Hybrid modelling of biotechnological processes using neural networks

L. Chen[a], O. Bernard[b], G. Bastin[a],*, P. Angelov[c]

[a]*Centre for Systems Engineering and Applied Mechanics (CESAME), Universite Catholique de Louvain, Batiment Euler, 4-6, avenue G. Lemaitre, 1348 Louvain la Neuve, Belgium*
[b]*Inria, Projet Comore, B.P. 93, 06902 Sophia Antipolis, France*
[c]*Loughborough University, Loughborough, Leicestershire LE11 3TU, UK*

Abstract

The hybrid modelling approach for bioprocesses combines a neural network representation of the reaction rates with a mass–balance description of the reactor. A procedure for the identification of hybrid models is proposed and illustrated with an experimental case-study. The key feature is a state transformation which allows to identify separately the kinetic models of the reaction rates even if they occur simultaneously in the reactor. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Biotechnology; Identification; Modelling; Neural networks

1. Introduction

Classical models of biotechnological processes in stirred tank bioreactors usually represent the specific reaction rates in the form of rational functions of the species concentrations. An interesting alternative is to use nonlinear black-box models (like neural networks) for the reaction rates. Their advantage is that they have a flexible structure and can be dependent on qualitative features as well as the quantitative values of the state variables.

A new approach was recently proposed to identify the so-called *hybrid models*, which combine the flexibility of a neural network representation of the reaction rates with a global mass–balance-based process description.

In the present paper, this hybrid modelling approach is combined with the general methodology for the structural identification of bioprocesses described in Bernard and Bastin (1998). The proposed methodology allows the separation of the identification of the pseudo-stoichiometric coefficients of the model from the identification of the reaction rates. The procedure is based on a state transformation which enables the kinetic functions to be completely decoupled from one another. This means that each biological reaction occurring in the

reactor can be represented separately by a neural net model as if it was the only one, although all the involved reactions take place simultaneously.

The procedure is illustrated with experimental data from an industrial pilot process.

2. Methodology

A general mass–balance model for biotechnological processes in stirred-tank bioreactors is written under the following form (Bernard & Bastin, 1998, Bastin & Dochain, 1990):

$$\frac{d\xi}{dt} = Kr(\xi) + u, \tag{1}$$

where the state $\xi = (\xi_1, \xi_2, \ldots, \xi_n)^T \in R^n$ is the vector of species concentrations in the reactor. The first term $Kr(\xi)$ represents the biological and biochemical conversions in the reactor. $K \in R^{n \times m}$ is the stoichiometric matrix and $r(\xi) = (r_1(\xi), r_2(\xi), \ldots, r_m(\xi))^T \in R^m$ is the vector of reaction rates. The second term $u = (u_1, u_2, \ldots, u_n) \in R^n$ represents the balance between inflows/outflows and dilution.

The modelling problem of a bioprocess described by this equation involves two sub-problems: (i) the estimation of the unknown stoichiometric coefficients which are entries of the matrix K; and (ii) the modelling of the reaction rates $r_i(\xi)$. This modelling problem is considered

* Corresponding author. Fax: + 3210-472-380.
 E-mail address: bastin@auto.ucl.ac.be (G. Bastin).

here under the assumption that (at least) one experiment is available where all the state variables ξ_i and all the "exogenous" signals u_i have been measured with a reasonable sampling frequency. The procedure described in Bernard and Bastin (1998) allows these two modelling sub-problems to be completely separated. It is based on a state transformation which is now presented.

It is assumed that the number of species n (the dimension of ξ) is larger than the number of reactions m (the dimension of $r(\xi)$) and that the stoichiometric matrix K is full rank. Under this assumption, two matrices $L \in R^{(n-m) \times n}$ and $M \in R^{m \times n}$ can be defined so that

$$LK = O_{(n-m)},$$

$$MK = I_m,$$

where M is a left inverse of K while L is orthogonal to K. Furthermore, these two matrices are chosen so that the $(n \times n)$ matrix

$$\begin{pmatrix} M \\ L \end{pmatrix}$$

is invertible. Then the following state transformation is defined:

$$\begin{pmatrix} y \\ z \end{pmatrix} = \begin{pmatrix} M \\ L \end{pmatrix} \xi.$$

In the new state variables (y, z), the state-space model is rewritten as two submodels

$$\frac{dy}{dt} = r(\xi) + Mu, \tag{2}$$

$$\frac{dz}{dt} = Lu. \tag{3}$$

The second part (3) of this transformed model does not explicitly involve the reaction rates and can be used to estimate the stoichiometric coefficients without modelling the reaction rates. The method is described in detail in Bernard and Bastin (1998) while the conditions under which the stoichiometric coefficients are identifiable with this method are described in Chen and Bastin (1996). An experimental application of this method can be found in Bernard, Bastin, Stentelaire, Lesage and Asther (1999).

This paper focuses on the first part (2) of the transformed model. Thus, it is assumed that the stoichiometric coefficients necessary to compute M are known or have been identified. In accordance with the minimal modelling principle (see e.g. Bastin & Dochain, 1990, Chapter 2), each reaction rate $r_i(\xi)$ is written in the form

$$r_i(\xi) = \varphi_i(\xi)\rho_i(\xi), \tag{4}$$

where $\varphi_i(\xi)$ is a *known* function and $\rho_i(\xi)$ (called *partial reaction rate*) represents the *unknown part* of the model. Obviously, when the reaction rate is assumed to be completely unknown, then $\varphi_i(\xi) = 1$ and $r_i(\xi) = \rho_i(\xi)$. The

interesting point is that the submodel (2) is written elementwise

$$\frac{dy_j}{dt} = \varphi_j(\xi)\rho_j(\xi) + (Mu)_j, \quad j = 1, \dots, m. \tag{5}$$

In this way, it appears that each reaction rate is completely decoupled from the other ones: the reaction rate r_j only enters the dynamics of y_j. It is then possible to try to identify a model for $\rho_j(\xi)$ *as if it were the only reaction occurring in the reactor!* The identification is thus much easier to perform than in the case where all the reaction rates are to be identified together.

Classical models of kinetic functions take the form of a rational fraction (like Monod, Contois or Haldane models, etc.). In this paper, we follow the hybrid modelling paradigm which has been proposed, e.g. in Psichogios and Ungar (1992), Chen et al. (1995) and Feyo de Azevedo, Dahm and Oliveira (1997): radial basis function neural networks (RBF-NN) are used as black-box non-linear models for the representation of the unknown part of the reaction rates as function of the species concentrations

$$\rho_j(\xi) = NN_j(\xi) = w_{0j} + \sum_{i=1}^{nh} w_{ij} N_i(\xi(j)), \tag{6}$$

where nh is the number of neurons in the hidden layer and w_{ij} are the weights of the output layer. The notation $\xi(j)$ represents a subset of the components of ξ which are the species that are supposed to influence the partial reaction rate ρ_j. The functions $N_i(.)$ are the radial basis functions defined as

$$N_i(v) = \exp\left(-\sqrt{-\log 0.5} \sum_k \frac{(v_k - \theta_{ik})^2}{\sigma^2}\right), \tag{7}$$

where θ_{ik} are the weights of the hidden layer and σ is the spread parameter.

The identification procedure is as follows.

(1) The data are first smoothed and interpolated with cubic splines denoted ξ^S and u^S. Non parametric estimates $\hat{\rho}_j(t)$ of the time evolution of the partial reaction rates are then computed by using any appropriate estimation algorithm (see e.g. Bastin & Dochain, 1990), for instance an observer-based estimator of the form

$$\frac{d\hat{y}_j}{dt} = \varphi_j(\xi^S)\hat{\rho}_j + (Mu^S)_j + \omega(y_j^S - \hat{y}_j), \tag{8}$$

$$\frac{d\hat{\rho}_j}{dt} = \lambda\varphi_j(\xi^S)(y_j^S - \hat{y}_j) \tag{9}$$

with $y_j^S = (L\xi^S)_j$ and λ, ω the design parameters.

(2) For each ρ_j separately, the training of the NN model (6) is then carried out with the MATLAB Neural

Net Toolbox. The weight coefficients w_{ij} and θ_{ij} are optimized by minimization of the mean-square error

$$J_1 = \int [\hat{\rho}_j(\tau) - NN_j(\xi^S(\tau))]^2 \, d\tau. \tag{10}$$

The number nh of hidden neurons and the spread parameter σ are selected by trial and error as it will be illustrated in the next section.

(3) A final refinement of the model is achieved by retuning the weights w_{ij} of the output layer in order to minimize the mean-square residual simulation error

$$J_2 = \int \|\xi^{sim}(\tau) - \xi^S(\tau)) \, d\tau\|^2, \tag{11}$$

where the simulated value $\xi^{sim}(\tau)$ is obtained from integrating the model equation (1):

$$\frac{d\xi^{sim}}{dt} = Kr(\xi^{sim}) + u^S$$

with each reaction rates defined as $r_j(\xi^{sim}) = \varphi(\xi^{sim})NN_j(\xi^{sim})$. This is easily done since the w_{ij} parameters enter the model (1) linearly.

Note that the quality of the interpolation using cubic splines obviously has an important influence on the quality of the subsequent estimations and the choice of the model structure. A trade-off has to be made between the smoothness and the accuracy of the approximation.

3. Case-study

In this section the methodology is illustrated with an experimental case-study. This paper considers the typical example of a fed batch process involving a single microbial population growing on a single limiting substrate with an enzyme catalysed production of a secondary metabolite. The process is described by the following two reactions:

Biomass growth: $S \to k_1 X$,

Production: $S \to k_2 P$,
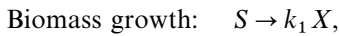
where S represents the substrate, X the biomass, P the product and k_1, k_2 (pseudo) stoichiometric coefficients.

The process takes place in a fed-batch stirred tank-bioreactor with a progressive substrate feeding and occasional withdrawals of the culture medium. The dynamics of the process are described by the mass–balance model

$$\frac{dS}{dt} = -\mu_1(S)X - \mu_2(S)X + \frac{F_{in}}{V}(S_{in} - S),$$

$$\frac{dX}{dt} = k_1 \mu_1(S)X - \frac{F_{in}}{V}X, \tag{12}$$

$$\frac{dP}{dt} = k_2 \mu_2(S)X - \frac{F_{in}}{V}X,$$

where S, X, P now denote the substrate, biomass and product concentrations in the reactor, S_{in} the inlet substrate concentration, F_{in} the volumetric inflow rate, V the volume of the culture medium, $\mu_1(S)$ the specific growth rate and $\mu_2(S)$ the specific production rate. The volume V is governed by the flow balance equation

$$\dot{V} = F_{in} - F_{out} \tag{13}$$

with F_{out} being the rate of withdrawals.

Model (12) is in the state-space form (1) with the following notations:

$$\xi = \begin{pmatrix} S \\ X \\ P \end{pmatrix}, \quad r = \begin{pmatrix} \mu_1(S)X \\ \mu_2(S)X \end{pmatrix},$$

$$K = \begin{pmatrix} -1 & -1 \\ k_1 & 0 \\ 0 & k_2 \end{pmatrix}, \quad u = \frac{F_{in}}{V} \begin{pmatrix} S_{in} - S \\ -X \\ -P \end{pmatrix}.$$

In the two reaction rates, there is a known part

$$\varphi_1 = X, \quad \varphi_2 = X \tag{14}$$

and an unknown part which is to be identified

$$\rho_1 = \mu_1(S), \quad \rho_2 = \mu_2(S). \tag{15}$$

A set of experimental data collected in an industrial pilot plant is used here as a benchmark for the case study (see Figs. 1 and 2). The experiment lasts about 230 h. The limiting substrate is a carbon source. The other nutrients and substrates needed for the growth are assumed to be non-limiting. The microorganism is a fungus and the excreted product is an antibiotic species. The sampled measurements of substrate S, biomass X and product P are represented in Fig. 1. These measurements are interpolated and smoothed with cubic splines also shown in the figure. The inflow rate F_{in} and the withdrawals $W_i = F_{out}/V$ are presented in Fig. 2. Observe that there are only two pulses of withdrawals while there is a permanent constant piecewise feeding of the reactor. The culture volume computed with Eq. (13) is also presented. The influent substrate concentration is $S_{in} = 300$ g/l. The yield coefficients identified from this data set are $k_1 = 0.44$ and $k_2 = 0.18$.

In this example, a straightforward choice for the matrix M is

$$M = \begin{pmatrix} 0 & k_1^{-1} & 0 \\ 0 & 0 & k_2^{-1} \end{pmatrix}.$$

The model development then follows the three-step procedure presented in the previous section.

Step 1: *Data smoothing and computation of non-parametric estimates of μ_1 and μ_2.* In the first step the offline data is smoothed and interpolated with cubic
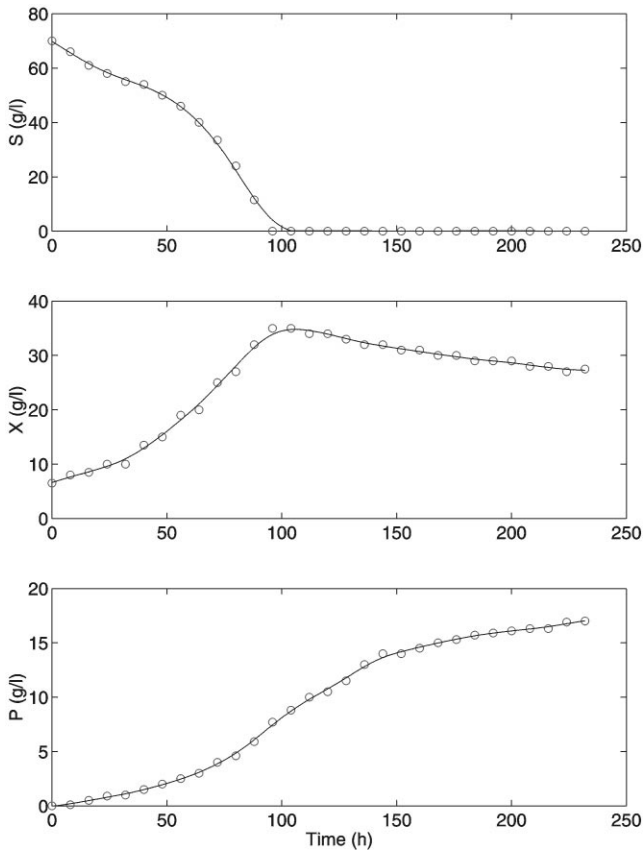
Fig. 1. Experimental data (○) and smoothed curves by cubic splines.
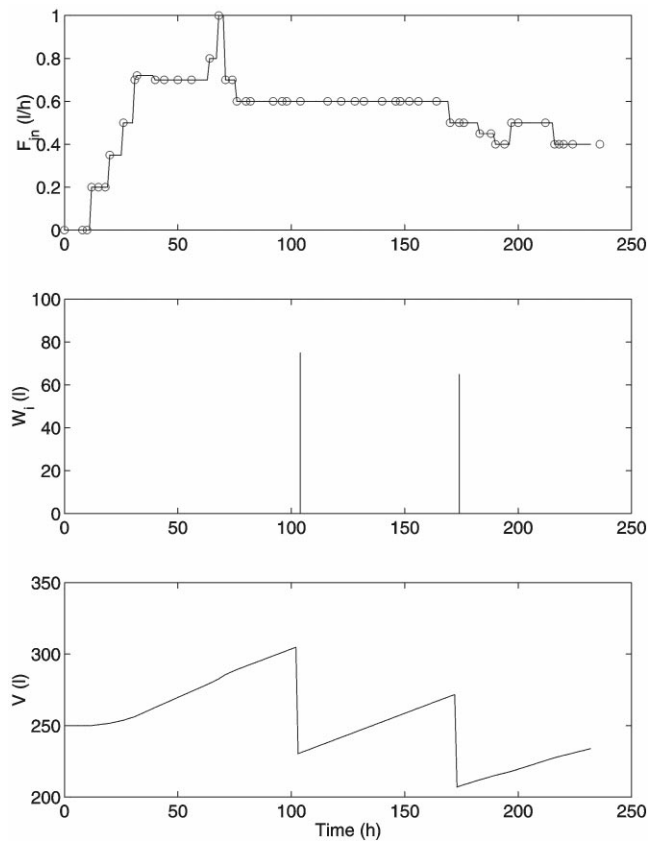

Fig. 2. Experimental data: inflow rate, withdrawals and volume.

splines. The interpolation is chosen such that the degree of smoothness is good enough to avoid unrealistic oscillations in the interpolated values while maintaining a satisfactory approximation to the original data (see Fig. 1).

Non-parametric estimates $\hat{\mu}_1$ and $\hat{\mu}_2$ are computed with the observer-based estimator (8) and (9). The result is shown in Fig. 5.

Step 2: *Identification of RBF-NN models*. In the second step, RBF-NN models $NN_1(S)$ and $NN_2(S)$ are identified for the two functions $\mu_1(S)$ and $\mu_2(S)$. For training the neural nets, the input is the spline-smoothed substrate concentration (Fig. 1) and the output is the non parametric estimate $\hat{\mu}_1$ or $\hat{\mu}_2$. As for any black-box model identification, the choice of the model is a trade-off between accuracy and complexity. In this paper, the criterion of accuracy is the mean-square error J_1

$$J_1 = \int [\hat{\mu}_j(\tau) - NN_j(S)]^2 \, d\tau, \quad j = 1, 2, \tag{16}$$

while the criterion for complexity is the condition number CN (the ratio between the smallest and the largest singular values) of the matrix made up of the radial basis

functions considered over the range of available input values. Two parameters are available to optimize this trade-off: the number of hidden nodes nh and the spread coefficient σ (see Eqs. (6) and (7)). For each choice of nh and σ, the determination of the other parameters w_{ij} and θ_{ij} is automatically performed with the MATLAB Neural Net toolbox.

The identification results for μ_1 are presented in Fig. 3. In Fig. 3(a) and (b) the mean square error J_1 and the condition number CN are plotted with respect to the number of hidden neurons nh and the spread parameter σ. A preliminary investigation has shown that values of nh in the range 3–5 and of σ in the range 0.02–0.04 are relevant. It was observed that the smallest mean-square error is achieved with models having $nh = 5$ neurons but with a bad condition number that indicates overparametrization. A good trade-off seems to be a model with parameters $nh = 4$ and $\sigma = 0.03$. The radial basis functions selected for this model are shown in Fig. 3(c). The parameters θ_{ij} and w_{ij} are given in Table 1. The quality of the fitting may be appreciated in Fig. 5.

In a similar way a model with parameters $nh = 4$ and $\sigma = 0.025$ is selected for μ_2. The parameters θ_{ij} and w_{ij} are given in Table 2 and the radial basis functions in Fig. 4.
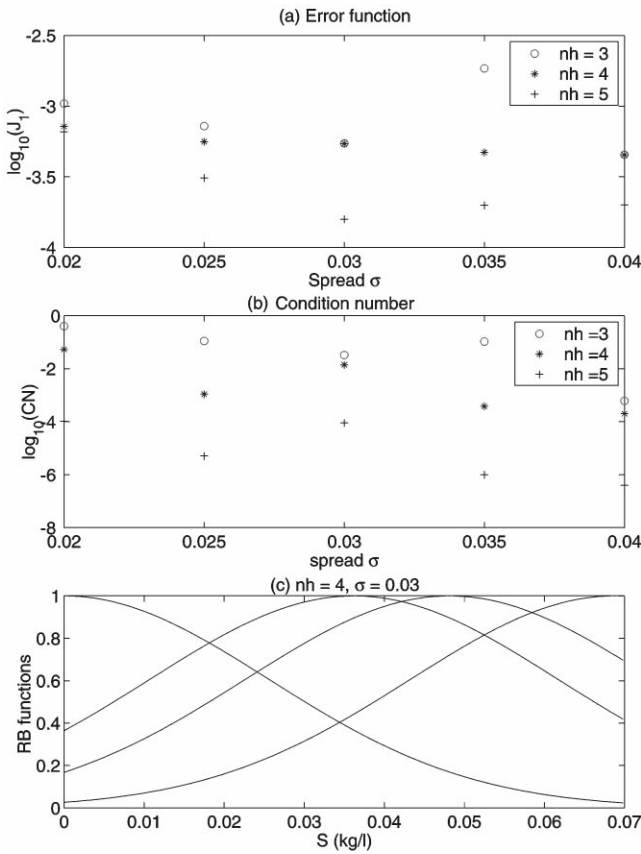
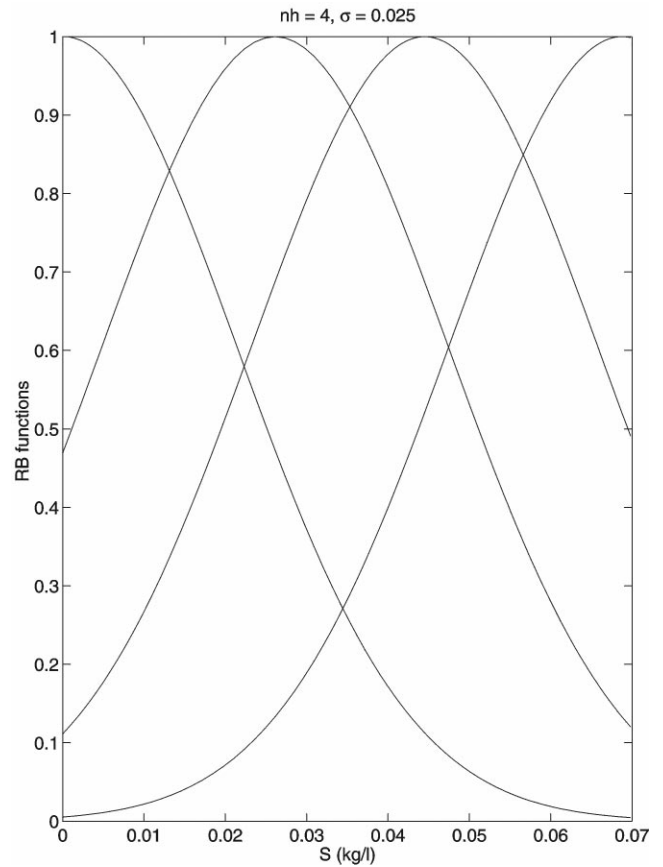Fig. 3. First identification of the hidden basis functions for μ_1.

Fig. 4. Radial basis functions for μ_2.

Step 3: *Final refinement.* A final refinement is performed by retuning the weights w_{ij} of the two models in order to minimize the mean-square simulation error for the substrate S, the biomass X and the product P as defined in (11). This is achieved using an iterative method. The simulation is compared to the experimental data in Fig. 6. The updated RBF-NN models of μ_1 and μ_2 are given in Fig. 5 while the refined values of w_{ij} are in Tables 1 and 2.

Comments. 1. In order to make a judicious choice of the model structure most of the data points after 100 h (see Fig. 1) are removed in step 2. Those repetitive points for S close to 0 would put a too high weight on the accuracy requirement for the model in that region which in turn requires a large number of neurons and implies numerical problems. However, the full data set is used in step 3 for the final refinement after the model structure has been chosen. One can see from Fig. 5 that the final refinement retunes the parameters and brings the model to lower values for small S and higher values elsewhere.

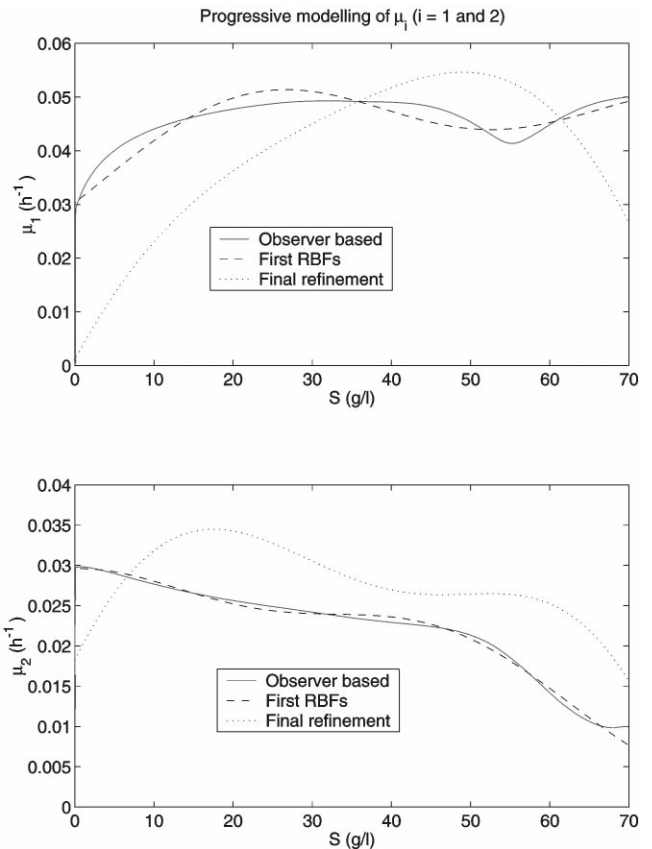2. The simulation quality of the state variables after the final refinement is considerably improved compared to

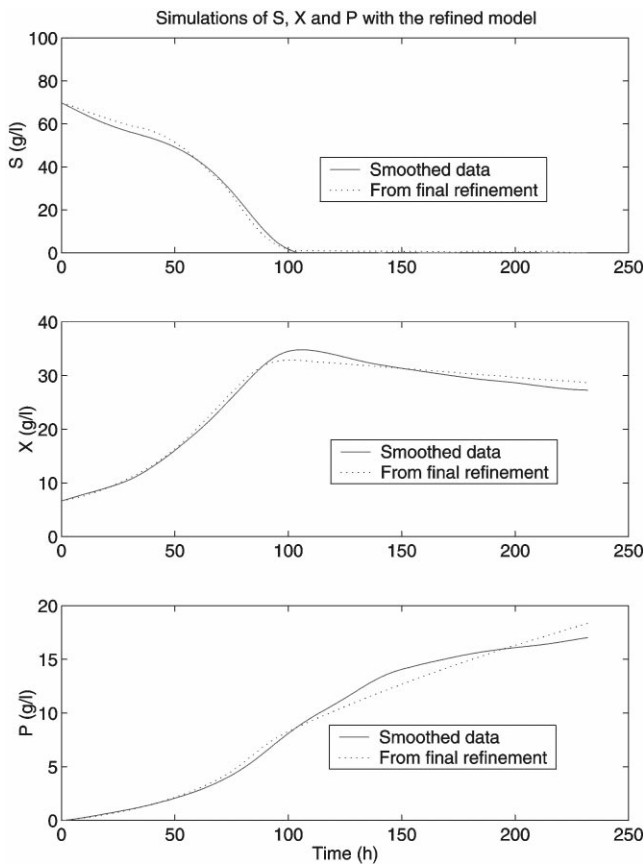Fig. 5. Successive models for the functions μ_1 and μ_2.

Fig. 6. Simulation results compared to the experimental data.

Table 1
Parameters for model μ_1

	μ_1	
	RBFs	Refinement
θ_{11}	3.6208e-02	
θ_{21}	6.8759e-02	
θ_{31}	4.8145e-02	
θ_{41}	1.0000e-04	
w_{11}	1.3857e-01	1.1706e-01
w_{21}	8.6517e-02	1.6212e-01
w_{31}	− 1.5073e-01	− 9.8384e-02
w_{41}	− 6.9741e-03	1.3616e-01
w_{01}	9.7893e-03	− 1.8037e-01

the simulation results with observer-based model and the first RBFs approximation (not shown in the paper). This can be easily explained in the context of the identification theory. Both models from the first approximations are obtained with an optimization criterion based on one-step ahead prediction errors while in the final refinement simulation errors or output errors are used to achieve a long-term optimization result.

Table 2
Parameters for model μ_2

	μ_2	
	RBFs	Refinement
θ_{12}	2.6119e-02	
θ_{22}	6.8759e-02	
θ_{32}	1.3220e-04	
θ_{42}	4.4496e-02	
w_{12}	− 1.0150e-02	6.0368e-02
w_{22}	− 7.9774e-03	8.9921e-02
w_{32}	2.7995e-02	7.8907e-02
w_{42}	2.7444e-02	2.1014e-02
w_{02}	3.3479e-03	− 9.1818e-02

4. Conclusions

A procedure for the identification of hybrid models has been proposed and illustrated with an experimental case-study. The key feature is a state transformation that allows the separate identification of the kinetic models of the reaction rates even if they occur simultaneously in the reactor. This identification procedure is divided into three steps. The first step consists of complementing data information by the interpolation of off-line data and building observer-based non-parametric estimation of the reaction rates that are modelled in the subsequent steps. In step 2, the model structure of each reaction rate is identified based on the separate target values provided by the observer-based estimates. The use of the radial basis functions and the explicit target for the function being modelled makes it easy in practice to choose a suitable model structure using qualitative and quantitative information. The procedure suggests the retuning of the output layer parameters by minimizing the simulation errors of the state variables, which is a more global optimization criterion. The case study has indeed shown the easy practice of this procedure and the model accuracy improvement through progressive modelling.

The procedure presented in this paper involves some empirical knowledge to achieve a good model result. For instance, the interpolation using cubic splines has an important effect on the choice of model structure. A trade-off between the smoothness and the accuracy of the approximation has been used in the paper. The choice of the number of hidden nodes (here number of radial basis functions) is an another example. Notice that according to the condition number criterion, a slightly better choice for *nh* was 3 but *nh* = 4 was chosen so that the model has more flexibility in the final refinement. Further analysis would be useful to derive more general guidelines for this procedure.

Acknowledgements

References

Bastin, G., & Dochain, D. (1990). *On line estimation and adaptive control of bioreactors*. Amsterdam: Elsevier.

Bernard, O., & Bastin, G. (1998). Structural identification of nonlinear mathematical models for bioprocesses. *Proceedings NOLCOS IFAC symposium*, Enschede, The Netherlands (pp. 448–453).

Bernard, O., Bastin, G., Stentelaire, C., Lesage, L., & Asther, M. (1999). Mass balance modelling of vanillin production from vanillic acid by cultures of the fungus *Pycnoporus cinnabarinus* in bioreactors. *Biotechnology and Bioengineering 65*(5) 558–571.

Chen, L., & Bastin, G. (1996). Structural identifiability of the yield coefficients in bioprocess models when the reaction rates are unknown. *Mathematical Biosciences*, *132*, 35–67.

Chen, L., Zhang, J., Morris, A. J., Montague, G. A., Kent, C. A., & Norton, J. P. (1995). Combining neural networks with physical knowledge in modelling and state estimation of bioprocesses. *Proceedings of the third European control conference*, Rome (pp. 2426–2431).

Psichogios, D. C., & Ungar, L. H. (1992). A hybrid neural network — first principles approach to process modelling. *A.I.Ch.E. Journal*, *10*, 1499–1511.

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