Chemical Engineering Journal 105 (2004) 61–69

Chemical Engineering Journal

www.elsevier.com/locate/cej

Application of steady-state and dynamic modeling for the prediction of the BOD of an aerated lagoon at a pulp and paper mill Part II. Nonlinear approaches

Karla Patricia Oliveira-Esquerre[a,*], Dale E. Seborg[b], Milton Mori[a], Roy Edward Bruns[c]

[a] *DPQ/FEQ/UNICAMP, P.O. Box 6066, 13081-970 Campinas, SP, Brazil*
[b] *Department of Chemical Engineering, University of California, Santa Barbara, CA 93106, USA*
[c] *IQM/UNICAMP, P.O. Box 6154, 13083-970 Campinas, SP, Brazil*

Received 3 March 2004; received in revised form 20 May 2004; accepted 27 June 2004

Abstract

Neural networks can provide effective predictive models for complex processes that are poorly described by first principle models, such as wastewater biological treatment systems. In this paper multilayer perceptron (MLP) and functional-link neural networks (FLN) are developed to predict inlet and outlet biochemical oxygen demand (BOD) of an aerated lagoon operated by International Paper of Brazil. In Part I, predictive models for both inlet and outlet BOD for the aerated lagoon were developed using linear multivariate regression techniques. For the current case study, MLP networks are the best choice for the prediction models. When only a relatively small number of samples is available, substantial improvement in inlet and outlet BOD prediction is shown for both FLN and MLP modeling using a reduced input variable set that was generated using partial least squares (PLS). Thus, this paper provides a novel approach for developing PLS–FLN model structures.

Keywords: Biochemical oxygen demand; Modeling; Artificial neural networks; Aerobic process; Bioprocess monitoring; Wastewater treatment

1. Introduction

Neural networks have been widely used for extracting information from data in the form of predictive input–output models, or as a way to efficiently represent the input space. One of their advantages is that they provide a very general framework which can, in principle, approximate any type of nonlinearity in the data [1]. Multivariate statistical methods provide alternative to neural networks. Nevertheless, multivariate statistical methods have mainly been used to establish linear relationships between variables, thus restricting their applicability and for small regions of low nonlinearity between predictor and predictive variables.

Although a neural network is an empirical model that is obtained by fitting data, some notable differences exist between neural networks and typical empirical models. As a result, neural networks offer distinct advantages in some areas but have some limitations in others, as shown in Section 2.

Developments over the last several years have provided significant insight into the nature of neural learning through proof of mathematical properties [2] and examination of the relationship between neural learning and mathematical approximation theory [3,4]. However, these efforts have focused on only neural or statistical methods, and the few cases for which there is cross-fertilization between the two involve applications to only analytical functions [5,6] or to simulated process data [7]. The present research provides a unique unifying view that evaluates both neural networks and linear multivariate regression techniques in a well-documented industrial case study.

* Corresponding author.
E-mail addresses: karla@feq.unicamp.br (K.P. Oliveira-Esquerre), seborg@engineering.ucsb.edu (D.E. Seborg), mori@feq.unicamp.br (M. Mori), burns@iqm.unicamp.br (R.E. Bruns).

1385-8947/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cej.2004.06.012

The primary goal of this research is to construct accurate models to predict inlet and outlet BOD of an aerated lagoon operated by International Paper of Brazil. Linear steady-state and dynamic models have been constructed and their results are presented in a companion paper [8]. Here, functional-link (FLN) and multilayer perceptron (MLP) neural networks are used as nonlinear modeling techniques for BOD prediction. Both techniques are attractive, due to their widely recognized abilities to learn nonlinear relations between input and output data. The potential improvement for the FLN and MLP techniques is also evaluated when partial least squares (PLS) is used to preprocess their input data.

The structure of this paper is as follows. In Section 2 some advantages and drawbacks of neural networks and statistical methods are presented. A brief description of the industrial data and the methodology used is given in Section 3. In Section 4 the results of neural network modeling are reported; in this section the sample prediction accuracy of linear models and neural networks are also compared for the case study. Some conclusions are presented in Section 5.

2. Neural network versus multivariate linear regression modeling approaches

Because neural networks are massively parallel, they have a better filtering capacity and generally perform better than traditional linear models with noisy or incomplete data [9]. The trained model may be continuously adapted to new data without the need of storing previous data. Neural networks can be designed to periodically update their input–output performances, resulting in continuous, on-line, self-correcting models. Most applications of linear statistical methods are for static, non-adaptive modeling, although modifications for recursive and adaptive modeling have been devised [5,9]. Neural networks can arbitrarily approximate any nonlinear input–output relationship and can map many independent variables with as many dependent variables as needed, while most traditional liner modeling tools map at most three dependent variables [9]. However, as neural networks function essentially as black boxes, the interpretation of these models is often difficult and gives limited physical insight into the data [10]. On the other hand, linear multivariate statistical methods provide physically interpretable models. The algorithms used for determining the model parameters for large data sets build the models in a stepwise manner and have guaranteed convergence.

Even though traditional linear multivariate statistical methods are unable to capture highly nonlinear behavior, the models are usually not continuously adapted to new data. However, based on the assumption that the underlying nonlinear relationship can be locally approximated by a linear model, multivariate regression techniques may be used to approximate complex relationships in small intervals of the predictor variables [11]. Partial least squares (PLS) regression has been shown to be a powerful linear regression technique for problems in which the data are noisy and highly correlated and in which there is a limited number of observations [12].

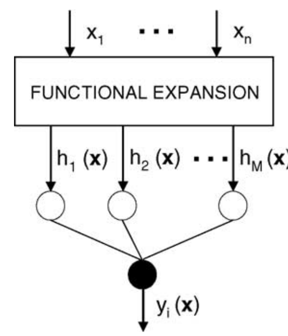

Fig. 1. General structure of a functional-link neural network.

The main, well-known disadvantages of neural network training are that it requires large quantities of experimental data and the training of the network can take too long to be practical. Furthermore, its nonlinear approximation function can cause local minimum problems.

Multilayer perceptron neural networks (MLP) have been successfully used in modeling biological wastewater treatment processes [12–20]. However, functional-link neural networks (FLN) [10] have received relatively little attention.

The FLN is a neural network with no hidden layers that is trained using supervised learning. The input is “enhanced” by generating additional terms via a transformation rule, such as a polynomial expansion. The idea is to increase the dimensionality of the feature space without requiring any additional information. These enhanced values are passed to a summation node (the output), which transforms these weighted values via a nonlinear activation function. Accordingly, the main advantages of the FLN are that its use not only simplifies the network architecture and the training algorithm but can also improve network performance [21].

A general structure of a FLN is shown in Fig. 1. x and y_i (x) are the input and output vectors, respectively, and $h(x)$ is the new vector generated by the functional expansion of the input space of dimension n_1 onto a new space of increased dimension, M ($M > n_1$). The input–output relationship of the FLN is shown in

$$y_i(x) = \sum_{j=1}^{M} w_{ij} h_j(x), \quad 1 \leq i \leq M \tag{1}$$

The theoretical foundations of FLN are described elsewhere [1,22,23]. Henrique [24] proposed a modification of the FLN structure, where the output given by Eq. (1) is transformed by an invertible nonlinear activation function. The merits of this network output transformation can be found in Costa et al. [21,22], Harada et al. [25] and Henriques et al. [26].

Table 1
Simple description of data sets 1 and 2

Parameters	Description	Data set 1	Data set 2
Predicted variables			
BOD_{in}	Inlet wastewater BOD (mg/L)	√	√
BOD_{out}	Outlet wastewater BOD (mg/L)	√	√
Predicted variables			
COD	Inlet wastewater COD (mg/L)	√	√
COL	Color (mg/L)	√	√
COND	Conductivity (μS/cm at 20 °C)	√	√
FR	Inlet flow rate (m^3/day)	√	√
NAM	Inlet ammonia concentration (mg/L)	–	√
NN	Inlet nitrate concentration (mg/L)	–	√
PAP	Paper production (t/day)	√	√
pH	pH	√	√
PULP	Pulp production (t/day)	√	√
RF	Rainfall (mm/day)	–	√
T	Temperature (°C)	√	√
TSS	Inlet total suspended solids (mg/L)	–	√

Data set 1 consists of 1094 samples (782 samples plus 312 test samples) and the ratios of data for learning, validation and test are 4:1:2. Data set 2 consists of a total of 79 samples and the ratio of learning data is 4:1.

3. Methodology

In this paper, FLN and MLP models are developed, validated and tested using both data sets 1 and 2, described in Part I [8]. Table 1 provides a simple description of these data sets. As mentioned in Part I [8], data set 1 was used to develop dynamic models. Two test sets, each with 156 samples, were obtained by linearly interpolating portions of the original industrial data that were initially excluded due to missing values.

The FLN inputs are expanded with a second-degree polynomial to generate nonlinear monomials, and the orthogonal least-squares estimator proposed by Billings et al. [27] is used to calculate the network weights and to eliminate the monomials which are not significant in explaining the output variance. This approach reduces the size and complexity of the neural network and avoids overfitting the data [21,24]. The sigmoid activation transfer function T (Eq. (2)) is used to transform the network output (Eq. (1)):

$$y = f(T) = \log\left(\frac{T}{1-T}\right) \quad \text{and} \quad T = \sum_{j=1}^{M} w_{ij} h_j(x) \qquad (2)$$

MLP training is carried out using the standard backpropagation algorithm. The sigmoidal function is used as the transfer function in both the hidden and output layers. The delta–bar–delta (DBD) technique [28] is used to determine the best ANN configuration, namely, the optimum number of nodes in the hidden layer. The relative performance of different MLP networks is compared by evaluating the mean squared error using an independent validation data set, and the network having the smallest error is selected. The performance of the selected network is then confirmed by measuring its performance on a third-independent set of data called the test set. A stop criterion based on the mean square error (MSE) for the validation data set, instead of the training data set, ensures model generalization.

The PLS technique is used to reduce the number of input variables in order to prune MLP and FLN modeling structures. NeuroSolutions Professional, a commercially available neural network software package, and a MATLAB computer program developed by Henrique [24] are used for MLP and FLN modeling, respectively.

4. Discussions and results

4.1. Modeling data set 1

The BOD prediction results are presented in terms of the multiple correlation coefficient (R^2, in %) and mean square error (MSE) for one validation and two test data sets, each with 156 points from data set 1.

First- and second-order FLN models were initially constructed for inlet BOD and their results are shown in Tables 2 and 3, respectively. The corresponding results for outlet BOD are shown in Tables 4 and 5. The number of input variables used to develop each model is given right after its name. These results indicate that using PLS to reduce the number of network inputs improves the accu-

Table 2
Validation and test results of the first-order FLN models and data set 1 (inlet BOD)

Model	Validation		Test				$\bar{R}^2 \pm s$	$\overline{\text{MSE}} \pm s$[a]
	R^2	MSE[a]	R^2	MSE[a]	R^2	MSE[a]		
Steady-state modeling								
FLN-8 (p = 9)	46.1	25.8	44.8	20.5	54.8	18.2	48.6 ± 5.4	21.5 ± 3.9
PLS–FLN-8 (p = 9)	46.4	22.1	45.7	23.0	54.5	13.6	48.9 ± 4.9	19.6 ± 5.2
PLS–FLN-3 (p = 4)	46.7	22.4	45.2	23.4	53.2	13.9	48.4 ± 4.3	19.9 ± 5.2
Dynamic modeling								
FLN-16 (p = 17)	44.7	24.6	45.0	19.9	55.4	15.5	48.4 ± 6.1	20.0 ± 4.6
PLS–FLN-16 (p = 17)	45.9	24.9	46.2	27.9	55.6	13.7	49.2 ± 5.5	22.2 ± 7.5
PLS–FLN-3 (p = 4)	45.2	24.3	44.5	28.4	54.8	13.8	48.2 ± 5.8	22.2 ± 7.5

p: number of parameters estimated during training.
[a] These results have to be multiplied by 10^{-4} to obtain MSE values for normalized data.

Table 3
Validation and test results of the second-order FLN models and data set 1 (inlet BOD)

Model	Validation		Test				$\bar{R}^2 \pm s$	$\overline{\mathrm{MSE}} \pm s$[a]
	R^2	MSE[a]	R^2	MSE[a]	R^2	MSE[a]		
Steady-state modeling								
FLN-8 (p = 45)	34.6	33.5	41.2	36.0	46.3	23.0	40.7 ± 5.9	30.8 ± 6.9
PLS–FLN-8 (p = 45)	37.6	39.0	47.5	24.2	38.0	18.8	41.0 ± 5.6	27.3 ± 10.5
PLS–FLN-3 (p = 10)	47.3	20.7	43.3	21.6	54.9	13.7	48.5 ± 5.9	18.7 ± 4.3
Dynamic modeling								
FLN-16 (p = 153)	26.6	62.7	10.8	92.4	22.9	33.6	20.1 ± 8.3	62.9 ± 29.4
PLS–FLN-16 (p = 153)	27.5	54.2	10.8	83.3	24.1	33.3	20.8 ± 8.8	56.9 ± 25.1
PLS–FLN-3 (p = 10)	47.4	23.3	43.2	27.4	53.9	13.9	48.2 ± 5.4	21.5 ± 6.9

p: number of parameters estimated during training.
[a] These results have to be multiplied by 10^{-4} to obtain MSE values for normalized data.

Table 4
Validation and test results of the first-order FLN models and data set 1 (outlet BOD)

Model	Validation		Test				$\bar{R}^2 \pm s$	$\overline{\mathrm{MSE}} \pm s$[a]
	R^2	MSE[a]	R^2	MSE[a]	R^2	MSE[a]		
Steady-state modeling								
FLN-8 (p = 9)	37.6	20.6	27.4	26.1	38.8	25.2	34.6 ± 6.3	24.0 ± 3.0
PLS–FLN-8 (p = 9)	42.2	21.7	30.1	26.9	39.7	33.2	37.3 ± 6.4	27.3 ± 5.8
PLS–FLN-3 (p = 4)	41.3	23.1	27.9	28.0	38.7	34.8	36.0 ± 7.1	28.6 ± 5.9
Dynamic modeling								
FLN-16 (p = 17)	39.2	20.6	29.8	25.2	44.8	25.0	37.9 ± 7.6	23.6 ± 2.6
PLS–FLN-16 (p = 17)	41.2	21.0	34.9	23.1	47.8	39.2	41.3 ± 6.5	27.8 ± 10.0
PLS–FLN-3 (p = 4)	41.7	21.6	30.6	22.6	41.6	21.3	38.0 ± 6.4	21.8 ± 0.7

p: number of parameters estimated during training.
[a] These results have to be multiplied by 10^{-4} to obtain MSE values for normalized data.

racy of the second-order FLN models. For first-order FLN, modest improvement in the accuracy of validation and test data sets is apparent using the PLS technique to preprocess inputs.

MLP models were developed next. The variation in MSE as a function of the number of hidden neurons was used to identify the best MLP and PLS–MLP topologies. The ratio of learning data to validation data was 4:1. Tables 6 and 7 show the best results and model structures for inlet and outlet BOD, respectively; the number of input variables used to develop each model is given right after its name. In this case, PLS did not improve the MLP nonlinear mapping.

Despite the use of the validation data set to tune the main MLP design parameter (number of nodes in the hidden layer), significant differences between validation and test results are only observed in MLP models and outlet BOD prediction.

A closer examination of Tables 2–7 reveals that for both FLN and MLP approaches, dynamic modeling gives the best results for outlet BOD prediction, in contrast to inlet BOD. The relationship between the outlet BOD and the inlet variables is indeed influenced by the biological complexity and physical structure of the aerated lagoon. These results agree with earlier results obtained using linear regression techniques [8].

Table 5
Validation and test results of the second-order FLN models and data set 1 (outlet BOD)

Model	Validation		Test				$\bar{R}^2 \pm s$	$\overline{\mathrm{MSE}} \pm s$[a]
	R^2	MSE[a]	R^2	MSE[a]	R^2	MSE[a]		
Steady-state modeling								
FLN-8 (p = 45)	36.3	26.0	28.2	36.1	36.0	32.9	33.5 ± 4.6	31.7 ± 5.2
PLS–FLN-8 (p = 45)	45.3	24.7	30.4	46.7	41.2	32.2	39.0 ± 7.7	34.5 ± 11.2
PLS–FLN-3 (p = 10)	47.1	22.1	34.3	34.2	40.2	40.4	40.5 ± 6.4	32.2 ± 9.3
Dynamic modeling								
FLN-16 (p = 153)	22.2	67.8	10.4	47.6	5.5	87.2	12.7 ± 8.6	67.5 ± 19.8
PLS–FLN-16 (p = 153)	18.6	64.8	21.1	55.0	14.5	118.0	18.1 ± 3.3	79.3 ± 33.9
PLS–FLN-3 (p = 10)	47.2	21.4	34.8	21.2	47.2	21.2	43.1 ± 7.2	21.3 ± 0.1

p: number of parameters estimated during training.
[a] These results have to be multiplied by 10^{-4} to obtain MSE values for normalized data.

Table 6
Validation and test results of the MLP models and data set 1 (inlet BOD)

Model	Validation		Test				$\bar{R}^2 \pm s$	$\overline{\mathrm{MSE}} \pm s$[a]
	R^2	MSE[a]	R^2	MSE[a]	R^2	MSE[a]		
Steady-state modeling								
MLP-8 (n = 10, p = 101)	60.1	9.0	59.8	11.1	65.3	10.3	61.7 ± 3.1	10.1 ± 1.1
PLS–MLP-8 (n = 4, p = 41)	46.9	12.6	44.4	16.7	54.5	13.8	48.6 ± 5.3	14.4 ± 2.1
PLS–MLP-3 (n = 10, p = 51)	44.2	13.6	40.9	22.3	53.0	17.2	46.0 ± 6.3	17.7 ± 4.4
Dynamic modeling								
MLP-16 (n = 8, p = 145)	52.0	11.1	55.5	12.3	55.6	13.4	54.4 ± 2.1	12.3 ± 1.2
PLS–MLP-16 (n = 4, p = 73)	51.0	13.4	47.2	15.5	42.1	17.8	46.8 ± 4.5	15.6 ± 2.2
PLS–MLP-3 (n = 10, p = 51)	46.6	12.4	37.1	18.2	56.0	13.2	46.6 ± 9.5	14.6 ± 3.1

p: number of parameters estimated during training; n: number of neurons in hidden layer.
[a] These results have to be multiplied by 10^{-4} to obtain MSE values for normalized data.

Table 7
Validation and test results of the MLP models and data set 1 (outlet BOD)

Model	Validation		Test				$\bar{R}^2 \pm s$	$\overline{\mathrm{MSE}} \pm s$[a]
	R^2	MSE[a]	R^2	MSE[a]	R^2	MSE[a]		
Steady-state modeling								
MLP-8 (n = 8, p = 81)	49.8	16.6	30.7	24.5	46.1	23.6	42.2 ± 10.1	21.6 ± 4.3
PLS–MLP-8 (n = 8, p = 81)	37.8	22.0	25.1	32.5	33.1	32.8	32.0 ± 6.4	29.1 ± 6.2
PLS–MLP-3 (n = 6, p = 31)	39.7	43.3	29.8	59.1	35.6	57.3	35.0 ± 5.0	53.2 ± 8.6
Dynamic modeling								
MLP-16 (n = 3, p = 55)	56.7	14.4	35.6	23.7	47.9	23.1	46.7 ± 10.6	20.4 ± 5.2
PLS–MLP-16 (n = 6, p = 109)	49.4	17.5	32.7	27.4	31.0	31.0	37.7 ± 10.2	25.3 ± 7.0
PLS–MLP-3 (n = 6, p = 31)	43.4	44.1	35.2	35.2	49.2	33.1	42.6 ± 7.0	37.5 ± 5.8

p: number of parameters estimated during training; n: number of neurons in hidden layer.
[a] These results have to be multiplied by 10^{-4} to obtain MSE values for normalized data.

Considering both validation and test results, the steady-state MLP model with 10 hidden nodes gives the best predictions for inlet BOD. For outlet BOD prediction, test results indicate that the dynamic MLP with three nodes and PLS–FLN with three LVs are the best models. However, validation results indicate that MLP gives the best outlet BOD prediction performance.

Comparisons of the best models for inlet and outlet BOD are shown in Fig. 2 for the validation and test sets. Both the slope and the correlation coefficient for the straight lines indicate good agreement between predictions and the data. Figs. 3 and 4 compare predicted and measured values for both models and for the validation and test sets. It can be seen that these models can reproduce the overall variation observed in the bioprocess for the period considered, especially for the case of inlet BOD prediction. More than 90 and 85%, for inlet and outlet BOD, respectively, of the data contained relative deviations of calculated from measured values smaller than 10% for both output variables. The time series plots of the residuals do not appear to have any systematic structure, indicating that the models fit the data well. There does seem to be a slight tendency for negative deviations in Fig. 2 for small

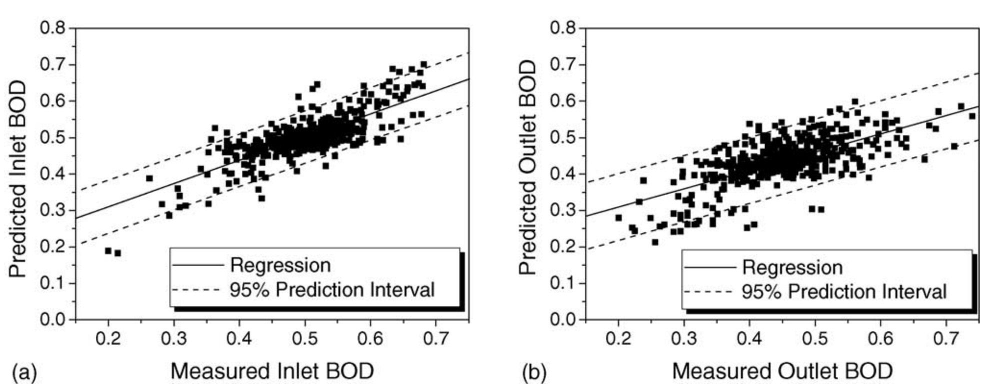

Fig. 2. Relation between predicted vs. measured BOD (solid line). Upper and lower dashed lines indicate the 95% prediction estimation interval according to (a) steady-state MLP model with 10 nodes for inlet BOD and (b) dynamic MLP model with three nodes for outlet BOD.

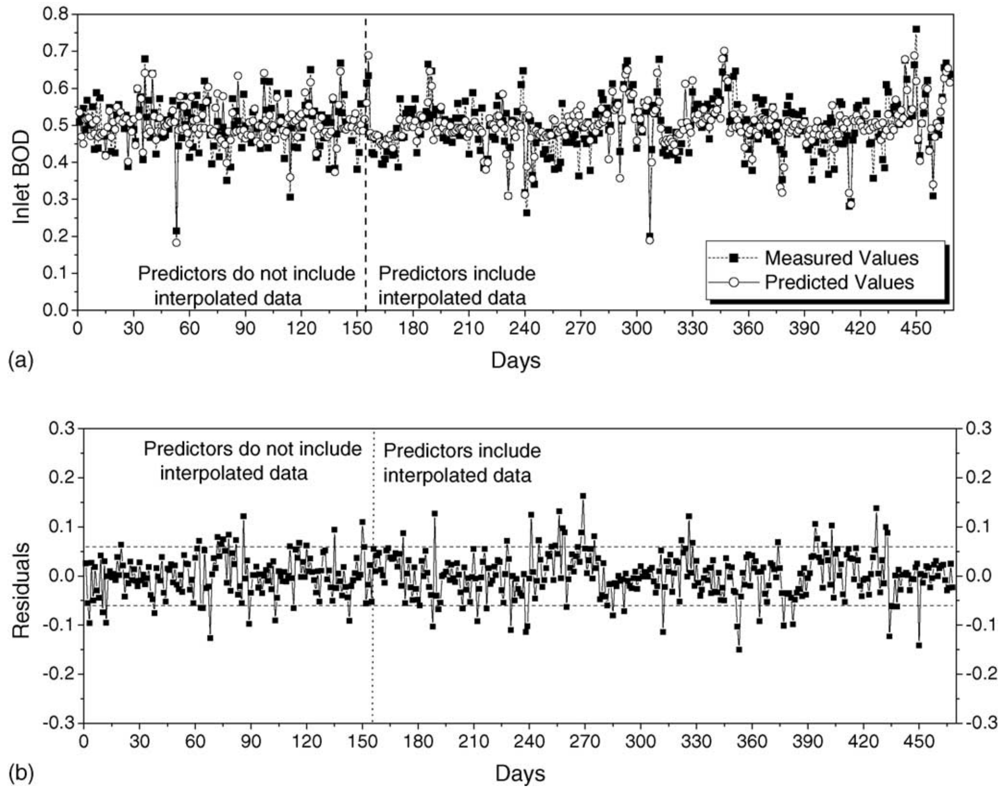

Fig. 3. Time series plot of (a) measured and predicted inlet BOD for steady-state MLP model with 10 nodes and (b) residuals—upper and lower dashed lines indicate the 95% confidence interval.

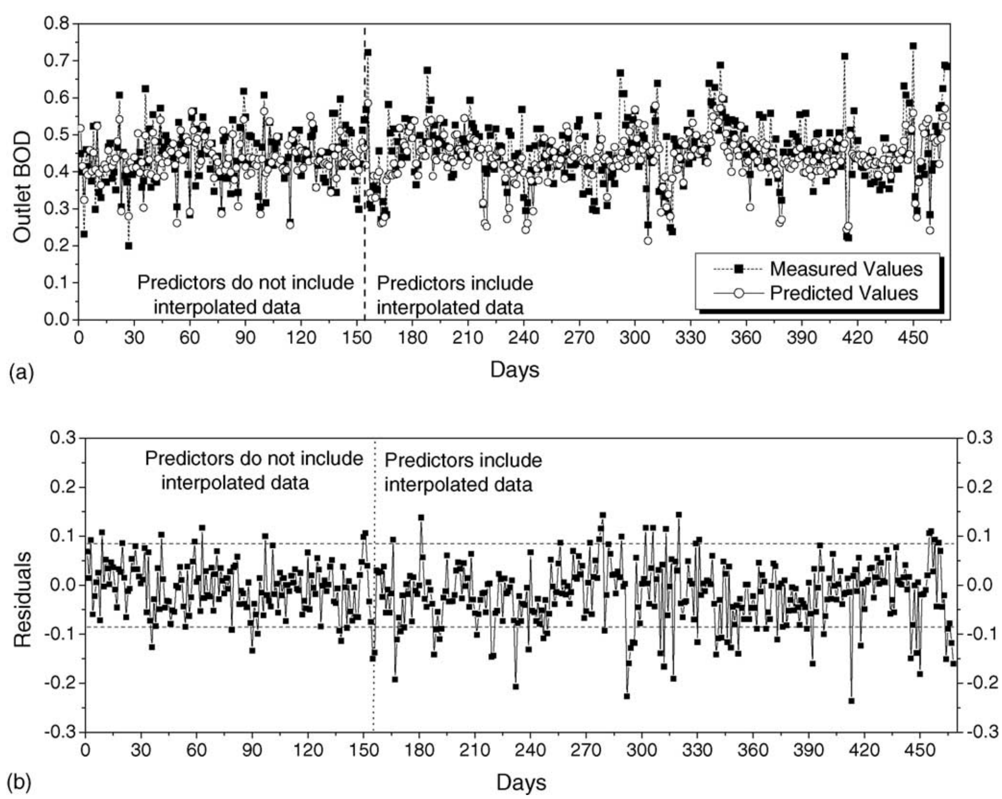

Fig. 4. Time series plot of (a) measured and predicted outlet BOD for dynamic MLP model with three nodes and (b) residuals—upper and lower dashed lines indicate the 95% confidence interval.

Table 8
Validation results of the FLN and MLP models and data set 2 (inlet and outlet BOD)

	Inlet BOD			Outlet BOD		
	Model	R^2	MSE[a]	Model	R^2	MSE[a]
Steady-state models from data set 2	FLN-12 (p = 13)[b]	48.7	15.2	FLN-12 (p = 13)[b]	57.7	18.3
	PLS–FLN-12 (p = 13)[b]	48.7	15.2	PLS–FLN-12 (p = 13)[b]	68.4	19.2
	PLS–FLN-4 (p = 5)[b]	48.2	14.6	PLS–FLN-4 (p = 5)[b]	71.3	21.2
	FLN-12 (p = 27)[c]	0.2	105.0	FLN-12 (p = 31)[c]	7.2	26.9
	PLS–FLN-12 (p = 29)[c]	16.2	62.8	PLS–FLN-12 (p = 31)[c]	9.9	87.2
	PLS–FLN-4 (p = 15)[c]	38.3	33.8	PLS–FLN-4 (p = 15)[c]	63.7	23.6
	MLP-12 (n = 3, p = 43)	66.0	10.5	MLP-12 (n = 3, p = 43)	66.0	4.1
	PLS–MLP-12 (n = 3, p = 43)	62.5	11.6	PLS–MLP-12 (n = 3, p = 43)	74.4	6.3
	PLS–MLP-4 (n = 2, p = 13)	67.5	9.7	PLS–MLP-4 (n = 6, p = 37)	74.8	5.3
Best model from data set 1	MLP-8 (n = 10, p = 101)	81.3	6.9	MLP-16 (n = 3, p = 55)	69.7	6.8

p: number of parameters estimated during training; n: number of neurons in hidden layer.
[a] These results have to be multiplied by 10^{-4} to obtain MSE values for normalized data.
[b] First-order FLN.
[c] Second-order FLN.

values of measured inlet and outlet BOD. This may indicate that a different model is necessary to provide more accurate predictions for small BOD values.

4.2. Modeling data set 2

The FLN and MLP results for inlet and outlet BOD are shown in Table 8. The number of input variables used to develop each model is given right after its name. The prediction results are presented in terms of the multiple correlation coefficient (R^2, in %) and mean square error (MSE) for one validation data set that consisted of 15 samples from data set 2. The prediction results for this reduced data set are also shown for the best inlet and outlet BOD models developed from data set 1.

The performance of both first-order FLN and MLP models is substantially improved for outlet BOD prediction using the reduced PLS input data set. On the other hand, first-order FLN and MLP models show modest improvement in inlet BOD prediction in the same situation. The most notable improvement in R^2 and MSE occurs for second-order FLN models. In this case, the FLN method is not very good when the original variables are its inputs, but its prediction performance is significantly improved when only the most significant LVs are used.

The best validation results are obtained using the PLS–MLP model with two and six hidden nodes for inlet and outlet BOD prediction, respectively. Nevertheless, the best model obtained for data set 1, MLP-8, results in better inlet BOD predictions than all of the models developed from data set 2.

For outlet BOD, the PLS–MLP model with six nodes for data set 2 shows better results than those obtained with the best model for data set 1. In any event, it is surprising that the models constructed with such a reduced number of samples can predict reliable values of inlet and outlet BOD, especially in this case in which only historical data are available.

4.3. Neural network versus multiple linear regression approaches

Tables 2 and 3 of Part I [8] show the results obtained for inlet and outlet BOD prediction, respectively, using multivariate regression techniques and data set 1, while Table 4 of Part I [8] shows the results for data set 2.

Considering the results obtained using data set 1, it can be observed that the MLR model performs slightly better than any FLN model for inlet BOD prediction. The same result occurs when the most important LVs are used for either PLS or FLN modeling. For outlet BOD prediction, no significant difference is observed between MLR and the FLN models; but the PLS–FLN models of first- and second-order perform slightly better than the PLS model, mainly for dynamic modeling.

The most significant improvement in inlet and outlet BOD prediction is observed when MLP modeling is used instead of MLR. PLS–MLP models with three LVs only perform better than PLS models for outlet BOD prediction.

For data set 2, the PLS approach for input reduction provided better results for neural networks, but was less successful for linear regression techniques.

It should be noted that, in theory, using an infinite number of independent variables to explain the change in a dependent variable would result in an R^2 of one for the modeling data set. In other words, the R^2 value can be manipulated and should be suspect. The adjusted R^2 [29] value can be used as an attempt to correct this shortcoming, because it will not always increase when additional model parameters are added. In contrast to R^2, the adjusted R^2 only increases if the additional model parameters improve the regression results significantly in order to compensate for the increase in regression degrees of freedom. Nevertheless, there is no similar statistical parameter to perform reliable comparative analyses of the predictive performances of neural network models, and the methods proposed in the literature usually lead to contradictory results. As is done in this paper, the comparative analysis of statistical and neural network models should be based on R^2 values estimated from the validation (and test) data sets.

5. Conclusions

This research was motivated by the complexity of modeling bioprocesses of industrial wastewater treatment. Artificial neural networks were developed through the use of universal approximators (sigmoidal MLP networks) or linear combinations of expanded input variables (monomers at FLN) for modeling the inner mapping between input and output variables.

Comparing the FLN and MLP approaches with the approaches proposed in this paper, PLS–FLN and PLS–MLP, respectively, when a considerable number of samples is available (data set 1), it can be seen that, although PLS did not improve the MLP nonlinear mapping, it did improve the second-order FLN for both outlet and inlet BOD prediction. No significant differences are observed between FLN and PLS–FLN results using first-order monomers.

One of the biggest challenges of this research was to deal with the high incidence of missing values in the recorded data, a common situation for industrial data records. When only a small number of data are available (data set 2), the combined use of PLS and neural networks has shown to provide prediction results that have statistical parameters significantly higher than those obtained using these techniques separately. For this case study, the PLS–MLP approach was the best of any method evaluated, but there still appears to be room for improvement.

The results reported in this paper show the improvement in modeling capabilities achieved on using MLP neural networks instead of just the linear multivariate regression techniques considered in Part I. Not surprisingly, improvements in outlet BOD prediction using dynamic instead of steady-state modeling, as opposed to the inlet BOD prediction, is clearly observed. These results agree with those obtained by the linear multivariate regression techniques of Part I.

Caution is required when neural networks models (or any other empirical models) are extrapolated beyond the range of the data analyzed or to other systems, because microbial activities and most wastewater quality parameters of biological treatment systems are site-specific variables.

Finally, it should be noted that the empirical modeling techniques developed in this research can be used to model other complex systems, including other types of industrial processes.

Acknowledgements

The authors would like to thank FAPESP (Proc. No. 99/10257-0) for their financial support, International Paper of Brazil for providing the industrial data, and Edson Guaracy Lima Fujita and Aline C. Costa (UNICAMP) and Fred Loquasto III (UCSB) for their valuable comments.

References

[1] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (1989) 359–366.
[2] K. Funahashi, On the approximate realization of continuous mappings by neural networks, Neural Networks 2 (1989) 183–192.
[3] U. Utojo, B.R. Bakshi, Neural Networks in Bioprocessing and Chemical Engineering, Appendix: Connections Between Neural Networks and Multivariate Statistical Methods: An Overview, Academic Press, 1995.
[4] T. Poggio, F. Girosi, Networks for approximation and learning, Proc. IEEE 9 (1990) 1481–1497.
[5] B.R. Bakshi, U. Utojo, A common framework for the unification of neural, chemometric and statistical modeling methods, Anal. Chim. Acta 384 (1999) 227–247.
[6] B.R. Bakshi, R. Chatterjee, Unification of neural and statistical methods as applied to materials structure–property mapping, J. Alloys Compd. 279 (1998) 39–46.
[7] L.A.C. Meleiro, Projeto e aplicação de controladores baseados em modelos lineares, neurais e nebulosos, Ph.D. Dissertation, School of Chemical Engineering, University of Campinas (UNICAMP), Campinas, SP, Brazil, 2002.
[8] K.P. Oliveira-Esquerre, D.E. Seborg, R.E. Bruns, M. Mori, Application of steady-state and dynamic modeling for the prediction of BOD for an aerated lagoon at a pulp and paper mill. I. Linear approaches, Chem. Eng. J., in press.
[9] D.R. Baughman, Neural Networks in Bioprocessing and Chemical Engineering, Academic Press, 1995.
[10] S. Chen, S.A. Billings, Neural networks for nonlinear dynamic system modelling and identification, Int. J. Contr. 56 (1992) 319–346.
[11] G. Baffi, E.B. Martin, A.J. Morris, Non-linear projection to latent structures revisited (the neural network PLS algorithm), Comp. Chem. Eng. 23 (1999) 1293–1307.
[12] M. Cote, B.P.A. Grandijean, P. Lessard, J. Yhibault, Dynamic modeling of the activated sludge process: improving prediction using neural networks, Water Res. 29 (1995) 995–1004.
[13] M. Häck, M. Köhne, Estimation of wastewater process parameters using neural networks, Water Sci. Technol. 33 (1996) 101–115.
[14] C.A. Gontarski, P.R. Rodrigues, M. Mori, L.F. Prenem, Simulation of an industrial wastewater treatment plant using artificial neural networks, Comp. Chem. Eng. 24 (2000) 1719–1723.
[15] M.F. Hamoda, I.A. Al-Ghusain, A.H. Hassan, Integrated wastewater treatment plant performance evaluation using artificial neural networks, Water Sci. Technol. 40 (1999) 55–65.
[16] D.S. Lee, J.M. Park, Neural network modeling for on-line estimation of nutrient dynamics in a sequentially operated batch reactor, J. Biotechnol. 75 (1999) 229–239.
[17] K.P. Oliveira-Esquerre, M. Mori, R.E. Bruns, Simulation of an industrial wastewater treatment plant using artificial neural networks and principal components analysis, Braz. J. Chem. Eng. 19 (2002) 365–370.
[18] H. Pu, Y. Hung, Use of artificial neural networks: predicting trickling filter performance in a municipal wastewater treatment plant, Environ. Manage. Health 6 (1995) 16–27.
[19] S.J. Wilcox, D.L. Hawkes, F.R. Hawkes, A.J. Guwy, A neural network, based on bicarbonate monitoring, to control anaerobic digestion, Water Res. 29 (1995) 1465–1470.

[20] H. Zhao, O.I. Hao, A.S.C.E. Fellow, T.J. McAvoy, C.H. Chang, Modeling nutrient dynamics in sequencing batch reactor, J. Environ. Eng. 123 (1997) 311–319.
[21] A.C. Costa, A.S.W. Henriques, T.L.M. Alves, R. Maciel Filho, E.L. Lima, A hybrid neural model for optimization of fed-batch fermentation, Braz. J. Chem. Eng. 16 (1999) 53–63.
[22] A.C. Costa, T.L.M. Alves, A.W.S. Henriques, R. Maciel Filho, E.L. Lima, An adaptative optimal control scheme based on hybrid neural modelling, Comp. Chem. Eng. 22 (Suppl.) (1998) S859–S862.
[23] Y.H. Pao, Adaptive Pattern Recognition and Neural Networks, Addison-Wesley, Reading, MA, 1989.
[24] H.M. Henrique, PhD Dissertation, Chemical Engineering Program/COPPE/UFRJ, Rio de Janeiro, RJ, Brazil, 1999.
[25] L.H.P. Harada, A.C. Costa, R. Maciel Filho, Hybrid neural modeling of bioprocesses using functional link networks, Appl. Biochem. Biotechnol. 98 (2002) 1009–1024.
[26] A.W.S. Henriques, A.C. Costa, T.L.M. Alves, E.L. Lima, Optimization of fed-batch processes: challenges and solutions, Braz. J. Chem. Eng. 16 (1999) 171–177.
[27] S.A. Billings, S. Chen, M.J. Korenberg, Identification of MIMO nonlinear systems using a forward-regression orthogonal estimator, Int. J. Contr. 49 (1989) 2157–2189.
[28] R.A. Jacobs, Increased rates of convergence through learning rate adaptation, Neural Networks 1 (1998) 295–307.
[29] D.C. Montgomery, E.A. Peck, Introduction to Linear Regression Analysis, Wiley, New York, 1992.