NON-LINEAR SYSTEM IDENTIFICATION USING UNCOUPLED STATE MULTIPLE-MODEL APPROACH

Rodolfo Orjuela, Didier Maquin and José Ragot

Centre de Recherche en Automatique de Nancy
UMR 7039, Nancy-Université, CNRS
2, Avenue de la Forêt de Haye, 54516
Vandœuvre-lès-Nancy Cedex France

Abstract: Multiple-model approach is an interesting alternative and a powerful tool for modelling complex processes. This paper deals with the off-line identification of non-linear systems employing the multiple-model approach. We use an uncoupled state multiple-model in opposition to the classically used coupled state multiple-model (*Takagi-Sugeno*). The use of this new mutiple-model structure reveals a new undesirable phenomenon, called *unhooking*, that deteriorates the quality of the obtained approximation. An original solution is proposed to avoid this phenomenon.

Keywords: Multiple-models, Non-linear system identification, Model structure

1. INTRODUCTION

Non-linear models are widely used in engineering science applications to describe the dynamic behaviour of real-world processes. Due to their mathematical complexity, they are not easily exploitable for designing a control law or setting up a diagnosis strategy, even if they give a good description of the considered process.

Assuming that the modelled process evolves around an operating point, a linearisation stage of the non-linear model is then possible and leads to the reduction of the mathematical complexity of the model. Hence, analysis tools available for linear systems can be used. However, in practice, the linearity assumption is not always respected and consequently the linearised model does not represent the whole behaviour of the process.

Moreover, it will be more interesting in practice to have a model that gives a good global characterisation of the dynamic behaviour of the process and that is easily usable by engineers for example with techniques for linear systems.

To fulfil these expectations, new modelling techniques have been developped, among them multiple-model approach. We propose in this paper an off-line identification procedure of non-linear systems using an uncoupled state multiple-model approach.

The outline of this paper is as follows. Two multiple-model structures using coupled and uncoupled states are discussed in section 2. Section 3 presents the parametric estimation procedure for uncoupled state multiple-model. The limits and problems encountered in the application of this procedure are illustrated through an academic example in section 4. A modification of the multiple-model structure is then introduced in section 5. The goal is to improve the approximation quality and performances of the identified multiple-model.

2. STRUCTURES OF MULTIPLE-MODELS

The *divide and conquer* strategy is the starting point of many modelling techniques of non-linear systems (N.L.S.). The basis of these techniques is the decomposition of operating space of a non-linear system into a number L of operating zones that are characterised by a sub-model which is often chosen as linear. According to the zone where the non-linear system evolves then, the output $\hat{y}_i$ of each sub-system is more or less requested in order to describe the whole behaviour of the non-linear system.

The multiple-model (M.M.) output $\hat{y}(k)$ is defined by:

$$\hat{y}(k) = \sum_{i=1}^{L} \mu_i(\xi(k))\hat{y}_i(k), \tag{1}$$

the sub-model contribution depends on the *weighting function* $\mu_i(\xi(k))$. A large choice of weighting functions is possible (it is not the purpose of the present text to survey these possibilities). Here, they are obtained from normalised Gaussian function:

$$\mu_i(\xi(k)) = \frac{\omega_i(\xi(k))}{\sum\limits_{j=1}^{L} \omega_j(\xi(k))}, \qquad \text{where} \tag{2}$$

$$\omega_i(\xi(k)) = \exp\left(-(\xi(k) - c_i)^2/\sigma^2\right), \tag{3}$$

c_i is the centre of the i$^{\text{th}}$ weighting function and σ is the dispersion for all weighting functions. Decision variable ξ of weighting functions can depend on the measurable state variables and/or input/output variable. Two notions of weighting functions are defined *strongly blended* and *not much blended* (see Figure 1). There are various

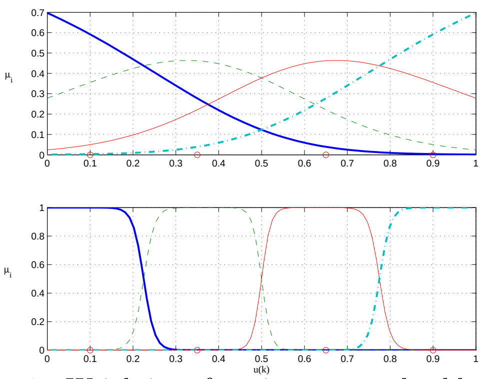

Fig. 1. Weighting functions strongly blended $\sigma = 0.35$ (up), weighting functions not much blended $\sigma = 0.08$ (down)

ways of connecting sub-models in order to generate the global output $\hat{y}(k)$. We can distinguish two multiple-model structures according the use of *coupled* or *uncoupled* states.

2.1 Coupled state multiple-model

The coupled state structure or Takagi-Sugeno (Takagi and Sugeno, 1985) is the more classically used in multiple-model analysis and synthesis (Murray-Smith and Johansen, 1997). The global output of this multiple-model is computed by:

$$\begin{aligned} \hat{x}(k+1) &= \tilde{A}(k)\hat{x}(k) + \tilde{B}(k)u(k) + \tilde{D}(k), \\ \hat{y}(k) &= \tilde{C}(k)\hat{x}(k), \end{aligned} \tag{4}$$

such that:

$$\tilde{A}(k) = \sum_{i=1}^{L} \mu_i(\xi(k))A_i, \; \tilde{B}(k) = \sum_{i=1}^{L} \mu_i(\xi(k))B_i,$$

$$\tilde{D}(k) = \sum_{i=1}^{L} \mu_i(\xi(k))D_i, \; \tilde{C}(k) = \sum_{i=1}^{L} \mu_i(\xi(k))C_i,$$

where $\hat{x} \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}^m$ the control, $\hat{y} \in \mathbb{R}^l$ the output vector and ξ the decision variable of weighting function μ_i.

The coupled state multiple-model is assimilated to a system whose parameters vary with time (notice that there is an only global state $\hat{x}$). Indeed, the sub-model parameters are blended in function of operating zones of the non-linear system.

2.2 Uncoupled state multiple-model

Filev (Filev, 1991) proposed another structure to connect the sub-models by an uncoupled structure:

$$\begin{aligned} \hat{x}_i(k+1) &= A_i\hat{x}_i(k) + B_iu(k) + D_i, \\ \hat{y}_i(k) &= C_i\hat{x}_i(k), \\ \hat{y}(k) &= \sum_{i=1}^{L} \mu_i(\xi(k))\hat{y}_i(k), \end{aligned} \tag{5}$$

where $\hat{x}_i \in \mathbb{R}^{n_i}$ is the state vector of the i$^{\text{th}}$ sub-model, $u \in \mathbb{R}^m$ is the control, $\hat{y} \in \mathbb{R}^l$ is the output vector. Here, the global output of the multiple-model is given by a weighted sum of the sub-system outputs (parameters of sub-models are not blended).

Notice that this multiple-model is given by parallel connection of Wiener-type sub-models (a linear sub-model in series with a non-linear function). Therefore each sub-model has and evolves independently in their own state space in function of the input control and their initial state. Let us be clear that the principal interest of this structure is the decoupling between the sub-models. Indeed, it is possible to think that:

- it is easier to transfer the analysis techniques of linear systems to multiple-model;
- different structures of sub-models can be employed. For example, linear and non-linear models with a different dimension of the state vector (of course the output vector dimension must be identical).

3. PARAMETRIC ESTIMATION

The coupled state multiple-model has been employed in several identification procedures of N.L.S in the last few years.

Johansen and Foss (Johansen and Foss, 1995) have developed an algorithm for operating zone decomposition of N.L.S. based on a priori process information and a heuristic. The sub-models are added at each iteration in function of non-linear system complexity and the wished accuracy (ascending approach). *Gasso* (Gasso, 2000) proposed another operating space decomposition technique of the N.L.S. based on a grid partition. The multiple-model complexity is reduced by elimination of irrelevant sub-models on one hand and merging redundant models on the other one (descending approach). *Gugaliya and al.* (Gugaliya *et al.*, 2005) suggest a modification of CART (Classification and Regression Tree) algorithm used in the piecewise linear model identification in order to constructing a multiple-model.

On the other hand, the identification procedure based on the uncoupled state multiple-model has been not much used. Some works in the control law synthesis of non-linear systems have used successfully this structure (Gawthrop, 1995), (Gregorcic and Lightbody, 2000). The M.M. is obtained by local linearisation of the non-linear system around different operating points.

Recently *Venkat et al.* (Venkat *et al.*, 2003) have proposed an identification methodology based on this multiple-model structure for a control application. The identification of sub-models is based on input/output data that are obtained in one small operating zone of the non-linear system thus assuring the linearity of each set of data. A great number of particular experiments of non-linear system are necessary to generate several sets of data. Indeed, one set of data is necessary for each sub-model identification.

We propose an off-line identification procedure of non-linear systems using an uncoupled multiple-model approach. The identification problem is given in the following statement: the aim is the parameter identification of L sub-models when the weighting functions $\mu_i(\xi)$ are fixed, on one hand and input data $u(k)$ and output data $y(k)$ of a SISO non-linear system are given on the other one.

3.1 Estimation criteria

Typically three estimation criteria: global, local and combined, are employed in the parameter identification of multiple-model. The choice of estimation criterion depends on the purpose (application perspective) of multiple-model.

3.1.1. Global criterion The global criterion is defined by:

$$J_G = \frac{1}{2}\sum_{k=1}^{N} \varepsilon(k)^2, \tag{6}$$

where N is the number of training data and $\varepsilon(k)$ is the global error between the M.M. output and the N.L.S. output, given by :

$$\varepsilon(k) = \hat{y}(k) - y(k). \tag{7}$$

This criterion encourages a good global characterisation between global behaviour of the non-linear system and the multiple-model, but the local behaviour is not considered. This criterion is interesting when the multiple-model is used in a *prediction* perspective.

3.1.2. Local criterion The local criterion is defined by:

$$J_L = \frac{1}{2}\sum_{i=1}^{L}\sum_{k=1}^{N} \mu_i(\xi(k))\varepsilon_i(k)^2, \tag{8}$$

where $\varepsilon_i(k)$ is the local error between the i^{th} sub-model output and the N.L.S. output, given by:

$$\varepsilon_i(k) = \hat{y}_i(k) - y(k). \tag{9}$$

This criterion favours the good local accurate approximation between local behaviour of the sub-models and local behaviour of the non-linear system. The multiple-model obtained may be used in a *phenomenon explication* perspective. Indeed the sub-models may be often comparable to a linearised models of the non-linear system. However, in comparison to global criterion, a number of sub-models more important is necessary to have a good global characterisation of the non-linear system (Gasso, 2000).

3.1.3. Combined criterion The combined criterion defined by (Yen *et al.*, 1998):

$$J_C = \alpha J_G + (1-\alpha)J_L, \tag{10}$$

provides an alternative to combining advantages of the two above criteria, then, it is possible to take advantages of each criterion in function of weighting scalar value of α.

3.2 Parametric estimation procedure

In this section, the parameter estimation base on the uncoupled state multiple-model approach is presented. See (Walter and Pronzato, 1997) for complements and general discussion about identification techniques.

The column vector θ is the vector of unknown parameters of the multiple-model represented in the form of a *partitioned vector* in *L column blocks* θ_p:

$$\theta = [\theta_1 \dots \theta_p \dots \theta_L]^T \tag{11}$$

where each column block θ_p is formed by the q_p unknown parameters of the particular sub-model p:

$$\theta_p = [\theta_{p,1} \dots \theta_{p,q} \dots \theta_{p,q_p}]^T \tag{12}$$

$\theta_{p,q}$ denoting the unknown scalar parameter of the sub-model p.

Here, the parametric estimation of $\theta_{p,q}$ is based on an iterative minimisation procedure of a quadratic criterion J (global, local or combined) with Gauss-Newton's algorithm:

$$\theta^+ = \theta - \mathrm{H}^{-1}\mathrm{G}, \tag{13}$$

where θ is the vector of parameters at a particular iteration, θ^+ this evaluated vector in the following iteration, $\mathrm{H} = \frac{\partial^2 \mathrm{J}}{\partial \theta \partial \theta^\mathrm{T}}$ is the Hessian matrix and $\mathrm{G} = \frac{\partial \mathrm{J}}{\partial \theta}$ the gradient vector. The calculus of the gradient vector and the Hessian matrix is based on the calculation of *sensitivity functions* of output multiple-model with respect to sub-models parameters.

Often, Hessian may not be invertible (singular matrix) or the inverse may not be definite positive, then the parameter update is not possible. To avoid these problems, the Marquardt's algorithm is used for ill-conditioned problems:

$$\theta^+ = \theta - \Delta(\mathrm{H} + \lambda \mathrm{I})^{-1}\mathrm{G}, \tag{14}$$

where Δ is the step size that minimise the criterion in the direction of vector $\mathrm{H}^{-1}\mathrm{G}$, λ is a scalar and I the identity matrix of appropriate dimension. If λ is small then the Gauss-Newton's algorithm is used. On the other hand, if λ is great then the method of gradient descent is used. Notice that if H is a positive definite matrix then there is always a $\Delta \leq 1$ that minimises the criterion in the direction of $\mathrm{H}^{-1}\mathrm{G}$.

The convexity of problem is not ensured then the algorithm solution may not be a global minimum. Consequently, the choice of an initial guess for the parameter vector is a difficult and delicate step in order to ensure the algorithm convergence. Sometimes, in difficult cases, several choices of initial parameters may be necessary.

3.3 Parametric estimation with global criterion

The gradient G_G of J_L is given by the simple derivation of the global criterion (6) with respect to parameters θ:

$$G_G = \frac{\partial J_G}{\partial \theta} = \sum_{k=1}^{N} \varepsilon(k) \frac{\partial \hat{y}(k)}{\partial \theta}, \tag{15}$$

where

$$\frac{\partial \hat{y}(k)}{\partial \theta} = \sum_{i=1}^{L} \mu_i(\xi(k)) \frac{\partial \hat{y}_i(k)}{\partial \theta}, \tag{16}$$

$\frac{\partial \hat{y}_i(k)}{\partial \theta}$ are the sensitivity functions of the i^{th} output with respect to unknown parameters of the multiple-model. The sensitivity functions are defined as follows:

$$\frac{\partial \hat{y}_i(k)}{\partial \theta_{p,q}} = \frac{\partial C_i}{\partial \theta_{p,q}} \hat{x}_i(k) + C_i \frac{\partial \hat{x}_i(k)}{\partial \theta_{p,q}}. \quad \begin{matrix} p = 1, 2, ..., L \\ q = 1, 2, ..., q_p \end{matrix} \tag{17}$$

The state is unknown at the present moment k, the sensitivity functions are calculated at the next moment $k+1$ by derivation of (5) with respect to each parameter $\theta_{p,q}$:

$$\begin{aligned} \frac{\partial \hat{x}_i(k+1)}{\partial \theta_{p,q}} &= \frac{\partial A_i}{\partial \theta_{p,q}} \hat{x}_i(k) + A_i \frac{\partial \hat{x}_i(k)}{\partial \theta_{p,q}} \\ &+ \frac{\partial B_i}{\partial \theta_{p,q}} u(k) + \frac{\partial D_i}{\partial \theta_{p,q}}. \end{aligned} \tag{18}$$

Hessian matrix H_G is obtained by a double derivation of the global criterion (6) with respect to parameters θ:

$$H_G = \frac{\partial^2 J_G}{\partial \theta \partial \theta^T} = \sum_{k=1}^{N} \underbrace{\varepsilon(k)}_{\to 0} \frac{\partial^2 \hat{y}(k)}{\partial \theta \partial \theta^T} + \sum_{k=1}^{N} \frac{\partial \hat{y}(k)}{\partial \theta} \frac{\partial \hat{y}(k)}{\partial \theta^T}. \tag{19}$$

Here, we have employed the Gauss-Newton's algorithm in order to reduce the costly in computational time. Indeed, the second order derivation term in equation (19) is neglected with the assumption that the error $\varepsilon(k)$ tends to zero. Hence, an approached Hessian is obtained by sensitivity functions used in the gradient vector calculus:

$$H_G \approx \sum_{k=1}^{N} \frac{\partial \hat{y}(k)}{\partial \theta} \frac{\partial \hat{y}(k)}{\partial \theta^T}, \tag{20}$$

3.4 Parametric estimation with local criterion

The calculation of the gradient vector G_L and the Hessian matrix H_L is performed as in the case of the global criterion. By noting that the sensitivity functions $\frac{\partial \hat{y}_i(k)}{\partial \theta}$ in the local criterion and global criterion are identical.

Assuming that the weighting functions are not much blended, the distinction between global criterion and local criterion disappears; the two approaches of identification offer similar results. On the other hand, if the weighting functions are strongly blended, we will be able to choose the combined criterion (10) in order to weight the sub-models interpretation compared to the quality of the global model.

It is well worth noting that the expressions (16), (17) and (18) are the generics forms often simplified in the algorithm implementation. Indeed, these equations may be rewritten as:

$$\frac{\partial \hat{y}_i(k)}{\partial \theta_p} = 0 \quad \text{when} \quad p \neq i \begin{matrix} i = 1, 2, ..., L \\ p = 1, 2, ..., L \end{matrix},$$

because the parameters of each sub-model are completely independent (uncoupled structure) of the parameters of the other sub-models.

4. IDENTIFICATION EXAMPLE

Let us use the uncoupled structure to mimic the behaviour of the non-linear system:

$$\begin{aligned} x(k+1) &= Ax(k) + \sin(\gamma u(k))(\beta - u(k)), \\ y(k) &= x(k), \end{aligned} \tag{21}$$

with $A = 0.95$, $\gamma = 0.8\pi$ and $\beta = 1.5$.

Here, the identification of the multiple-model is realised with a global criterion. We want a prediction multiple-model without interpretation of the local behaviours. The parametric estimation is based on relation (14), the gradient vector G_G

and the Hessian matrix H_G are defined by the relations (15) and (20). The multiple-model comprises arbitrarily $L = 6$ sub-models (this quantity may be optimised).

The parameters A_i, B_i, D_i and C_i of the sub-models are scalar type (first order sub-model). The weighting functions μ_i depend on the input signal $\xi(k) = u(k)$, the centres are: $c_i = [0, 0.2, 0.4, 0.6, 0.8, 1]$ and the dispersion $\sigma = 0.2$.

4.1 Linearized multiple-model

The multiple-model is obtained by local linearisation of the non-linear system (21) around different operating points c_i. The dynamic behaviour of the linearised multiple-model is shown in figure 2.

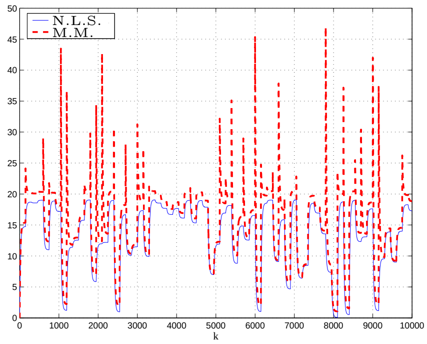

Fig. 2. Non-linear system output and linearised multiple-model output

This new multiple-model structure reveals a new undesirable phenomenon, called *unhooking*, that produces the "peaks" in the commutation zones. On the other hand, using weighting functions strongly blended produces, in this case, an important static error. Indeed, the local characterisation of the sub-models is not guaranteed by the weighting functions.

However, the parameters issued from linearisation may be employed as an initial guess for the parameters in the next identification procedure.

4.2 Identified multiple-model

The input $u(k)$ of the system is formed by the concatenation of piecewise constant signals with variable amplitude ($u \in [0, 1]$). A set of 5000 input/output data points is used to build the multiple-model. A set of identical dimension input/output data points is used to validate the multiple-model. We now employ as the initial parameters those obtained by linearisation of the non-linear system.

The identified multiple model yields a bad mimic of the dynamic behaviour of the non-linear system (21) (see figure 3). Indeed, the appearance of "peaks", in the commutation zones of the control signal, deteriorates considerably the quality of the obtained approximation. The undesirable *unhooking* phenomenon produces a noise effect in the identification procedure. Next section presents some explanations and an original solution in order to eliminate this undesirable phenomenon.

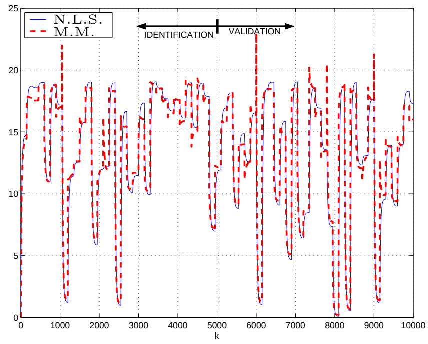

Fig. 3. Non-linear system output and identified multiple-model output

5. MODIFICATION OF MULTIPLE-MODEL STRUCTURE

The *unhooking* phenomenon is due to trade-off between the sub-models outputs at commutation moment and the consequence may be strong variations on the global output.

Indeed, let us remember that each sub-model involves independently in their own state space. Hence, jumps (at commutation moment) between outputs of the sub-models produce sometimes a discontinuous global output of the uncoupled state multiple-model.

Therefore an *unhooking* phenomenon is presented if the sub-model outputs are significantly different at contribution moment. If the sub-model outputs are close, at the contribution moment, then the *unhooking* phenomenon decreases and it is removed if the outputs are identical.

To sum up, the *unhooking* phenomenon is related to an initial condition problem of the sub-model outputs at the contribution moment (notice that this phenomenon does not exist in the coupled state multiple-model approach). Consequently, in the presence of a *unhooking* phenomenon the parametric estimation may be deteriorated and the multiple-model identified gives, in the unhooking zones, a bad approximation of the N.L.S. (see figure 3).

A first idea in order to avoid this phenomenon is to consider strongly blended weighting functions and a greater number of sub-models. On the other hand, *Gawthrop* (Gawthrop, 1995) has proposed the incorporation of the local feedback via classical observer theory to each of the sub-models.

We propose an original solution which consists to filter the control signal, the decision signal of μ_i and the multiple-model output, by adding three low-pass filters in the multiple-model structure. The purpose of this approach is to consider gradually the contribution of each sub-model and to soften commutation between sub-model outputs. These three filters are considered as the unknown parameters in the identification procedure of the multiple-model.

The new multiple-model is defined by:

$$\hat{x}_i(k+1) = A_i\hat{x}_i(k) + B_i\tilde{u}(k) + D_i,$$
$$\hat{y}_i(k) = C_i\hat{x}_i(k), \quad (22)$$
$$\tilde{y}(k) = \sum_{i=1}^{L} \mu_i(\xi(k))\hat{y}_i(k), \text{ with } \xi(k) = \hat{u}(k),$$
$$\tilde{u}(k) = F_1(q^{-1})u(k), \quad (23)$$
$$\hat{u}(k) = F_2(q^{-1})u(k), \quad (24)$$
$$\hat{y}(k) = F_3(q^{-1})\tilde{y}(k), \quad (25)$$

where $\hat{y}$ is the new multiple-model output.

New sensitivity functions are calculate as in the section 3.3, obviously taking into account the new parameters of the three low-pass filters (here, F_1, F_2 and F_3 are first order filters).

Figure 4 shows the new dynamic behaviour of the multiple-model with the three filters. We can notice the good approximation provided by the identified multiple-model. Let us be clear that the phenomenon of unhooking is rejected.

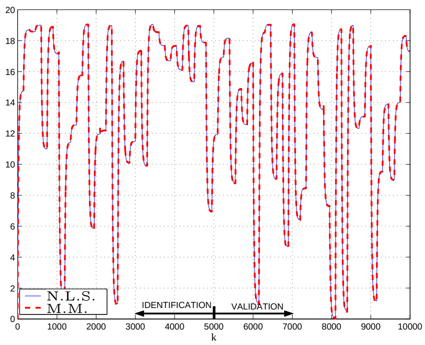

Fig. 4. Non-linear system output and new identified multiple-model output

6. CONCLUSION

We have investigated a non-linear system identification using an uncoupled state multiple-model in contrast to classically used coupled state multiple-model (Takagi-Sugeno).

The uncoupled state multiple-model provides a interesting alternative in the modelling, the control and the diagnosis of non-linear systems. Indeed, the particularity of this structure is to have the completely independent sub-models. Therefore, one can suppose that analysis tools available for linear systems can be a priori easily applied in contrast to coupled state multiple-model.

On the other hand, an undesirable phenomenon, called *unhooking*, that deteriorates the quality of the obtained approximation is revealed. An original solution is suggested to avoid this phenomenon. The solution is based on the incorporation of three low-pass filters in the identification procedure. An academic example is also provided to demonstrate the good efficiency of this solution.

The proposed identification procedure can be extended to include the optimisation of multiple-model dimension: sub-model orders and their number. It is possible also to consider an optimisation procedure of the weighting functions μ_i. The identification of MISO non-linear systems may be also contemplated.

REFERENCES

Filev, D. (1991). Fuzzy modeling of complex systems. *International Journal of Approximate Reasoning* **5**, 281–290.

Gasso, K. (2000). Identification des systèmes dynamiques non-linéaires : approche multi-modèle. PhD thesis. Institut National Polytechnique de Lorraine.

Gawthrop, P.J. (1995). Continuous-time local state local model networks. In: *Proceedings of IEEE Conference on Systems, Man & Cybernetics*. Vancouver, Canada. pp. 852–857.

Gregorcic, G. and G. Lightbody (2000). Control of highly nonlinear processes using self-tuning control and multiple/local model approaches. In: *2000 IEEE International Conference on Intelligent Engineering Systems, INES 2000*. pp. 167–171.

Gugaliya, J.K., R.D. Gudi and S. Lakshminarayanan (2005). Multi-model decomposition of nonlinear dynamics using a fuzzy-CART approach. *Journal of Process Control* **15**(4), 417–434.

Johansen, T. A. and B. A. Foss (1995). Identification of non-linear system structure and parameters using regime decomposition. *Automatica* **31**(2), 321–326.

Murray-Smith, R. and T.A. Johansen (1997). *Multiple model approaches to modelling and control*. Taylor & Francis.

Takagi, T. and M. Sugeno (1985). Fuzzy identification of systems and its applications to modelling and control. *IEEE Trans. on Systems Man and Cybernetics* **15**, 116–132.

Venkat, A.N., P. Vijaysai and R.D. Gudi (2003). Identification of complex nonlinear processes based on fuzzy decomposition of the steady state space. *Journal of Process Control* **13**(6), 473–488.

Walter, E. and L. Pronzato (1997). *Identification of parametric models: from experimental data*. Chap. 4. Springer-Verlag Berlin and Heidelberg GmbH & Co. K.

Yen, J., L. Wang and C.W. Gillespie (1998). Improving the interpretability of Takagi-Sugeno fuzzy models by combining global learning and local learning. *IEEE Trans. on Fuzzy Systems* **6**, 530–537.