Bioprocess hybrid parametric/nonparametric modelling based on the concept of mixture of experts

J. Peres[1*], **R. Oliveira**[2], **S. Feyo de Azevedo**[1]

[1]Department of Chemical Engineering, Faculty of Engineering, University of Porto,
Rua Dr. Roberto Frias, 4200-465 Porto, Portugal.

[2]REQUIMTE/CQFB, Departamento de Química, Faculdade de Ciências e Tecnologia,
Universidade Nova de Lisboa, P-2829-516 Caparica, Portugal.

Abstract

This paper presents a novel method for bioprocess hybrid parametric/nonparametric modelling based on mixture of experts (ME) and the Expectation Maximisation (EM) algorithm. The bioreactor system is described by material balance equations whereas the cell population subsystem is described by an adjustable mixture of parametric/nonparametric sub-models inspired in the ME architecture. This idea was motivated by the fact that cellular metabolism has an inherent "modular" structure, organised in metabolic reactions pathways, with complex interactions. As main conclusions it can be stated that MEs trained with the EM algorithm are able to systematically detect metabolic shifts with the individual experts developing expertise in describing the individual pathways. The MLP and the ME outperform systematically the RBF network in terms of the ratio model accuracy/number of parameters. The ME network outperforms the MLP network in its ability to describe metabolic switches.

1 Introduction

The development of optimal control strategies for bioprocesses is frequently constrained by the availability of sufficiently accurate mathematical models for supporting such developments. Bioprocesses may be generically characterised as complex systems exhibiting non-linear and time-varying dynamics. The main source of complexity arises from the intracellular phenomena. Model-based bioreactor performance optimisation studies rarely incorporate detailed descriptions of the intracellular phase. The method is simply too "expensive" for routine application in the biochemical industries. It has been pointed out by several authors that hybrid parametric/nonparametric modelling techniques may represent a cost-effective alternative for the analysis of bioprocesses (Schubert et al., 1994; Simutis et al., 1997). Hybrid parametric/nonparametric systems combine First Principles modelling with chemometric techniques for extracting knowledge hidden in process data. The most widely adopted hybrid structure for bioreactor systems combines macroscopic material and/or energy balance equations with artificial neural networks such as the MLP or the RBF networks (Psichogios & Ungar, 1992; Thompson & Kramer, 1994; Montague & Morris, 1994; Chen et al., 2000). The job for the neural networks in these structures is the nonparametric modelling of unknown reaction kinetics, which is normally the most challenging part of the process to be modelled in a mechanistic sense. One important feature of living cells is the fact that they process different substrates through different metabolic pathways. Diauxic growth on two carbon sources is one of such examples. Another is aerobic/anaerobic growth, depending on the presence or absence of dissolved oxygen in the medium. For example, the *S. cerevisiae* yeast grows through three different metabolic pathways for exploiting energy and basic material sources and is able to switch between a oxidative metabolic state and a oxido-reductive metabolic state (Sonnleitner & Kappeli, 1986). With high glucose supply or in conditions of oxygen limitation, *Echireschia coli* produces acetate via an alternative pathway. In case of glucose limitation, *E. coli* is able to alternatively metabolise acetate (e.g.

---

*Corresponding author. Tel. +351-22-5081693. E-mail: jperes@fe.up.pt.

Reiling et al., 1985). The biological systems exemplified have inherent non-linear discontinuous reaction kinetics due to switching between metabolic mechanisms. The popular MLPs and RBFs networks have some limitations for approximating discontinuous input-output systems. Eikens & Karim (1999) showed that modular neural networks, such as the ME (Jacobs et al., 1991), are well suited for the identification of processes that switch between different operating conditions. Such modular network architectures might be an interesting alternative for bioprocess hybrid modelling. A modular network architecture consists of two or more (small) network modules mediated by an integration unit, which decides how to combine their outputs to form the final system output and which modules should learn which training patterns (Haykin, 1994). This type of architecture performs task decomposition in the sense that it learns to partition a task into two or more functionally independent tasks and allocates distinct networks to learn each task (Jacobs et al., 1991). The main objective of this work is to develop a hybrid modelling method for bioprocesses accounting for the intracellular modular structure of the cells subsystem. The combination of First Principles modelling with modular network architectures is thus explored. The remaining of this paper is organised in 4 more sections. In Section 2 a mechanistic/ME hybrid structure for bioreactor systems is described and the ME network architecture is briefly reviewed. In Section 3 a case study are presented where hybrid models employing ME, MLP and RBF networks are compared. Finally, in Section 4, the main conclusions are presented.

2   Mechanistic/mixture of experts hybrid model for bioreactor systems

Hybrid model structures have been classified as parallel and/or serial (Thompson & Kramer, 1994; Psichogios & Ungar, 1992). In parallel structures a full mathematical model is available that however is not sufficiently accurate for model-based applications. A nonparametric modelling technique is then combined in parallel with the mathematical model, having access to the same input variables and correcting the mathematical model outputs. In the serial case, there is knowledge concerning the general structure of the system, but parts thereof are not known in a mechanistic sense. Such unknown sub-systems are modelled with nonparametric techniques, which feed with information to mechanistic parts. The hybrid model structures that naturally arise in bioreactor modelling problems tend to be simultaneously serial and parallel (Oliveira et al., 2005; Peres et al., 2005). The bioreactor system was described by material balance equations while a parallel neural network/mechanistic structure represented the cell population system. In this work a similar structure is adopted that however uses a ME architecture for modelling the unknown reaction kinetics term instead of the usual MLPs or RBFs networks (see Fig. 1) This model may be mathematically stated by the following two equations:

$$\frac{d\mathbf{c}}{dt} = \mathbf{r}(\mathbf{c}, \mathbf{w}) - D\mathbf{c} + \mathbf{u} \tag{1}$$

$$\mathbf{r} = \mathbf{K}\mathbf{r}_{mec}(\mathbf{c})\,\varphi_{ME}(\mathbf{c}, \mathbf{w}), \tag{2}$$

with $\mathbf{c}$ a vector of n concentrations, K a $n \times m$ yield coefficients matrix, $\mathbf{r}_{mec}$ a $m \times r$ matrix of mechanistically known kinetic expressions, $\varphi_{ME}(\mathbf{c}, \mathbf{w})$ a vector of r unknown kinetic functions modelled with ME networks, $\mathbf{w}$ a vector of parameters that must be estimated from data, D is the dilution rate, $\mathbf{u}$ is a vector of net volumetric input rates (possibly control inputs).

THE MIXTURE OF EXPERTS ARCHITECTURE

The ME architecture has strong statistical foundations since it was inspired by the concept of mixture models from Statistics field (Titterington et al., 1985). The ME architecture (Jacobs et al., 1991; Haykin, 1994) consists of a set of K expert networks and a gating network (see Fig. 1). Basically, the task of each expert j is to approximate a function $f_j : \mathbf{c} \to \varphi_j$ over a region of the input space. The task of the gating system is to assign an expert network to each input vector $\mathbf{c}$. The final system output φ_{ME} is a combination of the expert network outputs $- \varphi_{ME} = \sum_{j=1}^{K} g_j(\mathbf{c})\varphi_j(\mathbf{c})$ – where $g_j(\mathbf{c})$ are the gating outputs. The expert modules are simple linear functions for non-linear regression problems or linear functions with a single output non-linearity for classification problems. In some non-linear regression problems it may be necessary to use more complex non-linear experts. MLP networks with the tangent
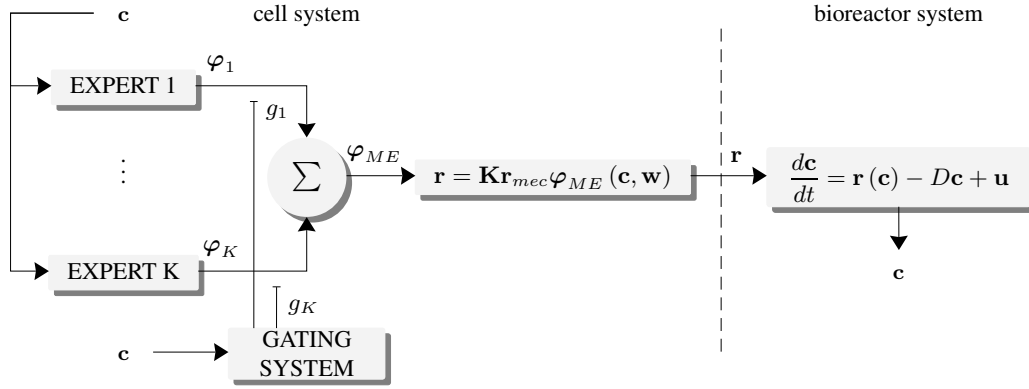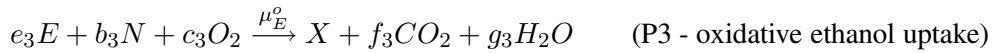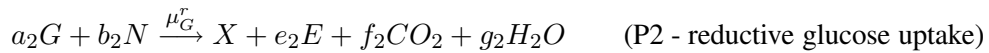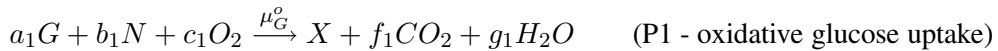
Figure 1: Proposed Mechanistic/Mixture of Experts hybrid structure

hyperbolic function in the hidden layers and linear functions in the output layer was the expert network structure adopted in this work. Also, different forms for the gating system have been reported. The softmax function suggested initially by Jacobs et al. (1991), is a normalised exponential function of the inputs $\mathbf{c}$ and provides a "soft" hyperplane divison (Ramamurti & Ghosh, 1999). A localised gating system based on Gaussian functions provides more flexible "soft" hyper-ellipsoids input space partitions (Ramamurti & Ghosh, 1999) and was adopted in this work. The parameters present in each expert and in the gating system were identified using the well known EM algorithm. The Reader is referred to the work of Jordan & Jacobs (1994) for details of this technique.

3   Results and discussion

The *S. cerevisiae* yeast cells metabolise glucose via two pathways under aerobic conditions: oxidative and/or reductive pathways, with ethanol the end product of the reductive pathway. The yeast cells are also able to use ethanol as an alternative substrate, but the ethanol can be metabolised oxidatively only. Sonnleitner & Kappeli (1986) developed a simple unstructured kinetic model based on the respiratory bottleneck concept (see Oliveira et al. 2005 for detailed description). The model defines 3 macroscopic reactions for carbon source utilisation with well-defined stoichiometry:

$$a_1 G + b_1 N + c_1 O_2 \xrightarrow{\mu_G^o} X + f_1 CO_2 + g_1 H_2O \qquad \text{(P1 - oxidative glucose uptake)}$$

$$a_2 G + b_2 N \xrightarrow{\mu_G^r} X + e_2 E + f_2 CO_2 + g_2 H_2O \qquad \text{(P2 - reductive glucose uptake)}$$

$$e_3 E + b_3 N + c_3 O_2 \xrightarrow{\mu_E^o} X + f_3 CO_2 + g_3 H_2O \qquad \text{(P3 - oxidative ethanol uptake)}$$

with X, G, E, N, O_2 and CO_2 the biomass, glucose, ethanol, ammonia, oxygen and carbon dioxide respectively suspended or dissolved in the liquid phase, a_i-g_i with i =1,2 or 3 are stoichiometric co-efficients, and μ_G^o, μ_G^r and μ_E^o specific growth rates associated with the three macroscopic reactions. The yeast cells may find themselves in one of two metabolic states: oxidative or oxido-reductive. In the oxidative metabolic state, only pathways (P1) and (P3) take place. In conditions of oxygen saturation, it may be shown that the oxidative metabolic state is only possible for glucose concentrations lower than 0.042 g/L. The oxido-reductive state, corresponding to pathways (P1) and (P2), occurs for glucose concentrations higher than 0.042 g/L, presuming again oxygen saturation. The switch between the two metabolic states is a crisp "if-then" switch. Consequently, pathways (P2) and (P3) are never simultaneous. The main objective in this study is to evaluate the proposed hybrid modelling technique and in particular to verify if a mixture of two experts trained with the EM algorithm is capable of distinguishing between the oxidative and oxido-reductive states, and if the individual experts develop expertise in describing the one or the other metabolic state.
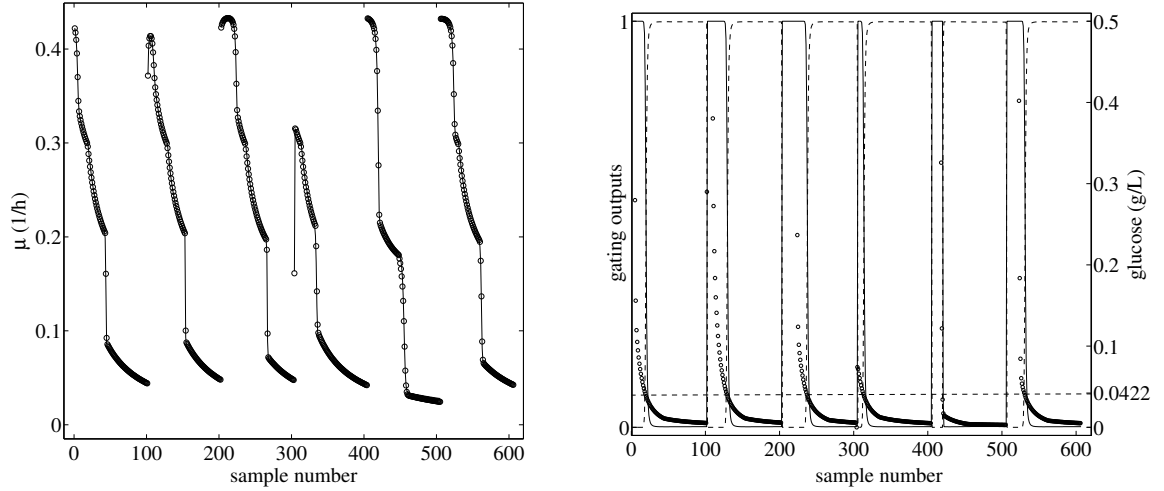
Figure 2: Results for six simulated batches: (a) specific growth rate estimates with a ME with 2 experts (18 parameters): measured values (o), estimated values (–).(b) Gaussian gating network outputs: g_1 ($\cdots$), g_2 (–). The true switch is when substrate is constant and equal to 0.0422 g/L.

In a first study, 6 batches were simulated using the model described in Oliveira et al. (2005) for distinct operating conditions of feed rate (F), glucose concentration in the feed (S_{G0}), and initial medium composition. To simplify the analysis, oxygen was assumed to be never limiting. In such circumstances the specific growth rate is a function of glucose and ethanol concentrations only. Data of specific growth rate (μ), $\mu_{exp} = \frac{1}{X}\frac{dX}{dt} + D = \mu_G^o + \mu_G^r + \mu_E^o = f(S_G, S_E)$, was collected at sampling intervals of 0.2 h along with the concentrations of glucose (S_G) and ethanol (S_E), yielding a total number of measured patterns of $np = 606$. Measurement noise was excluded from this analysis thus avoiding the issues of model overfitting. The main objective in this study is to compare the ME, MLP and RBF networks in mapping the relationship between (μ) and the two concentrations (S_G) and (S_E). The ME network was configured with $K = 2$ MLP experts with equal size $\{2, 2, 1\}$. The total number of parameters was 24 (9 parameters per expert and 6 parameters for the gating system - the gating had just one input, the glucose concentration). The ME was trained with the EM algorithm. The results obtained after 1000 iterations are shown in Figs. 2(a-b). It is not surprising the fact that this small ME network was able to model this system with almost negligible error for all 6 batches (the mean squared error (MSE) was 2.59×10^{-6}) as shown in Fig. 2a, given the powerful capacity for nonlinear function approximation of the experts. A more significant result is the fact that the ME network learned to distinguish between the two metabolic states: expert (1) developed expertise in describing the oxidative metabolic state whereas expert (2) developed expertise in describing the oxido-reductive state. Fig. 2b shows the gating outputs (g_1) and (g_2) over measured pattern along with the corresponding glucose measurements. The gating outputs intercept at glucose concentration of 0.042 g/L that is precisely where the true process switch occurs. A MLP network with 18 parameters was trained over the same data and the same structure as the experts in the ME. The network was trained with the same algorithm employed to solve the first K optimisations of the M step: quasi-Newton with a Conjugate Gradient method (CG) along with error backpropagation for the analytical evaluation of gradients. The objective function was defined in this case as a least squares problem. After 2000 iterations the MSE stabilised at 3.62×10^{-5}. This error is higher than that obtained with the ME but still very small (one order of magnitude below). In practical terms an almost perfect mapping is achieved indicating that there is no apparent advantage of using a ME network in this example maybe because it is too simple. The results produced by a RBF network with 2 inputs (S_G and S_E), a single hidden layer with 16 nodes and an output layer, trained over the same data, were however much worst (MSE=2.3×10^{-4}). The basis function units were symmetrical Gaussian density functions. The training algorithm follows the scheme proposed by Moody & Darken (1989).
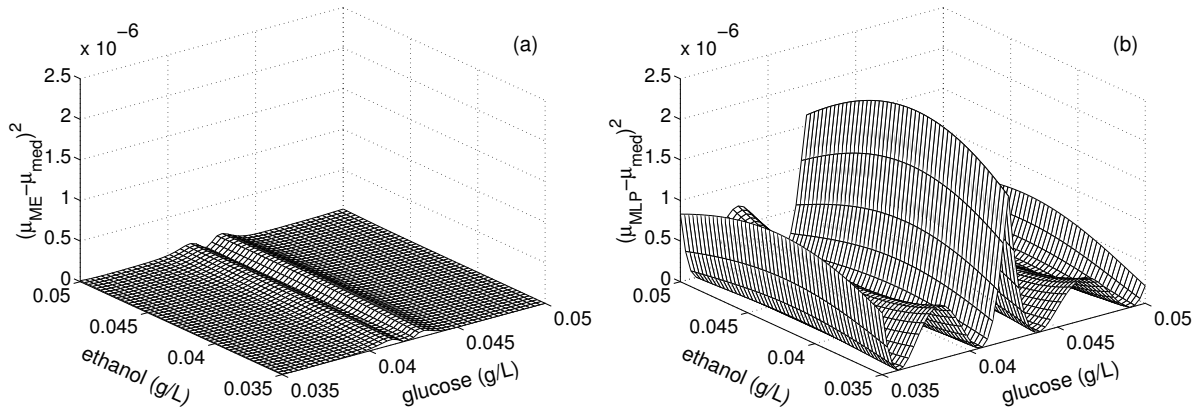
Figure 3: The square error of the specific growth rate estimates with: (a) a ME with 2 MLP experts (18 parameters); (b) a MLP network with 17 parameters.

In a second study, a grid of 2601 points in the 2D input space of (S_G) and (S_E) was generated using the same model of the first study. The objective was to assess the performance of the ME and MLP networks in the vicinity of the switch between metabolic states. The ME network was configured with 2 MLP expert networks with dimensions $\{2, 2, 1\}$ each. The output of the network was again the total specific growth rate and the inputs were $\mathbf{c} = \{S_G, S_E\}$. The gating system was a localised gating network as in the previous case. The results are presented in Figs. 3(a-b) and Fig. 4a. Figs. 3(a) and 3(b) show the modelling error for the ME and MLP respectively. Two differences are evident: (i) the mean sum error (MSE) is much smaller for the Mixture of Experts (ME) than for the MLP and (ii) the MSE is erratic in the case of the MLP. This is a relevant result though not totally unexpected because it is well known that MLPs have difficulties in mapping discontinuous systems and exhibit oscillatory behaviour at the extremities (Haykin, 1994). The use of a ME could represent a clear advantage for modelling processes that are run near metabolic switches. The case of recombinant *S. cerevisiae* or *E. coli* are such examples since they are facultative aerobic microorganisms and the build up of ethanol or acetate are associated with lower biomass and product yields. The more accurate results provided by the ME network arise from the ability to detect the switch of both metabolic states and to assign each expert to describe the individual metabolic states. Fig. 4(b) is a contour plot showing the true process switch. The metabolic switch is independent of the ethanol concentration and occurs for constant glucose concentration of S_G=0.042 g/L. The black surface signals the oxidative state while the white surface signals the oxido-reductive state. Fig. 4(a) shows a contour plot of the gating system outputs in the same 2D input space. The black colour indicates the g_1 output whereas white represents g_2 output. The ME transition occurs precisely for the true switch of S_G=0.042 g/L. The ME transition is however soft when compared to the true crisp switch shown in Fig. 4(b). Soft transitions are however characteristic of biological systems thus this result is not seen as a disadvantage of the Gaussian gating system.

4   Conclusions

The proposed mechanistic/mixture of experts hybrid model for bioreactor systems was compared with the conventional hybrid technique employing the Multi Layer Perceptron (MLP) and the Radial Basis Function (RBF) networks. As main conclusions it can be stated that MEs trained with the Expectation Maximisation algorithm are able to sistematically detect metabolic shifts with the individual experts developing expertise in describing the individual pathways. The MLP and the ME outperformed systematically the RBF network in terms of the ratio model accuracy/number of parameters. The ME network outperforms the MLP network in its ability to describe metabolic switches.
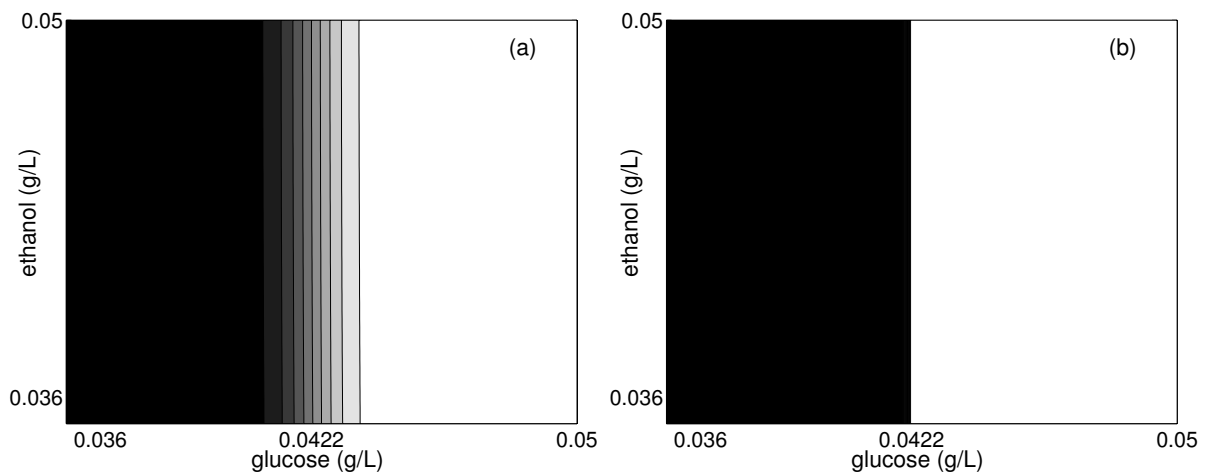
Figure 4: (a) Gaussian gating network outputs. (b) The true switch is when glucose is constant and equal to 0.0422 (g/L)

References

Chen, L., Bernard, O., Bastin, G. & Angelov, P. (2000), 'Hybrid modelling of biotechnological processes using neural networks', *Control Engineering Practice* **8**(7), 821–827.

Eikens, B. & Karim, M. N. (1999), 'Process identification with multiple neural network models', *International Journal Of Control* **72**(7-8), 576–590.

Haykin, S. (1994), *Neural Networks: A comprehensive foundation*, Macmillan College Publishing Company, Inc.

Jacobs, R. A., Jordan, M. I., Nowlan, S. J. & Hinton, G. E. (1991), 'Adaptive mixtures of local experts', *Neural Computation* **3**, 79–87.

Jordan, M. I. & Jacobs, R. A. (1994), 'Hierarchical mixtures of experts and the em algorithm', *Neural Computation* **6**(2), 181–214.

Montague, G. & Morris, J. (1994), 'Neural-network contributions in biotechnology', *Trends in Biotechnology* **12**(8), 312–324.

Moody, J. & Darken, C. J. (1989), 'Fast learning in networks of locally-tuned processing units', *Neural Computation* **1**, 281 – 294.

Oliveira, R., Peres, J. & Feyo de Azevedo, S. (2005), Hybrid modelling of fermentation processes using artificial neural networks: A study on identification and stability, *in* M. Pons & J. F. M. van Impe, eds, 'Computer Applications in Biotechnology 2004', Elsevier (ISBN: 0-08-044251-X), pp. 195 – 200.

Peres, J., Oliveira, R. & Feyo de Azevedo, S. (2005), Hybrid modelling of fermentation processes: A study on the use of modular neural networks for modelling cells reaction kinetics, *in* M. Pons & J. F. M. van Impe, eds, 'Computer Applications in Biotechnology 2004', Elsevier (ISBN: 0-08-044251-X), pp. 293 – 298.

Psichogios, D. C. & Ungar, L. H. (1992), 'A hybrid neural network-1st principles approach to process modeling', *AIChE Journal* **38**(10), 1499–1511.

Ramamurti, V. & Ghosh, J. (1999), 'Structurally adaptive modular networks for nonstationary environments', *IEEE Transactions on Neural Networks* **10**(1), 152–160.

Reiling, H. E., Laurila, H. & Fiechter, A. (1985), 'Mass-culture of escherichia-colimedium development for low and high-density cultivation of escherichia coli-b/r in minimal and complex media', *Journal of Biotechnology* **2**(3-4), 191–206.

Schubert, J., Simutis, R., Dors, M., Havlik, I. & Lubbert, A. (1994), 'Hybrid modeling of yeast production process-escombination of a-priori knowledge on different levels of sophistication', *Chemical Engineering & Technology* **17**(1), 10–20.

Simutis, R., Oliveira, R., Manikowski, M., de Azevedo, S. F. & Lubbert, A. (1997), 'How to increase the performance of models for process optimization and control', *Journal of Biotechnology* **59**(1-2), 73–89.

Sonnleitner, B. & Kappeli, O. (1986), 'Growth of *Saccharomyces-cerevisiae* is controlled by its limited respiratory capacity formulation and verification of a hypothesis', *Biotechnology and Bioengineering* **28**(6), 927–937.

Thompson, M. L. & Kramer, M. A. (1994), 'Modeling chemical processes using prior knowledge and neural networks', *AIChE Journal* **40**(8), 1328–1340.

Titterington, D. M., Smith, A. F. M. & Makov, U. E. (1985), *Analysis of Finite Mixture Distributions*, New York: Wiley.