Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Biotechnology 115 (2005) 317–328
JOURNAL OF Biotechnology
www.elsevier.com/locate/jbiotec

Parallel hybrid modeling methods for a full-scale cokes wastewater treatment plant

Dae Sung Lee[a,*], Peter A. Vanrolleghem[b], Jong Moon Park[a]

[a] *School of Environmental Science and Engineering, Advanced Environmental Biotechnology Research Center (AEBRC), Pohang University of Science and Technology, San 31, Hyoja-dong, Pohang, Kyungbuk 790-784, Republic of Korea*
[b] *BIOMATH, Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure Links 653, B-9000 Gent, Belgium*

Received 30 December 2003; received in revised form 11 June 2004; accepted 7 September 2004

Abstract

Parallel hybrid modeling methods are applied to a full-scale cokes wastewater treatment plant. Within the hybrid model structure, a mechanistic model specifies the basic dynamics of the relevant process and a non-parametric model compensates for the inaccuracy of the mechanistic model. First, a simplified mechanistic model is developed based on Activated Sludge Model No. 1 and the specific process knowledge of the cokes wastewater treatment process. Then, the mechanistic model is combined with five different non-parametric models – feedforward back-propagation neural network, radial basis function network, linear partial least squares (PLS), quadratic PLS and neural network PLS (NNPLS) – in parallel configuration. These models are identified with the same data obtained from the plant operation to predict dynamic behavior of the process. The performance of each parallel hybrid model is compared based on their ease of model building, prediction accuracy and interpretability. For this application, the parallel hybrid model with NNPLS as non-parametric model gives better performance than other parallel hybrid models. In addition, the NNPLS model is used to analyze the behavior of the operation data in the reduced space and allows for fault detection and isolation.

Keywords: Hybrid modeling; Industrial wastewater treatment plant; Mechanistic model; Non-parametric model; Partial least squares

1. Introduction

The activated sludge process is one of the most widespread wastewater treatment techniques for both domestic and industrial wastewater. Dynamic mathematical models enhance the understanding of the biological phenomena and provide the basis for design and operation of biological wastewater treatment systems. To date, the Activated Sludge Model No.1, or ASM1 (Henze et al., 1987), is accepted to be one of the most successful models for carbonaceous substrate and nitrogen removal processes in many applications.

* Corresponding author. Tel.: +82 54 279 8650; fax: +82 54 279 8659.
E-mail address: dslee@postech.ac.kr (D.S. Lee).

0168-1656/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2004.09.001

However, the high complexity of the ASM1 with its numerous processes and parameters often gives rise to identification problems. Furthermore, since the model is designed for domestic wastewater treatment only, there is a significant limitation in its application to industrial wastewater treatment processes. It is not easy or worthwhile to spend too much time and effort to simulate peculiarities and non-idealities of an industrial process using ASM1.

In recent years, hybrid neural network modeling approaches have received considerable attention (Psichogios and Ungar, 1992; Thompson and Kramer, 1994; van Can et al., 1997; Lee et al., 2002). These approaches are potentially very efficient to obtain more accurate predictions of process dynamics by taking the advantages of both the mechanistic model and the neural network model. Within the hybrid model structure, the mechanistic model specifies the basic dynamics of the relevant process variables. The neural network model, combined with the mechanistic model either in parallel or in serial configuration, accounts for unknown and non-linear parts of the mechanistic model. In case, the mechanistic model has a reasonably accurate structure, thereby reducing the identification problem to that of estimating unmeasured process parameters, the serial hybrid neural model gives good performance (Psichogios and Ungar, 1992). However, when the mechanistic model's structure is highly uncertain, or contains a large number of very complex process parameters like biological wastewater treatment process, then the parallel hybrid model provide a significant advantage over both the mechanistic model and the neural network model (Lee et al., 2002).

There has been a great preference in neural networks, especially feedforward back-propagation neural networks (FBNN) and radial basis function network (RBFN), as non-parametric model (Thompson and Kramer, 1994; Zhao et al., 1999). However, the approach is applicable to other non-parametric models available, each varying in complexity and ease of development. Alternative non-parametric models include linear partial least squares (PLS), quadratic PLS (QPLS) and neural network PLS (NNPLS).

In this study, a simplified mechanistic model was developed to simulate the dynamic behavior of a full-scale cokes wastewater treatment plant. Then, the above mentioned five different non-parametric methods were incorporated into the mechanistic model in parallel configuration. The non-parametric models were identified on the residuals between the mechanistic model and the operation data to compensate for non-linearity and uncertainty that arise from the inherent process complexity. The parallel hybrid models were tested on their prediction capability and compared based on their ease of model building, prediction accuracy and interpretability.

2. Materials and methods

2.1. Cokes wastewater treatment plant

The cokes wastewater treatment process (CWTP) at a steel-making 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 plant. Most of the chemical oxygen demand (COD) originates from phenol, which is a toxic inhibitory substrate but is also a carbon source for acclimatized microorganisms (Richards and Shieh, 1989). In addition to phenol, cyanides and other toxic aromatic hydrocarbons such as cresol, indole and toluene contribute to the wastewater COD. Since a high concentration of nitrogen compounds was found inhibitory to biodegradation, pre-treatment 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 days. Oxygen was introduced by submerged mechanical aerators. Concentrated sludge from the bottom of the clarifier was split into two streams: the first was recycled to the beginning of the first two aeration tanks 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 five months were collected at 8 h intervals. All samples were analyzed for mixed liquor suspended solids (MLSS), COD, suspended solids (SS), cyanide (CN) and phenol according to the Standard Methods (APHA, 1995). Dissolved oxygen concentration, pH, influent flow rate and

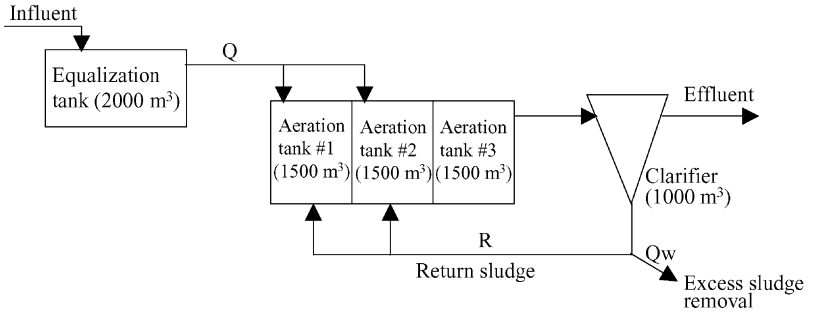


Fig. 1. Schematic diagram of full-scale cokes wastewater treatment plant.

recycle flow rate were also measured at each sampling time. This measuring campaign resulted in 466 operational data sets in total. The first 240 sets of data were used for training and the remaining 226 data sets for the validation of the developed models.

2.2. Simplified mechanistic model

A mechanistic model for the CWTP was developed on the basis of ASM1. For the reactor system shown in Fig. 1, the mass balances for process variables in the reactor were established to predict the concentrations of effluent components. For practical reasons, minor mechanisms were neglected and only known major mechanisms were taken into account. Therefore, the following assumptions and simplifications were made based on the process knowledge and understanding of the CWTP.

(a) All reactors were assumed to be well-agitated and aerated continuous-flow reactors. The kinetics of microorganisms were assumed to be the same in all reactors.
(b) Active biomass was divided into two types of organisms: heterotrophic (X_H) and cyanide-degrading organisms (X_{CN}).
(c) Biodegradable carbonaceous material was classified as readily biodegradable substrate (S_S), slowly biodegradable substrate (X_S) and cyanide compounds (S_{CN}). Phenolic compounds in the influent were assumed to be readily biodegradable organic matter (S_S). Inert particulate products (X_P) arising from biomass decay were also included.
(d) Nitrification reactions were neglected, since no nitrate or nitrite was detected. There have been several reports that nitrification is inhibited by phenol, most of the poly-nuclear aromatic hydrocarbons and cyanide, all of which were present as carbonaceous compounds in the cokes wastewater (Lee and Park, 1998).
(e) Since the dissolved oxygen concentration was maintained at 1.5 mg O_2/l, oxygen transfer was assumed not to be rate-limiting.
(f) Cyanide compounds are toxic to heterotrophs. The shock loading of cyanide during the measuring campaign caused a deterioration of the biological wastewater treatment process. To describe cyanide inhibition upon the growth rate of heterotrophs, the following rate expression was used:

$$\mu = \hat{\mu}_H \left(\frac{S_S}{K_S + S_S} \right) \left(\frac{K_I}{K_I + S_{CN}} \right) \quad (1)$$

(g) No reaction was assumed during the clarification and, therefore, the secondary settler was considered to be a simple separating point.

As a result, the mechanistic model consisted of eight components and five rate equations, as listed in Table 1. The parameter values used in the model were initially based on literature values. For heterotrophs, most default parameters provided by ASM1 were used. Kinetic parameters suggested by Gaudy et al. (1982) were used for cyanide-degrading organisms. From the results of sensitivity analysis, only three parameters μ_H, μ_{CN} and Y_H were optimized using the simplex method in order to minimize the deviations between the simulation values and the corresponding operational data.

Table 1
Simplified mechanistic model for the cokes wastewater treatment plant

j	Process	Component i								Process Rate ρ_j ($ML^{-3}\,T^{-1}$)
		1 S_I	2 S_S	3 S_{CN}	4 X_I	5 X_S	6 X_H	7 X_{CN}	8 X_P	
1	Aerobic growth of heterotrophs		$-\frac{1}{Y_H}$				1			$\hat{\mu}_H \left(\frac{S_S}{K_S+S_S}\right)\left(\frac{K_I}{K_I+S_{CN}}\right) X_H$
2	Aerobic growth of cyanide-degrading microorganisms			$-\frac{1}{Y_{CN}}$				1		$\hat{\mu}_{CN} \left(\frac{S_{CN}}{K_{CN}+S_{CN}}\right) X_{CN}$
3	Decay of heterotrophs					$1-f_P$	-1		f_P	$b_H X_H$
4	Decay of cyanide-degrading microorganisms					$1-f_P$		-1	f_P	$b_{CN} X_{CN}$
5	Hydrolysis of entrapped organics		1			-1				$k_h \frac{X_S/X_H}{K_X+X_S/X_H} X_H$
Observed conversion rates ($ML^{-3}\,T^{-1}$)		$r_i = \sum_j \nu_{ij}\rho_j$								

Stoichiometric papameters
Heterotrophic yield: Y_H
Cyanide-degrading microorganisms' yield: Y_{CN}
Fraction of biomass yielding particulate products: f_P

Kinetic parameters
Heterotrophic growth and decay: $\hat{\mu}_H$, K_S, K_I, b_H
Cyanide-degrading microorganisms' growth and decay: $\hat{\mu}_{CN}$, K_{CN}, b_{CN}
Hydrolysis: k_H, K_X

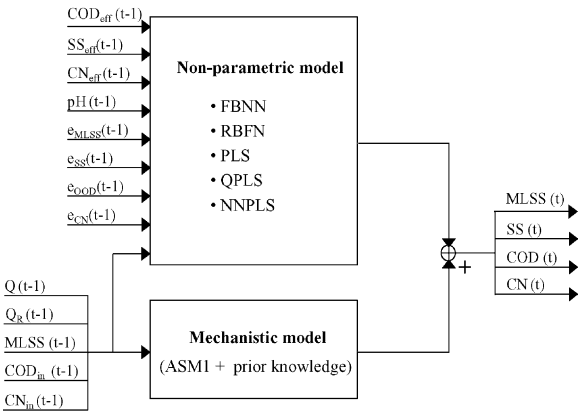


Fig. 2. Parallel hybrid modeling approaches.

2.3. Parallel hybrid model

In the parallel hybrid model structure, non-parametric models described below are combined with the mechanistic model in parallel configuration as shown in Fig. 2. The non-parametric models were used to estimate the difference between the mechanistic model predictions and the corresponding operational data (i.e., the residuals). The key to success in developing parallel hybrid models lies in the information content of the residuals. If non-parametric models are trained to extract some useful information from the residuals, the accuracy of the model would be improved. Since the mechanistic model was developed by making several assumptions and simplifications, we expected that the residuals would have dynamic information that was not contained in the mechanistic model. In addition, external disturbances such as composition variation, shock loadings of toxic compounds and temperature variation were fed into the non-parametric model so that it could capture the effects of these disturbances. The inputs fed to both the non-parametric model and the mechanistic model include influent flow rate $Q(t-1)$, sludge recirculation rate $Q_R(t-1)$, MLSS$(t-1)$ in the reactor and the influent concentrations of $COD_{in}(t-1)$ and $CN_{in}(t-1)$. The effluent concentrations of $COD_{eff}(t-1)$, SS_{eff} $(t-1)$ and CN_{eff} $(t-1)$; pH$(t-1)$ in the reactor; and the residuals $e(t-1)$ of MLSS, SS, COD and CN were only fed to the non-parametric models. The non-parametric outputs were the residuals $e(t)$ of MLSS, SS, COD and CN. All programs used in this work were implemented in MATLAB by using the Neural Network Toolbox (Demuth and Deale, 2001) and the PLS Toolbox (Wise and Gallagher, 2000).

2.4. Feedforward Back-propagation neural network

Neural networks have been successfully applied to various biochemical processes (Montague and Morris, 1994; Zhang et al., 1994; Lee and Park, 1999). They have a distinct ability to model non-linear dynamic systems without requiring a structural knowledge of the process to be modeled. Neural networks can map a set of input patterns onto a corresponding set of output patterns on the basis of historical data from any given system. However, neural network models have also been criticized for a lack of dependence upon physical relationships and a poor capacity for extrapolation.

A feedforward back-propagation neural network (FBNN) was employed in this study. The FBNN structure consists of one input layer, one hidden layer and one output layer. Each layer can have a number of neurons (processing elements), which are connected linearly by weights to the neurons in the neighboring layers. The number of neurons in the input and output layers are predetermined by the number of input and output variables. The hidden layer has the hyperbolic tangent as activation function and the output layer the linear function. Prior to training, all variables were scaled to the range −1 to 1. The Levenberg–Marquardt algorithm was used for the training process. The training process adjusts weights to minimize the error between the measured output and the output produced by the network. Through this adjustment, the neural network learns the input–output behavior of the system. The optimal structure of the neural network was determined by varying the number of neurons in the hidden layer. By comparing the performance of network configurations in the recall and generalization process (Baughman and Liu, 1995), we found that the hidden layer with six neurons gave the best results.

2.5. Radial basis function

A radial basis function network (RBFN) is a network structure that employs local receptive fields to perform functional mappings (Haykin, 1999). When an input vector is applied to the RBFN, each neuron in the hidden layer will output a significant non-zero response according to how close the input vector is to each neuron's weight vector. The most frequently used radial basis function is the Gaussian activation function:

$$\Psi j(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{u}_j\|^2}{2\sigma_j^2}\right) \tag{2}$$

where $\mathbf{x}$ is an input vector, $\mathbf{u}_j$ is a weight vector and σ_j is the spread of the jth basis function. The output of the RBFN is the weighted average of the output associated with each hidden unit:

$$y_k(\mathbf{x}) = \frac{\sum_j w_{jk}\Psi_j(\mathbf{x})}{\sum_j \Psi_j(\mathbf{x})} \tag{3}$$

where $y_k(\mathbf{x})$ is the kth output and w_{jk} are the weights between the hidden and output layers. A gradient descent algorithm was used for training the network parameters. The RBFN was designed by creating neurons one at a time. Initially, the hidden layer had no neurons. At each iteration, the input vector with the greatest error was used to create a neuron. The training algorithm continued until the new network met the specified error goal or the maximum number of neurons was reached. The spread parameter σ and the number of neurons in the hidden layer were optimized as 1.8 and 15, respectively, using the response surface methodology (Box and Draper, 1987; Draper and Lin, 1990).

2.6. Partial least squares

The partial least squares (PLS) method 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** (Wold et al., 2001). PLS reduces the dimension of the predictor variables by extracting factors or latent variables, which 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**.

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**):

$$\begin{aligned}\mathbf{X} &= \sum_{i=1}^{a} \mathbf{t}_i\mathbf{p}_i^{\mathrm{T}} + \mathbf{E}\\ \mathbf{Y} &= \sum_{i=1}^{a} \mathbf{u}_i\mathbf{q}_i^{\mathrm{T}} + \mathbf{F}\end{aligned} \tag{4}$$

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

the score vector **u**.

$$\mathbf{u}_i = b_i \mathbf{t}_i + h_i \quad (5)$$

where b is a regression coefficient which is determined by minimizing the residual h. There are several algorithms to calculate the PLS model parameters. In this work, the non-linear Iterative Partial Least Squares (NIPALS) algorithm was used with the exchange of scores (Geladi and Kowalski, 1986). 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. Based on the cross-validation results, six latent variables were included into the PLS model. It explained 87.70% of the variance of matrix **X** and 76.36% of matrix **Y**.

2.7. Non-linear partial least squares

In order to capture non-linear structures between the predictor block and the responses, the PLS model can be extended to non-linear partial least squares models (Baffi et al., 2000). Major approaches have been to incorporate non-linear functions within the linear PLS framework. Especially, quadratic functions and neural networks have been used to identify the non-linear inner mapping between the input and the output latent variables. Wold et al. (1989) proposed the QPLS method to make a polynomial fit for the PLS inner relation. QPLS works just like PLS and uses the NIPALS algorithm to calculate the latent variables. Once a pair of latent variables is calculated, polynomial functions are used to model the functional relationship between the pair of latent variables. In this study, the degree of the polynomial used was two and the number of latent variables in the model was set to six on the basis of the cross-validation analysis, which then explained 87.57% of the variance of matrix **X** and 76.53% of matrix **Y**.

Neural network PLS (NNPLS) is an integration of neural networks with PLS to model non-linear processes with input co-linearity (Qin and McAvoy, 1992). 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:

$$\mathbf{u}_i = F(\mathbf{t}_i) + \mathbf{v}_i \quad (6)$$

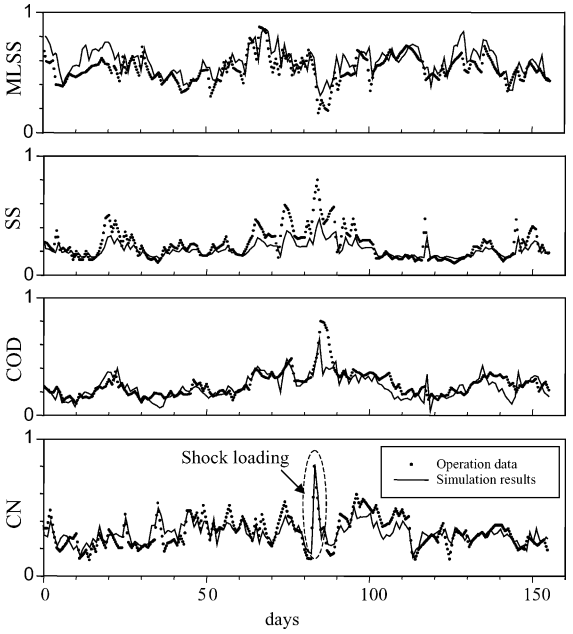


Fig. 3. Simulation results of the simplified mechanistic model.

where $F(\cdot)$ stands for the inner relation represented by a neural network and **v** is the residuals. The neural network is trained to capture the non-linearity 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 FBNN with sigmoid functions was used to identify the non-linear inner regression models for each of the six latent variables. The model explained 88.15% of **X**-variance and 78.33% of the **Y**-variance.

3. Results and discussion

3.1. Model identification results

Simulation results of the simplified mechanistic model are shown in Fig. 3. To maintain industrial confidentiality, the ordinates of all time-series simulation results were normalized to that of the variables of the CWTP. As can be seen in Fig. 3, the mechanistic model predicted some of the general characteristics of process behaviors, but there was a significant mismatch between the model prediction and actual plant data.

Table 2
Comparison of different parallel hybrid models

Non-parametric model	Structure	TNP[c]	$RSSE_{training}$	$RSSE_{validation}$	BIC
–	Only mechanistic model	12	0.143	0.201	−1591.6
FBNN	13-6-4	112	0.018	0.163	−1571.0
RBFN	13-15-4, sc[a] = 1.8	257	0.012	0.130	−1963.9
PLS	lv[b] = 6	80	0.056	0.184	−1640.5
QPLS	lv[b] = 6, order of polynomial = 2	98	0.021	0.045	−1483.7
NNPLS	lv[b] = 6, SISO FBNN	122	0.019	0.038	−1474.3

[a] Spread constant.
[b] Number of latent variables.
[c] Total number of parameters.

This discrepancy was especially severe when the process experienced the shock loading of toxic cyanide (from operation data 250 to 279) as described in detail in the next section. Normal cyanide concentration in the influent was about 12 mg l^{-1}, but during the upset period it increased abruptly up to 26 mg l^{-1} between days 84 and 87 (corresponding to data points 250 and 258). The system took the shock of the adverse loading with a leakage of organic carbon into the effluent and a rather severe decrease in biomass concentration (Fig. 3). After this shock loading period, the mechanistic model gave even poorer predictions than it did before the process damage. This can be attributed to changes in the activities of microorganisms or changes of kinetic parameters.

All parallel hybrid models were identified with the same operation data obtained from the plant. With the respective optimal parallel hybrid models, the results of the validation stage enabled the different modeling approach to be evaluated and compared. The performance of each model was evaluated in terms of the relative sum of squared error (RSSE) criterion. The RSSE performance index was defined as:

$$\mathrm{RSSE} = \frac{\sum \|t_{ij} - y_{ij}\|^2}{\sum \|y_{ij}\|^2} \quad (7)$$

where y_{ij} denotes the ith value of the output j and t_{ij} the associated prediction value. Table 2 shows the RSSE values of the training and validation data sets for five different hybrid modeling approaches. All models predict the dynamics of the wastewater treatment process with good accuracy compared with the mechanistic model. This is a clear indication that the residuals contained sufficient relevant information about the dynamic behavior of the system, but which was not considered in the mechanistic model.

Both the FBNN and RBFN modeling structures provided a very good fit to the training data. However, the hybrid FBNN model showed a relatively poor predicting ability for the validation data set. One disadvantage of such global methods like FBNN is that a new sample may change the model everywhere. Local interpolation methods like RBFN overcome this drawback by utilizing only the neighboring samples of the estimate at a lookup point (Carlin et al., 1994). However, the RBFN model required more neurons in the hidden layer than the FBNN. When the inputs vary over a wide range, the number of required neurons in the RBFN model becomes high. Initialization of weights and bias in the RBFN model is not a random process (Tettch et al., 1996). Therefore it was possible to optimize the RBFN model by a traditional factorial design for the spread parameter σ and the number of neurons in the hidden layer. For both the FBNN and the RBFN models, parameter optimization is an iterative process. Therefore, the iterative methods might not reach a unique minimum or be trapped in local minima. In addition, it is a disadvantage that the weights and biases in the network structures cannot be interpreted directly. These methods always give the output as a coupled dependency of all the input variables. Some researchers have tried to interpret the transformation made by the hidden layer, but the results are not very applicable for practical model interpretation (Garson, 1991; Gallinari et al., 1991).

It is clear from Table 2 that the parallel hybrid model with PLS as non-parametric model gave the worst prediction performance among the hybrid modeling approaches. This exemplified the fundamental weakness of the linear multivariate regression model. When

non-linear regression methods were used for modeling the inner relations in the PLS, the prediction capabilities were largely improved. The hybrid QPLS model gave a better prediction performance than the hybrid PLS model. However, quadratic models are still linear in their parameters and do not guarantee a proper solution for mapping non-linear relationships between the input and output variables (Bro, 1995). In this study, the parallel hybrid method with NNPLS as non-parametric model gave the best prediction performance, based on the RSSE of the validation data set (Table 2).

However, the goodness of fit for the parallel hybrid modeling approaches with different numbers of degrees of freedom cannot be assessed only by the RSSE. More complex models with larger numbers of parameters will improve the model fit to the data, as it reduces the RSSE of the residuals between the model predictions and the corresponding operation data. It is, therefore, necessary to have quantitative measures of model adequacy for deciding between competing model structure. For large data sets, the appropriate criterion to use is the Bayesian information criterion (BIC) proposed by Leonard and Hsu (1999):

$$\mathrm{BIC} = L(\hat{\vartheta}\,|\,y_{ij}) - \frac{p}{2}\ln\frac{N}{2\pi} \tag{8}$$

where L is the likelihood function of the model, $\hat{\vartheta}$ denotes the maximum likelihood estimates of the vector of unknown parameters, p the number of parameters and N the number of measurements. For the likelihood function L, Main et al. (1999) take the natural logarithm and maximize with respect to the unknown parameters, which gives the maximized logarithmic likelihood:

$$L(\hat{\vartheta}\,|\,y_{ij}) = -\frac{N}{2}\ln\left\{\sum[y_{ij} - f(\hat{\vartheta}\,|\,x_{ij})]^2\right\} \tag{9}$$

where $f(\hat{\vartheta}\,|\,x_{ij})$ is the model output at the ith value of the input x_j. From Eq. (8) and (9) a model with a high BIC is preferable to one with a lower value (Seher and Main, 2004). As shown in Table 2, the hybrid NNPLS model's BIC was much higher than the BIC of any other models, implying that the hybrid NNPLS model is the best model in this direct quantitative comparison. These results consistently showed that the hybrid NNPLS model outperformed the other models in the aspect of both model prediction and complexity.

Fig. 4 shows the simulation results of the hybrid NNPLS model. The prediction performance was greatly improved for both training and validation data sets, compared with the mechanistic model. Fig. 5 shows the first principal inner relations between the **X** and **Y**-block by both the PLS and NNPLS methods. It shows how the NNPLS method captures the non-linearity and thus outperforms the linear PLS model. This increased prediction performance can be

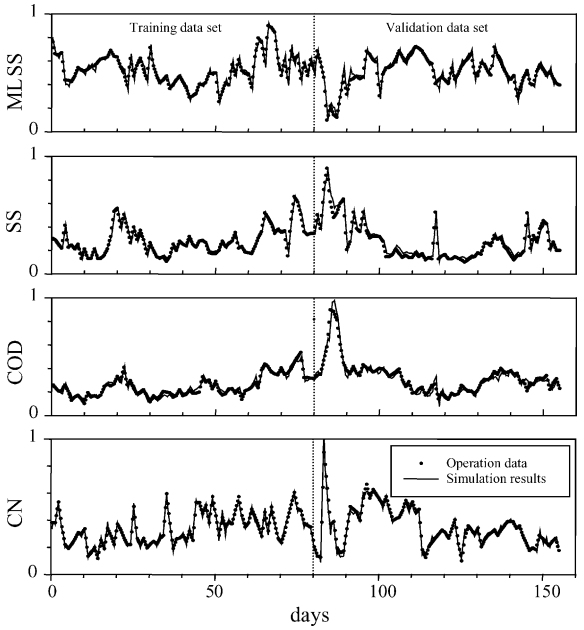


Fig. 4. Simulation results of the parallel hybrid model with NNPLS as non-parametric model.

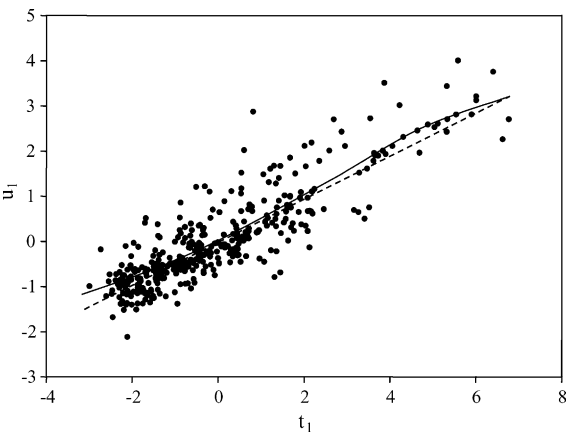
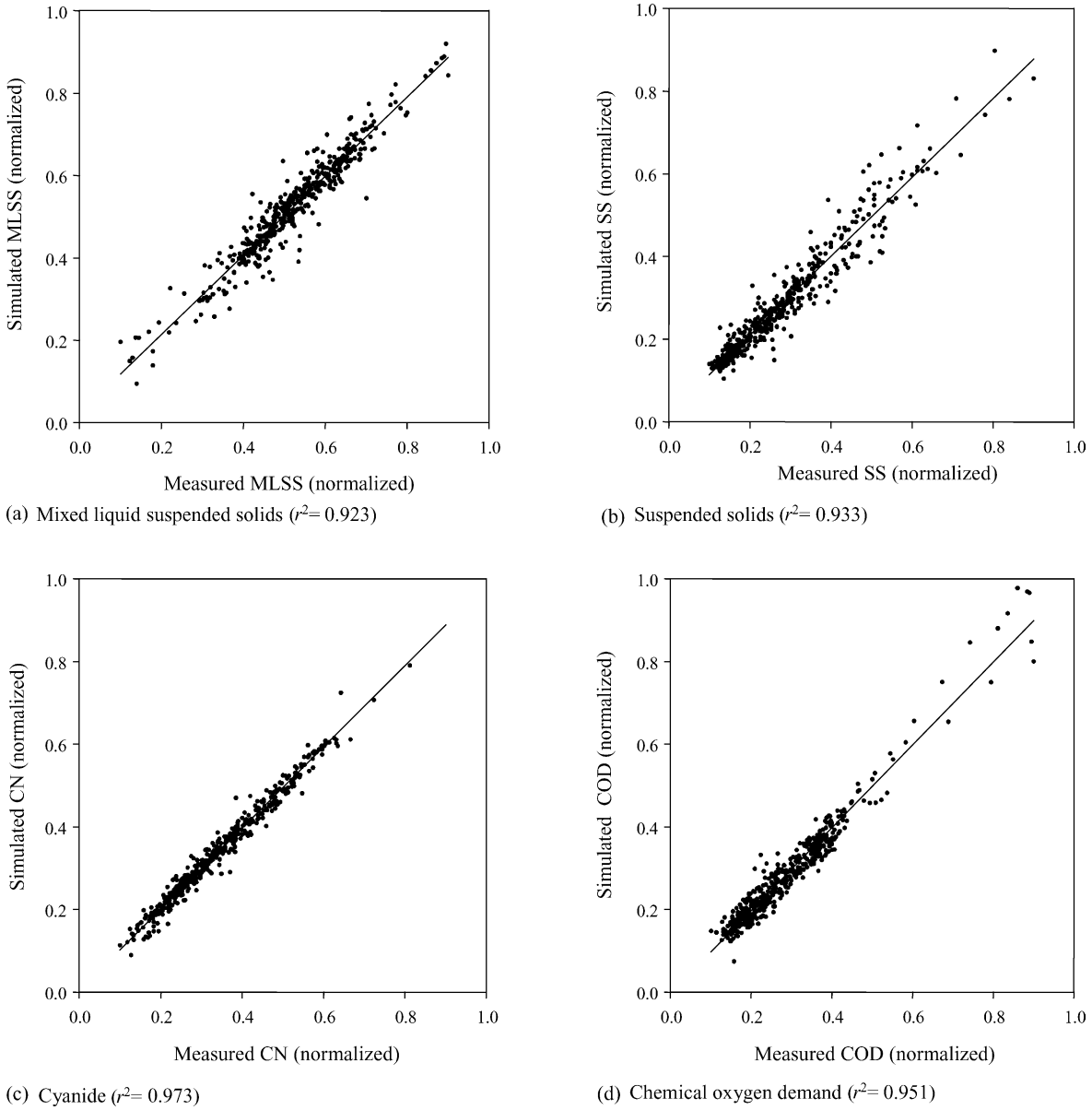


Fig. 5. Score plots of the first latent factor coefficients (dot: data points, dotted line: PLS inner model, solid line: NNPLS inner model).

explained by the fact that the CWTP, in this application, is an inherently non-linear system with time-varying reactions of the microorganisms and large variations in the incoming wastewater. The NNPLS could extract non-linear information from the residuals to compensate for the inaccuracy of the mechanistic model. The measured data versus the predicted values from the hybrid NNPLS model are shown in Fig. 6. Those plots visualize the performance of the models in an obvious way. The *r*-squared value (r^2) represents the fraction of the variance explained in the operation data by the model.



(a) Mixed liquid suspended solids (r^2= 0.923)

(b) Suspended solids (r^2= 0.933)

(c) Cyanide (r^2= 0.973)

(d) Chemical oxygen demand (r^2= 0.951)

Fig. 6. Measured vs. predicted values of the hybrid NNPLS model.

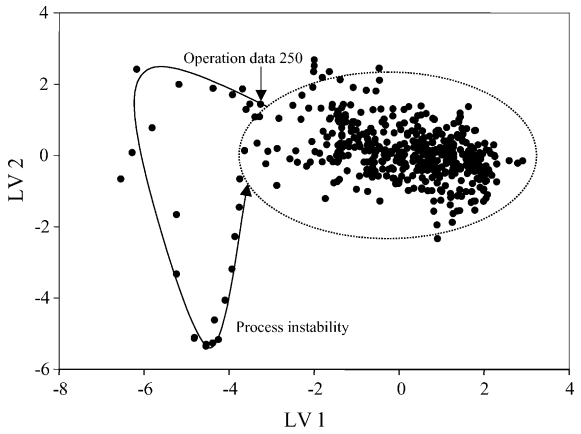


Fig. 7. NNPLS score plot of operation data (dotted ellipse: 95% confidence limit).

3.2. Process monitoring with NNPLS

PLS models have interesting capability of visualizing high-dimensional data through the lower-dimensional projections defined by the small number of latent variables. By examining the dynamic behavior of the process data in the reduced space, it is often possible to extract very useful information. Fig. 7 shows a score plot of the collected data in the space of the resulting first two latent variables from the NNPLS model in the parallel hybrid structure. The score plot provides an adequate representation of the process behavior. Most of the normal operating data were contained within the control contour of 95% confidence limit. However, when the process instability occurred, the corresponding process data points (from operation data 250 to 279) progressively moved outside the ellipsoidal boundary, and then returned to the inside the confidence limit, when the process was recovered from the damage. It is also straightforward to determine how much each of the input variables contributes to the variations in the output variables. Fig. 8 shows the contribution plot to the operation data 250, which clearly identifies that cyanide concentration in the influent contributed to the disturbance. From operation data analysis, we found that the process upset was caused by a shock loading of cyanide in the influent (Lee et al., 2002). Therefore, the additional advantage of the hybrid NNPLS model is to allow process monitoring and easier interpretation of process behavior, compared with the parallel hybrid model with FBNN and RBFN as non-parametric models.

4. Conclusions

Five different parallel hybrid modeling strategies were applied to a full-scale industrial wastewater treatment plant and the prediction performance of each model was evaluated and compared. First, a mechanistic model was developed based on ASM1 and the specific process knowledge of the cokes wastewater treatment plant. The mechanistic model could not adequately account for the dynamics of the process, but could predict only an approximation. Then, the mechanistic model was combined with non-parametric models in parallel configuration. The non-parametric

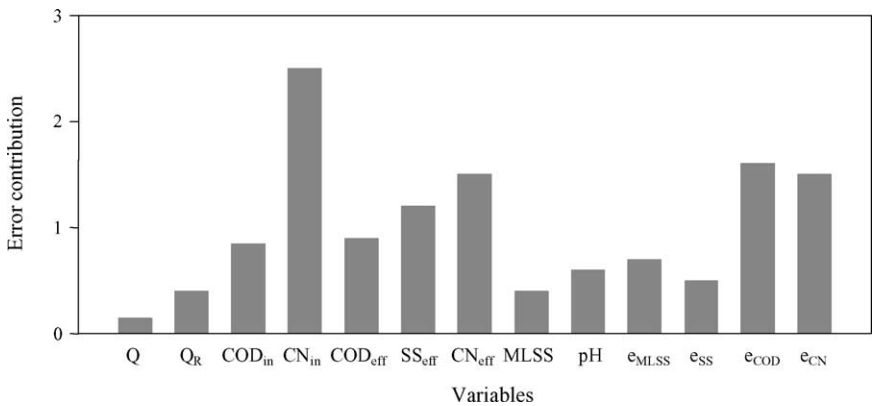


Fig. 8. Contribution plot to operation data point 250.

models evaluated include FBNN, RBFN, PLS, QPLS and NNPLS. All parallel hybrid models could improve the mechanistic model's prediction performance with available operational data. The parallel hybrid models with FBNN and RBFN method could fit the training data very well, but could not adequately predict validation data, compared with the hybrid models with QPLS and NNPLS. The hybrid PLS model was not able to sufficiently capture the inherently non-linear characteristics of the cokes wastewater treatment plant. In this study, the hybrid NNPLS model achieved the best prediction performance benefiting from the inclusion of a non-linear mapping between the input and output latent variables. In addition, the hybrid NNPLS method could analyze process behavior effectively and allow for fault detection and isolation. The proposed parallel hybrid modeling strategy is a cost-effective and accurate tool that can be applied to biological wastewater treatment processes in the absence of reasonably accurate process models.

Acknowledgements

The work was financially supported by both Brain Korea 21 Program and the Korea Science and Engineering Foundation through the Advanced Environmental Biotechnology Research Center at Pohang University of Science and Technology. This research was also supported by the Fund for Scientific Research – Flanders (F.W.O.) and the Ghent University Research Fund.

References

Baffi, G., Martin, E.B., Morris, A.J., 2000. Non-linear dynamic projection to latent structures modelling. Chemom. Intell. Lab. Syst. 52, 5–22.

Baughman, D.R., Liu, Y.A., 1995. Neural Networks in Bioprocessing and Chemical Engineering. Academic Press, San Diego.

Box, G.E.P., Draper, N.R., 1987. Empirical Model-Building and Response Surfaces. Wiely, New York.

Bro, R., 1995. Algorithm for finding an interpretable simple neural network solution using PLS. J. Chemom. 9, 423–430.

Carlin, M., Kavli, T., Lillekjendlie, B., 1994. A comparison of four methods for non-linear data modeling. Chemom. Intell. Lab. Syst. 23, 163–177.

Demuth, H., Deale, M., 2001. Neural network toolbox for use with MATLAB. The MathWorks, Massachusetts.

Draper, N.R., Lin, D.K.J., 1990. Small response-surface designs. Technometrics 32, 187–194.

Gallinari, P., Thiria, S., Badran, F., Fogelman-Soulie, F., 1991. On the relations between discriminant analysis and multi-layer perceptrons. Neural Networks 4, 349–360.

Garson, G.D., 1991. Interpreting neural-network connection weights. AI Expert 6, 46–51.

Gaudy A.F.Jr., Gaudy, E.T., Feng, Y.J., Brueggemann, G., 1982. Treatment of cyanide waste by the extended aeration process. J. Water Pollut. Cont. Fed. 54, 153–164.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal. Chim. Acta 185, 1–17.

Haykin, S., 1999. Neural Networks. Prentice-Hall, Englewood Cliffs, NJ.

Henze, M., Grady Jr., C.P.L., Gujer, W., Marais, G.V.R., Matsuo, T., 1987. Activated Sludge Model No. 1. IWA scientific and technical report No. 1, London.

Lee, D.S., Jeon, C.O., Park, J.M., Chang, K.S., 2002. Hybrid neural network modelling of a full-scale industrial wastewater treatment process. Biotechnol. Bioeng. 78, 670–682.

Lee, D.S., Park, J.M., 1999. Neural network modelling for on-line estimation of nutrient dynamics in a sequentially-operated batch reactor. J. Biotech. 75, 229–239.

Lee, M.W., Park, J.M., 1998. Biological nitrogen removal from coke plant wastewater with external carbon addition. Water Environ. Res. 70, 1090–1095.

Leonard, T., Hsu, J.S.J., 1999. Bayesian Methods. Cambridge University Press, New York.

Main, I.G., Leonard, T., Papasouliotis, O., Hatton, C.G., Meredith, P.G., 1999. One slope or two? Detecting statistically significant breaks of slope in geophysical data with application to fracture scaling. Geophys. Res. Lett. 26, 2801–2804.

Montague, G., Morris, J., 1994. Neural-network contributions in biotechnology. Trends Biotechnol. 12, 312–324.

Psichogios, D.C., Ungar, L.H., 1992. A hybrid neural network-first principles approach to process modeling. AIChE J. 38, 1499–1511.

Qin, S.J., McAvoy, T.J., 1992. Non-linear PLS modelling using neural networks. Comput. Chem. Eng. 23, 395–411.

Richards, D.J., Shieh, W.K., 1989. Anoxic-Oxic activated-sludge treatment of cyanides and phenols. Biotechnol. Bioeng. 33, 28–32.

Seher, T., Main, I.G., 2004. A statistical evaluation of a stress-forecast earthquake. Geophys. J. Int. 157, 187–193.

Standard Methods for Examination of Water and Wastewater. American Public Health Association/American Water Works Association/Water Environment Federation, Washington, DC.

Tettch, J., Metcalfe, E., Howells, S.L., 1996. Optimisation of radial basis and backpropagation neural networks for modelling auto-ignition temperature by quantitative-structure property relationships. Chemom. Intell. Lab. Syst. 32, 177–191.

Thompson, M., Kramer, M.A., 1994. Modeling chemical processes using prior knowledge and neural networks. AIChE J. 40, 1328–1340.

van Can, H.J.L., Te Braake, H.A.B., Hellinga, C., Luyben, K.C.A.M., Heijnen, J.J., 1997. An efficient model development strategy for

bioprocesses based on neural networks in macroscopic balances. Biotechnol. Bioeng. 54, 549–566.

Wise, B.M., Gallagher, N.B., 2000. PLS toolbox version 2. 1 for use with MATLAB^{TM}. Eigenvector Research, Washington DC.

Wold, S., Kettaneh-Wold, N., Skagerberg, B., 1989. Non-linear PLS modelling. Chemom. Intell. Lab. Syst. 7, 53–65.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemom. Intell. Lab. Syst. 58, 109–130.

Zhang, Q., Reid, J.F., Litchfield, J.B., Ren, J., Chang, S.-W., 1994. A prototype neural network supervised control system for Bacillus thuringiensis fermentations. Biotechnol. Bioeng. 43, 483–489.

Zhao, H., Hao, O.J., McAvoy, T.J., 1999. Approaches to modeling nutrient dynamics: ASM2, simplified model and neural nets. Wat. Sci. Tech. 39, 227–234.