

# A Multi-Modeling Strategy based on Belief Function Theory

Messaoud RAMDANI\*, Gilles MOUROT and José RAGOT

CRAN - INPL - 2, Avenue de la Forêt de Haye

54516 Vandoeuvre les Nancy, France

\*LASA Laboratory, Department of Electronics, University of Annaba

BP. 12, Annaba 23000, Algeria

E-mail: mes-ramdani@yahoo.com {gmourot, jragot}@ensem.inpl-nancy.fr

**Abstract**—In this paper, a multi-modeling strategy based on belief function theory is developed. The basic idea is to consider a fuzzy rule based system with a belief structure (BS) as output. The focal elements of each rule are formed by a subset of a collection of functional models. A particular attention is paid to the topic of combining global and local fuzzy models of Takagi-Sugeno type by a specific formulation of the proposed model. Some examples are given to show the validity of the approach.

**KeyWords:** belief structure, piece of evidence, functional models, multimodel.

## I. INTRODUCTION

The mathematical description of a priori unknown dynamic process from observed data and/or expert knowledge is a widely encountered problem in engineering, industry, time series analysis and other fields. In most cases, the models consist of a set of functional relationships between the elements of a set of variables. A common way consists of making an estimation by specifying a function with known structure but unknown parameters to be estimated. Although the obtained models might be acceptable in many cases, this approach does not seem satisfactory from the perspective of knowledge engineering and intelligent reasoning. More specifically, in many practical problems the prediction should be accompanied with some additional information that quantify their reliability, e.g., for example in the form of a confidence interval.

For the modeling of complex problems, the principle of divide-and-conquer is widely adopted in approximating unknown nonlinear mappings by a smooth combination of simple functional models. One of the most outstanding models is the TSK fuzzy model which has attracted a great attention due to its performances in many applications. Typically, fuzzy modeling involves structure and parameter identification. The latter is usually addressed by some gradient descent variant, e.g., the least squares algorithm or back-propagation. The former describes the inherent structure for a concrete problem by partitioning each input variable range into fuzzy sets. In general, the clustering is done in the input space of training data without integrating the interaction between the input variables and the output dynamics. In order to take into account the interaction

between input and output variables, the clustering is done in the product space of input and output variables instead of the input space in traditional algorithms of fuzzy modeling.

In the context of fuzzy modeling, the problem of identifying the parameters of the constituent local linear models of Takagi-Sugeno fuzzy models is of great importance because it is difficult to find a tradeoff between global model accuracy and interpretability [20], [21].

Here, we propose a multi-modeling strategy based on evidence theory whose main distinctive feature concerns the nature of the outputs, which consist in a set of belief masses. By using a particular formulation, we show that the local and global models can be combined in a single model.

This paper is organized as follows: section 2 describes briefly the basics of evidence theory. In section 3, the TSK model is presented and a special attention is paid to the issue of combining local and global models. The formulation of the functional approximation using the belief function theory and the description of the proposed model as well as further developments is also presented in section 4. Some numerical results are given in section 5. Conclusions are drawn in section 6.

## II. THE DEMPSTER-SHAFFER THEORY

### A. Knowledge model

In this section, we give a brief overview of certain aspects of D-S theory. For exposition on the D-S theory [1], see Shafer's seminal work [2] and [4]. This theory is based on the definition of a set of hypothesis  $\Omega$  called the frame of discernment, defined as follows:

$$\Omega = \{H_1, \dots, H_j, \dots, H_N\} \quad (1)$$

It is composed of  $N$  exhaustive and exclusive hypotheses  $H_j$  for  $j = 1, \dots, N$ . Let us denote  $2^\Omega$ , the power set composed with  $2^N$  propositions  $A$  of  $\Omega$ :

$$2^\Omega = \{\phi, \{H_1\}, \{H_2\}, \dots, \{H_N\}, \{H_1 \cup H_2\}, \{H_1 \cup H_3\}, \dots, \Omega\} \quad (2)$$

where  $\phi$  is the empty set. The concept of *basic belief assignment* (BBA) plays a central role in evidence theory. The mass of belief in an element of  $\Omega$  is quite similar to a

probability distribution, but differs by the fact that the unit mass is distributed among the element of  $2^\Omega$ , that is to say not only on singletons  $H_j$  in  $\Omega$  but on the composite hypotheses too. The belief  $m^i$  assigned to an information source  $S^i$  is thus defined by  $m^i : 2^\Omega \rightarrow [0, 1]$  such that:

$$\begin{aligned} m^i(\phi) &= 0, \\ \sum_{A \subseteq \Omega} m^i(A) &= 1. \end{aligned} \quad (3)$$

The mass  $m^i(A)$  represents how strongly the evidence supports  $A$ . Each subset  $A \subseteq \Omega$  such that  $m^i(A) > 0$  is called a focal element of  $m^i$ . Let us denote  $\mathcal{F}^i$  the set of focal elements associated to a belief function  $m^i$ . From this BBA, a belief function  $Bel^i$  and a plausibility function  $Pl^i$  are defined respectively, as

$$Bel^i(A) = \sum_{B \subseteq A} m^i(B) \quad (4)$$

and

$$Pl^i(A) = \sum_{A \cap B \neq \phi} m^i(B) \quad (5)$$

The quantity  $Bel^i(A)$  can be interpreted as measure of one's belief that hypothesis  $A$  is true. The plausibility  $Pl^i(A)$  can be viewed as the total amount of belief that could be potentially placed in  $A$ . Note that functions  $m^i$ ,  $Bel^i$  and  $Pl^i$  are three representation of the same piece of information. In evidence theory, one of the main difficulties concerns the modeling of the knowledge of the problem by a proper choice of the belief functions  $m^i$ .

### B. Dempster's rule of combination

In many cases, the combination or the fusion of different information sources is an interesting solution to obtain more relevant information. Evidence theory provides a coherent framework for integrating different information sources. In fact, for a given number of BBA's  $m^i$  obtained from different information sources  $S^i$ , the use of a combination rule provides combined masses summarizing the knowledge of the different sources. These belief masses can then be used for decision making with the advantage of the total knowledge contained in the belief functions given by each source. For two information sources  $S^1$  and  $S^2$ , the induced BBA's  $m^1$  and  $m^2$ , can be combined by the so-called Dempster's rule of combination to provide a new BBA  $m = m^1 \oplus m^2$ , called the orthogonal sum of  $m^1$  and  $m^2$ , and defined as:

$$m(A) = \frac{\sum_{B \cap C = A} m^1(B)m^2(C)}{1 - m(\phi)} \quad \forall A \subseteq \Omega \quad (6)$$

where the quantity in the numerator corresponds to the conjunctive rule of combination and the mass  $m(\phi)$  assigned to the empty set is defined by

$$m(\phi) = \sum_{B \cap C = \phi} m^1(B)m^2(C) \quad (7)$$

In the above equations, the mass  $m(\phi)$  reflects the conflict between the two sources  $S^1$  and  $S^2$ . Assuming the normality

of the BBA's ( $m(\phi) = 0$ ), the use of this rule is possible only if  $m^1$  and  $m^2$  are not totally conflicting, i.e., if there exist two focal elements  $B$  and  $C$  of  $m^1$  and  $m^2$  satisfying  $B \cap C \neq \phi$ . Let us denote the belief function resulting from the combination of  $K$  information sources as:

$$m = m^1 \oplus \dots \oplus m^i \dots \oplus m^K \quad (8)$$

where  $\oplus$  represents the operator of combination.

### III. THE TSK MODEL

The fuzzy rule based model proposed by Sugeno and his colleague's [14] is able to describe complex, nonlinear processes [17], [18], [16]. It is based on the fact that an arbitrary complex system is a combination of mutually inter-linked sub-systems. Let  $K$  regions, corresponding to individual sub-systems, be determined in the state space under consideration. The behavior of the system in these regions can then be described with simpler functional relationships. If the dependence is linear and if one rule is assigned to each sub system, the TSK fuzzy model can be represented with  $K$  rules of the following form:

$$\begin{aligned} R^i : & \text{IF } x_1 \text{ is } A_1^i \text{ and } \dots \text{ and } x_r \text{ is } A_r^i \\ & \text{THEN } y \text{ is } f^i(x) \quad \forall i = 1, \dots, K \end{aligned}$$

Of particular interest is the linear case of  $f^i(x)$ :

$$f^i(x) = \theta_{i0} + \theta_{i1}x_1 + \dots + \theta_{ir}x_r \quad (9)$$

where  $f^i(x)$  defines a locally valid model on the support of the cartesian product of fuzzy sets constituting the premise parts of the  $i$ th rule  $R^i$ , and  $\mathbf{x} = [x_1, \dots, x_r]^T \in \mathcal{X}$  is the vector of input (antecedent) variables.  $A_1^i, \dots, A_r^i$  are fuzzy sets defined in the antecedent space, and  $\theta_i^T = [\theta_{i0}, \dots, \theta_{ir}]$  is the vector of the consequent parameters of  $f^i(x)$ . The final output  $\hat{y} \in \mathcal{Y}$  is computed by taking the weighted average of the rule consequents

$$\hat{y} = \frac{\sum_{i=1}^K \beta^i(x) f^i(x)}{\sum_{i=1}^K \beta^i(x)} \quad (10)$$

where  $\beta^i(x)$  is the degree of activation of the  $i$ th rule.

$$\beta^i(x) = \prod_{j=1}^r \mu_{A_j^i(x_j)}, \quad i = 1, \dots, K \quad (11)$$

and  $\mu_{A_j^i(x_j)} : \mathcal{R} \rightarrow [0, 1]$  is the membership function of the fuzzy set  $A_j^i(x_j)$  in the antecedent of  $R^i$ .

The construction of the TS fuzzy model from data is solved in two steps: 1) structure identification 2) parameters estimation. In the first step, the antecedent and consequent variables of the model are determined. From the available training data that contain  $N$  input-output samples, a regression matrix  $X$  and an output vector  $\mathbf{y}$  are constructed

$$X = [\mathbf{x}_1, \dots, \mathbf{x}_N]^T, \quad \mathbf{y} = [y_1, \dots, y_N]^T \quad (12)$$

In the second step, the number of rules  $K$ , the antecedent fuzzy sets  $A_j^i$ , and the parameters of the rule consequents

$\theta_i$  for  $i = 1, \dots, K$  are identified. In order to capture the interaction between the input and output variables, the fuzzy clustering in the Cartesian product-space  $\mathcal{X} \times \mathcal{Y}$  is a useful method. This is based on the fact that the different clusters represent operating regions, where the system behaviour is approximated by local linear models. The data set  $Z$  to be clustered is formed by combining  $X$  and  $y$

$$Z = [X; y]^T \quad (13)$$

Once the training data  $Z$  and the number of clusters  $K$  are given, the Gustafson-Kessel (GK) clustering algorithm [19] is used to discover the potential regions of the rules.

From the the partition matrix  $U$ , whose  $ik$ th element  $\mu_k^i \rightarrow [0, 1]$  is the membership degree of the data  $z_k$ , the  $k$ th row of  $Z$  in cluster  $i$ , it is possible to extract the fuzzy sets in the antecedent parts. One-dimensional fuzzy sets  $A_j^i$  are obtained from multidimensional fuzzy clusters (given by  $U$ ) by point-wise projection onto the space of the input variable  $x_j$ :

$$\mu_{A_j^i(x_{kj})} = \text{proj}_j(\mu_k^i) \quad (14)$$

where  $\mu_k^i$  is the level of belonging of the  $k$ th sample (vector  $x_k$ ) to the  $i$ th cluster, while  $\mu_{A_j^i(x_{kj})}$  is the value of the membership of the  $j$ th input variable of the  $k$ th sample ( $j$ th co-ordinate of the vector  $x_k$ ) to the fuzzy set  $A_j^i$ . Since all the functions  $\mu_{A_j^i(x_{kj})}$  under consideration represent membership function, the condition  $\mu_{A_j^i(x_{kj})} : \mathcal{R} \rightarrow [0, 1]$  must hold true. Value of firing strength of the  $i$ th rule for each  $k$ th input sample is computed as *and-conjunction* by means of the product operator:

$$\beta_k^i = \beta^i(x_k) = \prod_{j=1}^r \mu_{A_j^i(x_{kj})} \quad k = 1, 2, \dots, N \quad (15)$$

The correct application of this equation requires an introduction of a threshold  $\xi$  ( $\xi = 0.05$ , for example) for the fulfillment of the condition:  $\mu_{A_j^i(x_{kj})} \leq \xi \Rightarrow \mu_{A_j^i(x_{kj})} = 0$ . The consequent parameters for each rule are obtained as a least squares estimate. Let  $X_e$  denote the matrix  $[1, X]$ ;  $W_i$  is a diagonal matrix of dimension  $N \times N$  having the normalized membership degree  $w^i(x_k) = \beta^i(x_k) / \sum_{j=1}^K \beta^j(x_k)$  as its  $k$ th diagonal element. Hence a matrix composition  $X'$  of dimension  $N \times K(r+1)$  is formed:

$$X' = [(W_1 X_e), (W_2 X_e), \dots, (W_K X_e)] \quad (16)$$

where matrix  $X_e = [1, X]$  contains rows  $[1, x_k^T]$ . Denote  $\theta'$ , the column vector of dimension  $K(r+1)$  given by

$$\theta' = [\theta_1^T, \theta_2^T, \dots, \theta_K^T]^T \quad (17)$$

where  $\theta_i^T = [\theta_{i0}, \dots, \theta_{ir}]$  for  $1 \leq i \leq K$ . The model  $y = X\theta' + \varepsilon$ , where  $\varepsilon$  is the approximation error, has the following least squares solution

$$\theta' = [(X')^T X']^{-1} (X')^T y \quad (18)$$

This identification strategy corresponds to a global learning. A second class of methods is known as local learning. As

opposed to global learning, where the local linear models cooperate for the minimisation of a global cost function, local learning strategies encourage the linear models to compete. In this case, the weighted least-squares method is used separately for each rule:

$$\theta_i = [X^T W_i X]^{-1} X^T W_i y \quad (19)$$

#### IV. FUZZY MODELING USING BELIEF FUNCTIONS

In the last years, many methods based on the Dempster-Shafer theory have been proposed in different areas including data fusion [3], [6], [7], classification [9], [11], [8] and regression [5], [12], [10]. Among the different methods, we mention only the two approaches that are relevant to our work. The first approach was proposed by Yager [5] in the context of fuzzy modeling. This strategy allows the integration of probabilistic uncertainty in fuzzy rule based systems. The output of the rules is a belief structure whose focal elements are fuzzy sets among the output variable linguistic terms.

The second approach was proposed by Denoeux [9] in the context of classification. This approach considers each neighbour of a pattern to be classified as an item of evidence supporting certain hypotheses concerning the class membership of that pattern. Based on this evidence, basic belief masses are assigned to each subset of the set of classes. Such masses are obtained for each of the  $k$ -nearest neighbours of the pattern under consideration and aggregated using the Dempster's rule of combination. Lately, the above approach has been improved and applied in regression analysis [10], [12]. For a given input query vector, the output variable is obtained in the form of a fuzzy belief assignment (FBA), defined as a collection of fuzzy sets of values with associated masses of belief. In [13], the output FBA is computed non-parametrically on the basis of the training samples in the neighbourhood of the query point. In this approach, the underlying principle is that the neighbours of the query point are considered as sources of partial information on the response variable; the bodies of evidence are discounted as a function of their distance to the query point, and pooled using the Dempster's rule of combination.

Based on the approach proposed by Yager [5] and the principle introduced by Denoeux [?], we propose a new model for approximating nonlinear functional mappings. This model is illustrated through a particular form of the functional models. More specifically, the functional models are linear models used in the TSK fuzzy model.

##### A. The proposed model

Let  $\mathcal{D} = (x_k, y_k)_{k=1}^N$  be a set of input-output data obtained from an unknown nonlinear process,  $\mathcal{X}$  and  $\mathcal{Y}$  are the domains of variation of the input of dimension  $r$  and the scalar output, respectively.

The principle underlying the fuzzy modeling based on Dempster-Shafer theory assumes the existence of a certain number  $c$  of functional relationships between the input variables and the output variable, denoted by  $f^j(x)$ ,  $j = 1, \dots, c$ .

In order to find a link between the evidence theory, which starts by the definition of a set of hypothesis as described in section 2, we shall consider that the above relationships models form the frame of discernment  $\Omega$ . Thus, we have:

$$\Omega = (\{f^1\}, \dots, \{f^j\}, \dots, \{f^c\}) \quad (20)$$

where  $\{f^j\}$  is the hypothesis that corresponds to the functional model  $f^j(x)$ .

In our modeling strategy, the model consists of  $K$  rules of the form:

$$R^i : IF \ x_1 \text{ is } A_1^i \text{ and } \dots \text{ and } x_r \text{ is } A_r^i \\ THEN \ y \text{ is } m^i$$

where  $m^i$  is a BBA whose focal elements are among the hypotheses of the frame of discernment  $\Omega$ . Let us denote by  $F_{ij}$ ,  $j = 1, \dots, J(i)$ ; the  $J(i)$  focal elements of  $m^i$  and denote by  $m^i(F_{ij})$  the weight (a mass of probability) associated to the  $j$ th focal element  $F_{ij}$  of  $m^i$ . This formulation which is quite general brings some aspects of the evidence theory to aggregate different BBA, where every BBA has its own elements. The firing strength of the  $i$ th rule is defined by the product of the membership degrees of the corresponding fuzzy sets:

$$\mu^i(x) = \prod_{j=1}^r \mu_{A_j^i}(x_j) \quad (21)$$

where  $\mu_{A_j^i}(x_j)$  is the membership function of the fuzzy set  $A_j^i(x_j)$ . In order to predict an output value  $y$  for an input vector  $x$ , every rule provides a piece of evidence concerning the value of the unknown output  $y$ . This item of evidence can be assimilated by a belief function  $m^i$ :

$$\begin{cases} m^i(\{f^j\} | x) = p_{ij} \phi_i(x), \quad j = 1, \dots, J(i) \\ m^i(\Omega | x) = 1 - \phi_i(x) \\ m^i(A | x) = 0 \quad \forall A \in \mathcal{F}^\Omega - \mathcal{F}^i \end{cases} \quad (22)$$

In the above equation  $\mathcal{F}^\Omega$  is the power set of  $\Omega$  and  $\mathcal{F}^i$  denote the focal elements of  $m^i$ . The quantities  $m^i(\{f^j\} | x)$ , and  $m^i(A | x)$  are the masses assigned to the subsets  $\{f^j\}$  and  $A$  after taking knowledge of  $x$ . The quantity  $m^i(\Omega | x)$  is the mass assigned to the frame of discernment after taking knowledge of  $x$ . The function  $\phi_i(x)$  is related to the input domain (domain of expertise) of the  $i$ th rule. In the present work, it is defined by:

$$\phi_i(x) = \alpha_i \mu^i(x) \quad (23)$$

where  $\mu^i(x)$  is given by (22) and  $\alpha_i$  is a weighting factor which verify  $(0 < \alpha_i < 1)$ . It is defined by a supplementary parameters  $\beta_i$  to control the importance of the rule  $R^i$  in the inference process:

$$\alpha_i = \frac{1}{1 + e^{-\beta_i}} \quad (24)$$

The coefficients  $p_{ij}$  which are application dependent must verify the following constraint:

$$\sum_{j=1}^{J(i)} p_{ij} = 1 \quad (25)$$

so that the piece of evidence provided by the  $i$ th rule is divided among the focal elements  $\mathcal{F}^i$  of the belief structure  $m^i$ . In order to make a decision, the outputs of the different rules which are belief structures, are combined using the Dempster's rule of combination. The final belief structure is then:

$$m = \oplus_{i=1}^K m^i \quad (26)$$

where  $\oplus$  represents the operator of combination. This belief structure  $m$  which is a vector of  $(c + 1)$  elements (masses of probabilities) should be normalized to give the following normalized belief structure:

$$m'_j = m_j / \sum_{q=1}^{c+1} m_q \quad j = 1, \dots, c + 1 \quad (27)$$

The overall multimodel is defined as a combination of the functional prototypes with a single model representing the frame of discernment  $\Omega$ , denoted by  $f^{(\Omega)}$ :

$$\hat{y} = \sum_{j=1}^c \omega_j(x) f^j(x) + \omega_{c+1}(x) f^{(\Omega)}(x) \quad (28)$$

where  $\omega_j$ ,  $j=1, \dots, c+1$  are the mixing coefficients, which verify  $\sum_{j=1}^{c+1} \omega_j(x) = 1$ . Taking the mixing coefficients as the normalized masses  $m'$  which is a vector of size  $(c + 1)$  gives the following model:

$$\hat{y} = \sum_{j=1}^c m'(\{f^j\}) f^j(x) + m'(\Omega) f^{(\Omega)}(x) \quad (29)$$

with  $m'(\{f^j\})$  is the  $j$ th element of  $m'$  and  $m'(\Omega)$  is the  $(c+1)$ th element of  $m'$ .

### B. Combining local and global fuzzy TSK models

In this subsection, we address the issue of combining local and global models derived from the TSK fuzzy models. To do that, we consider the case when the number of input prototypes (or rules) is equal to the number of functional prototypes; e.g  $K = c$  and we assume that the input region of every rule is dominated by the contribution of one model. In a first step, the Gustafson-Kessel (GK) clustering algorithm [19] is used to cluster the data in the Cartesian product space  $\mathcal{X} \times \mathcal{Y}$  and to define the antecedent parts of the rules. Thus, the natural choice of the coefficients  $p_{ij}$  introduced in equation (22) is  $p_{ij} = 1$  if  $i = j$  and  $p_{ij} = 0$  if  $i \neq j$  and the expression of the belief function  $m^i$  is reduced to the following form:

$$\begin{cases} m^i(\{f^i\} | x) = \phi_i(x) \\ m^i(\Omega | x) = 1 - \phi_i(x) \\ m^i(A | x) = 0 \quad \forall A \in \mathcal{F}^\Omega - \mathcal{F}^i \end{cases} \quad (30)$$

In order to obtain a compromise between local interpretability and global accuracy, we consider that the linear models  $f^i(x)$  of parameters  $\theta_i^T = [\theta_{i0}, \dots, \theta_{ir}]$  are identified by the weighted least squares (19). However, the model  $f^{(\Omega)}(x)$  is defined as a global TSK fuzzy model:

$$f^{(\Omega)} = \frac{\sum_{i=1}^K \mu^i(x) g^i(x; \Lambda_i)}{\sum_{i=1}^K \mu^i(x)} \quad (31)$$

with  $g^i(x; \Lambda_i)$  being local linear functions of parameters  $\Lambda_i^T = [\Lambda_{i0}, \dots, \Lambda_{ir}]$  identified globally by a single least-squares equation (18), where  $\mu^i(x)$  is the firing level of the  $i$ th rule defined by equation (21). It is worthwhile to notice that the induced mass of the belief structure  $m^i$  of the  $i$ th rule is divided among its local model  $f^i(x)$  and the global model  $f^\Omega(x)$ .

## V. EXPERIMENTAL RESULTS

In this section, two examples are tested to verify the validity of the proposed strategy: the first one is a univariate function. The other is the gas furnace data of Box and Jenkins [24], which is well known and frequently used as benchmark example for modeling and identification. The performance index is the mean squares error (MSE). The weighting factor  $\alpha_i = 0.99$  for  $i = 1, \dots, K$  so that all the rules have the same weight.

### A. Nonlinear static function approximation

To illustrate the performance of the proposed methodology, let us consider the univariate function taken from [22], [23]:

$$y(x) = 3e^{-x^2} \sin(\pi x) + \eta \quad (32)$$

where  $\eta$  is Gaussian noise with zero mean and  $\sigma^2 = 0.15$ . By using random inputs  $x$  uniformly distributed in the  $[-3, 3]$ , 300 samples of  $y(x)$  were obtained. This gives the identification  $Z = \{(x_k, y_k); k = 1, \dots, 300\}$ . The data  $Z$  are clustered by the GK clustering algorithm, with  $K = 7$  clusters. The different models are given by Table I. The local linear models  $f^j(x); j = 1, \dots, K$  and the mass assigned to the global model  $f^\Omega(x)$  are shown in Fig. 1. The actual and estimated values are shown in Fig. 2.

The TSK global model (29) obtained with GK+LS gives a  $MSE = 0.0058$  and the TSK global model obtained with GK+WLS (minimizing individual locally weighted prediction error criteria) gives a  $MSE = 0.0196$ . However, the combination scheme provides a  $MSE = 0.0181$ , which represents a compromise between the two models but the main advantage lies in the output belief structure as a powerful representation of different kinds of uncertainty and in the additional information of the mass  $m'(\Omega)$ . Depending of the specific formulation of the elementary basic belief assignments, the mass  $m'(\Omega)$  is relatively high in the overlapping regions of the local models (local experts) where the global model  $f^\Omega(x)$  is expected to give better results.

### B. Box and Jenkins gas furnace dataset

The Box and Jenkins gas furnace data are frequently used in performance evaluation of system identification methods. The data consist of 296 I/O measurements with a period of 9s: the input  $u(t)$  is gas flow rate and the output measurement is CO2 concentration in outlet gas. The instantaneous value of the output  $y(t)$  can be regarded as being influenced by ten variables  $y(t-1), \dots, y(t-4), u(t-1), \dots, u(t-6)$ . Following the recommendation of [15] and with the aim of achieving a comparison with other available models, are

TABLE I  
THE DIFFERENT MODELS FOR THE UNIVARIATE FUNCTION

Local linear models	The sub-models of $f^\Omega$
$f^1(x) = -4.0022x - 4.2350$	$g^1(x) = -4.2842x - 4.4591$
$f^2(x) = -0.0103x - 0.0349$	$g^2(x) = -0.0061x - 0.0465$
$f^3(x) = +7.1408x - 0.0649$	$g^3(x) = +7.4794x - 0.0649$
$f^4(x) = +0.6677x + 1.2226$	$g^4(x) = +0.7113x + 1.2915$
$f^5(x) = +0.6675x - 1.3520$	$g^5(x) = +0.7111x - 1.4210$
$f^6(x) = -4.0021x + 4.1051$	$g^6(x) = -4.2841x + 4.3293$
$f^7(x) = -0.0103x - 0.0950$	$g^7(x) = -0.0061x - 0.0834$

TABLE II  
BOX AND JENKINS DATA

Model	No inputs	No rules	MSE
Sugeno and Yasukawa [15]	3	6	0.190
Wang and Langari [16]	6	2	0.066
local modeling GK+WLS	2	5	0.1575
global modeling GK+LS	2	5	0.1288
GK+combined	2	5	0.1401
local modeling GK+WLS	10	3	0.0557
global modeling GK+LS	10	3	0.0386
GK+combined	10	3	0.0447

taken as input variables  $x = [y(t-1), u(t-4)]^T$ . Fig. 3 shows actual values and estimated values, obtained by means of the model with 5 rules.

The above results are obtained without any optimization of the parameters governing the behaviour of the model. The structure of the combined model which integrates explicitly some conflicting objectives is well suited to be optimized by a gradient descent technique or by a nonlinear optimization technique because the global model and the local models share the same antecedent membership functions. Finally, it is important to notice that the combination of the local and global models is introduced here to show the validity of the modeling strategy based on Dempster-Shafer theory and other models can be defined.

## VI. CONCLUSIONS

This paper describes a model for the modeling of the functional relationships using belief function theory and its validity is illustrated by addressing the issue of combining global and local learning in fuzzy modeling. The impact of the optimization of the different parameters on the model performance is under study. As a future work, we need also to examine in more details the behaviour of the model in the presence of different kinds of uncertainties.

## REFERENCES

- [1] A. Dempster, Upper and lower probabilities induced by multivalued mapping, *Annals of Mathematical Statistics*, AMS-38, 1967, pp. 325-339.
- [2] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, Princeton, New Jersey, 1976.
- [3] P. Smets and R. Kennes, The Transferable Belief Model, *Artificial Intelligence*, vol. 66, pp. 191-234, 1994.
- [4] R. R. Yager, M. Fedrizzi, and J. Kacprzyk, Advances in the Dempster-Shafer theory of evidence, New York: Wiley, 1994.
- [5] R. R. Yager and D. P. Filev, Including probabilistic uncertainty in fuzzy logic controller modeling using Dempster-Shafer theory, *IEEE Trans. Syst., Man, Cybern.*, vol. 25, pp. 1221-1230, 1995.
- [6] Bloch, I. Information combination operators for data fusion: A comparative review with classification, *IEEE Trans. Syst., Man, Cybern. A*, vol. 26, pp. 52-67, 1996.
- [7] E. Lefevre, O. Colot and P. Vannoorenberghe, Belief function combination and conflict management, *Information fusion*, vol. 3, pp. 149-162, 2002.
- [8] Binaghi, E., Gallo, I. and Madella, P., A neural model for fuzzy Dempster-Shafer classifiers, *International Journal of Approximate Reasoning*, vol. 25, pp. 89-121, 2000.
- [9] T. Denoeux, A k-nearest neighbour classification rule based on Dempster-Shafer theory, *IEEE Trans. Syst., Man, Cybern.*, vol. 25, pp. 804-814, 1995.
- [10] T. Denoeux, Function approximation in the framework of evidence theory: A connectionist approach, *Proceedings of the 1997 International Conference on Neural Networks (ICNN'97)*, vol. 1, pp. 199-203, Houston, June 1997. IEEE.
- [11] T. Denoeux, A Neural network classifier based on Dempster-Shafer theory, *IEEE Trans. Syst., Man, Cybern. A*, vol. 30, pp. 131-150, 2000.
- [12] S. Petit-Renaud and T. Denoeux, Regression analysis using fuzzy evidence theory, *Proceedings of FUZZ-IEEE'99*, vol. 3, pp. 1229-1234, Seoul, August 1999.
- [13] S. Petit-Renaud and T. Denoeux, Nonparametric regression analysis of uncertain and imprecise data using belief functions, *International Journal of Approximate Reasoning*, vol. 35, pp. 1-28, 2004.
- [14] T. Takagi and Sugeno, Fuzzy identification of systems and its applications to modeling and control, *IEEE Trans. Syst., Man, Cybern.*, vol. 15, pp. 116-132, Jan 1985.
- [15] M. Sugeno and T. Yasukawa, A fuzzy logic based approach to qualitative modeling, *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 7-31, Feb 1993.
- [16] L. Wang and R. Langari, Building Sugeno-type models using fuzzy discretization and orthogonal parameter estimation techniques, *IEEE Trans. Fuzzy Syst.*, vol. 3, pp. 454-458, Nov. 1995.
- [17] A. Fiordaliso, A nonlinear forecasts combination method based on Takagi-Sugeno fuzzy systems, *International Journal of Forecasting*, 14: 367-379, 1998.
- [18] A. Fiordaliso, Autostructuration of fuzzy systems by rules sensitivity analysis, *Fuzzy Sets and Systems*, 118, pp 281-296, 2001.
- [19] D. E. Gustafson and W. C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, in *Proc. IEEE CDC, San Diego, CA*, pp. 761-766, 1979.
- [20] John Yen, Improving the Interpretability of TSK fuzzy Models by Global Learning and Local Learning, *IEEE Trans. Fuzzy Syst.*, vol. 6, no. 4, pp. 530-537, Nov. 1998.
- [21] Tor A. Johansen and Robert Babuska, Multiobjective Identification of Takagi-sugeno Fuzzy Models, *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 6, pp. 847-859, Dec. 2003.
- [22] M. Setnes, R. Babuska, and H. B. Verbruggen, Rule-Based Modeling: Precision and Transparency, *IEEE Trans. Syst., Man., Cybern. C*, vol. 28, No. 1, pp. 165-169, 1998.
- [23] Junhong H. Nie and T. H. Lee, Rule-Based Modeling: Fast Construction and Optimal Manipulation, *IEEE Trans. Syst. Man., Cybern. A*, vol. 26, No. 6, pp. 728-738, 1996.
- [24] G. E. P. Box and G. M. Jenkins, Time Series analysis, Forecasting and control, Sna Francisco, CA: Holden Day, 1970.

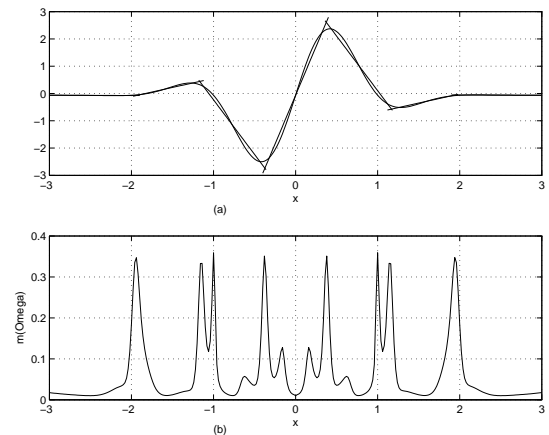


Fig. 1. (a) The local models (b) the mass of probability associated to the global model

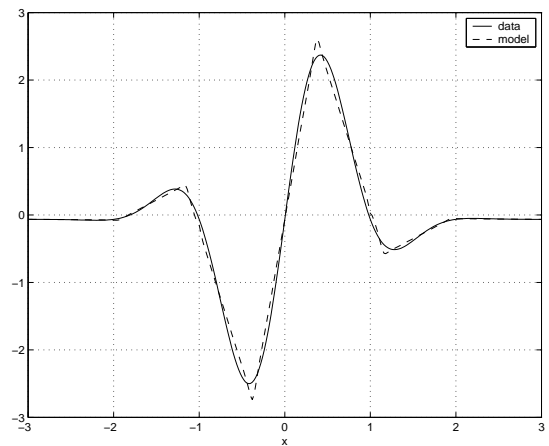


Fig. 2. Modeling of the univariate function.

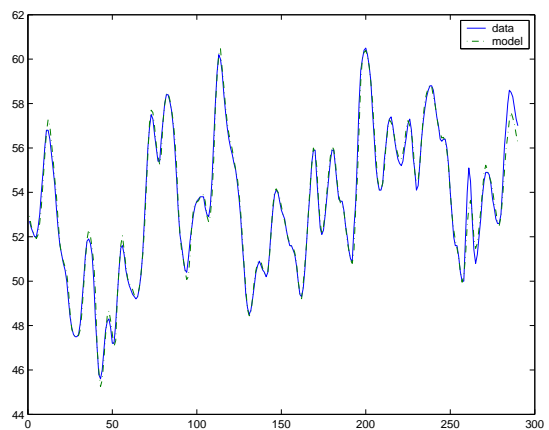


Fig. 3. Modeling of Box and Jenkins furnace data.