Brazilian Journal of Chemical Engineering

ISSN 0104-6632
Printed in Brazil

Vol. 19, No. 04, pp. 365 - 370, October - December 2002

SIMULATION OF AN INDUSTRIAL WASTEWATER TREATMENT PLANT USING ARTIFICIAL NEURAL NETWORKS AND PRINCIPAL COMPONENTS ANALYSIS

K.P.Oliveira-Esquerre[1], M.Mori[2*] and R.E.Bruns[3]

[1,2]Departamento de Processos Químicos, Faculdade de Engenharia Química, UNICAMP,
P.O.Box 6066, Barão Geraldo, 13081-970 Campinas - SP, Brazil
E-mail: karla@feq.unicamp.br, E-mail:mori@feq.unicamp.br,
[3]Instituto de Química, UNICAMP, P.O. Box 6154, Barão Geraldo,
13083-970, Campinas - SP, Brazil, E-mail:bruns@iqm.unicamp.br

(Received: March 5, 2002 ; Accepted: March 27, 2002)

Abstract - This work presents a way to predict the biochemical oxygen demand (BOD) of the output stream of the biological wastewater treatment plant at RIPASA S/A Celulose e Papel, one of the major pulp and paper plants in Brazil. The best prediction performance is achieved when the data are preprocessed using principal components analysis (PCA) before they are fed to a backpropagated neural network. The influence of input variables is analyzed and satisfactory prediction results are obtained for an optimized situation.
Keywords: Artificial neural networks, Principal components analysis, Wastewater treatment and Biochemical oxygen demand.

INTRODUCTION

Milling in the pulp and paper industry is a serious concern. Existing effluents may contain potentially harmful chemicals, introduced during the papermaking operation. Thus, without proper treatment, the wastewater may pollute the environment upon its return to the water source (Yang, 1996).

In recent years, computer-based methods have been applied to many areas of environmental chemistry. In the process industry the use of modern control strategies is required due to increasingly stringent regulation of effluent quality (Cohen et al., 1997; Lee and Park, 1999). Operational control of a biological wastewater treatment plant is often complicated because of variation in raw wastewater compositions, strengths and flow rates owing to the changing and complex nature of the treatment process (Hamoda et al., 1999). Moreover, a lack of suitable process variables limits the effective control of effluent quality (Harremoës et al., 1993, Lee and Park, 1999).

The modeling traditionally used in bioprocesses is based on balance equations together with rate equations for microbial growth, substratum consumption and formation of products, and since microbial reactions coupled with environmental interactions are nonlinear, time-variable and of a complex nature (Hamoda et al., 1999; Lee and Park, 1999), traditional deterministic and empirical modeling has shown some limitations (Cote et al., 1995; Hamoda et al., 1999).

Recently, some studies using artificial neural networks (ANNs) in modeling biological wastewater treatment processes have been published, providing

*To whom correspondence should be addressed

an alternative approach (Cote et al., 1995; Häck and Kohne, 1996; Gontarski et al., 2000; Hamoda et al., 1999; Lee and Park, 1999; Pu and Hung, 1995; Wilcox et al., 1995; Zhao et al., 1997).

Recent studies indicate that consideration of statistical principles in the ANN model building process may improve modeling performance (Maier and Dandy, 2000). For example, principal components analysis (PCA), which is a technique that shows an orthogonal variable transformation, can be used for pruning ANNs and improving nonlinear mapping (Kanjilal, 1995; Kompany-Zared, 1999). The use of ANNs in combination with PCA has been shown to have merits (Cancilla and Fang, 1996; Holcomb and Morari, 1992; Kompany-Zared et al., 1999). Principal component regression can be an alternative to ANN in modeling if the system shows linear behavior (Despagne and Massart, 1998).

The aim of this study is to develop an estimation model that can provide accurate predictions of the biochemical oxygen demands of the output stream of a biological wastewater treatment plant (BOD_{out}). There is a five-day delay in determination of BOD, and when this is added to the hydraulic residence time, it is often too late to make proper adjustments in the wastewater treatment process.

In this work, the predictive models presented for the estimation of BOD_{out} are calculated from ANN and Principal Component Regressions (PCR). The results show that neither principal component regression nor artificial neural network treatment is satisfactory when used separately in modeling and simulation. Neural networks present superior results for the training set but poorer ones for the test set relative to those obtained from PCR. Since there was a limited amount of experimental data available on the wastewater treatment plant at RIPASA S/A, an overfitting of the training set occurred.

The best prediction performance is achieved when the data are preprocessed using PCA before they are fed to a backpropagated neural network composed of five neurons in a hidden layer and the delta-bar-delta (DBD) learning algorithm. A correlation index between the predicted and actual effluent data using the best model is 0.81 for the training set as compared to 0.77 for the test set.

Other studies applying neural networks to actual data from chemical processes found a correlation index equal to 0.8 for a coke furnace (Blaesi and Jensen, 1992) and 0.82 for an acetic anhydride production plant (Nascimento et al., 2000). For modeling a wastewater treatment plant using neural networks, Hamoda et al. (1999) found a correlation index of 0.74 for BOD prediction, Belanche et al. (1999) found 0.504 for COD prediction and Häck and Köhne (1996) found 0.92 and 0.82 for COD and nitrate prediction, respectively.

Owing to incomplete information on the above modeling processes, the optimality of the results cannot be assessed, and it is difficult to draw meaningful conclusions about the performance of the different models. However, based on the results obtained, there is no doubt that ANNs have great potential as tools for the prediction of water resources.

PROCESS DESCRIPTION

The process wastewater is routed for primary treatment followed by biological treatment. Two parallel settling tanks, provided with mixing and flocculation chambers, constitute the primary treatment. Biological treatment consists of two aerated lagoons. The solids removal stream flows from the primary and secondary settling tanks to a drying system, where the solids are properly eliminated. A simplified process flowsheet for the wastewater treatment plant is shown in Figure 1.

Data Collection for Model Prediction

To construct the predictive model, five variables for the aerated lagoons and two for the milling process were chosen using engineering judgement regarding which ones might have an important effect on BOD_{out} prediction (see Table 1).

The average hydraulic residence time in the aerated lagoons is used to establish a data input/output relationship. The original database, obtained from the plant control system and from the laboratory, has much unusable information that must be eliminated, so the two-year daily record is reduced to 71.

ANALYSIS OF THE RESULTS

Using a random selection method, 80% of all data records were assigned to the training set, while the remaining 20% were relegated to the validation set.

As described above network 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 due to its suitable application, specially for continuous-value

input/output pairs. The quickprop (QP) and delta-bar-delta (DBD) techniques are used in determining the best ANN internal representation. The optimal hidden layer is determined by varying the total number of nodes from 1 to 10. The stop criteria are based on the Mean Square Error (MSE) for the validation set instead of that for the training set to ensure model generalization.

The best results were obtained for the ANN composed of five neurons in the hidden layer using the DBD technique, which has a correlation index of 0.60 for the validation set of predicted and observed BOD's. In order to improve this result and to prune the ANN structure, input variables are preprocessed, using the PCA technique, before they are fed to the backpropagated ANN. PCA seeks relevant directions for the input data that maximize variance.

The principal components analysis shows that 94% of the variance in the input data could be accounted for by the first five directions. This suggests the elimination of the last two principal components (PCs) in ANN modeling.

An analysis of the importance of each PC for the PCA-ANN model is carried out based on Garson's (1991) work. It also recommends exclusion of the sixth and seventh PCs owing to the low values of their calculated importance measures. As can be seen in Table 2, the best PCA-ANN performance is obtained by excluding the sixth PC from the input layer for the network composed of five hidden nodes. These results can be compared with those of multiple linear regression of the principal components (PCR), which have lower correlation indexes of 0.53 and 0.51 for the training and validation sets, respectively. Even if multiple quadratic regression of principal components (PCR^2) is performed, the correlation indexes are not as high as those for the PCA-ANN model results in Table 2.

Table 3 presents p-values for the correlation index and the F-test for analysis of variance, carried out in order to verify the significance of the ANN, PCA-ANN and PCR^2 models (Barros et al., 1995). Both tests assumed a 95% confidence level.

The statistical results show that the PCA-ANN model best adjusts to the aerated lagoons system data and is capable of predicting relatively accurate values for validation set data. This is particularly important when one takes into account the complexity of the wastewater treatment system and the large quantity of missing data in the training set.

Although the PCA-ANN model provides the most accurate results in all cases, this is not true of ANN using data that is not preprocessed by the principal component transformation. For the validation set, quadratic PCR gives slightly better results. This is probably due to overfitting of the training set by the ANN, since the validation set results are quite inferior to those of the training set. The overfitting effect appears to be less serious for the PCR results, as expected, but also for the PCA-ANN results. This is important for our application since our data set has quite a few missing values.

Figure 2 presents a graphical representation of the measured and predicted data of BOD_{out}, using the best modeling structure, i.e., PCA-ANN without the sixth PC. The most important features are well reproduced except for the large peak at data number 46 and some smaller ones, specially the one at 61 in the validation set.

The Statistica and Minitab computer programs were used for statistical analysis, PCA data pretreatment and PCR modeling. NeuroSolutions Profesional, a commercially available neural network, was used for ANN modeling.

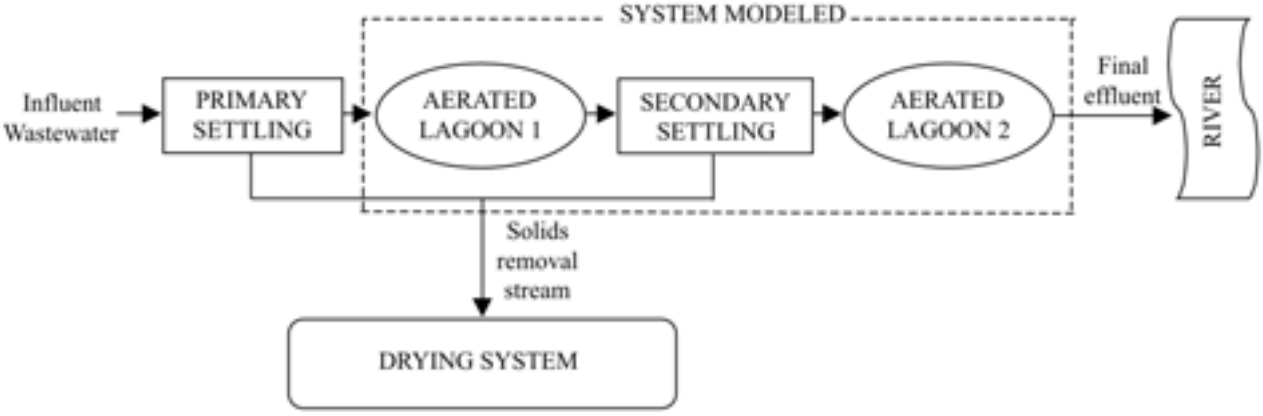

Figure 1: Process flowsheet for wastewater treatment plant at RIPASA S/A.

Table 1: Basic statistics descriptors for selected variables.

Parameter	Description	Average	Minimum	Maximum	Standard Deviation
BOD_{in}	Inlet wastewater BOD [mg/L]	547.9	106.0	814.0	117.0
COD_{in}	Inlet wastewater COD [mg/L]	1352.0	225.0	1690.0	236.3
$Flow_{in}$	Inlet flow rate [m^3/day]	1729.4	1473.0	2146.0	125.1
BOD_{out}	Outlet wastewater BOD [mg/L]	43.2	17.0	102.0	16.3
COD_{out}	Outlet wastewater COD [mg/L]	432.3	284.0	646.0	72.5
$Flow_{out}$	Outlet flow rate [m^3/day]	1621.5	1327.0	1965.0	150.5
Pulp	Pulp production [ton/day]	731.0	250.2	829.0	104.4
Paper	Paper production [ton/day]	752.6	422.0	954.0	104.3

in: input of aerated lagoons system
out: output of aerated lagoons systems

Table 2: Results of BOD_{out} prediction by PCA-ANN structure model.

	Training Data Set		Validation Data Set	
Network PCA-ANN inputs	**MSE**	**Correlation Index**	**MSE**	**Correlation Index**
All PCs	0.0072	0.76	0.0071	0.72
Excluding 6^{th} PC	0.0061	0.80	0.0065	0.77
Excluding 7^{th} PC	0.0071	0.76	0.0086	0.67
Excluding 6^{th} and 7^{th} PC	0.0068	0.77	0.0078	0.70

Table 3: Statistical results of the ANN, PCA-ANN and PCR^2 models.

	Training Data Set				Validation Data Set			
Model	**MSE**	**Correlation Index**	**p-value**	**F-test***	**MSE**	**Correlation Index**	**p-value**	**F-test***
ANN	0.0077	0.74	0.00	69.32	0.0097	0.60	0.02	7.47
PCA-ANN	0.0062	0.80	0.00	96.77	0.0065	0.77	0.00	17.09
PCR^2	0.0105	0.61	0.00	32.65	0.0088	0.64	0.01	8.57

* F-test for 1 degree of freedom in the numerator, and 12 and 55 degree of freedom in the denominator for the training and validation data set, respectively.

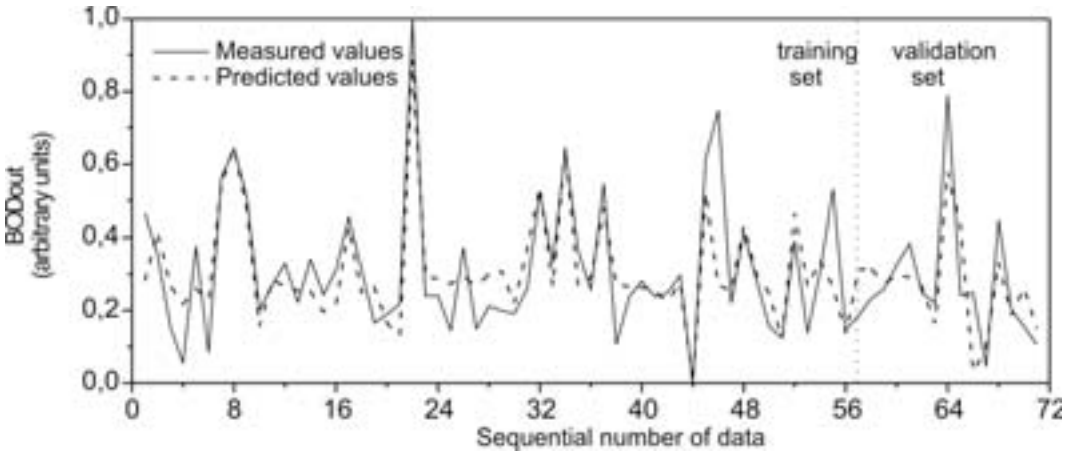

Figure 2: Measured and predicted BOD_{out}, using the PCA-ANN structure model (without sixth PC).

CONCLUSIONS

Recent studies showed that neural network models are capable of modeling a wastewater treatment system (Maier and Dandy, 2000). However, a simple feedforward back-propagation network with only a hidden layer gave an unsatisfactory performance for the simulation and prediction of the BOD_{out} for the data set treated here.

In order to improve network performance, the PCA technique was applied for data preprocessing. The combined use of PCA and ANN has been shown to provide prediction results that have statistical parameters significantly superior to those obtained using these techniques separately.

This work also shows the advantage of neural networks in their ability to represent highly nonlinear relationships, even for a system that presents operational data limitations (imprecision associated with measured variables, a limited range of variables, a large number of missing values, etc.), as long as the input data have been orthogonalized. The PCA technique helps the nonlinear ANN mapping by its orthogonal transformation of variables and reduction of system dimensionality.

Compared to the statistical methods, ANNs do provide a more general framework for determining relationships between data and do not require the specification of any functional form.

NOMECLATURE

BOD_{in}	Biochemical oxygen demand in the inlet wastewater [mg/L]
BOD_{out}	Biochemical oxygen demand in the outlet wastewater [mg/L]
COD_{in}	Chemical oxygen demand in the inlet wastewater [mg/L]
COD_{in}	Chemical oxygen demand in the outlet wastewater [mg/L]
$Flow_{in}$	Inlet flow rate [m^3/day]
$Flow_{out}$	Outlet flow rate [m^3/day]
Pulp	Pulp production [ton/day]
Paper	Paper production [ton/day]

ACKNOWLEDGMENTS

The authors are grateful to Vitaly Félix Rodríguez Esquerre, Dr. Manoel Telhada and professors Dr. Ronei Jesus Poppi and Dr. Fernando Von Zuben for fruitful discussions during the course of this work, to RIPASA S/A for providing the data set and to CNPq and FAPESP for their financial support.

REFERENCES

Barros, B. N., Scarminio, I. S., Bruns, R. E. (1995). Planejamento e otimização de experimentos. UNICAMP Press. Campinas, São Paulo, Brazil.

Belanche, L. A., Valdés, J. J., Comas, J., Roda, I. R., Poch, M. (1999). Towards a model of input-output behavior of wastewater treatment plants using soft computing techniques. Environ. Modelling & Software, 14, 409.

Blaesi, J., Jensen, B. (1992). Can neural networks compete with process calculations? Intech Applying Technology, 34.

Cancilla, D. A., Fang, X. (1996). Evaluation and quality control of environment analytical data from the Niagara River using multiple chemometric method. Journal Great Lagoons Res., 22, 241.

Cohen, A., Janssen, G., Brewster, S. D., Seeley, R., Boogert, A. A., Graham, A. A., Mardani, M. R., Clarke, N., Kasabov, N. K. (1997). Application of computational intelligence for on-line control of a sequencing batch reactor (SBR) at Morrinsville sewage treatment plant. Wat. Sci. Tech., 35, 63.

Cote, M., Grandijean, B. P. A., Lessard, P., Yhibault, J. (1995). Dynamic modeling of the activated sludge process: improving prediction using neural networks. Wat. Res., 29 (4), 995.

Despagne F., Massart, D. L (1998). Neural networks in multivariate calibration. Analyst, 123, 157.

Garson, G. D. (1991). Interpreting neural-network connection weight(s). AI Expert, 47.

Gontarski, C. A., Rodrigues, P. R., Mori, M., Prenem, L. F. (2000). Simulation of an industrial wastewater treatment plant using artificial neural networks. Comput. Chem. Eng., 24, 1719.

Häck, M., Köhne, M. (1996). Estimation of wastewater process parameters using neural networks. Water Sci. Technol., 33 (1), 101.

Hamoda, M. F., Al-Ghusain, I. A., Hassan, A. H. (1999). Integrated wastewater treatment plant performance evaluation using artificial neural networks. Water. Sci. Tech., 40 (7), 55.

Harremoës, P., Capodaglio, A. G., Hellstrom, B. G., Henze, M., Jensen, K. N., Lynggaaard-Jensen, A.,

Otterpohl, R., Soeborg, H. (1993). Wastewater treatment plants under transient loading-performance, modeling and control. Wat. Sci. Tech., 27 (12), 71.

Holcomb, T. R., Morari M. (1992). PLS/ Neural Networks. Comput. Chem. Eng., 16 (4), 393.

Kanjilal, P. P. (1995). On the application of orthogonal transformation for design and analysis of feedforward networks. IEEE Trans. Neural Networks, 6, 1061.

Kompany-Zared, M. (1999). A. Massoumi and Sh. Pezeshk-Zadeh, Simultaneous spectrophotometric determination of Fe and Ni with xylenol orange using principal component analysis and artificial neural networks in some industrial samples. Talanta, 48, 283.

Lee, D. S., Park, J. M. (1999). Neural network modeling for on-line estimation of nutrient dynamics in a sequentially-operated batch reactor. Journal of Biotechnology, 75, 229.

Maier, H. R., Dandy, G. C. (2000). Neural networks for prediction and forecasting of water resources variables: a review of modelling issues and applications. Environ. Modelling & Software, 15, 101.

Nascimento, C. A. O., Giudici, R., Guardani, R. (2000). Neural network based approach for optimization of industrial chemical process. Comput. Chem. Eng., 24, 2303.

Pu, H., Hung, Y. (1995). Use of artificial neural networks: Predicting trickling filter performance in a municipal wastewater treatment plant. Envir. Manag. Health, 6 (2), 16.

Yang, G. B. O. (1996). Managing secondary treatment systems. TAPPI Journal, 12, 52.

Wilcox, S. J., Hawkes, D. L., Hawkes, F. R., Guwy, A. J. (1995). A neural network, based on bicarbonate monitoring, to control anaerobic digestion. Wat. Res., 29 (6), 1465.

Zhao, H., Hao, O. I., Fellow, A. S. C. E., McAvoy, T. J., Chang, C. H. (1997). Modeling nutrient dynamics in sequencing batch reactor. Journal of Environ. Eng., 123 (4), 863.