Pergamon

Computers chem. Engng, Vol. 21, Suppl., pp. S751-S756, 1997

Printed in Great Britain

PII:S0098-1354(97)00140-3

0098-1354/97 $17.00+0.00

Hybrid Modelling of Biochemical Processes: A comparison with the conventional approach

S. Feyo de Azevedo[1], B. Dahm[1], F.R. Oliveira[2]

[1]Departamento de Engenharia Quimica, Faculdade da Universidade do Porto,
Rua dos Bragas, 4099 Porto Codex, PORTUGAL, E-mail - sfeyo@fe.up.pt
[2]Universidade do Minho, Electrónica Industrial, Largo do Paço, 4700 Braga, PORTUGAL

Abstract - This paper addresses attitudes and forms of process modelling in biochemical engineering. Baker's yeast production in a fed-batch fermenter, at laboratory scale, is employed as case-study. Three modelling approaches are described and compared, viz. - the conventional mechanistic approach, formulations based on different artificial neural network (ANN) topologies and a hybrid mechanistic-ANN structure. A standard 2-step procedure of model development, estimation (training) and validation with two independent sets of experiments, has been carried out. The mechanistic model, using reaction kinetic schemes from the literature, fine tuned by classical non-linear regression, gave smooth predictions for the validation data runs, but showed limited ability in predicting the test data.. The ANN were able to describe experiments at the training stage, but failed the validation (i.e. extrapolation) procedure, giving oscillatory predictions of the process state. Additionally, this approach suffers from a strong influence of the net parameters, which must be chosen by trial and error. The hybrid model predictions are good with the training and very satisfactory with the experimental test data. The indication is that the latter is a powerful tool for process modelling in biochemical engineering, particularly when limited theoretical knowledge of the process is available.

INTRODUCTION

The building and use of mathematical models based on observed data has been long accepted as a basic scientific methodology. Models may be of a more or less formal character, but they have the basic feature that they attempt to link observations together into some pattern (Ljung, 1989). With the progress in digital technology, and thinking of bringing the theory into practice, computational modelling and model-based applications have emerged and are today recognised as areas of great priority for the future (Edgar, 1996). The questions are - which models and which applications?

The classical feeling of Chemical (and Biochemical) Engineering is with the use of models based on the so-called *first principles*, which, as pointed out by Villermaux (1996), are the same of 100 years ago. Today, much for reasons of the difficulties experienced in the analysis of biochemical processes, (and because there are the technical means for other directions of studies) it is openly accepted that the quantitative, mechanistic knowledge of the present may be insufficient and may have to be compensated or complemented with other forms of knowledge - statistical, qualitative, fuzzy or eminently heuristic.

The conventional approach for process modelling is based on the balance equations for mass, energy, and, if necessary, momentum and population. This form of modelling requires further knowledge about reaction kinetics, thermodynamic, transport and physical properties. For fermentation processes the predictive ability of conventional models is quite limited. This is mainly and simply due to the intrinsic non-linear time-varying characteristics of the cell metabolism, with kinetic structures which often are only partially known, or even completely unknown.

In recent years a new attitude for modelling was introduced to biochemical engineering, based on artificial neural networks (Scott and Ray, 1993, Ye *et al.*, 1994, Montague and Morris, 1994). ANN are able, in many instances, to represent multivariable relationships, particularly those that occur in highly non-linear dynamic systems, without any knowledge of the underlying process. The network parameters are estimated by training the net with the *a priori* knowledge of the process. This appealing mathematical tool may however lead to predictions which may conflict with (violate) fundamental constraints represented by the conservation principles, particularly when outside the domain of training.

A natural form of process modelling, trying to overcome disadvantages of the two pure approaches presented, has then emerged, by combining them in a hybrid formulation (Lübbert and Simutis, 1994).

Hybrid modelling aims at including all available knowledge of the process. The foundations of the hybrid model are on the conservation principles. The poorly known or unknown properties of a process, such as the reaction kinetics, are modelled with the aid of artificial intelligence methodologies, including the a priori knowledge of the process.

The present paper is about these three approaches for process modelling, which will be presented in relation to the fed-batch production of baker's yeast at laboratory scale.

MODELLING APPROACHES

Baker's yeast fermentation and the conventional approach

Yeast growth may be characterised by three metabolic pathways -

Respiratory growth on glucose (oxidative pathway):

$$S + O \xrightarrow{\mu_s^o} X + C \tag{1}$$

Fermentative growth on glucose (reductive pathway):

$$S \xrightarrow{\mu_s^r} X + C + E \tag{2}$$

Respiratory growth on ethanol (oxidative pathway):

$$E + O \xrightarrow{\mu_e^o} X + C \tag{3}$$

where S represents glucose; O: oxygen; X: biomass; E: ethanol; C: carbon dioxide and μ_s^o, μ_s^r, μ_e^o: specific growth rates for the three pathways.

The mechanistic model for the fed-batch fermentation in fed-batch regime is obtained from mass balances for all components, considering that the reactor is well mixed. Furthermore it is assumed that the yield coefficients (Y) are constant and the dynamics of the gas phase can be neglected. Then the set of model equations is -

$$\frac{dX}{dt} = \left(\mu_s^o + \mu_s^r + \mu_e^o - D\right)X \tag{4}$$

$$\frac{dS}{dt} = \left(-\frac{\mu_s^o}{Y_{X/S}^o} - \frac{\mu_s^r}{Y_{X/S}^r}\right)X + \left(S_f - S\right)D \tag{5}$$

$$\frac{dE}{dt} = \left(\frac{\mu_s^r}{Y_{X/E}^r} - \frac{\mu_e^o}{Y_{X/E}^{oe}}\right)X - DE \tag{6}$$

$$\frac{dO}{dt} = \left(-\frac{\mu_s^o}{Y_{X/O}^o} - \frac{\mu_e^o}{Y_{X/O}^{oe}}\right)X - DO + OTR \tag{7}$$

$$\frac{dC}{dt} = \left(\frac{\mu_s^o}{Y_{X/C}^o} + \frac{\mu_s^r}{Y_{X/C}^r} + \frac{\mu_e^o}{Y_{X/C}^{oe}}\right)X - DC - CTR \tag{8}$$

where D is the dilution rate (= *F/V*, ratio feedrate/volume) and S_f is the substrate concentration in the feed.

The accumulation of the working volume during the fed-batch process is represented by -

$$\frac{dV}{dt} = DV \tag{9}$$

The gas transfer rates are given by -

$$OTR = K_L^O a\left(O^* - O\right) \tag{10}$$

$$CTR = K_L^C a\left(C - C^*\right) \tag{11}$$

where $K_L^i a$ are overall mass transfer coefficients for oxygen and carbon dioxide and O* and C* are the corresponding equilibrium concentrations.

The kinetic scheme employed is based on the model proposed by Sonnleitner and Käppeli (1996). Details on the reaction scheme, as well as the relevant kinetics data, including the yield coefficients, are presented and summarised elsewhere (Feyo de Azevedo *et al.*, 1996).

Artificial neural networks

Feedforward networks and partly recurrent networks are the two types of ANN mostly employed in chemical and biochemical engineering applications (Yet-Pole *et al.*, 1996; Cheng *et al.*, 1995).

In feedforward networks only connections are allowed that point from a subordinate layer to a layer above. The disadvantage of this network structure is the static behaviour of the net, as it only gives the combination between current input data and the related output values. Time dependencies related to the dynamics of the process can only be included by using time dependent input variables.

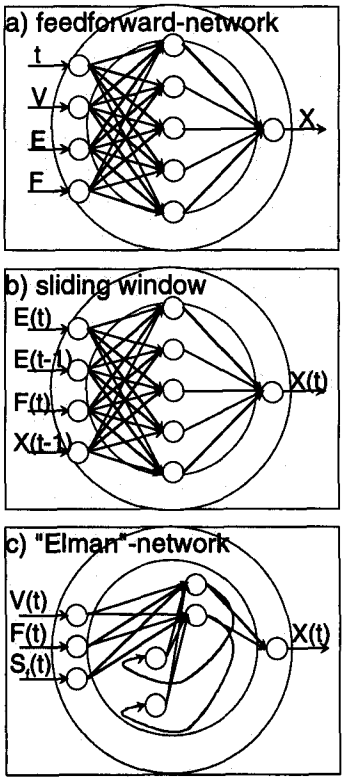

Fig.1.: Neural network structures applied in this work

Fig. 1a. shows the feedforward network used in this study. Possible input variables are feedrate, initial component concentrations, reactor volume, ethanol concentration and time. The optimum net structure, i.e. the number of input units, number of hidden layers and

number of units in the hidden layer, depends strongly on the number of available training data and must be found by trial and error.

To overcome the disadvantage of the static character of the feedforward network, partly recurrent networks (Cheng *et al.*, 1995, You and Nikolaou, 1993) were also considered. In this network structure, recurrent connections between units are allowed. Fig. 1b. shows a network structure, where the output of previous time steps is used as input to the net. In this way a time window is moved over the training data set. This structure gives a correlation between the data inside the time window and the output. But it does not consider the complete history of the data.

A more complex partly recurrent network is the Elman-net (Hagan et al., 1996) shown in Fig. 1c. Here the recurrent structure is implemented inside the net. The output of the hidden units is used as input to the so called context units, whose output serve as additional input to the hidden units. In this way it is possible to represent the dynamics of the process and the complete history of the data within the weights of the net. Furthermore the latter network type is known to have advantages for training with noisy training data, because the recurrent structure does not depend directly on the output values of the network (Zell, 1994).

All networks were trained with a modified backpropagation rule. The calculations for the feedforward network and the network with the sliding window technique were performed in MATLAB™. The calculations for the Elman-network were carried out with the neural network simulator SNNS (Zell *et al.*, 1995).

Hybrid model

For the prediction of biomass in the baker's yeast fermentation a hybrid model was developed which combines the known fundamental constraints, i.e. the mass balance, with the a priori knowledge of the process, i.e. the available measurements. The latter are included through an artificial neural network.

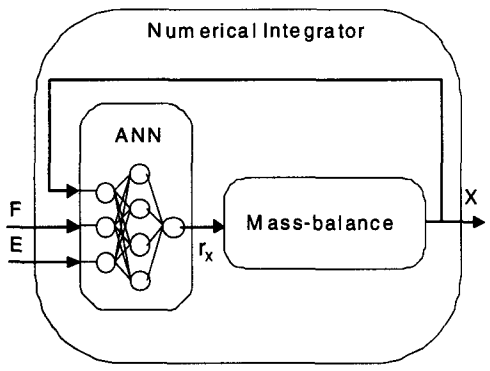

Fig. 2.: Structure of the hybrid model

The fundamental part of the hybrid model is the mass balance for the biomass, corresponding to Eq.(4) -

$$\frac{dX}{dt} = r_x X - DX, \tag{12}$$

where r_x describes the kinetics. As the reaction kinetic is not well known, this relevant information is obtained as the output of an artificial neural network The input data to the ANN can be chosen from the available measurements. Fig.2 shows the structure of the hybrid model developed.

In comparison with the model presented by Schubert *et al.* (1994), who employ a time delayed biomass input, here the actual biomass is used as input to the artificial neural net. This makes it necessary to solve the ANN simultaneously with the numerical integration of the balance equation

The main problem related to the hybrid model is on the training of the net. As the values for r_x are not known from experiments, the usual training methods such as backpropagation fail. Instead the net must be trained with the output of the balance equation, the biomass X.

The usual way of estimating the weights between the units, is to minimise the sum squared error

$$J = 0.5\sum_{k=1}^{N_p}\left(X_k - X_{k,\exp}\right)^2 \tag{13}$$

by changing the weights $w_{ij,n}$ in the direction of gradient $\partial J/\partial w_{ij,n}$

$$w_{ij,n+1} = w_{ij,n} - \eta\frac{\partial J}{\partial w_{ij,n}} \tag{14}$$

The derivation of J, Eq.(13), with respect to $w_{ij,n}$ gives

$$\frac{\partial J}{\partial w_{ij,n}} = \sum_{k=1}^{N_p}\left(X_k - X_{k,\exp}\right)\frac{\partial X_k}{\partial w_{ij,n}} \tag{15}$$

Hence it is necessary to determine the derivatives $\partial X_k/\partial w_{ij,n}$.

One possibility suggested by Schubert *et al.* (1994) is the sensitivity approach. In this approach the weights are introduced as additional variables to the differential equation -

$$\frac{\partial X}{\partial t} = f\left(X(t), y(t), w\right) \tag{16}$$

Differentiating eq. (16) and rearranging the LHS, leads to:

$$\frac{d}{dt}\left(\frac{\partial X}{\partial w_{ij,n}}\right) = \frac{\partial f}{\partial X}\frac{\partial X}{\partial w_{ij,n}} + \frac{\partial f}{\partial w_{ij,n}} \tag{17}$$

This differential equation can be solved, if the function f, the derivatives of f and the boundary conditions are known. From Eq. (12) one obtains -

$$\frac{\partial f}{\partial X} = r_X + X\frac{\partial r_X}{\partial X} - D \tag{18}$$

$$\frac{\partial f}{\partial w_{ij,n}} = X\frac{\partial r_X}{\partial w_{ij,n}} + r_X\frac{\partial X}{\partial w_{ij,n}} - D\frac{\partial X}{\partial w_{ij,n}} \tag{19}$$

The first derivatives of $r_x(X)$ in order to X and to $w_{ij,n}$ are the remaining relationships required to solve the system of ordinary differential equations. These derivatives can be obtained from the equations for the net. Therefore, the transfer-functions of every unit in the net must be differentiable for the whole range.

In the present work the ANN employed was a feedforward net with three input units, one hidden layer and one output unit. A linear transfer function for the units in the input and the output layer were chosen and a hyperbolic tangent function was chosen for the hidden layer.

RESULTS

Experimental data available

The experimental information employed in this analysis was obtained from runs carried out in a laboratory rig with a 5-litre fermenter (Oliveira, 1997), adequately equipped with analytical instruments and with a computer-based data acquisition and control unit. Biomass measurements were obtained off-line. All runs were performed in fed-batch regime (starting volumes of 2.5 litre) for between 16 and 19 hours. The strain employed was *Saccharomyces cerevisiae* H1022 (ATCC32167), operating temperature was of 30 °C and the carbon feed was pure glucose syrup with concentrations ranging from 30-50 g/l. These may be considered typical conditions at laboratory scale. 2 sets of 4 runs each were employed for the 2-step procedure of estimation (training for the nets) and validation.

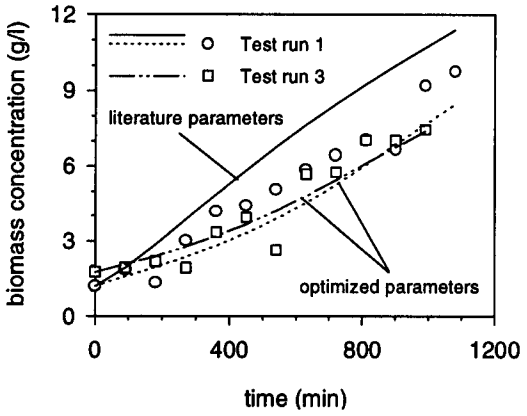

Fig.3.: Mechanistic model

The mechanistic model

Simulation results, employing the mechanistic model with parameters directly from the literature, are shown in Fig. 3 for one of the runs. Predictions are smooth but unsatisfactory. A step of parameter estimation was then performed with the 'training' data set, fine-tuning by non-linear regression the kinetic parameters (maximum specific glucose and oxygen up-take rates) to which the system output exhibits more pronounced sensitivity. Simulation was subsequently performed for the test runs. Results (Fig. 3) have somehow improved, but not significantly. When the kinetic structure is reasonably known it can be expected that classical modelling with first principle models, with adequately tuned parameters allow for smooth prediction of trends, but not for accurate predictions.

Artificial neural networks

Fig. 4 shows the predictions of the pure artificial neural network structures for two different training data sets. The agreement with the experiments is globally good for all three different ANN. The following net parameters were employed:

Feedforward net: 4 input units (time, feedrate, ethanol concentration, volume), 1 hidden layer with 6 units, one output unit.

Partly recurrent net (sliding window): 7 input units (feedrate at time t and t-1, ethanol concentration at time t and t-1, volume at time t and t-1, biomass concentration at time t-1 (previous net-output)), 1 hidden layer with 6 units, 1 output unit.

Partly recurrent net (Elman-net): 3 input units (feedrate, ethanol concentration, volume), 1 hidden layer with 6 units, 1 output unit.

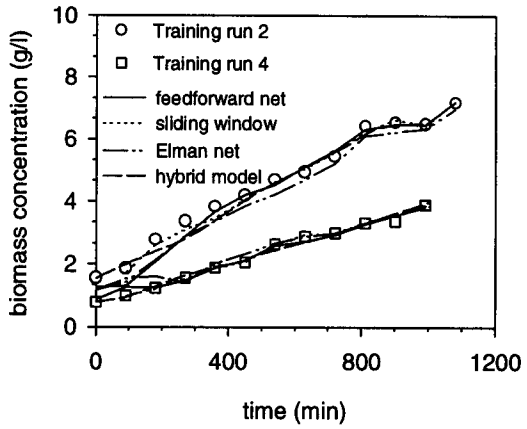

Fig. 4.: Results of ANN and hybrid modelling for training data

Fig. 5 show the results of the ANN for a typically bad test run. The ANN are not able to describe the experiment satisfactorily. It is noticed that the predictions exhibit a type of oscillatory behaviour, induced, it seems, by conditions which differ from those of the training domain, even if only for a relatively short period of time. This problem may possibly be overcome by increasing the number of training data presented to the net. It is also worth mentioning (not seen in the figure) that the results are strongly influenced by the number of parameters, i.e. the number of weights, which must be adjusted during training. It does not seem possible to give a rule for the right choice of the optimum number of units in the net, which must be found by trial and error. As a rule of thumb the number of adjustable parameters should be equal or smaller than the number of available training data. For the feedforward net and the partly recurrent net with the sliding window technique the number of

units strongly influence the extrapolation capacity of the trained net. With respect to this number of units, the Elman-net exhibited less sensitivity. Furthermore the initial weights of the net before training were seen to strongly influence the results both for the feedforward net and for the sliding window structure. This is an indication of a large number of local optima on the response surface of these two types of structures.

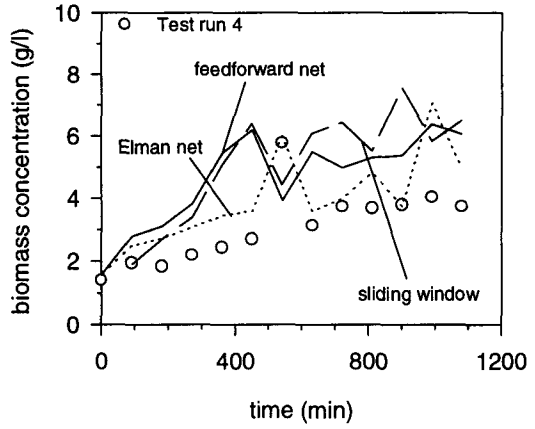

Fig. 5.: Testing biomass prediction with ANN

Hybrid modelling

The hybrid structure experiences no difficulty in reproducing the experimental information employed for training (Fig. 4). Also, in comparison with the predictions of the pure ANN the hybrid approach shows always superior performances, as illustrated in Fig. 6.

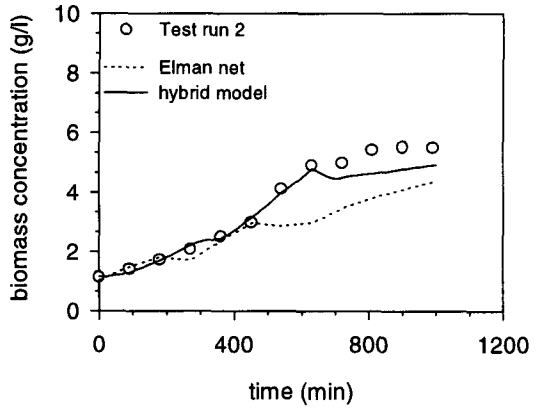

Fig. 6.: ANN and Hybrid model predictions of process behaviour

In the absence of any information about the kinetics, the hybrid approach seems to be able to predict smoothly the process behaviour, with significant good levels of accuracy (Fig. 7).

The number of units in the ANN-part of the hybrid model, as well as the initial values of the weights in the net have less influence on the accuracy of the predictions, comparatively to that observed for the pure ANN approaches. These parameters are only responsible for reaching the predefined error goal. A too small number of units may limit the error goal level which can be attained. A disadvantage of the hybrid model is on the comparatively longer time required for training, as for every training step the set of differential equations must be solved. As such, it is most relevant, for practical applications, to look for an optimum (sensible) combination of number of units and acceptable error goal.

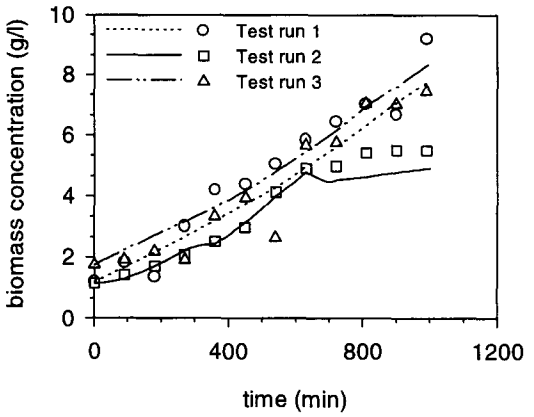

Fig. 7 : Hybrid model predictions of process behaviour

CONCLUSION AND OUTLOOK

The main purpose of this work is insight, not numbers. With relation to the case study presented, the hybrid modelling approach reveals clear advantages when compared both to the conventional and to the pure ANN approaches . It gives good results for the training data and very satisfactory predictions for the test data, considering that only a few training data sets were used. The ANN structures employed do not lead to satisfactory results for the test data. Furthermore the hybrid model is far less sensitive to internal parameters. The main disadvantage of the hybrid model is on the relatively longer computation time for training the net.

The mechanistic model, employing kinetic values from the literature is not able to describe the real process. With some parameter fine tuning, by classical non-linear regression, there is an improvement in the quality of the predictions. The message is that all theoretical information should be included in any modelling exercise.

Both for the ANN and for the hybrid approaches, results are seen to improve with the number of data runs employed for training. In particular, it is clear that results obtained from any modelling approach employing at some stage ANN, will need careful consideration particularly when outside the domain of training. This only confirms the need for developing hybrid algoritms which include forms of measuring the quality of the predictions given by ANN structures.

ACKNOWLEDGEMENTS

This work was carried out at the *Instituto de Sistemas e Robótica, Universidade do Porto (Systems and Robotics Instituto, University of Porto)* and was partially supported by the Human Capital and Mobility Program, Network in Chemical Process Control, contract no. ERB CHRX CT94-0672

NOTATION

arabic symbols

C	carbon dioxide concentration	(g/l)
CTR	carbon dioxide transfer rate	(g/(l min))
D	dilution rate	(1/min)
E	ethanol concentration	(g/l)
J	sum squared error	
$K_L^C a$	carbon dioxide transfer coefficient	(1/min)
$K_L^O a$	oxygen transfer coefficient	(1/min)
N_p	number of patterns	
O	oxygen concentration	(g/l)
OTR	oxygen transfer rate	(g/(l min))
r	reaction rate	(1/min)
S	substrate concentration	(g/l)
t	time	(min)
V	volume	(l)
w	weight	
X	biomass concentration	(g/l)
Y	yield coefficient	(g/g)
y	system input vector	

greek symbols

η	learning rate	
μ	specific biomass growth rate	(1/min)

superscripts

o	oxidative reaction	
oe	oxidative reaction for ethanol	consumption
r	reductive reaction	
*	saturation	

subscripts

C	carbon dioxide
E	ethanol
e	on ethanol
f	in feed
i	indice for unit in preceding layer
j	indice for unit in following layer
n	update step
O	oxygen
S	substrate
s	on substrate
X	biomass

REFERENCES

Cheng, Y., Karjala, T. W. and Himmelblau, D. M., 1995, *Identification of Nonlinear Dynamic Processes with Unknown and Variable Dead Time Using an Internal Recurrent Neural Network.* Ind. Eng. Chem. Res., **34**, 1735-1742

Edgar, T. F., 1996, *Modelling and Control - Back to the Future, Part I.* CAST Communications **19** (1), 7-12

Feyo de Azevedo, S., P. C. Pimenta, F. R. Oliveira, E. C. Ferreira, 1996, *Computer Based Studies on Bioprocess Engineering, II - Tools for Process Operation.* in 'Computer and Information Science Applications in Bioprocess Engineering' , Eds A. R. Moreira, K. Wallace, NATO-ASI Series E , Applied Sciences, Vol. 305, p. 27-50, Kluwer Academic Pub., Dordrecht.

Hagan, M.T., Demuth, H.B. and Beale, M., 1996, *Neural Network Design*, PWS Pub. Co, Boston.

Ljung, L., 1987, *System Identification - Theory for the User.* Prentice-Hall, New Jersey.

Lübbert, A. and Simutis, R., 1994, *Using Measurement Data in Bioprocess Measurement and Control.* Tibtech **12**, 304-311

Montague, G., Morris, J., 1994, *Neural network contribution in biotechnology.* TIBTECH **12**, 312-323

Oliveira, F. R., 1997, Ph.D. Thesis (in preparation).

Schubert, J., Simutis, R., Dors, M.; Havlik, I. and Lübbert, A., 1994, *Bioprocess optimization and control: Application of hybrid modelling.* Journal of Biotechnology **35**, 51-68

Scott, G. M., Ray, W.H., 1993, *Creating efficient nonlinear neural networkprocess models that allow model interpretation.* J. Proc. Cont. 1993, **3**, 163-178

Sonnleitner, B., Käppeli, O., 1986, *Growth of Saccharomyces cerevisiae is controlled by its Limited Respiratory Capacity: Formulation and Verification of a Hypothesis.* Biotech. Bioneg.,. **28**, 927-937

Villermaux, J., 1996, *Future Prospects for Chemical Engineering Research and Technology.* Chem. Tech. Europe, January/February, p. 21-23.

Ye, K., Fujioka, K., Shimizu, K.., 1994, *Efficient control of fed-batch baker's yeast cultivation based on neural network.* Process Control and Quality **5**, 245-250

Yet-Pole, I., Wen-Tengu, W. and Yun-Chuan, L. , 1996, Neural network modelling for on-line state estimation in fed-batch culture of l-Lysine production, Chem. Engng. Journal, **61**, 35-40

You, Y., Nikolaou, M., 1993, *Dynamic process modelling with recurrent neural networks.* AIChE Journal, **39**, 1654-1667

Zell, A., 1994, *Simulation Neuronaler Netze*, Addison-Wesley (Deutschland) Gmbh

Zell, A., et al. 1995, *SNNS Stuttgarter Neural Network Simulator, User Manual, Version 4.1.* University of Stuttgart, Institute for parallel and distributed high performance systems, Report No. 6/95.