{\vartheta}|x_{ij})]^2\right\} \tag{12}$$

where $f(\hat{\vartheta}|x_{ij})$ is the model output at the ith value of the input x_j. From Eqs. (11) and (12), a model with a high BIC is preferable to one with a lower value [32]. The KPLS model's BIC value was higher than those of the linear PLS and NNPLS models (Table 2), implying that the KPLS model is even a much better model in this direct quantitative comparison. These results consistently showed that the KPLS model outperformed other linear and nonlinear PLS models in the aspect of both model prediction and complexity.

4. Conclusions

Although the conventional PLS modeling methods give a linear model from a lot of collinear measurements, it is not capable of modeling nonlinear systems. In this work, a KPLS method was applied to the CWTP. The KPLS mapped the nonlinear input space into a high-dimensional feature space where the data structure is likely to be linear. Principal components in the feature space were calculated by means of integral operators and nonlinear kernel functions. It required only linear algebra to develop a process modeling system compared to other nonlinear methods that involve nonlinear optimization. The KPLS gave a much better prediction performance than the linear and nonlinear PLS methods. The increased prediction performance of KPLS could be explained by the fact that the biological wastewater treatment plant is an inherently nonlinear process, and the KPLS model could capture the nonlinearities in the original data space benefiting from the linear data structure in the feature space. However, in this application, the model's results should not be extrapolated without care as the training and validation data sets showed rather little dynamics. Since the developed model was derived from every 8 h, it might not be appropriate for control purposes. But it could give operators and process engineers a reliable and accurate estimation of the key process variables in real time that would allow them to arrive at the optimum operational strategy for the CWTP. The successful application of the KPLS method to the industrial wastewater treatment plant has demonstrated the feasibility and effectiveness of the kernel-based algorithm. The methodology is fairly general and is applicable to most chemical and biological wastewater treatment plants.

Acknowledgements

This work was supported by Korea Ministry of Environment as "the Eco-Technopia 21 Project" (Grant No. 071-071-106). The work was also financially supported by the ERC program of MOST/KOSEF (R11-2003-006-01001-1) through the Advanced Environmental Biotechnology Research Center at POSTECH.

References

[1] R.L. Cooper, J.R. Catchpole, The biological treatment of carbonization effluents-IV, Water Res. 7 (1973) 1137–1153.
[2] M. Zhang, J.H. Tay, Y. Qian, X.S. Gu, Coke plant wastewater treatment by fixed biofilm system for COD and NH_3-N removal, Water Res. 32 (1998) 519–527.
[3] M.W. Lee, J.M. Park, Biological nitrogen removal from coke plant wastewater with external carbon addition, Water Environ. Res. 70 (1998) 1090–1095.
[4] M.S. Kumar, A.N. Vaidya, N. Shivaraman, A.S. Bal, Performance evaluation of a full-scale coke oven waste water treatment plant in an integrated steel plant, Ind. J. Environ. Health 45 (2003) 29–38.
[5] D.J. Richards, W.K. Shieh, Anoxic-oxic activated-sludge treatment of cyanides and phenols, Biotechnol. Bioeng. 33 (1989) 32–38.
[6] S. Ebbs, Biological degradation of cyanide compounds, Curr. Opin. Biotechnol. 15 (3) (2004) 231–236.
[7] D. Park, Y.-S. Yun, D.S. Lee, S.-R. Lim, J.M. Park, Column study on Cr(VI) removal using the brown seaweed Ecklonia biomass, J. Hazard. Mater. 137 (3) (2006) 1377–1384.
[8] Y.-S. Yun, M.W. Lee, J.M. Park, C.-I. Lee, J.-S. Huh, H.-D. Chun, Reclamation of wastewater from a steel-making plant using an airlift submerged biofilm reactor, J. Chem. Technol. Biotechnol. 73 (1998) 162–168.
[9] D. Park, D.S. Lee, Y.M. Kim, J.M. Park, Bioaugmentation of cyanide-degrading microorganisms in a full-scale cokes wastewater treatment facility, Bioresour. Technol. 99 (6) (2008) 2092–2096.
[10] R.G. Luthy, Biological treatment of coal coking and coal gasification wastewaters, J. Water Pollut. Cont. Fed. 53 (3) (1981) 325–339.
[11] N. Shivaraman, P. Kumaran, R.A. Pandey, P. Choudhary, S.K. Chatergee, N.M. Parahad, Microbial degradation of thiocyanate, phenol and cyanide in completely mixed aeration system, Environ. Pollut. Contam. (Series A) 39 (1985) 141–149.
[12] D.S. Lee, C.O. Jeon, J.M. Park, K.S. Chang, Hybrid neural network modelling of a full-scale industrial wastewater treatment process, Biotechnol. Bioeng. 78 (2002) 670–682.

[13] Y.M. Kim, D. Park, D.S. Lee, J.M. Park, Instability of biological nitrogen removal in a cokes wastewater treatment facility during summer, J. Hazard. Mater. 141 (1) (2007) 27–32.
[14] A. Akcil, Destruction of cyanide in gold mill effluents: biological versus chemical treatments, Biotechnol. Adv. 21 (2003) 501–511.
[15] Y.L. Paruchuri, N. Shivaraman, P. Kumaran, Microbial transformation of thiocyanate, Environ. Pollut. 68 (1990) 15–28.
[16] G. Stephanopoulos, K.Y. San, Studies on on-line bioreactor identification I—theory, Biotechnol. Bioeng. 26 (1984) 1176–1188.
[17] W. Bae, B.E. Rittman, A structured model of dual limitation kinetics, Biotechnol. Bioeng. 49 (1996) 683–689.
[18] G. Baffi, E.B. Martin, A.J. Morris, Non-linear dynamic projection to latent structures modeling, Chemom. Intell. Lab. Syst. 52 (2000) 5–22.
[19] S. Wold, N. Kettaneh-Wold, B. Skagerberg, Nonlinear PLS modeling, Chemom. Intell. Lab. Syst. 7 (1989) 53–65.
[20] S.J. Qin, T.J. McAvoy, Non-linear PLS modelling using neural networks, Comput. Chem. Eng. 23 (1992) 395–411.
[21] D.S. Lee, P.A. Vanrolleghem, J.M. Park, Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant, J. Biotechnol. 115 (2005) 317–328.
[22] B. Schölkopf, A.J. Smola, K. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (5) (1998) 1299–1319.
[23] R. Rosipal, L.J. Trejo, Kernel partial least squares regression in reproducing kernel Hilbert space, J. Mach. Learning Res. 2 (2001) 97–123.
[24] S. Haykin, Neural Networks, Prentice-Hall, New Jersey, 1999.
[25] APHA, Standard Methods for the Examination of Water and Wastewater, 20th ed., APHA, Washington, DC, 1998.
[26] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemom. Intell. Lab. Syst. 58 (2001) 109–130.
[27] P. Geladi, B.R. Kowalski, Partial least-squares regression: a tutorial, Anal. Chim. Acta 185 (1986) 1–17.
[28] U. Kruger, D. Antory, J. Hahn, G.W. Irwin, G. McCullough, Introduction of a nonlinearity measure for principal component models, Comp. Chem. Eng. 29 (2005) 2355–2362.
[29] S.J. Qin, Partial least squares regression for recursive system identification, in: Proceedings of the 32nd Conference on Decision and Control, San Antonio, Texas, 1993, pp. 2617–2622.
[30] T. Leonard, J.S.J. Hsu, Bayesian Methods, Cambridge University Press, New York, 1999.
[31] I.G. Main, T. Leonard, O. Papasouliotis, C.G. Hatton, P.G. Meredith, One slope or two? Detecting statistically significant breaks of slope in geophysical data with application to fracture scaling, Geophys. Res. Lett. 26 (1999) 2801–2804.
[32] T. Seher, I.G. Main, A statistical evaluation of a 'stress-forecast' earthquake, Geophys. J. Int. 157 (2004) 187–193.