amcs

NONLINEAR SYSTEM IDENTIFICATION USING HETEROGENEOUS MULTIPLE MODELS

RODOLFO ORJUELA *, BENOÎT MARX **,***, JOSÉ RAGOT **,***, DIDIER MAQUIN **,***

* Modelling Intelligence Process and System Laboratory, MIPS, EA 2332
University of Haute-Alsace, 12 rue des frères Lumière, F-68093 Mulhouse, France
e-mail: rodolfo.orjuela@uha.fr

**CRAN, UMR 7039
University of Lorraine, 2 Av. de la forêt de Haye, F-54516 Vandœuvre-lès-Nancy, France
e-mail:{benoit.marx,jose.ragot,didier.maquin}@univ-lorraine.fr

***CRAN, UMR 7039
CNRS, France

Multiple models are recognised by their abilities to accurately describe nonlinear dynamic behaviours of a wide variety of nonlinear systems with a tractable model in control engineering problems. Multiple models are built by the interpolation of a set of submodels according to a particular aggregation mechanism, with the *heterogeneous multiple model* being of particular interest. This multiple model is characterized by the use of *heterogeneous submodels* in the sense that their state spaces are not the same and consequently they can be of various dimensions. Thanks to this feature, the complexity of the submodels can be well adapted to that of the nonlinear system introducing flexibility and generality in the modelling stage. This paper deals with off-line identification of nonlinear systems based on heterogeneous multiple models. Three optimisation criteria (global, local and combined) are investigated to obtain the submodel parameters according to the expected modelling performances. Particular attention is paid to the potential problems encountered in the identification procedure with a special focus on an undesirable phenomenon called the *no output tracking effect*. The origin of this difficulty is explained and an effective solution is suggested to overcome this problem in the identification task. The abilities of the model are finally illustrated via relevant identification examples showing the effectiveness of the proposed methods.

Keywords: nonlinear system identification, multiple models, heterogeneous submodels.

1. Introduction

Effective dealing with many problems in control engineering, such as modelling, control and diagnosis, generally needs the use of nonlinear models instead of the standard Linear Time-Invariant (LTI) models. The latter can be satisfactorily used to represent a mildly nonlinear system in some situations as proposed by Mäkilä and Partington (2003). However, they have fundamental limitations in their abilities to deal with highly nonlinear behaviours and complex interactions in a large operating range, for example, when an extended working range of the system (global behaviour) must be considered in preference to a reduced operating range in the neighbourhood of an operating point (local behaviour).

Nonlinear models are frequently obtained from theoretical modelling on the basis of *a priori* knowledge on the nature and the intrinsic mechanisms of the systems. These white-box models are often extensive and complex, and their usability for real-world applications (control, state estimation, etc.) often requires some simplifications. Besides, the *a priori* knowledge on relatively complex systems is partially or totally unavailable in many practical situations, and consequently the theoretical modelling fails. In this quite common case, experimental modelling (identification) from measurements must be used as an alternative to theoretical modelling. Nonlinear system identification has been the focus of a great deal of attention in the past decades and several models with adequate algorithms have been proposed, for instance,

temporal Volterra series, kernel estimators, block oriented models, radial basis function networks, artificial neural networks, fuzzy models, neuro-fuzzy models, multiple models, etc. (Sjöberg *et al.*, 1995; Babuška, 1998; Ljung, 1999; Nelles, 2001). Nevertheless, adequate structure selection in this multitude of existing structural nonlinear models is particularly critical for the success of the identification task as well as model usability. Ideally, as mentioned by Edwards and Hamson (2001), *the success of a model depends on how easily it can be used and how accurate are its predictions. Note also that any model will have a limited range of validity and should not be applied outside this range.* In summary, accurate modelling in a large domain of validity must be accomplished by models as simple as possible to cope with real-world problems.

Multiple models, also known as *local model networks*, are recognised by their abilities to capture highly nonlinear behaviours in a wide operational range of the system with an exploitable model (Johansen and Foss, 1993; Murray-Smith and Johansen, 1997; Leith and Leithead, 1999). Indeed, multiple models offer a good trade-off between accuracy, complexity and usability. In this modelling approach, the operating space of the system is decomposed in a finite number of possibly overlapping operating zones. The closeness to each of them is quantified by a *weighting function*. Linear models (*submodels*) are then able to describe the dynamics of the system inside each operating region. The global modelling of the system is finally performed by considering the contribution of each submodel according to the operating point of the system. In this way a single complex model whose parameters are not easily identifiable is replaced by a set of submodels judiciously interconnected via an interpolation mechanism. Different structures of multiple models can be derived according to the interpolation mechanism. However, as stated by Filev (1991), two main structures can be clearly distinguished based on the use of *homogeneous* or *heterogeneous submodels*. A comparison between these multiple models has recently been proposed by Gregorčič and Lightbody (2008).

In the first structure, all the submodels share the same state space and are consequently called *homogeneous submodels*. A good example of this kind of multiple model is given by the well-known *Takagi–Sugeno* (TS) multiple model proposed by Takagi and Sugeno (1985), in a fuzzy modelling framework, and by Johansen and Foss (1993), in a multiple model modelling framework. This multiple model structure, called here the *homogeneous multiple model*, has been largely adopted for nonlinear modelling and its abilities of dealing with highly nonlinear systems are unquestionable (Murray-Smith and Johansen, 1997; Babuška, 1998; Abonyi and Babuška, 2000; Verdult *et al.*, 2002; Kiriakidis, 2007). That kind of model has also been largely studied to cope with fault detection and estimation (Xu *et al.*, 2012) or fault-tolerant control (Rodrigues *et al.*, 2007; Ichalal *et al.*, 2012). From a structural point of view, the aggregation of submodels is achieved via a time varying weighted sum of the submodel parameters (Gregorčič and Lightbody, 2008). Hence, system complexity inside each operating zone must be quite similar because the submodels have the same dimension and in some cases the number of parameters needed for system description is inevitably increased (overparametrization). This fact is known as the *curse of dimensionality*, where the number of parameters needed for an accurate representation increases extremely rapidly along with nonlinear system complexity (Leith and Leithead, 1999).

In the second multiple model structure, the submodels do not share the same state space and are consequently called *heterogeneous submodels*. This feature offers some degrees of freedom of particular interest to cope with the curse of dimensionality. Indeed, the dimensions of the submodels can be different and they can be adjusted to fit to system complexity inside each operating zone. In this way, the flexibility and generality of this multiple model structure are undoubtedly increased. This kind of multiple model, initially proposed by Filev (1991), is reported in the literature under several designations, such as the local-state local model network (Gawthrop, 1995), the multiple local models (Gatzke and Doyle III, 1999; Venkat *et al.*, 2003; Vinsonneau *et al.*, 2005), local model networks by blending the outputs (Gregorčič and Lightbody, 2000; 2008), multiple model for models with a non-common state (Kanev and Verhaegen, 2006), the neuro-fuzzy decoupling multiple model scheme (Uppal *et al.*, 2006) and the recently decoupled multiple model (Orjuela *et al.*, 2006; 2008; 2009). Despite their different names, these approaches share a similar multiple model structure. In the sequel, this multiple model is called the *heterogeneous multiple model*. The previously quoted works have illustrated successful implementations of this structure for modelling (Venkat *et al.*, 2003; Vinsonneau *et al.*, 2005; Orjuela *et al.*, 2006), control (Gawthrop, 1995; Gatzke and Doyle III, 1999; Gregorčič and Lightbody, 2000) or state estimation and diagnostic (Kanev and Verhaegen, 2006; Uppal *et al.*, 2006; Orjuela *et al.* 2008; 2009) and have shown its relevance. Hence, this kind of multiple model can be used as an interesting alternative to the homogeneous multiple model.

The main contribution of this paper is to provide a supervised off-line identification algorithm for nonlinear systems on the basis of the heterogeneous multiple model structure. The proposed algorithm is based on a specific analysis (sensitivity function computation) of parameter identification of submodels using different optimisation criteria (local, global or combined local/global). Indeed, according to the expected model performances, three cost

criteria to be minimised are presented and the relationship between them is discussed. A particular attention is paid to practical implementation and potential problems encountered in the identification procedure. Indeed, an undesirable identification phenomenon, called the *no output tracking effect*, is clearly revealed and a modification of the multiple model structure is proposed to reduce its impact on the identification quality. These topics seem poorly investigated in the literature related to this kind of multiple model.

The remainder of the paper is organised as follows. In Section 2, the heterogeneous multiple model is presented. The parametric identification problem is stated in Section 3, where three cost criteria are proposed. The identification procedure using these criteria is exposed in Section 4, where the computation of the sensitivity functions is detailed. Section 5 is devoted to the explanation of the *no output tracking effect* and a way to overcome this problem is proposed. Finally, the abilities of this multiple model to cope with nonlinear system identification are assessed in Section 6 by means of different identification examples. Some concluding remarks and directions for future research are presented in Section 7.

2. Heterogeneous multiple model structure

This section is devoted to the description of the heterogeneous multiple model structure. A mathematical formulation of this multiple model is firstly detailed. The weighting function definition employed in this paper is afterwards proposed.

2.1. Heterogeneous multiple model structure. The state space structure of the heterogeneous multiple model used in this work is given by

$$x_i(k+1) = A_i(\theta_i)x_i(k) + B_i(\theta_i)u(k) \quad \text{(1a)}$$
$$+ D_i(\theta_i),$$
$$y_i(k) = C_i(\theta_i)x_i(k), \quad \text{(1b)}$$
$$y(k) = \sum_{i=1}^{L} \mu_i(\xi(k))y_i(k), \quad \text{(1c)}$$

where $x_i \in \mathbb{R}^{n_i}$ and $y_i \in \mathbb{R}^p$ are respectively the state vector and the output of the i-th submodel, $u \in \mathbb{R}^m$ is the input and $y \in \mathbb{R}^p$ is the output of the multiple model. The matrices $A_i \in \mathbb{R}^{n_i \times n_i}$, $B_i \in \mathbb{R}^{n_i \times m}$, $D_i \in \mathbb{R}^{n_i}$ and $C_i \in \mathbb{R}^{p \times n_i}$ of each submodel are constant matrices whose entries θ_i must be determined.

The number L of submodels is intuitively given by the *granularity* of the operating space decomposition needed for an accurate representation of the system under investigation. A complete partition of the operating space of the system into L operating zones is performed

using a characteristic variable $\xi(k)$ of the system called the *decision variable* that is assumed to be known and real-time available. The distance to each operating zone is quantified with the help of the *weighting functions* $\mu_i(\xi(k))$ (cf. Section 2.2). The output $y_i(k)$ of the submodels is unmeasurable internal signals of the model (1); only $u(k)$ and $y(k)$ are available for parameter identification.

It is worth noting from (1) that each submodel evolves independently in its own state space depending on the input control and its initial condition. Hence, the state space dimension n_i of each submodel can be different from one another and it can be adjusted to correctly describe the system behaviour in each operating zone. It can be expected from this feature that some flexibility in the representation of nonlinear systems will be introduced. This multiple model is then suited for black-box modelling of complex systems with a variable structure in the operating space such as biological and chemical systems whose state dimension (e.g., the number of products or species) may vary according to the operating conditions. This feature also offers the possibility to apply model order reduction techniques to the submodels to reduce the global model complexity. It can be noticed that other multiple model structures can be obtained by blending the inputs of the submodels instead of the outputs (Vinsonneau *et al.*, 2005) or by blending both of them.

2.2. Weighting function definition. The weighting functions $\mu_i(\xi(k)) \in [0, 1] : \mathbb{R}^w \to \mathbb{R}^1$ quantify the relative contribution of each submodel to the global model according to the current operating point of the system by means of the decision variable $\xi \in \mathbb{R}^w$. This last variable is a real-time available characteristic variable of the system, e.g., the system input.

Here, the weighting functions are obtained from normalised multidimensional Gaussian basis functions:

$$\omega_i(\xi(k)) = \prod_{j=1}^{w} \exp\left(-\frac{(\xi_j(k) - c_{i,j})^2}{\sigma_j^2}\right), \quad \text{(2a)}$$
$$\mu_i(\xi(k)) = \frac{\omega_i(\xi(k))}{\sum_{j=1}^{L} \omega_j(\xi(k))}, \quad \text{(2b)}$$

where ξ_j is the j-th component of the decision variable vector ξ, the two parameters $c_{i,j}$ and σ_j are respectively employed to place the centre and to choose the overlapping of weighting functions in the operating space. It can be noted that other mathematical definitions can be used. In each case, the weighting functions $\mu_i(\xi(k))$ must satisfy the following convex sum properties:

$$\sum_{j=1}^{L} \mu_j(\xi(k)) = 1, \qquad \forall k, \quad \text{(3a)}$$
$$0 \leq \mu_j(\xi(k)) \leq 1, \qquad \forall j = 1, \dots, L, \quad \forall k. \quad \text{(3b)}$$

According to the operating space decomposition, the weighting functions (2) can be qualified as *strongly overlapped* for smooth decomposition or *weakly overlapped* for hard decomposition, as depicted in Figs. 1 and 2. The relationship between the functions $\omega_i(\xi(k))$ and $\mu_i(\xi(k))$ is also shown in this figure. The overlapping between these functions can be viewed as the overlapping between the validity regions of the submodels. As will be pointed out in Section 3.3, these notions play an important role in the potential interpretation given to the identified submodels. Throughout this paper and for the sake of simplicity, the weighting functions $\mu_i(\xi(k))$ are written as $\mu_i(k)$.
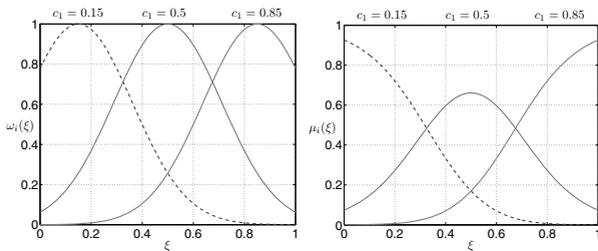

Fig. 1. Strongly overlapped weighting functions: $\omega_i(\xi)$ (left) and $\mu_i(\xi)$ (right) for $\sigma_i = 0.3$.
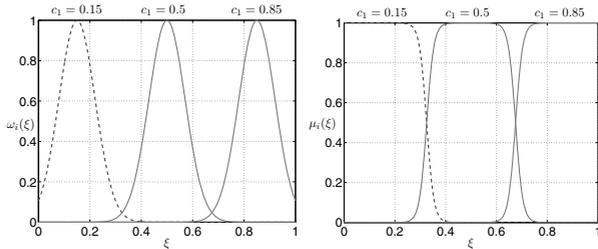

Fig. 2. Weakly overlapped weighting functions: $\omega_i(\xi)$ (left) and $\mu_i(\xi)$ (right) for $\sigma_i = 0.1$.

3. Parametric identification problem

The multiple model parametric identification generally deals with three main issues: (i) decision variable $\xi(k)$ selection, (ii) the choice of the weighting functions $\mu_i(k)$ for operating space decomposition and (iii) submodel parameter estimation for a given multiple model structure. Model validation constitutes the final stage of the identification procedure, for example, with the help of appropriate correlation tests (Billings and Zhu, 1994). Unfortunately, these issues are not straightforward and their simultaneous resolution leads to a very hard nonlinear problem. Hence, a systematic procedure for multiple model identification is not available despite the great effort made in this direction.

Here, the three mentioned identification problems are tackled as follows. The chosen decision variable $\xi(k)$ is an entry or the complete input vector $u(k)$ of the system. Indeed, the evolution of the current operating point of the system is *a priori* driven by the input system signal. This choice provides a solution to the first problem without loss of generality. Concerning the second problem, it is assumed that the operating space decomposition is *a priori* known (supervised case). Several ways can be employed to obtain this decomposition. The use of the static regime characteristic of the system is, for example, a simple way to perform this partition. In fact, this characteristic provides very helpful information about the local linear behaviours as well as the operating points (Venkat *et al.*, 2003). Clustering algorithms (Gustafson–Kessel algorithm, fuzzy c-means, fuzzy c-varieties, etc.) are also very useful to deal with the operating space decomposition in an efficient way (Babuška, 1998; Dumitrescu *et al.*, 2000). The last, but not the least, quoted problem is investigated in the sequel.

The basis of heterogeneous multiple model identification by minimising the error between the system and the model outputs is exposed by Gray *et al.* (1996). In the work of McLoone and Irwin (2003) the identification is achieved using a heterogeneous structure with a particular class of submodels known as *velocity-based* local models. The use of these submodels provides transparency of the overall network because the submodels are always affine. In the work of Venkat *et al.* (2003) the input/output data for submodel identification are generated in one particular small operating zone of the nonlinear system; the linearity of the submodel associated with each set of data is ensured in this way. Hence, submodel identification can be independently accomplished via standard linear identification techniques. However, a large number of experiments are necessary to obtain independent data sequences for submodel identification. Vinsonneau *et al.* (2005) perform the identification of heterogeneous multiple model with blended inputs with a global total least square technique. In this section, multiple model identification mainly deals with submodel parameter estimation θ_i in (1). Different optimisation criteria to accomplish the identification task are proposed and compared in the following sections.

3.1. Submodel parametric identification problem.
The set of parameters to be identified can be gathered in a partitioned vector θ as follows:

$$\theta = [\theta_1 \ldots \theta_i \ldots \theta_L]^T \in \mathbb{R}^q, \tag{4}$$

where each block θ_i is formed by the unknown parameters of the i-th submodel:

$$\theta_i = [\theta_{i,1} \ldots \theta_{i,q} \ldots \theta_{i,q_i}]^T \in \mathbb{R}^{q_i}, \tag{5}$$

where $\theta_{i,k}\,(k = 1, \ldots, q_i)$ is a scalar parameter to be estimated, q_i the number of parameters of the i-th submodel and $q = \sum_{i=1}^{L} q_i$ denotes the total number of parameters of the multiple model.

The statement of the identification problem, addressed in this paper, can be formulated as follows: *For given weighting functions and for a given set of input/output data of a MISO system, the parameters of the L submodels must be determined.* The relative accuracy of the identified multiple model is evaluated and quantified via a cost function $J(\theta)$ to be minimised, in other words

$$\hat{\theta} = \arg\min_{\theta} J(\theta). \tag{6}$$

The optimisation criterion $J(\theta)$ generally quantifies the output model quality with respect to the system output. According to the choice of this criterion and thanks to the flexibility of multiple models, some modelling specifications can be introduced (for example, a global representation of the systems and/or its local representation) in the identification procedure, as proposed in the next section.

3.2. Optimisation criteria. Three optimisation criteria (global, local and combined) to be minimised can be employed in parameter identification of multiple models (Yen *et al.*, 1998; Abonyi and Babuška, 2000; Johansen and Babuška, 2003). Obviously, the choice of a criterion is related to the expected performances of the model as well as its future applications.

Global criterion. The global criterion is defined by

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

$$= \frac{1}{2} \sum_{k=1}^{N} (y(k) - y_S(k))^2, \tag{7}$$

where N is the number of training data and $\varepsilon(k)$ is the *global error* between $y(k)$ the output of the identified multiple model (1c) while $y_S(k)$ stands for the system output given by the data. This criterion encourages the global approximation between the nonlinear system and the multiple model behaviour across the operating space. This criterion is interesting when the multiple model is used for *predictions* without any interpretation of the submodels. Indeed, the local behaviours of submodels inside each operating zone are not considered by (7).

Local criterion. The following weighting local criterion is

firstly introduced for each submodel:

$$J_{L,i}(\theta_i) = \frac{1}{2} \sum_{k=1}^{N} \mu_i(k) \varepsilon_i^2(k),$$

$$= \frac{1}{2} \sum_{k=1}^{N} \mu_i(k)(y_i(k) - y_S(k))^2, \tag{8}$$

where $\varepsilon_i(k)$ is the *local error* between the i-th submodel output $y_i(k)$ given by (1b) and the system output $y_S(k)$. The classic least-squares algorithm can be applied for each submodel identification when a hard operating space decomposition is considered, i.e., when $\mu_i(k)$ are not overlapped. In this particular case, the multiple model output $y(k)$ is given by the submodel output $y_i(k)$ according to the operating point. However, the outputs of the submodels are simultaneously available and a more general local criterion can be defined by considering the contribution of all submodels:

$$J_L(\theta) = \frac{1}{2} \sum_{i=1}^{L} J_{L,i}(\theta_i). \tag{9}$$

This criterion takes into account the local approximation provided by the submodels inside each operating zone and some interpretation can eventually be accorded, e.g., a local linearisation of the nonlinear system around operating points. Hence, each submodel is identified by only considering data inside its associated validity region according to the weighting term $\mu_i(k)$ in (8). In comparison with the global criterion (7), a large number of submodels can be necessary to provide a good global characterisation of the system.

Combined criterion. A trade-off of global and local criteria can be obtained by the combined criterion defined by (Yen *et al.*, 1998)

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

The use of this criterion makes the so-called multi-objective optimisation possible in which the two above criteria are more or less taken into account according to the weight given by the scalar $\alpha \in [0, \ 1]$. The influence of α on the trade-off between local model interpretability and the accurate global approximation was investigated by Johansen and Babuška (2003). Some modelling conflicts between the local and the global objectives are also pointed out in that paper and the way to detect and solve these conflicts is analysed.

3.3. Link between the criteria and the operating space decomposition. The weighting functions are used for zoning the operating space of the system. As previously mentioned, the operating space decomposition can be performed in a first step from *a priori*

knowledge of the system. The expected results of the identification procedure according to criterion selection are consequently conditioned by this partition.

Notice, in particular, that even if a local criterion is used, the local behaviour interpretation of the multiple model cannot be accomplished when the operating zones are strongly overlapped. Indeed, some submodels have a large validity zone and consequently a local interpretation inside a very delimited validity region is not well adequate. In opposition, a global interpretation can always be made because the multiple model is built for accomplishing this objective. The local and global features of the submodels are conditioned by the appropriate choice of the weighting functions, i.e., by the operating space decomposition.

4. Submodel parametric identification

An analytical solution of the discussed identification problem given by (6) is not available because the optimisation issue remains nonlinear with respect to the submodel parameters. Hence, the proposed parametric estimation of the unknown parameter vector θ is based on an iterative minimisation procedure of a criterion J (global, local or combined) according to the Gauss–Newton algorithm associated with the Levenberg and Marquardt method given by the following recurrence relation (Walter and Pronzato, 1997; Ljung, 1999; Nelles, 2001):

$$\theta^{(m+1)} = \theta^{(m)} \qquad (11)$$
$$- \Delta^{(m)} (\mathrm{H}(\theta^{(m)}) + \lambda^{(m)} I)^{-1} \mathrm{G}(\theta^{(m)}),$$

where $\theta^{(m)}$ is the estimated parameter vector at the m-th iteration, $\theta^{(m+1)}$ is the evaluated vector at the next iteration, $\mathrm{H}(\theta^{(m)}) = \frac{\partial^2 J}{\partial \theta \partial \theta^T}\big|_{\theta = \theta^{(m)}}$ is the Hessian matrix and $\mathrm{G}(\theta^{(m)}) = \frac{\partial J}{\partial \theta}\big|_{\theta = \theta^{(m)}}$ the gradient vector at the current iteration. In the sequel we shall simply write $\mathrm{H}(\theta) = \mathrm{H}(\theta^{(m)})$ and $\mathrm{G}(\theta) = \mathrm{G}(\theta^{(m)})$. The computation of $\mathrm{G}(\theta)$ and $\mathrm{H}(\theta)$ is based on the calculation of *sensitivity functions* of the multiple model output with respect to the submodel parameters. The step size $\Delta^{(m)}$ and the regularisation parameter $\lambda^{(m)}$ enhance the velocity and convergence avoiding some numerical problems, for example, a bad numerical conditioning of the Hessian matrix (Walter and Pronzato, 1997; Ljung, 1999).

4.1. Computation of $\mathrm{G_G}(\theta)$ and $\mathrm{H_G}(\theta)$ with a global criterion. The global gradient $\mathrm{G_G}(\theta)$ is calculated by differentiating the global criterion (7) with respect to the parameter vector θ:

$$\mathrm{G_G}(\theta) = \frac{\partial J_G(\theta)}{\partial \theta},$$
$$= \sum_{k=1}^{N} \varepsilon(k) \frac{\partial y(k)}{\partial \theta}. \qquad (12)$$

By using (1c), it follows that

$$\frac{\partial y(k)}{\partial \theta} = \sum_{i=1}^{L} \mu_i(k) \frac{\partial y_i(k)}{\partial \theta}. \qquad (13)$$

Here $\partial y_i(k)/\partial \theta$ are the first order sensitivity functions of the i-th submodel output with respect to the unknown parameters of the multiple model.

The Hessian matrix $\mathrm{H_G}(\theta)$ is obtained by differentiating the gradient vector (12) with respect to θ as follows:

$$\mathrm{H_G}(\theta) = \frac{\partial^2 J_G(\theta)}{\partial \theta \partial \theta^T},$$
$$= \sum_{k=1}^{N} \epsilon(k) \frac{\partial^2 y(k)}{\partial \theta \partial \theta^T} + \sum_{k=1}^{N} \frac{\partial y(k)}{\partial \theta} \frac{\partial y(k)}{\partial \theta^T}. \qquad (14)$$

Remark that the computation of the Hessian matrix (14) becomes very arduous due to the required computations of both first and second order sensitivity functions. However, the second order sensitivity functions can be neglected thanks to the Gauss–Newton method (Walter and Pronzato, 1997; Ljung, 1999; Nelles, 2001). In this case, the Hessian is computed by only considering the first order sensitivity functions already employed in the gradient computation. Finally, the approximate Hessian $\tilde{\mathrm{H}}_G(\theta)$ is given by

$$\tilde{\mathrm{H}}_G(\theta) = \sum_{k=1}^{N} \frac{\partial y(k)}{\partial \theta} \frac{\partial y(k)}{\partial \theta^T}. \qquad (15)$$

4.2. Computation of $\mathrm{G_L}(\theta)$ and $\mathrm{H_L}(\theta)$ with a local criterion. The gradient vector $\mathrm{G_L}(\theta)$ and the Hessian matrix $\mathrm{H_L}(\theta)$ computations considering the local criterion (9) are similar to the previous case. The local gradient $\mathrm{G_L}(\theta)$ is given by

$$\mathrm{G_L}(\theta) = \frac{\partial J_L(\theta)}{\partial \theta},$$
$$= \sum_{i=1}^{L} \sum_{k=1}^{N} \mu_i(k) \varepsilon_i(k) \frac{\partial y_i(k)}{\partial \theta}, \qquad (16)$$

and the approximation of its associated Hessian matrix $\tilde{\mathrm{H}}_L(\theta)$ by

$$\tilde{\mathrm{H}}_L(\theta) = \sum_{i=1}^{L} \sum_{k=1}^{N} \mu_i(k) \frac{\partial y_i(k)}{\partial \theta} \frac{\partial y_i(k)}{\partial \theta^T}. \qquad (17)$$

4.3. Computation of the sensitivity functions. Using (1b), the sensitivity functions are defined as follows:

$$\frac{\partial y_i(k)}{\partial \theta_{p,q}} = \frac{\partial C_i}{\partial \theta_{p,q}} x_i(k) + C_i \frac{\partial x_i(k)}{\partial \theta_{p,q}},$$
$$i = 1, \ldots, L, \quad p = 1, \ldots, L, \quad q = 1, \ldots, q_i, \quad (18)$$

where $\theta_{p,q}$ is a scalar parameter to be estimated, already introduced in Section 3.1. The sensitivity of the local state x_i with respect to the parameters at the time $k + 1$ is performed by the derivation of (1a) with respect to each parameter $\theta_{p,q}$ to be identified:

$$\frac{\partial x_i(k+1)}{\partial \theta_{p,q}} = \frac{\partial A_i}{\partial \theta_{p,q}} x_i(k) + A_i \frac{\partial x_i(k)}{\partial \theta_{p,q}}$$
$$+ \frac{\partial B_i}{\partial \theta_{p,q}} u(k) + \frac{\partial D_i}{\partial \theta_{p,q}}, \quad (19)$$
$$i = 1, \ldots, L.$$

A very interesting feature in heterogeneous multiple model identification is that the same sensitivity functions $\partial y_i(k)/\partial \theta$ are used for parametric estimation with a global or a local criterion reducing in this way the computation effort.

4.4. Parameter estimation algorithm. The identification procedure described so far can be summarised by the following identification algorithm.

Some practical implementation aspects of this algorithm are discussed in the sequel. It is worth noting that the sensitivity functions (18) and (19) are generic forms. They can be simplified in algorithm implementation because the parameters of each submodel are completely independent of the parameters of the other submodels. Consequently, the number of sensitivity functions to be computed can be considerably reduced as follows:

$$\frac{\partial y_i(k)}{\partial \theta_p} = 0 \quad \text{for} \quad p \neq i, \quad (20)$$
$$i = 1, \ldots, L, \quad p = 1, \ldots, L.$$

The practical implementation complexity of the proposed identification algorithm is largely reduced thanks to these simplifications.

Remark that two classic problems can be encountered when using iterative nonlinear algorithms (Ljung, 1999; Nelles, 2001; Walter and Pronzato, 1997). The first one concerns the *sufficiently rich* input signal. In the multiple model framework the entire operating zones of the system must be excited by the input on the one hand, while the decision variable $\xi(k)$ must be selected to trigger all submodels on the other. Let us remember that here the decision variable is the input of the system and therefore the identification can be well

accomplished by sufficient rich input. The second classic problem is related to the appropriate choice of an initial parameter guess θ_0 to ensure algorithm convergence towards a global optimum solution or to the best local optimum. Indeed, several local optimal solutions are often available for nonlinear optimisation problems. In this particular case, several choices of initial parameters may be necessary to reach the best local optimum solution ensuring in this way the best identitication of the nonlinear system under investigation. An interesting procedure to judiciously choose the initial parameters is to perform a first identification considering a hard operating space decomposition by means of the classic least-square algorithm using the local criterion (8). The submodel parameters obtained in this way can be employed as a starting point in the identification algorithm.

Algorithm 1. Heterogeneous multiple model identification algorithm.

Require: Weighting function definition
Require: Initial parameter vector $\theta^{(0)}$, initial state vector $x(0)$, initial state sensitivities $\partial x_i(0)/\partial \theta_{p,q}$
Require: Maximal number of iterations $m_{\max}$
Require: Maximal value of the gradient norm $\epsilon > 0$
Require: $0 \leq \alpha \leq 1$ in the criterion $J_C(\theta)$ given by (10)
Require: $m = 1$, $\Delta^{(1)} \approx 0$ and $\lambda^{(1)} \ll 0$ in (11)
  1: **repeat**
  2:     Multiple model simulation from evaluation of (1)
  3:     Sensitivity function computation from (18) and (19)
  4:     Gradient vector G computation from (12) and (16)
  5:     Hessian matrix H computation from (15) and (17)
  6:     Update the vector $\theta^{(m)}$ from (11)
  7:     Evaluate $J_C^{(m)}$
  8:     **if** $J_C^{(m-1)} < J_C^{(m)}$ **then**
  9:        Go to Stage 6, increase $\lambda^{(m)}$ and decrease $\Delta^{(m)}$
 10:     **else**
 11:        $m \Leftarrow m + 1$
 12:        Increase $\Delta^{(m)}$ and decrease $\lambda^{(m)}$
 13:     **end if**
 14: **until** $\|G\| < \epsilon$ or $m_{\max} \leq m$

4.5. Link between local and global approaches. The aim of this section is to investigate the similarities of the identification approaches using either a local or a global approach. Some algebraic manipulations are needed to highlight the similarities between the gradient vectors $G_G(\theta)$ and $G_L(\theta)$. Firstly, the global error $\varepsilon(k)$ can be rewritten as follows:

$$\varepsilon(k) = \sum_{j=1}^{L} \mu_j(k) \left(y_j(k) - y_S(k) \right), \quad (21)$$

$$= \sum_{j=1}^{L} \mu_j(k)\varepsilon_j(k),$$

using the convex properties (3) of the weighting functions. Secondly, the expression of $\varepsilon(k)$ given by (21) can be replaced in Eqn. (12) of $\mathrm{G_G}(\theta)$ as follows:

$$\mathrm{G_G}(\theta) = \sum_{k=1}^{N}\sum_{i=1}^{L}\sum_{j=1}^{L} \mu_i(k)\mu_j(k)\varepsilon_j(k)\frac{\partial y_i(k)}{\partial\theta}. \quad (22)$$

The main difference between the gradient vectors (12) and (16) only comes from the discussed estimation error. A global error $\varepsilon(k) = y(k) - y_S(k)$ is considered in the first case whereas a local error $\varepsilon_i(k) = y_i(k) - y_S(k)$ is used in the second one. However, thanks to algebraic manipulations the same local error as well as the same sensitivity functions $\partial y_i(k)/\partial\theta$ appear in the global gradient vector (22) and in the local one (16).

Now, the limit case where the weighting functions are not overlapped (i.e., very hard operating space decomposition) is considered for the sake of clarity. In this limit case, these two gradient vectors are identical because

$$\begin{cases} \mu_i(k)\mu_j(k) = 1 & \text{if } i = j, \\ \mu_i(k)\mu_j(k) = 0 & \text{otherwise.} \end{cases} \quad (23)$$

Hence, it can be expected that $\mathrm{G_L}(\theta) \approx \mathrm{G_G}(\theta)$ when weakly overlapped weighting functions are employed. The same comparison between Hessian matrices $\mathrm{H_G}(\theta)$ and $\mathrm{H_L}(\theta)$, respectively given by (15) and (17), allows us to show that $\mathrm{H_L}(\theta) \approx \mathrm{H_G}(\theta)$.

In conclusion, it can be expected that both optimisation criteria provide similar results when weakly overlapped weighting functions are considered because the gradient vectors and the Hessian matrices are similar. In fact, the optimisation directions are also very similar in these two cases. In other words, local properties of the submodels can be obtained using a global criterion when the operating zones are weakly overlapped (cf. Example 1), but global properties of the submodels cannot be expected to be found when weighting functions are weakly overlapped. For strongly overlapped weighting functions, it is advisable to choose the combined criterion (10) to weigh the submodels interpretation with respect to the quality of the global model.

5. On the no output tracking effect

In some particular modelling situations, undesirable discontinuities arise in the multiple model output (cf. Example 2). Indeed, an abrupt change in the decision variable $\xi(k)$ may cause an abrupt jump from one output submodel directly to another output, introducing a discontinuity in the multiple model output (Gatzke and Doyle III, 1999). Let us remember that the multiple model output is given by the weighted sum of the submodel outputs which are not necessarily close in the operating space. Therefore, their distance may produce a discontinuity in the multiple model output because the submodel outputs do not respond instantly. This phenomenon, called here the *no output tracking effect*, happens when the outputs of the submodels taken into consideration are far apart. The impact of this phenomenon decreases when the outputs are close and it completely vanishes if all the outputs are identical at the transition time.

This undesirable phenomenon unavoidably reduces the quality of the identified multiple model. It acts in the identification procedure as an internal perturbation strongly distorting the submodel identification results. The chosen decision variable, the operating space decomposition and the dynamics of the submodels play a more or less important role in the *no output tracking effect*. It is important to notice that this phenomenon is not systematic and only appears in particular cases.

Considering strongly overlapped weighting functions and a large number of operating zones is not always an efficient way to avoid this phenomenon. These constraints undoubtedly increase the complexity of the multiple model and reduce some degrees of freedom in the modelling stage, e.g., local modelling with strongly overlapped weighting functions cannot be well achieved (see Section 3.3). Furthermore, the previous solution does not provide a systematic reaction to the *no output tracking effect*, as shown in Example 2. Here, a modified heterogeneous multiple model, depicted in Fig. 3, is proposed to deal with this problem.
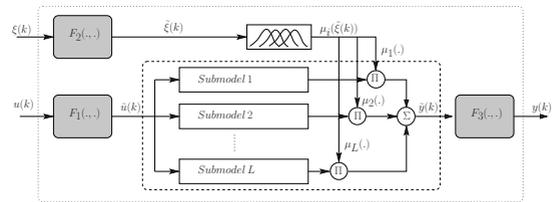

Fig. 3. Modified heterogeneous multiple model structure.

The main idea is to find a way to progressively take into consideration the contribution of each submodel. Three filters (with transfer functions F_1, F_2 and F_3) are introduced in the multiple model structure. These filters respectively act on the input, the decision variable $\xi(k)$ and the output of the multiple model. The proposed *modified heterogeneous multiple model* is given by

$$\tilde{u}(k) = F_1(z^{-1}, \theta_{F_1})u(k), \quad (24a)$$

$$\tilde{\xi}(k) = F_2(z^{-1}, \theta_{F_2})\xi(k), \quad (24b)$$

$$x_i(k+1) = A_i(\theta_i)x_i(k) + B_i(\theta_i)\tilde{u}(k) \quad (24c)$$
$$+ D_i(\theta_i),$$

$$y_i(k) = C_i(\theta_i)x_i(k), \tag{24d}$$

$$\tilde{y}(k) = \sum_{i=1}^{L} \mu_i(\tilde{\xi}(k))y_i(k), \tag{24e}$$

$$y(k) = F_3(z^{-1}, \theta_{F_3})\tilde{y}(k), \tag{24f}$$

where $y(k)$ is the modified multiple model output, z^{-1} is the one step delay operator, $F_1(z^{-1}, \theta_{F_1})$, $F_2(z^{-1}, \theta_{F_2})$ and $F_3(z^{-1}, \theta_{F_3})$ are the three additional filters and θ_{F_i} is the parameter vector of the i-th filter to be identified.

These three filters, often first or second order low-pass ones, are integrated in the multiple model structure and their parameters should be estimated in the identification procedure. The new augmented unknown parameter vector θ is then given by

$$\theta = \begin{bmatrix} \theta_1^T & \dots & \theta_i^T & \dots & \theta_L^T & \theta_{F_1}^T & \theta_{F_2}^T & \theta_{F_3}^T \end{bmatrix}^T, \tag{25}$$

where each block θ_i has already been defined in (5) and the blocks θ_{F_1}, θ_{F_2} and θ_{F_3} are formed by the parameters of the three additional filters to be identified. The identification procedure is achieved as proposed in Section 4, including the additional sensitivity functions of the multiple model output with respect to the filter parameters.

It can be noticed that a systematic guideline for selection of these three filters cannot be provided in a straightforward way. Indeed, a systematic use of the additional filters is not always needed to overcome the *no output tracking effect*, as shown in Section 6. These filters must be considered as supplementary degrees of freedom to enhance the identification task with few parameters when the *no output tracking effect* is detected. In many cases, the *no output tracking effect* can be avoided by only using the filter F_2 (i.e., using filtered decision variables $\tilde{\xi}(k)$) because abrupt jumps of the weighting functions are avoided.

6. Identification examples

In this section, the developed method is applied to three nonlinear system identification problems allowing to assess the effectiveness of the method as well as its limits. The performance of the models is assessed using the *Mean Square Error* (MSE) and the *Variance-Accounted-For* (VAF) indicators:

$$\text{MSE} = \frac{1}{N} \sum_{k=1}^{N} (y_S(k) - y(k))^2, \tag{26}$$

$$\text{VAF} = \max \left\{ 1 - \frac{\text{var}(y_S(k) - y(k))}{\text{var}(y_S(k))}, 0 \right\} \times 100, \tag{27}$$

where $y_S(k)$ and $y(k)$ are respectively the system and the model output, and $\text{var}(\cdot)$ denotes the variance of a signal (Verdult *et al.*, 2002). These indicators are computed using the validation data in the proposed examples.

Example 1. (*Identification case*) The following nonlinear system is considered:

$$y(k+1) = (0.6 - 0.1a(k))y(k) + a(k)u(k), \tag{28a}$$

$$a(k) = \frac{0.6 - 0.06y(k)}{1 + 0.2y(k)}. \tag{28b}$$

The operating space is decomposed into $L = 2$ operating zones according to the static characteristic of (28). The decision variable is the system input $\xi(k) = u(k)$. The parameters of the weighting functions (2) are $c_1 = -0.9$, $c_2 = 0.9$ and $\sigma_i = 0.9$. This choice ensures the blending of submodels but is not unique and the parameters of the weighting functions could be optimised. The structure of the submodels is

$$x_i(k+1) = a_i x_i(k) + b_i u(k), \tag{29}$$

where a_i and b_i are scalar parameters to be identified considering a global criterion (7). The identification is carried out using a data set of 750 samples generated by a smooth input signal with random magnitude $u(k) \in [-0.9, 0.9]$. A second data set of 750 samples is employed for the validation purpose.

The identification results are displayed in Fig. 4. In this figure, the nonlinear system is denoted by NLS and the identified multiple model by MM. The upper part of this figure shows a comparison between the nonlinear system and the identified submodels. Clearly, each submodel provides a local representation of the system behaviour. In fact, the first submodel provides a good approximation for negative input values. In contrast, the second submodel provides good approximation for positive input values. The global modelling of the nonlinear system (bottom of the figure) is finally obtained by appropriately considering the contribution of each submodel via middle overlapped weighting functions. The proposed multiple model represents local as well as global behaviour of the nonlinear system and the performance indicators are $MSE = 0.0012$ and $VAF = 99.72\%$. ♦

Example 2. (*On the no output tracking effect*) The goal of this identification example is to show the *no output tracking effect*. The following system is considered:

$$x(k+1) = ax(k) + \sin(\gamma u(k))(\beta - u(k)), \tag{30}$$

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

The operating space is intentionally decomposed in $L = 6$ operating zones using the system input as the decision variable, i.e., $\xi(k) = u(k)$. The parameters of the weighting functions (2) are $c_1 = 0$, $c_2 = 0.2$, $c_3 = 0.4$, $c_4 = 0.6$, $c_5 = 0.8$, $c_6 = 1$ and $\sigma_i = 0.2$. The submodel structure is given by
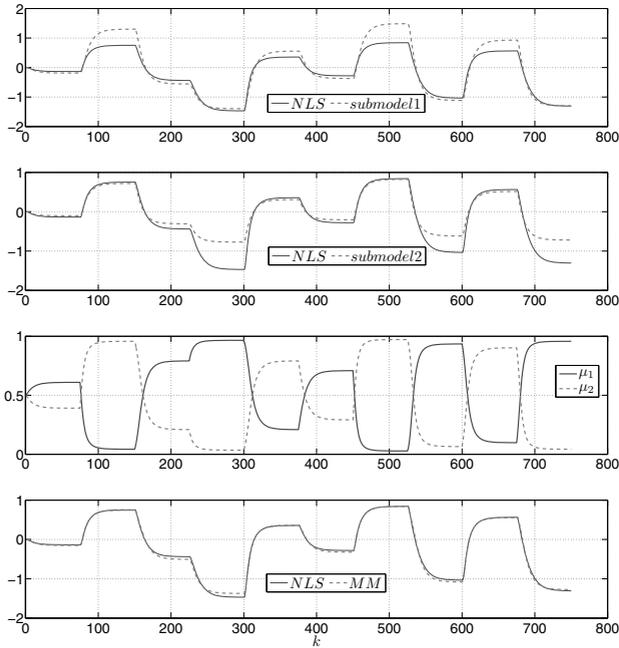
$$x_i(k+1) = a_i x_i(k) + b_i u(k) + d_i, \tag{31}$$

Fig. 4. System identification results using validation data.

where a_i, b_i and d_i are scalars to be identified. A global criterion (7) can be used for the identification task because the weighting functions are strongly overlapped. The identification is carried out using a data set of 5000 samples generated by a piecewise input signal with random duration and magnitude $u(k) \in [0, 1]$; a second data set of 5000 samples is employed for the validation purpose.

The comparison between the nonlinear system and the multiple model behaviours is depicted in Fig. 5. Clearly, the system behaviour is not well represented by the identified multiple model. The performance indicators are $MSE = 9.66$ and $VAF = 52.22\%$. Indeed, the system output is not well tracked by the multiple model output, in particular due to "picks" causing a loss of quality. These "picks" are undesirable and result in the *no output tracking effect* previously presented in Section 5. One can see in Fig. 5 that the *no output tracking effect* remains important despite the large number of submodels and the strongly overlapped weighting functions voluntarily employed in this example.

The modified heterogeneous multiple model (24) proposed in Section 5 is now employed to improve the identification, i.e., to attenuate the *no output tracking effect*. Here, the structures of the filters F_1, F_2 and F_3 are given by

$$\tilde{u}(k+1) = \alpha_1 \tilde{u}(k) + (1 - \alpha_1)u(k+1), \quad (32a)$$

$$\tilde{\xi}(k+1) = \alpha_2 \tilde{\xi}(k) + (1 - \alpha_2)\xi(k+1), \quad (32b)$$

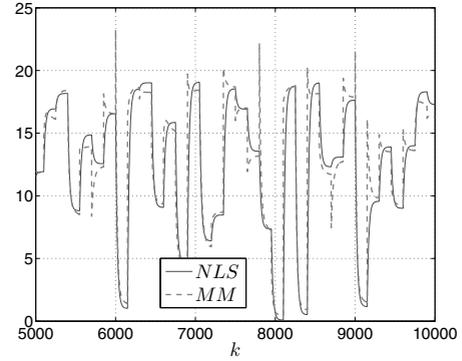$$y(k+1) = \alpha_3 y(k) + (1 - \alpha_3)\tilde{y}(k+1), \quad (32c)$$

Fig. 5. No output tracking effect: system and multiple model output validation.

where α_1, α_2 and α_3 are supplementary parameters to be estimated based on the ideas proposed in Section 5. As seen in Fig. 6, the modified heterogeneous multiple model provides an accurate representation of the system dynamics. The performance indicators are largely improved: $MSE = 0.0029$ and $VAF = 99.98\%$. The proposed strategy is an effective way to eliminate the no output tracking effect and to consequently recover the approximation abilities of the multiple model.
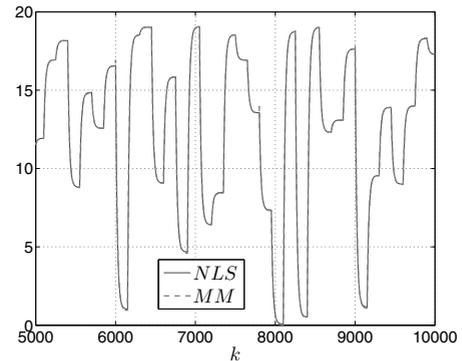

Fig. 6. No output tracking effect: system and modified multiple model output validation.

♦

Example 3. (*Nonlinear system identification*) The following system, firstly proposed by Narendra and Parthasarathy (1990) for neural networks identification, is considered a benchmark for black-box identification technique evaluation (Verdult *et al.*, 2002; Nie, 1994; Boukhris *et al.*, 1999; Wen *et al.*, 2007):

$$y(k+1)$$
$$= \frac{u(k)}{1 + y^2(k-1) + y^2(k-2)}$$
$$+ \frac{y(k)y(k-1)y(k-2)u(k-1)\,(y(k-2) - 1)}{1 + y^2(k-1) + y^2(k-2)}.$$
$$(33)$$

The identification is accomplished by considering a global criterion and an operating space uniformly partitioned in $L = 4$ operating zones. The parameters of the weighting functions (2) are $c_1 = -1$, $c_2 = -0.33$, $c_3 = 0.33$, $c_4 = 1$ and $\sigma_i = 0.4$. The identification task is accomplished using a partially modified multiple model. Indeed, the decision variable $u(k)$ is filtered by the low-pass filter (32b), and consequently only the filter F_2 is employed. The modified multiple model is composed of four second-order submodels.

In this benchmark, the identification is carried out using a data set of 800 samples generated by a piecewise input signal with random duration and magnitude $u(k) \in [-1, 1]$. The model validation is assessed using a second data set of 800 samples generated by an input signal given by

$$u(k) = \begin{cases} \sin\left(\dfrac{2\pi}{250}k\right) & \text{if } k \leq 500, \\ 0.8\sin\left(\dfrac{2\pi}{250}k\right) + 0.2\sin\left(\dfrac{2\pi}{25}k\right) & \text{if } k > 500. \end{cases}$$

The identification is carried out by successively considering the noise-free output case and the noisy output case. In the latter, a normally distributed random signal with zero mean and unit standard deviation is employed to disturb the system output. The obtained identification results in these two cases are respectively shown in Figs. 7 and 8.
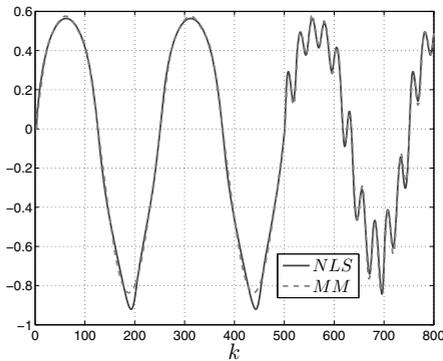

Fig. 7. Model validation for identification in the noise-free case.

Table 1 shows a comparison between our results and those found in the literature using other models. Remark that the two performance indicators MSE and VAF are not always computed in these works and the noise case is not considered. It can be pointed out that the identified multiple model yields performances comparable to those obtained with the other methods with few parameters (four second-order submodels). The abilities of this multiple model to cope with nonlinear system identification are clearly shown.
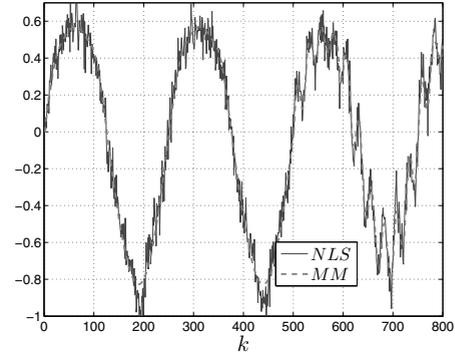
♦

Fig. 8. Model validation for identification in the noise case.

Table 1. Performance indicator comparisons (TS MM: Takagi–Sugeno multiple model (homogeneous model), BP-WARX: piecewise-affine basis function autoregressive eXogenous models).

Model name and reference	MSE	VAF
Neuro-fuzzy network with 34 rules (Nie, 1994)	0.00028	– –
TS MM, with seven 3rd order models (Boukhris *et al.*, 1999)	0.0003	– –
TS MM, with four 3rd order models (Verdult *et al.*, 2002)	0.0002	99.9%
BPWARX, with 10 BPWA functions (Wen *et al.*, 2007)	0.112	97.9%
Proposed method (noise-free case)	0.00067	99.7%
Proposed method (noise case, $SNR = 17$ dB)	0.0053	97.7%

7. Conclusion and further directions

In this work, the abilities of the heterogeneous multiple model to deal with the identification of nonlinear systems are investigated. With respect to classic multiple model structures, the heterogeneous multiple model enables to use heterogeneous submodels with different dimensions according to the complexity of the system inside each operating region. The main advantage of this model is the flexibility degree in the operating space decomposition of the system.

An identification algorithm was presented and three cost criteria (local, global or combined global/local) were provided according to the expected model performance. The relationship between these criteria was highlighted. An effective solution to avoid the undesirable *no output tracking effect* was also proposed and the results obtained showed the effectiveness of the proposed identification algorithm.

The suggested identification algorithm can be extended in order to optimise the operating space decomposition. Indeed, the weighting functions associated to each operating region are supposed to be

known in this work, even if they can also be optimised as the submodel parameters using the same optimisation procedure. In this way, the approximation abilities of the multiple model can be enhanced but the complexity of the optimisation problem increases because the total number of the parameters to be identified also increases. A solution to this problem can, however, be found via a *two level iterative algorithm*. In the first level, the proposed parameter identification can be used assuming that the weighting functions are known. Weighting function identification is performed in the second level by assuming that the submodels are known. Finally, the perspectives of this study are the extension of the proposed identification procedure to include the optimisation of the operating space decomposition according to the multiple model dimension, i.e., submodel orders and their number.

References

Abonyi, J. and Babuška, R. (2000). Local and global identification and interpretation of parameters in Takagi–Sugeno fuzzy models, *9th IEEE International Conference on Fuzzy Systems, FUZZ-IEEE, San Antonio, CA, USA*, pp. 835–840.

Babuška, R. (1998). *Fuzzy Modeling for Control*, Kluwer Academic Publishers, London.

Billings, S.A. and Zhu, Q.M. (1994). Nonlinear model validation using correlation test, *International Journal of Control* **60**(6): 1107–1120.

Boukhris, A., Mourot, G. and Ragot, J. (1999). Non-linear dynamic system identification: A multiple-model approach, *International Journal of Control* **72**(7/8): 591–604.

Dumitrescu, D., Lazzerini, B. and Jain, L.C. (2000). *Fuzzy Sets and Their Application to Clustering and Training*, CRC Press Taylor & Francis, Boca Raton, FL.

Edwards, D. and Hamson, M. (2001). *Guide to Mathematical Modelling*, 2nd Edn., Basingstoke, Palgrave, Chapter 1, p. 3.

Filev, D. (1991). Fuzzy modeling of complex systems, *International Journal of Approximate Reasoning* **5**(3): 281–290.

Gatzke, E.P. and Doyle III, F.J. (1999). Multiple model approach for CSTR control, *14th IFAC World Congress, Beijing, China*, pp. 343–348.

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

Gray, G.J., Murray-Smith, D.J., Li, Y. and Sharman, K.C. (1996). Nonlinear system modelling using output error estimation of a local model network, *Technical Report CSC-96005*, Centre for Systems and Control, Glasgow University, Glasgow.

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

Gregorčič, G. and Lightbody, G. (2008). Nonlinear system identification: From multiple-model networks to Gaussian processes, *Engineering Applications of Artificial Intelligence* **21**(7): 1035–1055.

Ichalal, D., Marx, B., Ragot, J. and Maquin, D. (2012). New fault tolerant control strategies for nonlinear Takagi–Sugeno systems, *International Journal of Applied Mathematics and Computer Science* **22**(1): 197–210, DOI: 10.2478/v10006-012-0015-8.

Johansen, T.A. and Babuška, R. (2003). Multi-objective identification of Takagi–Sugeno fuzzy models, *IEEE Transactions on Fuzzy Systems* **11**(6): 847–860.

Johansen, T.A. and Foss, A.B. (1993). Constructing NARMAX using ARMAX models, *International Journal of Control* **58**(5): 1125–1153.

Kanev, S. and Verhaegen, M. (2006). Multiple model weight estimation for models with no common state, *6th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes, SAFEPROCESS, Beijing, China*, pp. 637–642.

Kiriakidis, K. (2007). Nonlinear modelling by interpolation between linear dynamics and its application in control, *Journal of Dynamics Systems, Measurement and Control* **129**(6): 813–824.

Leith, D.J. and Leithead, W.E. (1999). Analytic framework for blended multiple model systems using linear local models, *International Journal of Control* **72**(7): 605–619.

Ljung, L. (1999). *System Identification: Theory for the User*, 2nd Edn., Prentice Hall PTR, London.

Mäkilä, P.M. and Partington, J.R. (2003). On linear models for nonlinear systems, *Automatica* **39**(1): 1–13.

McLoone, S. and Irwin, G.W. (2003). On velocity-based local model networks for nonlinear identification, *Asian Journal of Control* **5**(2): 309–315.

Murray-Smith, R. and Johansen, T.A. (1997). *Multiple Model Approaches to Modelling and Control*, Taylor & Francis, London.

Narendra, K.S. and Parthasarathy, K. (1990). Identification and control of dynamical systems using neural networks, *IEEE Transactions on Neural Networks* **1**(1): 4–27.

Nelles, O. (2001). *Nonlinear System Identification*, Springer-Verlag, Berlin/Heidelberg.

Nie, J. (1994). A neural approach to fuzzy modeling, *American Control Conference, ACC, Baltimore, MD, USA*, pp. 2139–2143.

Orjuela, R., Maquin, D. and Ragot, J. (2006). Nonlinear system identification using uncoupled state multiple-model approach, *Workshop on Advanced Control and Diagnosis, ACD'2006, Nancy, France*.

Orjuela, R., Marx, B., Ragot, J. and Maquin, D. (2008). State estimation for nonlinear systems using a decoupled multiple mode, *International Journal of Modelling Identification and Control* **4**(1): 59–67.

Orjuela, R., Marx, B., Ragot, J. and Maquin, D. (2009). On the simultaneous state and unknown inputs estimation of complex systems via a multiple model strategy, *IET Control Theory & Applications* **3**(7): 877–890.

Rodrigues, M., Theilliol, D., Aberkane, S. and Sauter, D. (2007). Fault tolerant control design for polytopic LPV systems, *International Journal of Applied Mathematics and Computer Science* **17**(1): 27–37, DOI: 10.2478/v10006-007-0004-5.

Sjöberg, J., Zhang, Q., Ljung, L., Benveniste, A., Delyon, B., Glorennec, P., Hjalmarsson, H. and Juditsky, A. (1995). Nonlinear black-box modeling in system identification: A unified overview, *Automatica* **31**(12): 1691–1724.

Takagi, T. and Sugeno, M. (1985). Fuzzy identification of systems and its applications to model and control, *IEEE Transactions on Systems, Man, and Cybernetics* **15**(1): 116–132.

Uppal, F.J., Patton, R.J. and Witczak, M. (2006). A neuro-fuzzy multiple-model observer approach to robust fault diagnosis based on the DAMADICS benchmark problem, *Control Engineering Practice* **14**(6): 699–717.

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

Verdult, V., Ljung, L. and Verhaegen, M. (2002). Identification of composite local linear state-space models using a projected gradient search, *International Journal of Control* **75**(16/17): 1385–1398.

Vinsonneau, B., Goodall, D. and Burnham, K. (2005). Extended global total least square approach to multiple-model identification, *16th IFAC World Congress, Prague, Czech Republic*, p. 143.

Walter, E. and Pronzato, L. (1997). *Identification of Parametric Models: From Experimental Data*, Springer-Verlag, Berlin.

Wen, C., Wang, S., Jin, X. and Ma, X. (2007). Identification of dynamic systems using piecewise-affine basis function models, *Automatica* **43**(10): 1824–1831.

Xu, D., Jiang, B. and Shi, P. (2012). Nonlinear actuator fault estimation observer: An inverse system approach via a T–S fuzzy model, *International Journal of Applied Mathematics and Computer Science* **22**(1): 183–196, DOI: 10.2478/v10006-012-0014-9.

Yen, J., Wang, L. and Gillespie, C.W. (1998). Improving the interpretability of Takagi–Sugeno fuzzy models by combining global learning and local learning, *IEEE Transactions on Fuzzy Systems* **6**(4): 530–537.

Rodolfo Orjuela received an M.Sc. in artificial intelligence, pattern recognition and robotics from the University of Paul Sabatier (France) in 2004, and an M.Sc. in automatic control from the University of Bordeaux (France) in 2005. In 2008, he received a Ph.D. degree in automatic control and signal processing from the National Polytechnic Institute of Lorraine (France). Since 2009, he has been as assistant professor of automatic control at the University of Haute-Alsace (France) with the MIPS Laboratory, MIAM-Team. His research interests include design of hierarchical control laws applied to automated guidance of on-ground vehicles (automotive and aircraft). Control law reconfiguration and fault tolerant control are also of interest for him. These works are mainly based-on predictive control, control allocation, linear parameter varying systems and multiple-models.

Benoît Marx was born in 1977, obtained an engineering degree and an M.Sc. degree, both in automatic control, in 2000 from Centrale Nantes (France), and the Ph.D. degree from the Grenoble Institute of Technology (France) in 2003. Since September 2004 he has been an assistant professor of automatic control at the University of Lorraine (France) and a member of the Research Centre for Automatic Control of Nancy. His research interests include state estimation, fault diagnosis and fault tolerant control of singular and/or nonlinear systems, with particular attention paid to nonlinear systems represented by multiple-models.

José Ragot is a researcher in the Research Center for Automatic Control of Nancy. His major research fields include data validation and reconciliation, process diagnosis, fault detection and isolation, fault tolerant control. A part of his activities is devoted to modeling and state estimation adapted to process diagnosis and mainly in the field of multi-models. He has successfully advised 70 Ph.D. and 3 HDR (French D.Sc.) theses and published about 500 refereed technical communications, including 120 papers in international journals, 340 communications in international conferences, 4 books and 14 chapters in collective volumes. Applications have been in various fields such as mineral and metallurgical processing, chemical engineering, water treatment, aerospace, environmental processes.

Didier Maquin is a professor of automatic control at the University of Lorraine, France. He teaches automatic control and applied mathematics in various engineering schools. He is a member of the Research Center for Automatic Control of Nancy and the scientific leader of a research group called Dependability and System Diagnosis, which gathers about twenty teachers and researchers, and as many Ph.D. students. From a national point of view, Didier Maquin is involved in the French Automatic Control Research Group. He has been on its directorate staff since 2001 and currently serves as the head of the theme *Dependability, Supervision and Maintenance*. Didier Maquin has co-authored around 50 journal articles and 140 conference communications.