Chemical Engineering Journal 104 (2004) 73–81

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 I. Linear approaches

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

[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 30 December 2003; received in revised form 17 February 2004; accepted 6 May 2004

Abstract

Accurate well-timed measurement of quality variables is essential to the successful monitoring and controlling of wastewater treatment systems. Because the measurements of these variables are difficult and often involve large time delays, predictive models for target quality variables have been widely considered. However, many microbial reactions and their interactions with the environment result in time dependent processes, making the development of bioprocess models difficult and time-consuming. In this paper, steady-state and dynamic predictive models based on multiple linear regression (MLR) and partial least squares (PLS) regression are presented. Water quality measurements and process information are used to develop models to predict biochemical oxygen demand (BOD) at the inlet and outlet of an aerated lagoon of a pulp and paper mill operated by International Paper of Brazil (IPB). The results show that linear steady-state and dynamic models are able to predict inlet and outlet BOD even for a complex process that has operational data limitations (imprecise measurements, a large number of missing values, etc.). A companion paper [Chem. Eng. J., submitted for publication] reports static and dynamic nonlinear models that were developed from the same 4 years of data using a neural network approach. Together, the two papers provide a well-documented application of linear and nonlinear empirical modeling techniques to an industrial case study. The modeling techniques are also valid for other types of industrial applications.

Keywords: Biochemical oxygen demand; PLS modeling; Linear multivariate regression techniques; Aerobic process; Bioprocess monitoring; Wastewater treatment

1. Introduction

Environmental concerns, manifested in changing market demands and more stringent environmental regulations, are among the most important incentives for technological change in the pulp and paper industry. As environmental restrictions tighten, many industrial wastewater treatment plant operators are being required to comply at levels that seriously challenge the capabilities of their plants.

In aerobic treatment systems, aerobic bacteria use oxygen to degrade organic compounds. For the system to function properly, many variables must be controlled. Dissolved oxygen levels (biochemical oxygen demand (BOD) and chemical oxygen demand (COD)), pH and nutrient levels (ammonia and phosphorus) are among the most critical. Although wastewater quality parameters can be measured by laboratory analyses, a significant time delay in the range of minutes to a few days is usually unavoidable. This lack of suitable process variable information in a timely manner limits the effective control of effluent quality [2,3].

To overcome these problems, deterministic and empirical models have been developed to estimate hard-to-measure process variables. Not surprisingly, the modeling traditionally used for bioprocesses, based on mass balance equations together with rate equations for microbial growth, substratum consumption and formation of products, has been shown to be inefficient for describing these mechanisms in wastewater treatment processes [3–6]. Therefore, in these cases when models based on first principles are not available, or requires excessive computation time, empirical models become attractive alternatives [7]. Linear empirical models are

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

1385-8947/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cej.2004.05.011

Nomenclature

BOD_{in}	inlet wastewater BOD (mg/L)
BOD_{out}	outlet wastewater BOD (mg/L)
COD	inlet wastewater COD (mg/L)
COL	color (mg/L)
COND	conductivity (μS/cm at 25 °C)
FR	inlet flow rate (m^3/day)
NAM	inlet ammonia concentration (mg/L)
NN	inlet nitrate concentration (mg/L)
pH	pH
PAP	paper production (t/day)
PULP	pulp production (t per day)
RF	rainfall (mm per day)
T	wastewater temperature (°C)
TSS	inlet total suspended solids (mg/L)

usually obtained by applying modeling techniques such as linear multivariate regression. These techniques have been successfully used to approximate complex relationships over small intervals of the predictor variables [8]. The underlying assumption is that the nonlinear behavior can be locally approximated by a linear model.

The main objective of this research is to develop an estimation model that provides accurate predictions of the BOD of inlet and outlet streams of an aerated lagoon at a pulp and paper mill operated by International Paper of Brazil (IPB). Steady-state and dynamic predictive models have been developed based on both multiple linear regression (MLR) and partial least squares (PLS) regression approaches. Here, the advantage of both techniques to model a complex and multivariate process is verified. The development of nonlinear models based on neural networks is the subject of a comparison paper [1].

This paper is organized in five sections. In Section 2 the process and the available process data are described. In Section 3 some concepts of the MLR and PLS approaches are briefly reviewed. The results obtained are given in Section 4. Finally, Section 5 presents the conclusions.

2. An aerated lagoon case study

Wastewater from International Paper of Brazil is routed for preliminary treatment followed by biological treatment. Two parallel settling tanks, provided with mixing and flocculation chambers, constitute the primary treatment. Biological treatment consists of one aerated and five nonaerated lagoons. As a secondary treatment facility, the aerated lagoon is used to remove organic load and suspended solids contained in the wastewater from the milling process. After treatment in a drying system, the solids removed are used as fertilizer.

Ten process variables for the aerated lagoon and two of the pulp and paper mill were chosen based on engineering judgment regarding which variables might be important for BOD prediction. Time series plots of the variables on a daily basis and their basic statistical properties are shown in Fig. 1 and Table 1, respectively. The symbols and their units are defined in nomenclature.

The original data base covered a period of 1427 consecutive days, about a 4-year daily record. However, the extremely high incidence of missing values for many variables is a relevant problem, specially for the TSS, NAM and NN variables, where missing values are more frequent than not.

Throughout the 4-year period, a gradual decrease in flow rate (FR) followed by an increase in COD can be observed. By contrast, the BOD_{in} and BOD_{out} values exhibit a large amount of variability that makes it difficult to detect any general relationships between these variations and the COD and flow rate patterns.

The color measurement is an indirect indication of the amount of lignin compounds in the effluent. The greater the amount of lignin compounds, the darker the effluent resulting in a greater tendency to produce foam. The color values of the inlet wastewater decreased abruptly after 2 January 2000 (see Fig. 1). This decrease might have been caused by a special change in the pulp and paper manufacturing or in the primary treatment. No related changes were observed for the other wastewater quality variables.

The optimal pH value for biological wastewater treatment lies between 6.5 and 8.0. However, it can be noted in Table 1 that, even though pH average and standard deviation were 7.45 and 1.21, the measured pH varied from 0.85 to 12.53. The extremely low pH value of 0.85 occurred on 6 April 1997, and appears to be an outlier because none of the other inlet or outlet variables have unusual values. Therefore, it was concluded that this value might have been recorded incorrectly (e.g. a typing error). On the other hand, the microbial consumption seems to be affected by high pH values. For example, pH values of 12.53 and 11.56 on 17–18 February 1998, respectively, were followed by high values of outlet BOD, 167 and 187 mg/L, on 17–18 February 1998, respectively. Alkalinity variations may be caused by the presence of various inorganic and organic chemicals, like dyes, heavy metals, detergents, starches, etc., that are normally present in pulp and paper effluents. The data for the day the pH value was recorded to be 0.85 was not used to construct the BOD predictive models.

The seasonal effects of the wastewater temperature can be verified in Fig. 1. It is expected that higher temperatures are associated with higher microbial growth rates, as will be seen later. Conductivity measures water's ability to conduct an electric current and is directly related to the total amount of dissolved salts (ions) in the wastewater. Although conductivity data are temperature sensitive, the data available for this research were automatically corrected and standardized to 25 °C.

Two data sets were then constructed for inlet and outlet BOD prediction; the 2-day hydraulic residence time was used to select appropriate input/output structures for the dy-

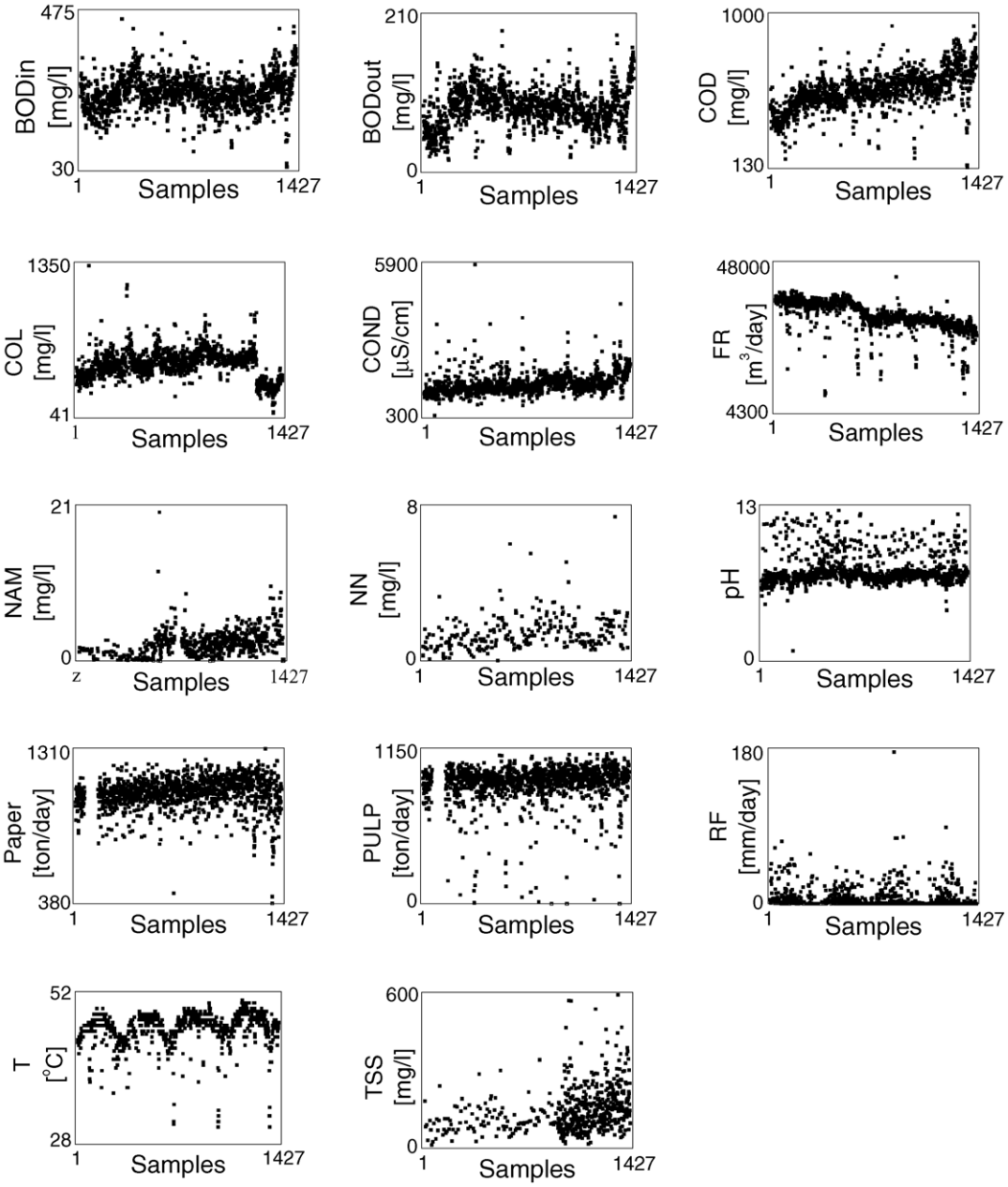

Fig. 1. Data of an aerated lagoon of a pulp and paper mill over a 4-year period (September 1996–July 2000).

namic models for outlet BOD. Data Set 1 contains only the most frequently measured variables, i.e. FR, COD, COND, COL, pH, T, PULP and PAP, whereas Data Set 2 contains all 12 variables in Table 1. Rainfall was not included in Data Set 1 owing to its high incidence of zero values. Because both data sets include only those days with actual values for the predicted and predictor variables, the 4-year daily records are reduced to 782 and 79 samples, respectively, for Data Sets 1 and 2, with each day considered as a new sample. It should be noted that, after exclusion of some samples, the percentage of missing samples of wastewater temperature decreased to 10%. Therefore, temperature values were also included in Data Set 1. Recent and past Data Set 1 values were used for dynamic modeling. For simplicity, linear interpolation was used to estimate the missing values instead of alternative, more complicated techniques [9].

Using a random selection method, about 80% of all data records of each data set were used to construct the multivariate model, while the remaining 20% were used for validation.

In order to maximize the amount of data used to test model performance on Data Set 1, the data initially excluded for having missing values were pretreated using linear interpolation. Then, two new validation data sets, each with 156 samples, were obtained and are called here as test sets. It must

Table 1
Basic statistics for the predicted and predictor variables

Parameter	Average	S.D.	Minimum	Maximum	Skewness	Kurtosis	Missing data (%)
BOD_{in}	245	46.2	41	449	0.08	1.56	6.2
BOD_{out}	85.1	25.4	16	187	0.15	0.43	6.0
COD	561	104	136	925	−0.16	1.05	6.2
COL	467	123	41	1317	0.50	3.41	3.6
COND	1529	377	379	5810	2.68	18	3.9
FR	67392	11582	4474	47850	−1.54	4.93	0
NAM	2.45	1.76	0	20	2.42	16.4	54
NN	1.44	0.88	0.03	7.38	2.42	11.0	80
PAP	1043	94.1	382.4	1304	−1.57	5.78	6.4
pH	7.45	1.21	0.85	12.53	1.79	4.17	3.7
PULP	886	155	0	1112	−3.31	14.2	7.8
RF	4.83	11.5	0	175	5.26	5	17
T	45.4	3.07	28	50.5	−2.31	8.57	33
TSS	149	85.9	12	591	1.58	4.09	60

be emphasized, however, that no more than two consecutive missing values were estimated using linear interpolation.

3. Modeling using multiple linear regression techniques

The standard multiple linear regression technique has been extended in a number of ways to address more sophisticated data analysis problems. When the regressors are few in number, are not significantly redundant (collinear) and have a well-understood relationship to the responses, then MLR can be a good way to transform data into information. However, if any of these three conditions does not hold, MLR can be inefficient or inappropriate. In these so-called soft modeling applications, the researcher is faced with many variables and poorly understood relationships, and the objective is merely to construct a good predictive model. Stepwise variable selection procedures can be valuable tools in data analysis, particularly in the early stages of building a model. However, this procedure can examine many variables and select those, which, by pure chance, have a good fit. Thus, important information may be lost with the excluded variables. Furthermore, when the stepwise procedure is automatic, it cannot take into account special knowledge the analyst may have about the data. Therefore, the model determined might not be the best from a practical point of view.

Partial least squares regression has been widely used for process monitoring and is probably the least restrictive of the various multivariate extensions of multiple linear regression. Its flexibility allows applications in situations where the use of traditional multivariate methods is severely limited, such as when there are fewer observations than predictor variables. Furthermore the predictor latent variables (LVs) are orthogonal to each other and modeling problems occurring in conventional MLR are avoided. Also, the PLS method is optimized to maximize the proportion of variance of the LVs, that is, explained by the predictor. Therefore, there is less risk that useful predictive information might be discarded or minimized than in stepwise regression.

4. Methodology

For this research, all data sets were normalized from 0.2 to 0.8 prior to analysis. Stepwise regression was used to add variables and/or remove them from the MLR model for the purposes of identifying a useful subset of predictors. Correlated variables were added and removed from the model at the $\alpha = 0.10$ (90% confidence) and $\alpha = 0.15$ (85% confidence) levels [10–12].

Outliers were identified graphically by plotting principal components analysis (PCA) and PLS scores [13,14]; newer robust multivariate methods of outlier identification are proposed by Hoo et al. [15]. The obtained results were analyzed together with the IPB process specialists who suggested that only the pH outlier data collected on 6 April 1997 should not be used to construct the models.

Unfortunately, there is no universal, automatic criterion for selection of the number of LVs for PLS models, and discrepancies between different criteria are common [16]. In this research, the appropriate number of LVs was selected by the increase in predicted accumulated variance (AV), which has to be at least 2% for adding a new LV [17].

The mean square error and coefficient of multiple determination (R^2) were calculated for calibration and validation results. *P*-values for R^2 were calculated in order to verify the significance of the multivariate regression. This test assumed a 95% confidence level.

Minitab, Matlab and Statistica computer programs were used for statistical analysis, data pretreatment, and MLR and PLS modeling.

5. Results

5.1. Modeling Data Set 1

For inlet BOD prediction using Data Set 1 as the calibration data, three LVs are optimal for both the steady-state and the dynamic models. The steady-state model explains 55.8

Table 2
$R^2$and MSE for inlet BOD and validation Data Set 1

	Model	Actual data		Interpolated data				$\bar{R}^2 \pm s$	$\overline{\text{MSE}} \pm s$[a]
		R^2	MSE[a]	R^2	MSE[a]	R^2	MSE[a]		
Steady-state models from Data Set 2	MLR-8	47.6	11.8	47.0	14.9	54.1	14.0	50.6 ± 3.9	13.5 ± 1.6
	MLR-5	45.8	12.1	46.9	14.9	54.7	13.9	49.2 ± 4.9	13.6 ± 1.4
	PLS-3	48.0	11.6	46.6	15.0	52.8	14.3	49.2 ± 3.3	13.6 ± 1.8
Dynamic models from Data Set 2	MLR-16	47.1	11.9	47.4	14.8	55.0	13.8	49.8 ± 4.5	13.5 ± 1.5
	MLR-6	47.0	11.9	47.0	14.9	53.4	14.3	46.2 ± 3.7	13.7 ± 1.6
	PLS-3	45.8	12.2	46.2	15.1	54.2	14.0	48.8 ± 4.7	13.8 ± 1.5

[a] These results have to be multiplied by 10^{-4} to obtain MSE values for normalized data.

and 42.0% of the predictor and predicted accumulated variance, while the dynamic model has predictor and predicted AV values of 44.0 and 46.7%. For outlet BOD prediction for Data Set 1, the optimal number of LVs is also three for both steady-state and dynamic models. The steady-state model explains 53.0 and 31.5% while the dynamic model explains 41.1 and 37.5% of the predictor and predicted AVs, respectively. Less than 1.6 and 1.3% increases in AVs would be obtained by adding one more LV to the inlet and outlet BOD models, respectively.

As expected, because COD is the variable best correlated with inlet and outlet BOD for both steady-state and dynamic models, the first LV of each PLS model is almost exclusively related to COD and this variable is the first one included in MLR by the stepwise method.

A variety of MLR models were considered containing as many as eight regressors. In preliminary screening tests, the most influential regressors were selected from the process variables in Table 1 using standard statistical techniques such as stepwise regression. As many as three lags were considered for the dynamic models. Only the best models are reported in this paper.

The multivariate regression models were then constructed and their prediction performances are shown in Tables 2 and 3 for inlet and outlet BOD, respectively. The *p*-value for the coefficient of multiple determination (R^2) is essentially zero for all models, indicating the statistical significance of the regressions. *s* denotes the sample standard deviation of the test and validation data sets.

Analysis of model performance was hampered by the large standard deviations for R^2 and MSE. No substantial difference between steady-state and dynamic modeling performances in inlet BOD prediction is clearly observable. Therefore, the simplest model, i.e. the steady-state MLR model with five predictors, is chosen as the best one and shown in the following equation:

$$\text{BOD}_{\text{in}}(t) = 0.02 + 0.61\text{COD}(t) + 0.23\text{FR}(t) - 0.15\text{T}(t) + 0.12\text{pH}(t) + 0.10\text{PULP}(t) \quad (1)$$

For outlet BOD, better performance is obtained using dynamic modeling. The MLR model with six predictors gives slightly better predictions for outlet BOD and is the simplest dynamic model obtained. It is shown in the following equation:

$$\text{BOD}_{\text{out}}(t+2) = -0.365 + 0.42\text{COD}(t) + 0.35\text{COD}(t-1) + 0.45\text{FR}(t) + 0.25\text{FR}(t+1) - 0.232\text{T}(t+1) + 0.381\text{COND}(t) \quad (2)$$

As expected, the COD terms have the largest positive coefficients in both inlet and outlet BOD models, followed by flow rate. Although the COD and FR data are autocorrelated both current and past values, are important for outlet BOD predictions, that is, these terms have *p*-values lower than 0.05 for the statistical tests of the regressors.

Not surprisingly, because microbial activity slows when the temperature declines, the regression models indicate negative relations between the predictors and both inlet and outlet BOD. As mentioned before, the microbial population is also pH sensitive, that is, microbial activity decreases when pH is not near the neutral range. Nevertheless, although pH

Table 3
$R^2$and MSE for outlet BOD and validation Data Set 1

	Model	Actual data		Interpolated data				$\bar{R}^2 \pm s$	$\overline{\text{MSE}} \pm s$[a]
		R^2	MSE[a]	R^2	MSE[a]	R^2	MSE[a]		
Steady-state models from Data Set 2	MLR-8	37.8	20.4	28.4	26.2	38.7	25.8	34.9 ± 5.7	24.1 ± 3.2
	MLR-4	39.0	20.0	28.5	26.2	38.9	25.5	35.5 ± 6.0	23.9 ± 3.4
	PLS-3	36.6	20.8	26.8	26.8	37.7	26.4	33.7 ± 6.0	24.6 ± 3.4
Dynamic models from Data Set 2	MLR-16	39.6	20.0	30.8	26.7	46.4	22.8	39.0 ± 7.8	23.2 ± 3.4
	MLR-6	40.6	19.8	31.4	26.5	46.7	22.3	39.6 ± 7.7	22.9 ± 3.4
	PLS-3	39.9	19.8	28.4	28.1	44.4	23.7	37.6 ± 8.3	23.9 ± 4.2

[a] These results have to be multiplied by 10^{-4} to obtain MSE values for normalized data.

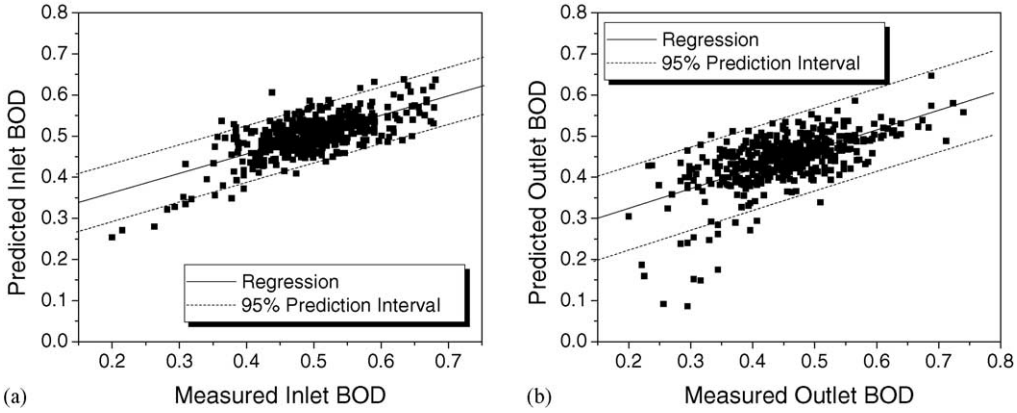

Fig. 2. Relation between predicted vs. measured BOD (solid line). Upper and lower dashed lines indicate the 95% prediction estimation interval according to (a) MLR model 8 and inlet BOD and (b) MLR model 5 and outlet BOD.

information seems to be necessary for inlet BOD predictions this is not so for outlet BOD. PULP and COND are the other variables included in the inlet and outlet BOD predictive models, respectively.

The negative constant in the BOD equation would correspond to a negative biological oxygen demand when all the variables in Eq. (2) were equal to zero. This solution is not physically reasonable. All data falls within the normalized range of about 0.2–0.8 for BOD. This indicates that a linear model will not be valid when BOD approaches the zero limit. In other words it is not possible to extrapolate this model for these small BOD values.

It should be noted that, despite the fact that PLS models contain only three LVs as predictors, estimation of their scores requires the measurement of all eight original variables, which could result in higher implementation costs when compared with the use of models for which variables have been removed. Here, it is difficult to identify the most important variables of the model because their loadings vary considerably from one LV to another (data not shown). Only

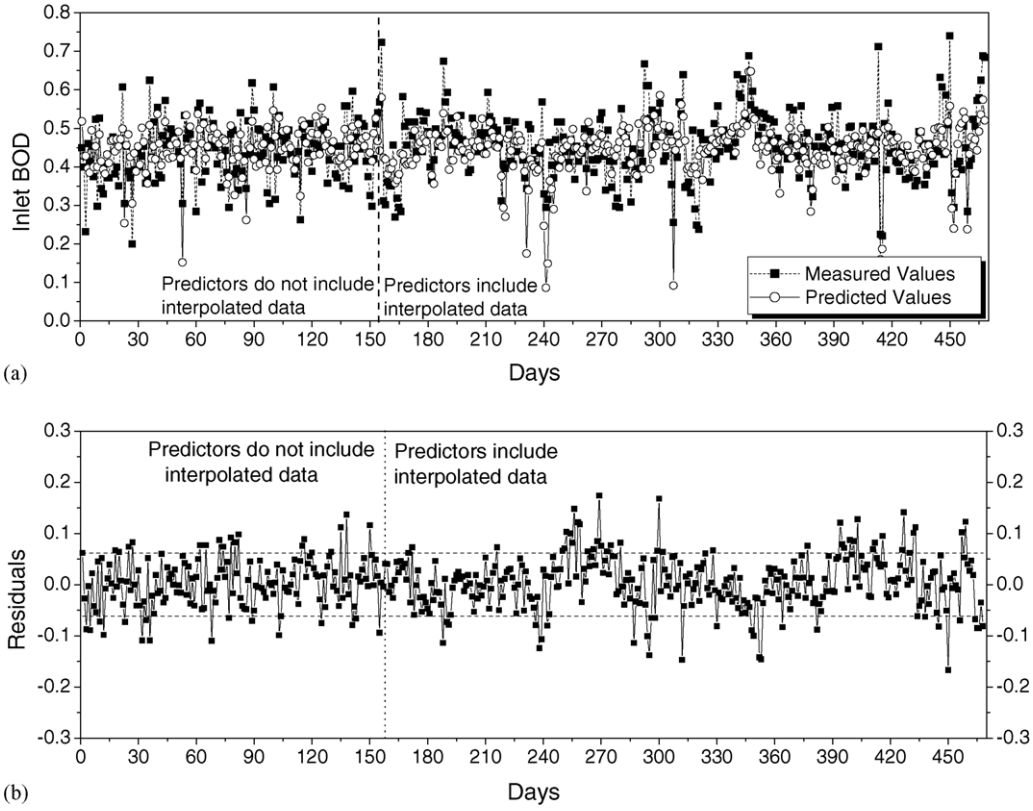

Fig. 3. Time series plot of (a) measured and predicted inlet BOD for MLR model 5 and (b) residuals—upper and lower dashed lines indicate the 95% confidence interval.

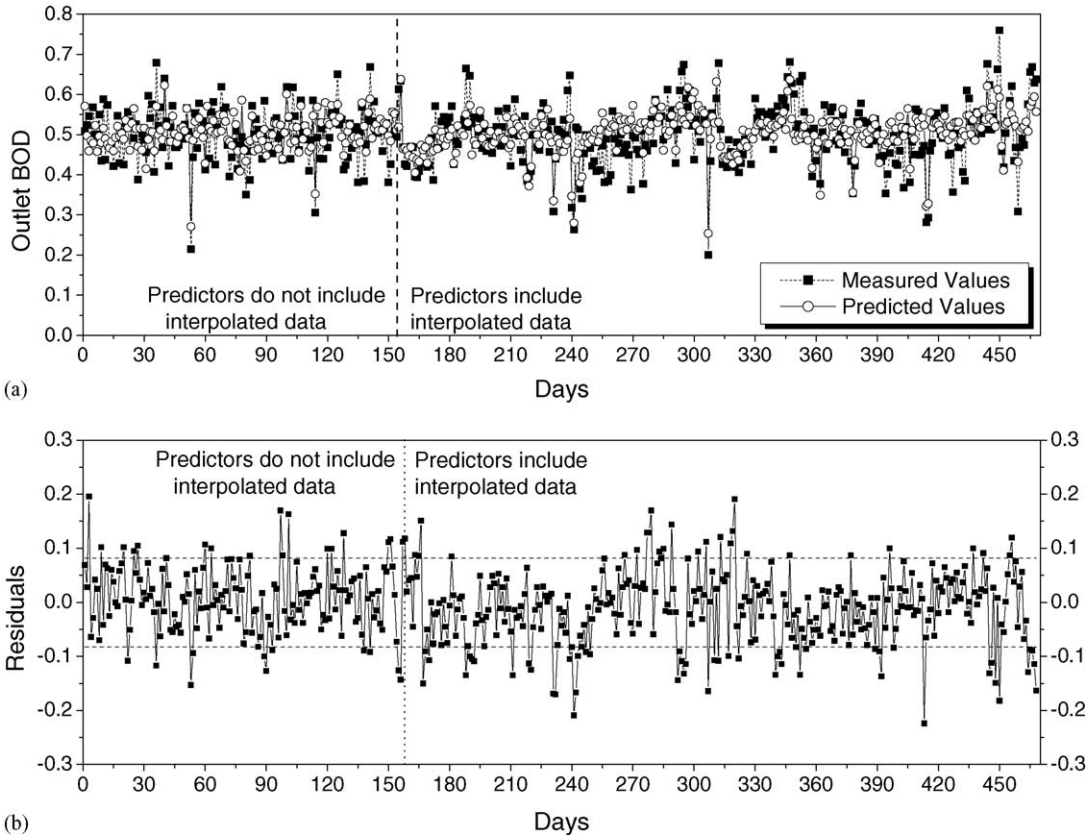

Fig. 4. Time series plot of (a) measured and predicted outlet BOD for dynamic MLR model 6 and (b) residuals—upper and lower dashed lines indicate the 95% confidence interval.

the variable color does not seem to be important to the PLS model. For problems with larger dimensionality, the use of specific approaches for identification of variable importance [14] is essential.

The plots of the predicted and measured BOD in Fig. 2 show that the best predictive models for both inlet and outlet BOD provide predictive capability and, as expected, predict most outputs within 95% prediction bands. Of course, linear models are local approximations and appear to be valid here only for measured inlet and outlet BOD above about 0.35. Below this value, the predictions in Fig. 2 exhibit large negative deviations below the 95% confidence band indicating that an alternative model passing through the origin is more appropriate for small inlet and outlet BOD values. In Figs. 2–4, the BOD residuals have been normalized, as discussed earlier.

Figs. 3 and 4 display the time series for the measured and predicted data for inlet and outlet BOD, respectively, using the best models obtained. Both inlet and outlet BOD are well reproduced even when interpolated data are used to develop the MLR models. Not surprisingly, inlet BOD validation data are slightly better reproduced than outlet BOD data. This result is accounted reasonable due to the complexity of the wastewater treatment system and the relationship between outlet BOD and the aerated lagoon inlet variables. The results indicate that more than 85% of the residuals are <0.1 for both output variables.

5.2. Modeling Data Set 2

The optimal number of LVs for both inlet and outlet BOD is four, which corresponds to 53.1 and 54.1% of the predictor AV and 66.15 and 62.77% of the predicted AV for the inlet and outlet BOD models, respectively. Only 0.18 and 0.67% increases in AV would be obtained for the inclusion of one more LV in the inlet and outlet BOD models, respectively.

As was the case for Data Set 1, the COD variable is most important for the first LV and is the first one included in MLR models by the stepwise regression method. The MLR and PLS modeling results for inlet and outlet BOD are shown in Table 4.

The R^2 and MSE results in Table 4 appear to be somewhat contradictory. While MSE values indicate that the PLS models give the best prediction results for both inlet and outlet BOD, the R^2 values indicate that the stepwise and MLR models are the best ones for inlet and outlet BOD, respectively. R^2 cannot be used as a criterion for model quality here because more model parameters are used in the MLR models than in the PLS ones.

Table 4
R^2 and MSE for inlet and outlet BOD models and validation Data Set 2

	Inlet BOD			Outlet BOD		
	Model	R^2	MSE[a]	Model	R^2	MSE[a]
Steady-state models from Data Set 2	MLR-12	57.4	13.8	MLR-12	67.6	11.8
	MLR-2	71.8	8.6	MLR-6	61.8	13.5
	PLS-4	64.8	8.0	PLS-4	65.4	10.9
Steady-state models from Data Set 1	MLR-8	70.3	10.8	MLR-8	73.9	5.7
	MLR-5	66.7	11.7	MLR-4	67.4	7.4
	PLS-3	69.1	10.5	PLS-3	77.1	4.7

[a] These results have to be multiplied by 10^{-4} to obtain MSE values for normalized data.

Comparing model performances for Data Sets 1 and 2, indicates that steady-state models for Data Set 1 give significantly better results for both inlet and outlet BOD, except for the stepwise models of inlet BOD prediction.

6. Summary and conclusions

Determination of an appropriate model structure for biological treatment systems of industrial wastewater is a formidable task. In this paper, steady-state and dynamic linear regression approaches have been evaluated for the purpose of developing models for prediction of inlet and outlet BOD of an aerated lagoon. It is not always clear whether the poor fit to the data owes to the structure of the model or to the estimation of model parameters. For this application, different models structures gave about the same performance.

To summarize, it was found that steady-state models give better prediction results for inlet BOD than for outlet BOD. Because the relationship between the outlet BOD and the inlet variables is influenced by the biological complexity and physical structure of the aerated lagoon, information on the aerated lagoon dynamics was found to be indeed important and necessary for the predictive models for outlet BOD. No significant differences were observed between MLR, stepwise and PLS model performances. Nevertheless, stepwise models were chosen as the best ones for inlet and outlet BOD prediction because of their simplicity.

Finally, there is no doubt that multiple linear regression techniques provide physically interpretable models and can be satisfactorily used to monitor inlet and outlet BOD behavior in relation to other aerated lagoon variables, as long as enough calibration data are available. It should be emphasized that specific chemical compounds in the wastewater, which may act in either a stimulatory or an inhibitory manner, can influence the microbial activity in the aerated lagoon, and the quality wastewater parameters can be strongly influenced by environmental conditions. Consequently, extrapolations beyond the range of the data analyzed here must be done with considerable caution because the resulting predictions are likely to be inaccurate. However, the empirical modeling techniques employed in this research are also applicable to other types of complicated industrial processes.

Acknowledgements

The authors wish to acknowledge FAPESP (Proc. No. 99/10257-0) for financial support, International Paper of Brazil for providing the industrial data, and especially Jeremy Conner (UCSB) for fruitful discussions during the development of this research.

References

[1] 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. Part II. Nonlinear approaches, Chem. Eng. J., submitted for publication.
[2] P. Harremoës, A.G. Capodaglio, B.G. Hellstrom, M. Henze, K.N. Jensen, A. Lynggaard-Jensen, R. Otterpohl, H. Soeborg, Wastewater treatment plants under transient loading-performance, modeling and control, Water Sci. Technol. 27 (1993) 71–115.
[3] 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.
[4] 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.
[5] 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.
[6] J.P. Steyer, D. Rolland, J.C. Bouvier, R. Moletta, Hybrid fuzzy neural network for diagnosis—application to the anaerobic treatment of wine distillery wastewater in a fluidized bed reactor, Water Sci. Technol. 36 (1997) 209–217.
[7] S. Park, C. Han, A nonlinear soft sensor based on multivariate smoothing procedure for quality estimation in distillation columns, Comp. Chem. Eng. 24 (2000) 871–877.
[8] 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.
[9] A.C. Atkinson, T.-C. Cheng, On robust linear regression with incomplete data, Comput. Stat. Data Anal. 33 (2000) 361–380.
[10] N.R. Draper, H. Smith, Applied Regression Analysis, 3rd ed., Wiley, New York, 1998.

[11] D.C. Montgomery, E.A. Peck, Introduction To Linear Regression Analysis, Wiley Press, New York, 1992.
[12] M. Sjostrom, S. Wold, A multivariate calibration problem in analytical chemistry solved by partial least-squares models in latent variables, Anal. Chim. Acta 150 (1983) 61–70.
[13] H. Martens, T. Naes, Multivariate Calibration, Wiley, New York, 1979.
[14] Z.I. Stefanov, K.A. Hoo, Hierarchical multivariate analysis of cockle phenomena, J. Chemometr. 17 (2003) 550–568.
[15] K.A. Hoo, K.J. Tvarlapati, M.J. Piovoso, R. Hajare, A method of robust multivariate outlier replacement, Comp. Chem. Eng. 26 (2002) 17–39.
[16] O. Ortiz-Estarelles, Y. Martín-Biosca, M.J. Medina-Hernández, S. Sagrado, E. Bonet-Domingo, On the internal multivariate quality control of analytical laboratories. A case study: the quality of drinking water, Chem. Intell. Lab. Syst. 56 (2001) 93–103.
[17] MATLAB PLS_Toolbox 2.0, Eigenvector Research Inc., Manson, WA, 1998.