ELSEVIER

Computers and Chemical Engineering 25 (2001) 783–791

Computers & Chemical Engineering

www.elsevier.com/locate/compchemeng

Knowledge based modular networks for process modelling and control

J. Peres, R. Oliveira, S. Feyo de Azevedo *

Department of Chemical Engineering, University of Porto, Rus Dr Roberto Frias, 4200-465 Porto, Portugal

Received 10 May 2000; accepted 5 January 2001

Abstract

This paper addresses methods of knowledge engineering for chemical and biochemical process modelling and control. The concept of Knowledge Based Modular (KBM) Networks is presented. KBM networks represent a method of expressing and combining different types of knowledge, usually available for modelling chemical and biochemical processes: mechanistic, heuristic and knowledge hidden in process data records. The Expectation Maximisation (EM) algorithm is employed to optimally combine the modules within the KBM network. The concepts are illustrated with an application to a baker's yeast production process. The results show that it is possible to obtain more accurate process description when all available sources of knowledge are incorporated in the process model.

Keywords: Process modelling; Knowledge engineering; Chemical/biochemical processes

Nomenclature

B	subset of clusters where no measurements are available
C	set of N_C clusters distributed equidistantly through S
$\{C\}_i$	subset of clusters associated to expert module i
$\boldsymbol{C}_j$	centre of cluster j in the input space S
$\boldsymbol{D}$	desired target vector
D	output space with $\dim(D) = m$
F	input feed rate into the fermenter
F_s	input glucose feed rate per unit volume
G_{in}	glucose concentration in the input feed rate F
g_I	ith output of the gating system
$h_i^{(t)}$	posterior probability for expert i and measured pattern t
$\boldsymbol{I}$	integer parameter vector for the gating system
K	number of expert modules in the KBM network
M	dimension of the output space $m = \dim(D)$
MSE	mean-square-error
N	dimension of the input space $n = \dim(S)$
N_B	number of clusters in subset B
N_C	number of clusters in set C
N_T	number of clusters in subset T
N_P	number of measured patterns

* Corresponding author. Fax: +351-2-2000808.
E-mail address: sfeyo@fe.up.pt (S. Feyo de Azevedo).

0098-1354/01/$ - see front matter
PII: S0098-1354(01)00665-2

$P(i \mid \boldsymbol{x})$	conditional probability of picking expert i given input $\boldsymbol{x}$
$P(\boldsymbol{y} \mid \boldsymbol{x}, i)$	conditional probability of output pattern $\boldsymbol{y}$ given the input $\boldsymbol{x}$ and given the expert module i
$q_{c,max}$	maximum oxygen uptake rate
r_I	value of the ith Gaussian function
S	input space
T	subset of clusters within the input region where measurements are available
V	working volume
$\boldsymbol{w}_I$	parameters vector of expert module i
$\boldsymbol{X}$	input vector
$\{\boldsymbol{x}, \boldsymbol{d}\}$	set of P measured patterns
X_I	biomass predicted by expert I
X	biomass predicted by the KBM network
X_m	measured biomass
$\boldsymbol{Y}$	output vector
$\boldsymbol{y}_I$	ith element of vector $\boldsymbol{y}$

Greek letters

μ	specific growth rate
θ_i, Λ_i	θ_i is the nearest cluster centre to input $\boldsymbol{x}$ belonging to subset $\{C\}_i$; Λ_i is the corresponding covariance matrix
σ_{ij}	standard deviation for the ith dimension of the Gaussian function j
Σ_j	covariance matrix of the Gaussian function function j

1. Introduction

The practice of industrial operations is in the process of absorbing changes and methods induced by the developments in digital instrumentation and computer technology. Automatic integration of knowledge and model-based methods are still, to a significant extent, young concepts in the history of the process industries. Process models represent today, more than ever before, the basis for advanced methodologies concerning chemical and biochemical process monitoring, optimisation and control. The concept and scope associated to the term *Model* is also changing. This paper deals with such evolution.

Throughout the years process behaviour has been modelled essentially by employing the so-called *chemical engineering science* or *first principles* approach, coupling mass, heat, momentum and population balances with basic kinetic and transport phenomena mathematical relationships. Not questioning the relevance of the role of such route of work, theoreticians and practicioners alike of the Chemical and Biochemical Process Industries (CBPI) recognise that for a non-negligible number of complex and/or ill-defined processes it is not reasonable to expect good results, at reasonable costs, from the sole use of *first principles* knowledge (Feyo de Azevedo, Dahm, & Oliveira, 1997; Simutis, Oliveira, Manikowski, Feyo de Azevedo, & Lübbert, 1997).

The situation in the biochemical industries is worth to be analysed in some more detail. In these industries several factors concurrently lead to benefit/costs ratios that do not favour the investment in classical model-based bioprocess operation (Royce, 1993; Simutis, Havlik, & Lübbert, 1993). Often key process variables are not measured, due either to sterility problems or to the unavailability of reliable measuring techniques or still because they are not measurable at all; also, the complexity of mechanisms for microorganisms growth and product formation and/or the complex uncharacterised reaction media hinder or preclude the development of reliable models for supporting both the indirect estimation of those unmeasured quantities and the use of model-based methodologies of process operation. Just as example, a major challenge for the future remains to be the development of reliable on-line estimation of intracellular components in biological systems (Sonnleitner, 1999).

Hence, in an environment where the *first-principles* knowledge available is scarce and the development times are compulsorily short, alternative methodologies for chemical and biochemical process modelling should be find wide application. The motivation is there for seeking new routes for modelling.

Methods and solutions employing black-box modelling approaches and in particular the application of Artificial Neural Networks (ANNs) for non-linear pro-

cess modelling and control are currently accepted as workable alternative or complementary approaches for the overall goal of process behaviour representation. As shown by Bhat and McAvoy (1990), the non-linear and time-varying nature of chemical processes may be modelled accurately and with short development times with ANNs provided there are enough measurements covering the whole process operating region.

The other important source of knowledge is heuristic knowledge. In industry, many decisions concerning process operation have for many years been made on the basis of the heuristic knowledge of engineers and operators supervising the system. As such one should expect that heuristic knowledge is available in large quantities in the industry. Such form of knowledge can be expressed directly with expert systems and should equally be considered as a valuable resource for process modelling and control (Hitzmann, Lübbert, & Schügerl, 1992; Sterbacek & Votruba, 1993).

Modelling may thus be viewed as an exercise of knowledge expression and representation in a compact form. Since different types of knowledge are usually available for the same process, two conceptual approaches for modelling are essentially possible:

1. to select one of the available sources of knowledge and to adopt and explore the corresponding modelling technique; or
2. to incorporate and express all the available sources of knowledge in a single hybrid process model.

In a knowledge engineering perspective the second alternative is more rational since the accuracy of the model depends essentially on the quantity and quality of the knowledge available. As shown in the papers of Schubert, Simutis, Doors, Havlik, and Lübbert (1994a); Schubert, Simutis, Doors, Havlik, and Lübbert (1994b); Feyo de Azevedo et al. (1997); Simutis et al. (1997), when *first-principles* knowledge is lacking mathematical models may be complemented with other modelling techniques that incorporate unused sources of knowledge, i.e with ANNs and/or Fuzzy/Expert systems.

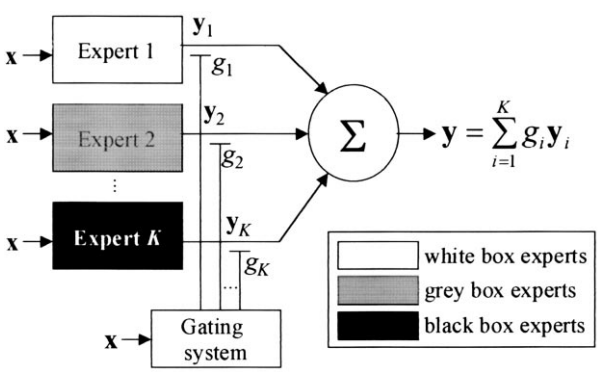

Fig. 1. Block diagram of a KBM network composed by K expert modules the outputs of the expert modules are mediated by a gating system. The expert modules express different types of knowledge (white-box, grey-box and black-box modules).

Several authors have been proposing hybrid model structures for chemical and biochemical processes. Psichogios and Ungar (1992) and Thompson and Kramer (1994) proposed hybrid models that combine mathematical models with ANNs for modelling chemical and biochemical reactors. Simutis et al. (1993) suggest the combination of ANNs with Fuzzy systems. In Glassey, Igonova, Ward, Montague, and Morris (1997) model is suggested that combines ANNs with expert systems for bioprocess supervision.

Hybrid modelling is not a structurally defined problem, this meaning that a theory is still missing. Normally the applications reported are problem oriented. In this paper we propose a methodology that encapsulates hybrid modelling in a more broaden knowledge engineering perspective.

2. Modelling through knowledge integration

The backbone of this methodology of modelling through knowledge integration consists of a network structure (connectionist system) with specialised nodes (modules), able to represent knowledge of different sources, with different levels of sophistication. The network, termed as Knowledge Based Modular Network, is thus able to incorporate and mix different types of knowledge and modelling techniques. A probabilistic interpretation is given to the structure and a *Maximum Likelihood* technique is used for parameter identification within modules and in the connections between modules.

In what follows the methodology is presented in detail.

2.1. Knowledge-Based Modular (KBM) networks

Jacobs, Jordan, and Barto (1991a), Jacobs, Jordan, Nowlan, and Hinton (1991b), Jordan and Jacobs (1994) developed a class of modular connectionist architectures termed as 'mixture of experts' and subsequently as 'hierarchical mixture of experts'. The main idea was to develop a connectionist architecture, which is able to learn how to partition a task into two or more functionally independent tasks and to allocate distinct networks to learn each task. The same concepts can be applied for developing a network of different types of models with the capability of becoming specialists in describing different parts of the process. This is the main idea behind the KBM network.

The KBM network is depicted in Fig. 1. It consists of a set of K expert modules mediated by a gating system. The task for the network is to map an input space S into a target output space D. Both the expert and the gating system have access to the inputs $x \in S$, which can be the same for all of them, but not necessarily the

same. The task of each expert i is to approximate a function f_i: $S \rightarrow D$ over a region of the input space S. The task of the gating system is to assign an expert network to each input vector $\boldsymbol{x}$. The final output $\boldsymbol{y}$ is a linear combination of the expert outputs $\boldsymbol{y}_i$ weighted by the gating outputs g_i. The details of the expert modules and of the gating system are described in Sections 2.2 and 2.3.

2.2. The expert modules

In Jacobs et al. (1991a) and Jordan and Jacobs (1994) the expert modules were linear relationships. Weigend, Mangeas, and Srivastava (1995) explored the case of non-linear experts (the architecture was termed as 'gated mixture of experts').

In the KBM structures expert modules should or are generally expected to be more elaborated in the sense that they should be able to express different forms of knowledge. We formalise the expert modules in the following way: the measured patterns $\{\boldsymbol{x}, \boldsymbol{d}\} \in \{S, D\}$, denoted by the input vector $\boldsymbol{x}$ and desired response (target vector) vector $\boldsymbol{d}$, are assumed to be generated by a set of K different regressive processes that are continuous non-linear and dynamical in nature. As such the expert modules should be of the form:

$$\frac{\mathrm{d}\boldsymbol{y}_i}{\mathrm{d}t} = f_i(\boldsymbol{y}_i, \boldsymbol{x}, \boldsymbol{w}_i), \tag{1}$$

being $\boldsymbol{y}_i$ the output of expert module i, $\boldsymbol{x}$ the input vector, $\boldsymbol{w}_i$ the parameters vector of expert i and f_i a continuous non-linear function.

The functional relationships f_i are not the same for all the expert modules. We assume that the expert modules may include a priori knowledge and that this knowledge may be available on different forms and levels of theoretical complexity. As such we have three main types of expert modules:

1. white-box modules, which express physical knowledge by means of mathematical equations;
2. grey-box modules, which express heuristic knowledge (e.g. by means of Fuzzy Inferential Systems); and
3. black-box modules that are able to capture information from process data (e.g. ANNs).

We assume that each form of knowledge will have some region of the input space where conditions are such that it represents the process better. The job for the KBM network is to partition the input space in several sub-regions in such a form that knowledge expression is maximised, through expert module specialisation (wheighing) in each of those sub-regions. With this strategy we are possibly extracting the best that each type of knowledge has to offer.

2.3. The gating system

There are two main types of gating systems mentioned in the literature: (i) based on 'softmax' functions (e.g. Jacobs et al. (1991a)); and (ii) based on Gaussian functions (e.g. Jacobs et al., 1991b; Ramamurti and Ghosh 1999).

It is important to mention at this point the role of the relationship between the nature of the experts and that of the gating system. With very simple expert modules, say linear modules, the network must be configured with many experts in order to achieve an acceptable mapping. With more sophisticated experts only a few experts are required. In the situation of few and complex experts, the gating system must allow a flexible partition of the domain of influence of each expert module. Having these ideas in mind, we propose a gating system that is formalised in two steps:

Step 1*:* Hyperspherical clustering of the input space S with a set C of N_C clusters distributed equidistantly through the input space S. The clusters are defined as Gaussian functions:

$$r(\boldsymbol{x}, \boldsymbol{c}_j, \Sigma_j) = (2\pi)^{-n/2} \times |\Sigma_j|^{-1/2} \exp\left\{ -1/2(\boldsymbol{x} - \boldsymbol{c}_j)^T \sum_j^{-1} (\boldsymbol{x} - \boldsymbol{c}_j) \right\}, \tag{2}$$

being $\boldsymbol{c}_j$ the cluster centres and Σ_j diagonal covariance matrices

$$\Sigma_j = \mathrm{diag}\{\sigma_{ij}^2\}.$$

Cluster set C is divided into two subsets: subset T of N_T clusters within the region where process measurements are available; and subset B of N_B clusters outside this region. Fig. 2a illustrates the concept, showing a projection of regions for a case of two inputs, x_1 and x_2.

Step 2*:* Association of each of the K experts to a subset of clusters $\{C\}_i$ $(i = 1, \ldots, K)$. Within the training region T, N_T/K clusters are randomly associated to each specific expert module i. Additionally, all N_B clusters in region B are associated with the mechanistic expert (among those defined) with better extrapolation properties.

For the universe of N_C clusters, the association between clusters and expert modules is formalised through an integer parameter vector $\boldsymbol{I}$, defined as:

$$\boldsymbol{I}_j = i \quad \text{for} \quad j = 1, \ldots, N_C, \tag{3}$$

where $\boldsymbol{I}_j = i$ means that cluster j is associated to expert module i.

Example (*illustrated in* Fig. 2b)*:* one input x_1; two experts ($K = 2$); a total of 20 clusters ($N_C = 20$); subset B with 10 clusters ($N_B = 10$); subset T with 10 clusters ($N_T = 10$); expert 1 (mechanistic) associated to 15 clus-

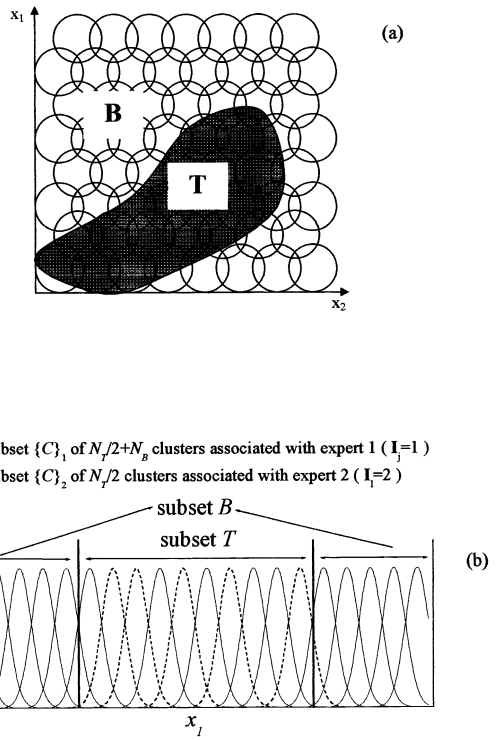

Fig. 2. Gating system (a) Schematic representation of a set of hyperspherical clusters distributed equidistantly in a two-dimensional input space. The region in grey defines a sub-space T where measurements are available; (b) example of association between clusters ($N_C = 20$) and experts ($K = 2$) for a case of one-dimension input space.

ters ($N_T/2 = 5$, randomly associated clusters from T plus all clusters from B); expert 2 associated to five clusters ($N_T/2 = 5$ randomly associated clusters from T).

The evaluation of the relative weight g_i of each expert module i is directly related to the cluster functions, as defined by Eq. (2), being performed as follows:

1. given the input pattern $\boldsymbol{x}$ and given the integer parameter vector $\boldsymbol{I}$, for each expert i pick the nearest cluster centre θ_i and corresponding covariance matrix Λ_i such that $\boldsymbol{I}_j = \boldsymbol{I}$,
2. evaluate the clusters value r_i using Eq. (2):

$$r_i = r(\boldsymbol{x}, \theta_i, \Lambda_i), \quad i = 1, \ldots, K, \tag{4}$$

3. compute the normalised gating system outputs g_i, from

$$g_i = \frac{r_i}{\sum_{l=1}^{K} r_l} \quad i = 1, \ldots, K. \tag{5}$$

2.4. Parameter identification

The KBM Network can be given a probabilistic interpretation and the Expectation Maximisation (EM) method may be used for parameter identification. The probabilistic interpretation is based on the assumption that the process that generates patterns $\boldsymbol{d}$ from patterns $\boldsymbol{x}$ obeys some probabilistic model $P(\boldsymbol{d}|\boldsymbol{x})$.

For a given expert i, we assume that the conditional probability of pattern $\boldsymbol{d}$ (measured values) given the input vector $\boldsymbol{x}$ and given the expert module i is given by a Gaussian probability function:

$$P(\boldsymbol{d}|\,\boldsymbol{x}, i) = (2\pi)^{-m/2} \times |\Sigma_i|^{-1/2} \exp\left\{-1/2(\boldsymbol{d}-\boldsymbol{y}_i)^T \sum_i^{-1} (\boldsymbol{d}-\boldsymbol{y}_i)\right\}, \tag{6}$$

being Σ_i the covariance matrix for expert i and $\boldsymbol{y}_i$ the output of expert module, which will determine the centre of the Gaussian function (Eq. (6)). Notice that $\boldsymbol{y}_i$ is a function of parameters $\boldsymbol{w}_i$ as defined in Eq. (1).

Also, the outputs of the gating system are interpreted as the conditional probability of peeking expert i given the input vector $\boldsymbol{x}$:

$$P(i|\,\boldsymbol{x}) = g_i(\boldsymbol{x}, \boldsymbol{I}). \tag{7}$$

With the above probabilistic interpretation, it is possible to formulate a likelihood function and to employ a maximum likelihood parameter estimator for the simultaneous estimation of parameter vectors $\boldsymbol{w}_i$ within modules and of parameter vector $\boldsymbol{I}$ of the gating system.

For the simpler case of the 'mixture of experts' neural network, Jacobs et al. (1991b) applied a gradient ascendant weights updating algorithm for maximising the likelihood function. Later on, Jordan and Jacobs (1994) suggested the use of the EM algorithm, which proved to converge much faster than the gradient ascendant algorithm.

The EM algorithm was adopted in the present work. It is a two-step iterative procedure (Jordan & Jacobs, 1994; Xu & Jordan, 1996), which, for each iteration, can be summarised as follows:

1. the *E-step*, where the posterior probabilities h_i are computed from:

$$h_i^{(t)} = \frac{g_i(\boldsymbol{x}^{(t)}, \boldsymbol{I}) P(\boldsymbol{d}^{(t)} \mid \boldsymbol{x}^{(t)}, i)}{\sum_{j=1}^{K} g_i(\boldsymbol{x}^{(t)}, \boldsymbol{I}) P(\boldsymbol{d}^{(t)} \mid \boldsymbol{x}^{(t)}, i)} \quad t = 1, \ldots, N_P, \tag{8}$$

where the superscript index (t) refers to measured pattern and N_P is the number of measured patterns.

2. the *M-Step*, where $K+1$ separate maximisation problems are solved

$$\boldsymbol{w}_i = \text{argmax}\left(\sum_{t=1}^{N_P} h_i^{(t)} \ln P(\boldsymbol{d}^{(t)} | \boldsymbol{x}^{(t)}, i) \right) \quad i = 1, \ldots, K, \tag{9a}$$

$$I = \text{argmax}\left(\sum_{t=1}^{N_P} \sum_{j=1}^{K} h_j^{(t)} \ln g_j^{(t)} \right). \tag{9b}$$

3. The K optimisations of Eq. (9a) evaluate new expert module parameters $\boldsymbol{w}_i$, whereas the optimisation of Eq. (9b) evaluates new parameters $\boldsymbol{I}$ for the gating system (this corresponding to 'cluster-to-expert' re-association).

The algorithm is stopped when no re-associations between clusters and experts occur between two consecutive iterations.

3. Case study: prediction of biomass concentration in a Baker's yeast fermentation process

3.1. Problem formulation and solution

In the present case study we illustrate the application of a simple KBM Network for predicting the evolution of biomass concentration in time as a function of the carbon source feed rate profile in a baker's yeast fed-batch process. Data of five fermentation runs performed in a 10 l Lab-scale fermenter (Oliveira, 1997) are available for developing the KBM network.

Pomerleau and Perrier (1990) proposed a dynamical model for the fed-batch baker's yeast process, which is based on a mechanistic description of the process. In the present case study the main goal is to combine this mechanistic model with another model that incorporates unused sources of knowledge. As such, two main sources of knowledge are explored:

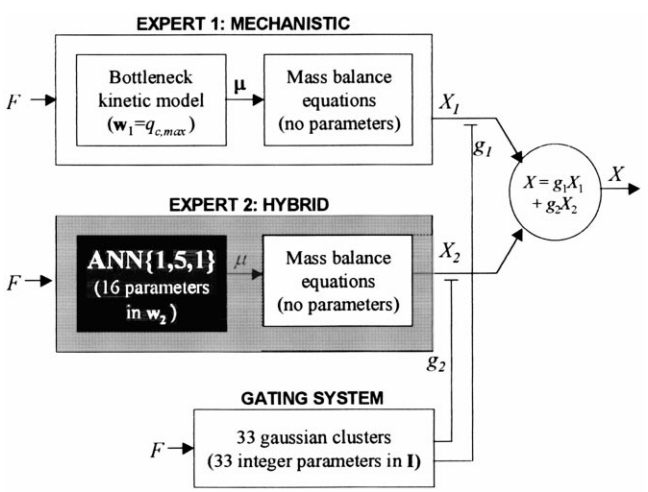

Fig. 3. KBM Network for biomass prediction in a baker's yeast fermentation process with $K=2$ expert modules. Expert 1 is a mechanistic model (white-box) whereas expert 2 is a hybrid model (grey-box).

1. *Mechanistic knowledge:* We consider that the mechanistic knowledge available is properly expressed in the mechanistic model of Pomerleau and Perrier (1990)
2. *Information hidden in process data:* we consider that important process features not incorporated in the mechanistic model are reflected in the set of measurements available.

The strategy is to activate the two sources of knowledge in a KBM Network for biomass prediction. As such, the KBM network will be composed by two expert modules mediated by a gating system according to the configuration depicted in Fig. 3.

Expert 1 is the dynamical mechanistic model described in Pomerleau and Perrier (1990). The model has 7 kinetic parameters involved. The parameter $q_{c,\ max}$ (maximum oxygen uptake rate) is one of most sensitive parameters in the model and is a characteristic of a given strain (Sonnleitner & Käppeli, 1986). For simplicity we consider that the parameter $q_{c,max}$ is the only parameter that must be identified for the actual process conditions. Thus, expert 1 has only one parameter involved $\boldsymbol{w}_1 = q_{c,\ max}$.

Expert 2 is a simple hybrid model (Fig. 3) based on the mass balance equation on biomass (X_2),

$$\frac{dX_2}{dt} = (\mu(F_s) - F/V)X_2, \tag{10}$$

being μ the specific growth rate, V the working volume and F_s the glucose feed rate per unit volume defined as $F_s = F\ G_{in}/V$, where G_{in} is the glucose concentration in the feed F. The specific growth rate is considered as an unknown nonlinear function of F_s. The idea is to employ an ANN for approximating the function $\mu(F_s)$. A small feedforward 3-layer network with sigmoidal activation functions is employed in order to avoid overfitting. The network has one input node, five hidden nodes and one output node, this corresponding to 16 parameters in the connections between nodes ($\boldsymbol{w}_2$).

The gating system was configured with 55 Gaussian clusters distributed equidistantly in the interval $F_s \in [0, 10]$ g l^{-1} h^{-1} with a standard deviations of $\sigma = 0.05$. Cluster subset T is composed of 33 clusters in sub-interval $F_s \in [0, 6]$. Cluster subset B is composed of 22 clusters in the complementary sub-interval $F_s \in [6, 10]$. This configuration is illustrated in Fig. 4. The parameter vector **I** is initialised randomly in order to associate experts 1 and 2 with 16 and 17 clusters within subset T respectively. Additionally, all 22 clusters in subset B were associated with expert 2. A total of 55 integer parameters are involved in the gating system.

The KBM network has thus 17 real parameters and 55 integer parameters that must be identified, employing the EM algorithm (Eqs. (8), (9a) and (9b)). The two optimisations in Eq. (9a) are performed employing a

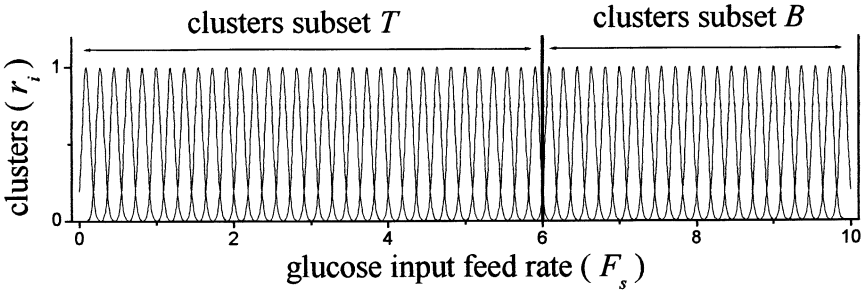

Fig. 4. Cluster distribution in the gating system of the baker's yeast KBM network. Total number of clusters is $N_C = 55$; $N_T = 33$ clusters belonging to subset T; $N_B = 22$ clusters belonging to subset B.

quasi-Newton algorithm (from the NAG library). The optimisation in Eq. (9b) was performed employing a simulated annealing algorithm (Cardoso, Salcedo, & Feyo de Azevedo (1994)).

4. Results and discussion

The results produced by the EM algorithm are rather sensitive to the initial parameter values. If initially one expert describes more accurately the dynamics of the process in the whole input space, the EM algorithm will not converge to the optimal combination of the two experts. The result will be such that only one expert is used for the mapping. As such, in a first step, we adapt independently the two experts to the available process measurements. The biomass prediction results, after this first step, are plotted together with the measured values in Fig. 5 a–e. The statistics used for comparing the models are the Mean-Square-Error (MSE) defined as:

$$\mathrm{MSE} = \frac{1}{P}\sum_{t=1}^{P}(Xm^{(t)} - X^{(t)}) \tag{12}$$

being $Xm^{(t)}$ measured biomass.

The plots in Fig. 5 a–e show that, qualitatively, both experts predict reasonably well the biomass measurements of runs (F1) to (F5), with exception of runs (F3) and (F4) where particularly the predictions of the mechanistic expert are rather bad. The initial MSE is 15.00 and 1.28 for the mechanistic and hybrid experts respectively. A qualitative analysis of the plots shows that for some measured points the mechanistic expert provides better predictions than the hybrid expert. This is a good starting point for applying the EM-algorithm.

In a second step the EM algorithm is employed for parameter identification. The results obtained after 10 iterations are plotted in Fig. 5 a–f. The final MSE is 0.35, which represents a significant improvement in relation to the initial predictions of both experts. Only three out of the 33 clusters in subset T are assigned to the mechanistic expert. This means that the hybrid expert is able to predict more accurately the current set of biomass measurements than the mechanistic expert. In Fig. 5f the gating system outputs are plotted as a function of the glucose feed rate. In the range $F_s \in [0, 6]$ g l^{-1} h^{-1}, the gating output corresponding to the hybrid expert (g_2) is almost always one whereas g_1 is zero. In the range $F_s \in [6, 10]$ g l^{-1} h^{-1} only the mechanistic expert is used. This is in agreement with the initial design condition that clusters in subset B are assigned to a mechanistic expert, which is supposed to be more reliable than other modelling techniques in extrapolation conditions.

5. Conclusion

In this paper a method is presented for improving the quality of process modelling through knowledge integration. The main concept is that of Knowledge Based Modular (KBM) Networks, which aims at providing a general framework for combining those different types of knowledge usually available both in chemical and biochemical processes: mechanistic, heuristic and knowledge hidden in process data records. The method searches for the optimal combination among the available models/sources of knowledge, using well-established statistical methods.

The concept was illustrated and tested with an application to baker's yeast production at laboratory level. Two experts were accepted and trained for the given experimental range of operation. The combined use of such experts has led to improvements in prediction capability as compared to the individual performances of each expert.

Work on the application of this methodology to other type of industrial processes, specifically to industrial crystallisation, is in progress.

A possible difficulty which may hinder the spreading of this type of approach may be on its complexity. With the existing computational capability of modern computers, such difficulty can be easily overcome by developing dedicated, flexible and friendly software tools.

Acknowledgements

This work was developed at the Instituto de Sistemas e Robótica — Porto (www.fe.up.pt/isrp), a Research Unit of the Faculty of Engineering, University of Porto, finantially supported by the Portuguese Foundation for Science and Technology — Fundação para a Ciência e a Tecnologia.

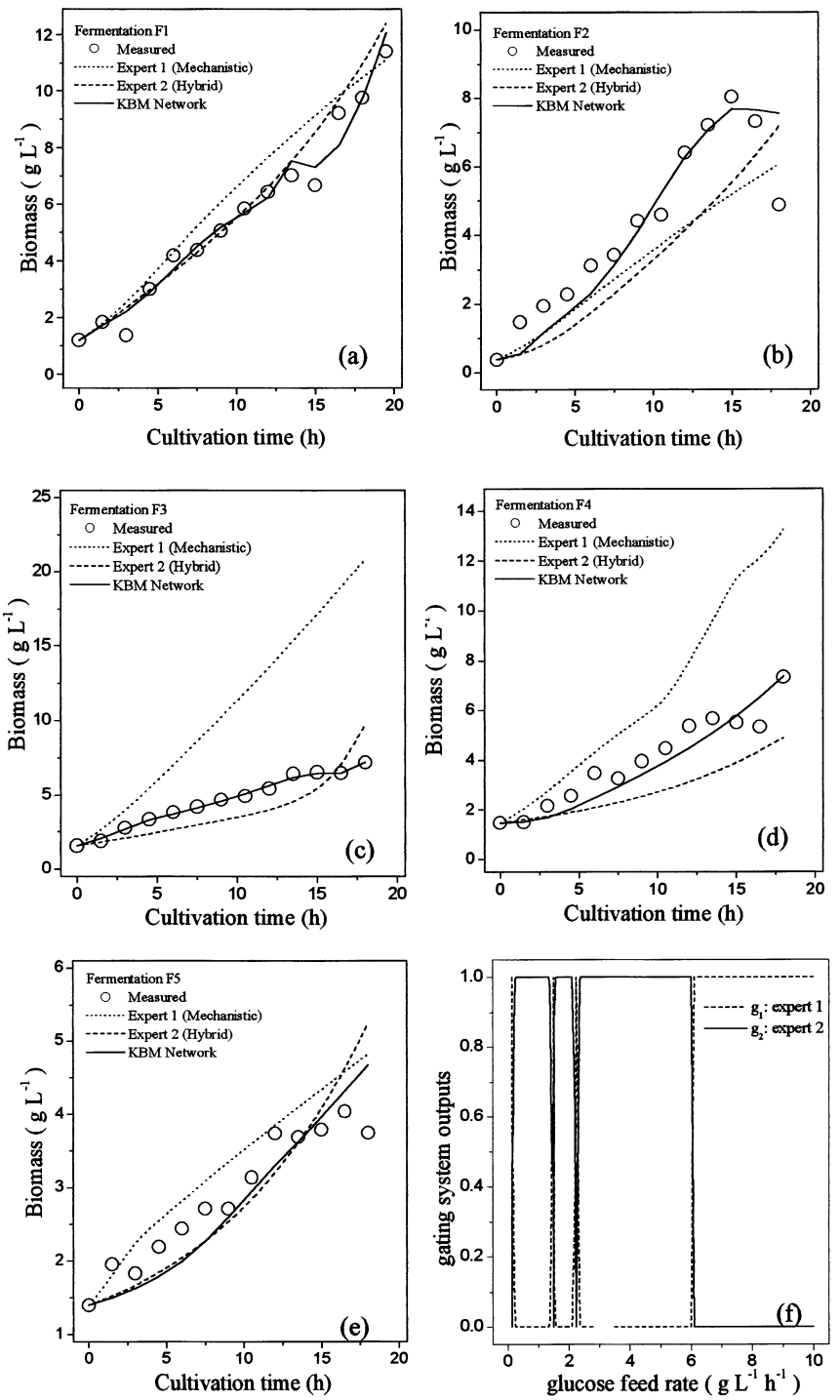

Fig. 5. Biomass prediction curves as a function of cultivation time in a fed-batch baker's yeast fermentation process. (a–e) Results for 5 fermentation runs. The circles represent measured biomass values in a 10 l fermenter. The dotted and dashed lines represent prediction results of Experts 1 and 2 before applying the EM algorithm. The full line represents the prediction results of the KBM Network after applying the EM algorithm. (*f*) Gating system outputs as a function of glucose feed rate.

References

Bhat, N., & McAvoy, T. (1990). Use of neural nets for dynamic modeling and control of chemical processes. *Computers and Chemical Engineering*, *14*, 573–583.

Cardoso, M. F., Salcedo, R. L., & Feyo de Azevedo, S. (1994). Nonequilibrium simulated annealing: a faster approach to combinatorial minimization. *Industrial Engineering and Chemical Research*, *33*, 8.

Feyo de Azevedo, S., Dahm, B., & Oliveira, F. R. (1997). Hybrid modeling of biochemical processes: a comparison with the conventional approach. *Computers and Chemical Engineering*, *21*, S751–S756.

Glassey, J., Igonova, M., Ward, A. C., Montague, G. A., & Morris, J. (1997). Bioprocess supervision: neural networks and knowledge based systems. *Journal of Biotechnology*, *52*, 201–205.

Hitzmann, B., Lübbert, A., & Schügerl, K. (1992). An expert system approach for the control of a bioprocess. I: knowledge representation and processing. *Biotechnology and Bioengineering*, *39*, 33–43.

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

Jacobs, R. A., Jordan, M. I., & Barto, A. G. (1991a). Task decomposition through competition in a modular connectionist architecture: the what and where vision tasks. *Cognitive Science*, *15*, 219–250.

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

Oliveira, F. M., (1997), Fermentation monitoring and control: application to baker's yeast production, Ph.D. thesis, Faculty of Engineering — University of Porto (in Portuguese)

Pomerleau, Y., & Perrier, M. (1990). Estimation of multiple specific growth rates in bioprocesses. *American Institute of Chemical Engineering Journal*, *36*(2), 207–215.

Psichogios, D. C., & Ungar, L. H. (1992). A hybrid neural network-first principles approach to process modelling. *American Institute of Chemical Engineering 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.

Royce, P. N. C. (1993). A discussion of recent developments in fermentation monitoring and control from a practical perspective. *Critical Reviews in Biotechnology*, *13*(2), 117–149.

Schubert, J., Simutis, R., Doors, M., Havlik, I., & Lübbert, A. (1994b). Bioprocess optimization and control: application of hybrid modelling. *Journal of Biotechnology*, *35*, 51–68.

Schubert, J., Simutis, R., Doors, M., Havlik, I., & Lübbert, A. (1994a). Hybrid modelling of yeast production processes — combination of a priori knowledge on different levels of sophistication. *Chemical Engineering and Technology*, *17*, 10–20.

Simutis, R., Havlik, I., & Lübbert, A. (1993). Fuzzy-aided neural network for real-time state estimation and process prediction in the alcohol formation step of production-scale beer brewing. *Journal of Biotechnology*, *27*, 203–215.

Simutis, R., Oliveira, R., Manikowski, M., Feyo de Azevedo, S., & Lübbert, A. (1997). How to increase the performance of models for process optimisation and control. *Journal of Biotechnology*, *59*, 73–89.

Sonnleitner, B., & Käppeli, O. (1986). Growth of *Saccharomyces Cerevisiae* is controlled by its limited respiratory capacity: formulation and verification of a hypothesis. *Biotechnology and Bioengineering*, *28*, 927–937.

Sonnleitner, B. (Ed.) (1999), Bioanalysis and biosensors for bioprocess monitoring. *Advances in Biochemical Engineering/Biotechnology* 66

Sterbacek, Z., & Votruba, J. (1993). An expert system applied to the control of an industrial-scale bioreactor. *The Chemical Engineering Journal*, *51*, B35–B42.

Thompson, M. L., & Kramer, A. (1994). Modeling chemical processes using prior knowledge and neural networks. *American Institute of Chemical Engineering Journal*, *40*(8), 1328–1340.

Weigend, A. S., Mangeas, M., & Srivastava, A. N. (1995). Nonlinear gated experts for time series: discovering regimes and avoiding overfitting. *International Journal of Neural Systems*, *6*, 373–399.

Xu, L., & Jordan, M. I. (1996). On convergence properties of the EM algorithm for Gaussian mixtures. *Neural Computation*, *8*, 129–151.