

# Compact and Transparent Fuzzy Models and Classifiers Through Iterative Complexity Reduction

Hans Roubos, *Student Member, IEEE*, and Magne Setnes, *Member, IEEE*

**Abstract**—In our previous work we showed that genetic algorithms (GAs) provide a powerful tool to increase the accuracy of fuzzy models for both systems modeling and classification. In addition to these results, we explore the GA to find redundancy in the fuzzy model for the purpose of model reduction. An aggregated similarity measure is applied to search for redundancy in the rule base description. As a result, we propose an iterative fuzzy identification technique starting with data-based fuzzy clustering with an overestimated number of local models. The GA is then applied to find redundancy among the local models with a criterion based on maximal accuracy and maximal set similarity. After the reduction steps, the GA is applied with another criterion searching for minimal set similarity and maximal accuracy. This results in an automatic identification scheme with fuzzy clustering, rule base simplification and constrained genetic optimization with low-human intervention. The proposed modeling approach is then demonstrated for a system identification and a classification problem. Results are compared to other approaches in the literature. Attractive models with respect to compactness, transparency and accuracy, are the result of this symbiosis.

**Index Terms**—Fuzzy classifier, genetic algorithm (GA), Iris data, rule base reduction, Takagi–Sugeno (T–S) fuzzy model, transparency and accuracy.

## I. INTRODUCTION

**F**UZZY SETS and fuzzy logic, introduced in 1965 by Zadeh [2], are applied in a wide variety of disciplines. Fuzzy modeling is one of those disciplines that is often used in systems identification and control, fault diagnosis, classification and decision support systems [3], [4]. Like many nonsymbolic modeling methods such as neural networks, fuzzy models are universal approximators [5]. However, fuzzy models differ from nonsymbolic methods, mainly in that they can represent knowledge in an inspectable manner using fuzzy if-then rules [6]. This facilitates validation and correction by human experts and provides a way of communicating with the users. Fuzzy models can be built by encoding expert knowledge into linguistic rules, giving a transparent system with knowledge that can be maintained and expanded by human experts. However, knowledge acquisition is not a trivial task. Experts are not always available and their knowledge is often incomplete, episodic and time-varying. Hence, there is an interest in data-driven fuzzy modeling.

Manuscript received November 10, 2000.

H. Roubos is with the Delft University of Technology, Faculty of Information Technology and Systems, Systems and Control Laboratory, 2600 GA, Delft, The Netherlands (e-mail: hans@ieee.org).

M. Setnes is with Heineken Technical Services, Research and Development, 2382 PH Zoeterwoude, The Netherlands (e-mail: magne@ieee.org).

Publisher Item Identifier S 1063-6706(01)06536-5.

Different approaches have been proposed to obtain fuzzy models from data. Most approaches, however, utilize only the function approximation capabilities of fuzzy systems and little attention is paid to the qualitative aspects. This makes them less suited for applications in which emphasis is not only on accuracy, but also on interpretability, computational complexity, and maintainability [7]–[11].

We focus on the design of interpretable fuzzy rule based models and classifiers from data with low-human intervention. A modeling scheme is presented that combines three previously studied tools for rule based modeling: 1) fuzzy clustering; 2) similarity driven simplification; and, 3) evolutionary optimization. Moreover, we show that different tools to initialize, tune, and manipulate fuzzy models can be favorably combined to obtain compact fuzzy rule based models of low complexity with still good approximation accuracy.

First, fuzzy clustering [3] is applied to obtain an initial rule based model from sampled data. Since rules obtained and tuned by data-driven techniques often contain redundancy in terms of similar (overlapping) fuzzy sets, similarity driven rule base simplification is applied to detect and merge compatible fuzzy sets in the model and to remove “don’t-care” terms [12]. Finally, since these methods are based on the separate identification and manipulation of the models premise and consequent parts, a constrained real-coded genetic algorithm (GA) is applied to simultaneously fine-tune (optimize) all parameters in the resulting rule base. In our previous work [1], it was shown that such a GA was able to strongly improve the models performance by small alterations to the rules.

By combining these tools in an iterative loop with a redundancy search, we propose a powerful fuzzy modeling scheme with complexity reduction. The algorithm starts with an initial model, obtained here by means of fuzzy clustering in the product space of measured inputs and outputs. Successively, rule base simplification and GA-based optimization are applied in an iterative manner. The GA performs a multi-criterion search for model accuracy while trying to exploit the possible redundancy in the model. In the next iteration, this redundancy will be used by the rule reduction and rule base simplification tools to reduce and simplify the rule base. The result is a compact fuzzy-rule base of low complexity with high accuracy. When the iterations terminate, a final GA-based optimization is performed to increase accuracy and transparency as opposed to the GA objective in the iterative loop, which tries to exploit redundancy.

The next section discusses data-driven fuzzy modeling. In Section III, rule base simplification tools are described and the proposed iterative complexity reduction scheme is introduced. In Section IV, the method is demonstrated on a nonlinear dy-

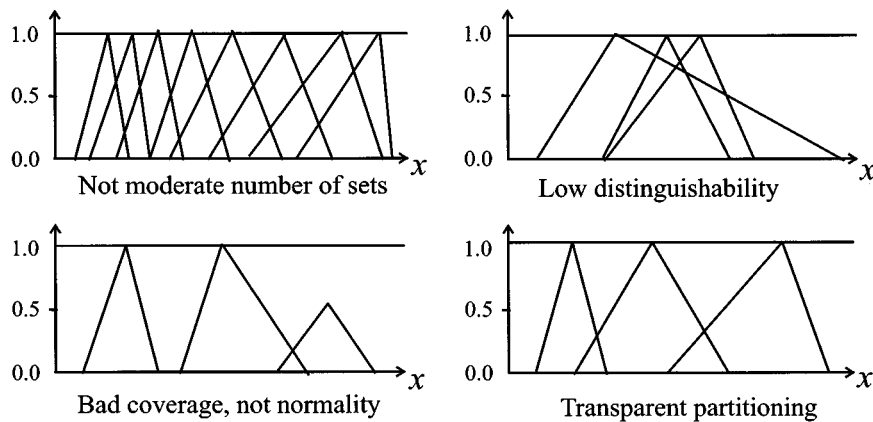


Fig. 1. Transparency of the fuzzy rule base premise.

dynamic systems model and the Iris classification problem—both known from the literature—and the results are compared to other published methods. Section V concludes the paper.

## II. DATA-DRIVEN MODELING

### A. The Takagi–Sugeno Fuzzy Model

Rule based models of the TS type [13] are especially suitable for the approximation of dynamic systems. In addition, accurate and interpretable TS fuzzy classifiers can be made as was discussed in [1]. The rule consequents are usually taken to be constant or linear functions of the inputs

$$R_i: \text{If } x_1 \text{ is } A_{i1} \text{ and } \dots x_n \text{ is } A_{in}, \text{ then} \\ \hat{y}_i = \zeta_{i1}x_1 + \dots, \zeta_{in}x_n + \zeta_{i(n+1)}, \quad i = 1, \dots, M. \quad (1)$$

Here  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$  is the input vector,  $\hat{y}_i$  is the output of the  $i$ th rule, and  $A_{i1}, \dots, A_{in}$  are fuzzy sets defined in the antecedent space by membership functions  $\mu_{A_{ij}}(x_j) : \mathbb{R} \rightarrow [0, 1]$ .  $\zeta_{ij}$  are the consequent parameters and  $M$  is the number of rules. The overall output of the fuzzy rule base model is computed by aggregating the individual contributions of the rules

$$\hat{y} = \sum_{i=1}^M p_i(\mathbf{x}) \hat{y}_i \quad (2)$$

where  $p_i(\mathbf{x})$  is the normalized firing strength of the  $i$ th rule

$$p_i(\mathbf{x}) = \frac{\prod_{j=1}^n A_{ij}(x_j)}{\sum_{i=1}^M \prod_{j=1}^n A_{ij}(x_j)}, \quad i = 1, 2, \dots, M. \quad (3)$$

In the following, we will apply the frequently used triangular membership functions  $\mu_{ij}$  to describe the fuzzy sets  $A_{ij}$  in the rule antecedents

$$\mu(x; a, b, c) = \max\left(0, \min\left(\frac{x-a}{b-a}, \frac{c-x}{c-b}\right)\right). \quad (4)$$

This choice was made for reasons of simplicity. The proposed approach, however, is as well applicable for other type of membership functions, e.g., trapezoids, Gaussians, or piecewise exponentials.

### B. Identification from Data

Given  $K$  input-output data pairs  $\{\mathbf{x}_k, y_k\}$ , the typical identification of the TS model is done in two steps: 1) the fuzzy rule antecedents are determined and 2) then least squares parameter estimation is applied to determine the consequents [3], [14]. In the examples in this paper, the antecedents of the initial fuzzy rule bases are obtained from fuzzy  $c$ -means clustering in the product space of the sampled input-output data. Following the approach in [3], [15], each cluster represents a certain region in the systems input-output state-space and corresponds to a rule in the rule base. The fuzzy sets in the rule antecedent are obtained by projecting the cluster onto the domain of the various inputs.

### C. Transparency and Accuracy

The initial rule base constructed by fuzzy clustering typically fulfills many criteria for transparency and good semantic properties [8]: 1) moderate number of rules; 2) distinguishable sets; 3) normal fuzzy sets; and 4) coverage of input patterns (see Fig. 1). The approximation capability of the rule base as obtained from fuzzy clustering, however, remains suboptimal. The projection of the clusters onto the input variables and their approximation by parametric functions like, e.g., triangular fuzzy sets, introduces a structural error since the resulting premise partition differs from the cluster partition matrix. Also, the separated identification of the rule antecedents and the rule consequents prohibits interactions between them during modeling. To improve the approximation capability of the initial model, a GA based optimization method is applied.

The transparency and compactness of the initial rule base are often also subject to improvement. The distinguishability of the rules and the terms (fuzzy sets) resulting from the projection depends, among others, on the difficult determination of the correct number of clusters (rules) in the data and their position in the product space. To reduce the rule base complexity, we iteratively seek for rule base redundancy and apply rule base simplification, as explained in the next section.

## III. RULE-BASE REDUCTION AND SIMPLIFICATION

### A. Similarity Driven Rule Base Simplification

The similarity driven rule base simplification method was proposed in [12]. A similarity measure is used to quantify the

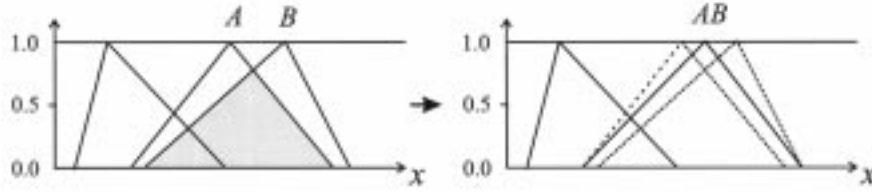


Fig. 2. Similarity driven rule base simplification.

redundancy among the fuzzy sets in the rule base. Similar fuzzy sets, representing compatible concepts are merged in order to obtain a generalized concept represented by a new fuzzy set that replaces the similar ones in the rule base. This reduces the number of different fuzzy sets (linguistic terms) used in the model. The similarity measure is also used to detect “don’t care” terms, i.e., fuzzy sets in which all elements of a domain have a membership close to 1. Similarity driven simplification differs from rule reduction in that it is driven by the similarity among fuzzy sets defined on the domain of the same antecedent variable and not in the product space of the inputs. Thus, the models term set can be reduced without necessarily any rules being removed.

where

$A, B$  fuzzy sets to be compared;  
 $|\cdot|$  denotes the cardinality of a set;  
the  $\cap$  and  $\cup$  operators represent the intersection and union, respectively.

We apply a similarity measure based on the set-theoretic operations of intersection and union

$$S(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (5)$$

where  $A$  and  $B$  are the fuzzy sets to be compared,  $|\cdot|$  denotes the cardinality of a set, and the  $\cap$  and  $\cup$  operators represent the intersection and union, respectively. For discrete domains  $X = \{x_j | j = 1, 2, \dots, m\}$ , this can be written as

$$S(A, B) = \frac{\sum_{j=1}^m [\mu_A(x_j) \wedge \mu_B(x_j)]}{\sum_{j=1}^m [\mu_A(x_j) \vee \mu_B(x_j)]} \quad (6)$$

where  $\wedge$  and  $\vee$  are the minimum and maximum operators, respectively.  $S$  is a symmetric measure in  $[0, 1]$ . If  $S(i, j) = 1$ , then the two membership functions  $A_i$  and  $A_j$  are equal and  $S(i, j)$  becomes 0 when the membership functions are nonoverlapping.

Similar fuzzy sets are merged when their similarity exceeds a user defined threshold  $\theta \in [0, 1]$ . In the examples,  $\theta = 0.5$  is applied. The optimal choice of this parameter depends on the number and type of membership functions and is, therefore, problem specific to some extent; however, generally we noticed to values in the range  $[0.4-0.7]$  may be a good choice. Different settings result in more or less conservative fuzzy-set merging. Merging reduces the number of different fuzzy sets (linguistic terms) used in the model and, thereby, increases the transparency. If all the fuzzy sets for a feature are similar to the universal set, or if merging led to only one membership function for a feature, then this feature is eliminated from the model. The method is illustrated in Fig. 2.

## B. Genetic Multicriteria Optimization

To improve the approximation capability or classification capability of the rule base, we apply a GA optimization method introduced in [1]. Also other model properties can be optimized by applying multicriteria functions, like, e.g., search for redundancy [16].

The model accuracy is measured as the mean square error for system approximation (8) and in terms of the number of misclassifications for a classifier (9). To reduce the model complexity, the accuracy objective is combined with a similarity measure in the GA objective function. Similarity is rewarded during the iterative process, that is, the GA tries to emphasize the redundancy in the model (see Fig. 3). This redundancy is then used to remove unnecessary fuzzy sets in the next iteration. In the final step, fine-tuning is combined with a penalty for similar fuzzy sets in order to obtain a distinguishable term set for linguistic interpretation. The GA seeks to minimize the following multicriteria function

$$J = (1 + \lambda S^*) \cdot J^* \quad (7)$$

where  $J^*$  is either the mean squared error (MSE) for system approximation problems

$$J^* = \text{MSE} = \frac{1}{K} \sum_{k=1}^K (y_k - \hat{y}_k)^2 \quad (8)$$

where  $y$  is the true output and  $\hat{y}$  is the model output, or for classification problems

$$J^* = \frac{1}{K} \left( \sum_{k=1}^K (y_k - \hat{y}_k)^2 + \sigma \cdot \sum_{k=1}^K (c_k \neq \hat{c}_k) \right) \quad (9)$$

where the classifications error is included, with  $c$  the class,  $\hat{c}$  the predicted class, and  $\sigma$  a weight factor. The MSE was needed in previous optimization schemes without redundancy measures to differentiate between various solutions with the same number of classification errors; and it was found to speed up the convergence of the GA [1]. Moreover, it helps to find fuzzy rules with consequents in the neighborhood of the class labels which improves the interpretability and prevents the optimization for making a black-box model based on interpolation between rules only.

Finally,  $S^* \in [0, 1]$  is the average of the maximum pair-wise similarity that is present in each input, i.e.,  $S^*$  is an aggregated similarity measure for the total model

$$S^* = \frac{1}{n} \sum_{i=1}^n \left( \frac{\max(S(A_{ij}, A_{ik}))}{\eta_i - 1} \right) \quad j, k \in 1, 2, \dots, \eta_i, \quad j \neq k \quad (10)$$

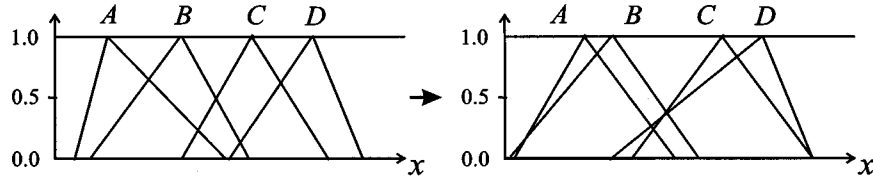


Fig. 3. Search for redundancy.

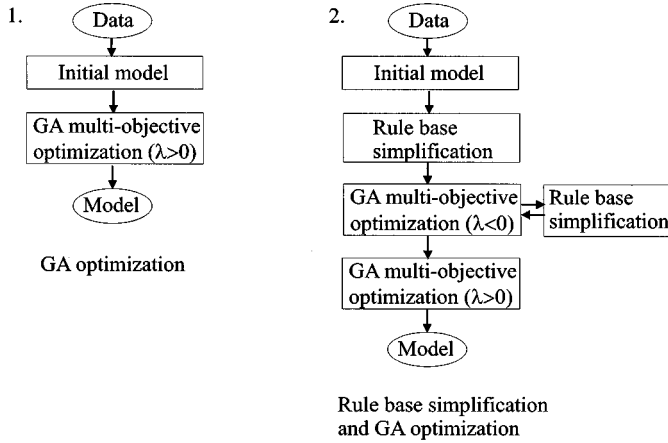


Fig. 4. Two modeling schemes resulting from a combination of tools.

where  $n$  is the number of inputs and  $\eta_i$  the number of sets for each input variable. The weighting function  $\lambda$  determines whether similarity is rewarded ( $\lambda < 0$ ), or penalized ( $\lambda > 0$ ). The absolute value of  $\lambda$  determines the trade-off between the similarity objective and the accuracy. Normally, some experience is necessary to decide about a good value, however, the final results seems to be not highly sensitive for the exact value. For the examples, generally good results were obtained with  $|\lambda|$  values in the range  $[0, 2]$ .

### C. Proposed Fuzzy Modeling Scheme

When an initial fuzzy model has been obtained from data, it is successively simplified and optimized in an iterative fashion. Combinations of the GA with the rule base simplification described above can lead to different modeling schemes. The two different approaches shown in Fig. 4 will be compared.

The first scheme consist of the initial model construction and fine-tuning based on the multi-objective criterion while the second scheme includes the iterative complexity reduction. The second scheme is as follows:

- 1) *Initialization*: Obtain an initial fuzzy model.
- 2) *Complexity reduction*: Repeat until termination:
  - a) similarity driven rule base simplification;
  - b) GA optimization with redundancy objective: *model accuracy while exploiting redundancy*.
- 3) GA fine tuning with transparency objective: *model accuracy with well separated fuzzy sets*.

Step 2 terminates when the rule base can not be further reduced or simplified.

A real-coded GA [17] is applied for the simultaneous optimization of the parameters of the antecedent membership functions and the rule consequents. The complete GA and fuzzy

model coding is given in [1], however, a few aspects of the fuzzy model coding are now explained for clearness of the examples that will follow.

The complete fuzzy model is coded on a chromosome that consist of a vector of  $N$  elements  $[v_1, v_2, \dots, v_k, \dots, v_N]$ . With a population size  $L$ , we encode the parameters of each fuzzy model (solution) in a chromosome  $\mathbf{s}_l, l = 1, \dots, L$ , as a sequence of elements describing the fuzzy sets in the rule antecedents followed by the parameters of the rule consequents. For a model of  $M$  fuzzy rules, triangular fuzzy sets (each given by three parameters), a  $n$ -dimensional premise and  $n + 1$  parameters in each consequent function, a chromosome of length  $N = M(3n + (n + 1))$  is encoded as

$$\mathbf{s}_l = (\text{ant}_1, \dots, \text{ant}_M, \theta_1, \dots, \theta_M) \quad (11)$$

where  $\theta_i$  contains the consequent parameters  $p_{iq}$  of rule  $R_i$ , and  $\text{ant}_i = (a_{i1}, b_{i1}, c_{i1}, \dots, a_{in}, b_{in}, c_{in},)$  contains the parameters of the antecedent fuzzy sets  $A_{ij}, j = 1, \dots, n$ , according to (4). In the initial population  $S^0 = \{\mathbf{s}_1^0, \dots, \mathbf{s}_L^0\}$ ,  $\mathbf{s}_1^0$  is the initial model, and  $\mathbf{s}_2^0, \dots, \mathbf{s}_L^0$  are created by random variation (uniform distribution) around  $\mathbf{s}_1^0$  within the defined search space.

The upper and lower limits of the search space are determined by two user defined bounds  $\alpha_1$  and  $\alpha_2$ , that determine the maximum allowed variation around the initial chromosome  $\mathbf{s}_1^0$  for the antecedent and the consequent parameters, respectively. The first bound,  $\alpha_1$  is intended to maintain the distinguishability of the models term set (the fuzzy sets) by allowing the parameters describing the fuzzy sets  $A_{ij}$  to vary only within a bound of  $\pm\alpha_1|\mathcal{X}_j|$  around their initial values, where  $|\mathcal{X}_j|$  is the length (range) of the domain on which the fuzzy sets  $A_{ij}$  are defined. By a low value of  $\alpha_1$ , one can avoid the generation of domain-wide and multiple overlapping fuzzy sets, which is a typical feature of unconstrained optimization. The second bound,  $\alpha_2$ , is intended to maintain the local-model interpretation of the rules by allowing the  $q$ th consequent parameter of the  $i$ th rule,  $\zeta_{iq}$ , to vary within a bound of  $\pm\alpha_2(\max_i(\zeta_{iq}) - \min_i(\zeta_{iq}))$  around its initial value.

## IV. EXAMPLES

### A. Example: Nonlinear Plant

We consider the 2nd order nonlinear plant studied by Wang and Yen in [19], [18], [20]

$$y(k) = g(y(k-1), y(k-2)) + u(k) \quad (12)$$

with

$$g(y(k-1), y(k-2)) = \frac{y(k-1)y(k-2)(y(k-1) - 0.5)}{1 + y^2(k-1)y^2(k-2)}. \quad (13)$$

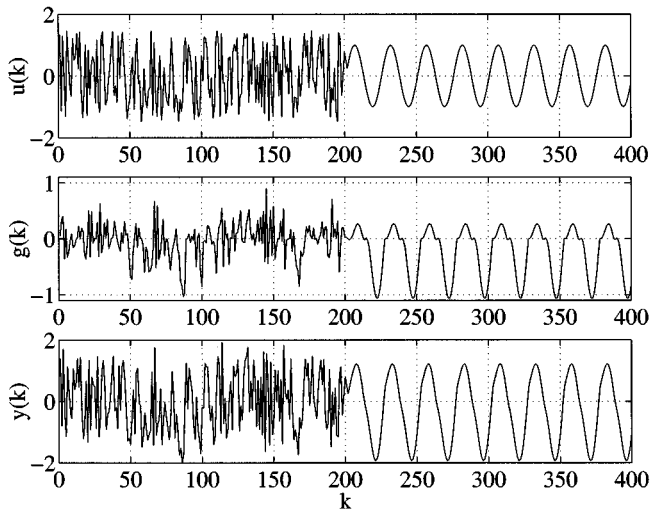


Fig. 5. Input  $u(k)$ , unforced system  $g(k)$ , and output  $y(k)$  of the plant in (12).

The goal is to approximate the nonlinear component  $g(y(k-1), y(k-2))$  of the plant with a fuzzy model. In [19], 400 simulated data points were generated from the plant model (12). 200 samples of identification data were obtained with a random input signal  $u(k)$  uniformly distributed in  $[-1.5, 1.5]$ , followed by 200 samples of evaluation data obtained using a sinusoid input signal  $u(k) = \sin(2\pi k/25)$  (Fig. 5).

This example was also used in [1] and a comparison with the results of [19], [18], [20] was made. Here we apply the newly proposed modeling scheme on the original data (not in [1]), show the results for the linear TS model and compare these with the results in [18]. Here, various information criteria were used to *successively* pick rules from a set of 36 rules in order to obtain a compact, but accurate model. The initial rule base was obtained by partitioning each of the two inputs  $y(k-1)$  and  $y(k-2)$  by six equally distributed fuzzy sets. The rules were picked in an order determined by an orthogonal transform. The best results obtained in each case are summarized in Table I. For reasons of completeness, we also added the results of [1] in this table.

We applied both the modeling approach proposed in Section III-C. The GA was applied with a population size:  $L = 40$ ; number of chromosomes  $n_C = 10$ ; domain parameters  $\alpha_1 = 25\%$  and  $\alpha_2 = 25\%$  and number of generations  $T = 400$  in the final optimization; and  $T = 200$  in the complexity reduction step. The threshold  $\lambda = 1$  for redundancy searches and  $\lambda = -1$  in the final optimization. The threshold for set merging was  $\theta = 0.5$  and  $\theta = 0.8$  for removing sets similar to the universal set (“don’t care” terms).

First, a TS model with linear consequents was considered, based on scheme 1 (Fig. 4). An initial model of only five rules was constructed by clustering. The MSE for both training and validation data were comparable with earlier results in [1]. Moreover, the result on the validation data (low-frequency signal) is twice as good as on the identification data, indicating the generality of the obtained model. By GA optimization, the MSE was reduced by 71% from  $4.9 \cdot 10^{-3}$  to  $1.4 \cdot 10^{-3}$  on the training data and by 80% from  $2.9 \cdot 10^{-3}$  to  $5.9 \cdot 10^{-4}$  on the evaluation data.

TABLE I  
LINEAR TS FUZZY MODELS FOR THE DYNAMIC PLANT TYPE

Ref.	No. of rules	No. of sets	MSE train	MSE eval
[18] <sup>†</sup>	36 (initial)	12 <i>B</i> -splines	$1.9 \cdot 10^{-6}$	$2.9 \cdot 10^{-3}$
	24 (optimized)	12 <i>B</i> -splines	$2.0 \cdot 10^{-6}$	$6.4 \cdot 10^{-4}$
[1] <sup>‡</sup>	5 (initial)	10 triangulars	$5.8 \cdot 10^{-3}$	$2.5 \cdot 10^{-3}$
	4 rules (optimized)	4 triangular	$1.2 \cdot 10^{-3}$	$4.7 \cdot 10^{-4}$
This paper	5 (initial)	10 triangulars	$4.9 \cdot 10^{-3}$	$2.9 \cdot 10^{-3}$
Fig. 6	5 rules (optimized)	10 triangulars	$1.4 \cdot 10^{-3}$	$5.9 \cdot 10^{-4}$
Fig. 6	5 rules (optimized)	5 triangulars	$8.3 \cdot 10^{-4}$	$3.5 \cdot 10^{-4}$

<sup>†</sup> The low MSE on the training data is in contrast to the MSE for the evaluation data which indicates overtraining.

<sup>‡</sup> These results are not based on the original data, but on newly generated data.

Second, a TS model with linear consequents was obtained by scheme 2. The initial model was obtained with five clusters, resulting in a model with five rules and ten-fuzzy sets. The model was reduced in two steps: 1) simplification reduces from 5+5 to 3+5 fuzzy sets and 2) simplification reduces to 2+3 sets. The resulting TS model with linear consequents has only five rules using 2+3 fuzzy sets Fig. 6. The identification and validation results as well as the prediction error are presented in Fig. 7.

The approximation properties are very similar to those in [18], but much fewer rules and fuzzy sets were necessary which resulted in a very compact and transparent model. The accuracy is also similar to the four-rule TS model in [1]. This model has one rule less but is not based on the original data, which may explain the difference. The linear consequent TS model extrapolates well and the difficult part in the low region is nicely approximated. The local submodels and the overall model output are shown in Fig. 8. The submodels approximate the local behavior well. In addition, the “real” surface as given by the original model is shown and it is clear that the TS model approximates the regions where data is supplied (system regions). Once again, the reduced and optimized TS model with five rules and five sets is comparable in accuracy to the initial TS model with five rules and ten fuzzy sets.

From the results summarized in Table I, we see that the proposed modeling approach is capable of obtaining good results using fewer rules and fuzzy sets than other approaches reported in the literature. Moreover, simple triangular membership functions were used as opposed to cubic *B*-splines in [19], [18], [20], and Gaussian-type basis functions in [19], [20]. By applying the GA after each rule base simplification step, not only accurate, but also compact and transparent rule based models were obtained.

### B. Example: Iris Data

The Iris data is a common benchmark in classification and pattern recognition studies [21], [22], [11], [1]. It contains 50 measurements of four features from each of the three species *Iris setosa*, *Iris versicolor*, and *Iris virginica* [23]<sup>1</sup>. The measurements are shown in Fig. 9.

There is a large variety of methods applied to this data set and the accuracy and complexity varies a lot. Ishibuchi et al. [22] reviewed nine fuzzy classifiers and ten nonfuzzy classifiers from the literature, giving between three and 24 misclassifications for the Iris classification problem for leaving-one-out validation.

<sup>1</sup>The original Iris data was recently republished in [24].

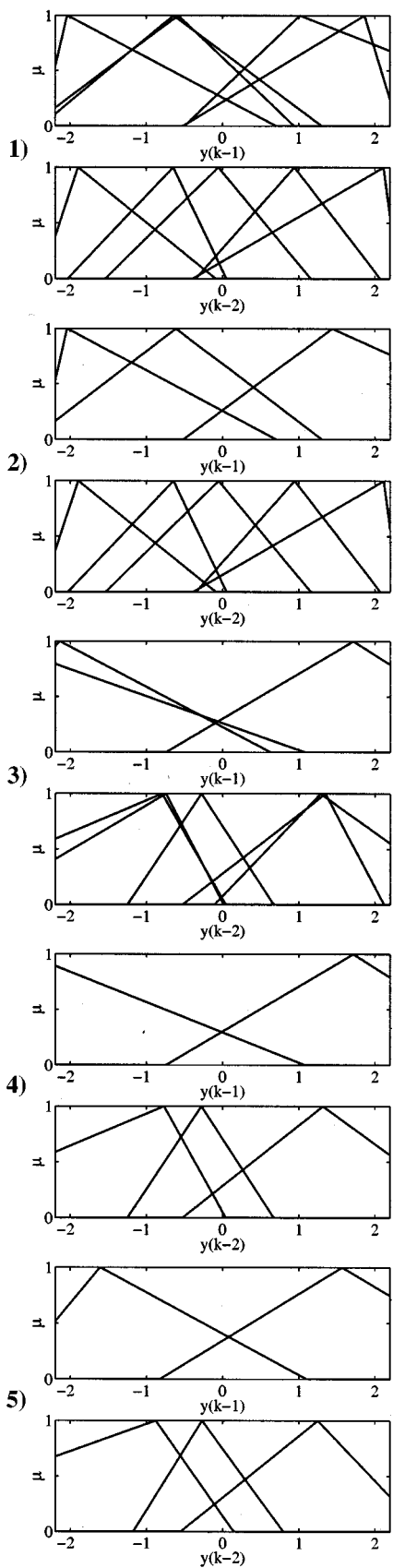


Fig. 6. Local linear TS model derived in five steps: 1) initial model with five clusters (5 rules + 10 sets), 2) set merging (5 rules + 8 sets), 3) GA-optimization (5 rules + 8 sets), 4) set-merging (5 rules + 5 sets), 5) final GA-optimization (5 rules + 5 sets).

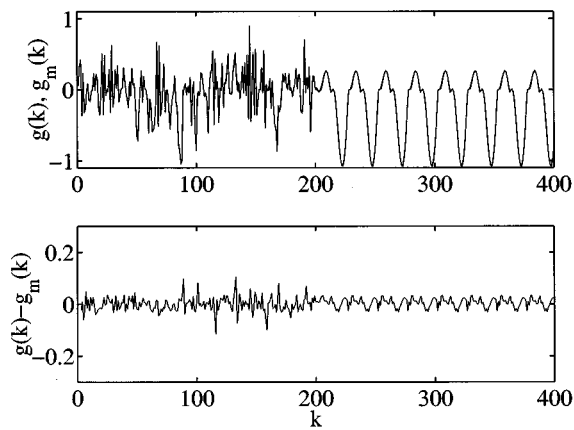


Fig. 7. Simulation of the five-rule TS model (top) and the error in the estimated output (bottom).

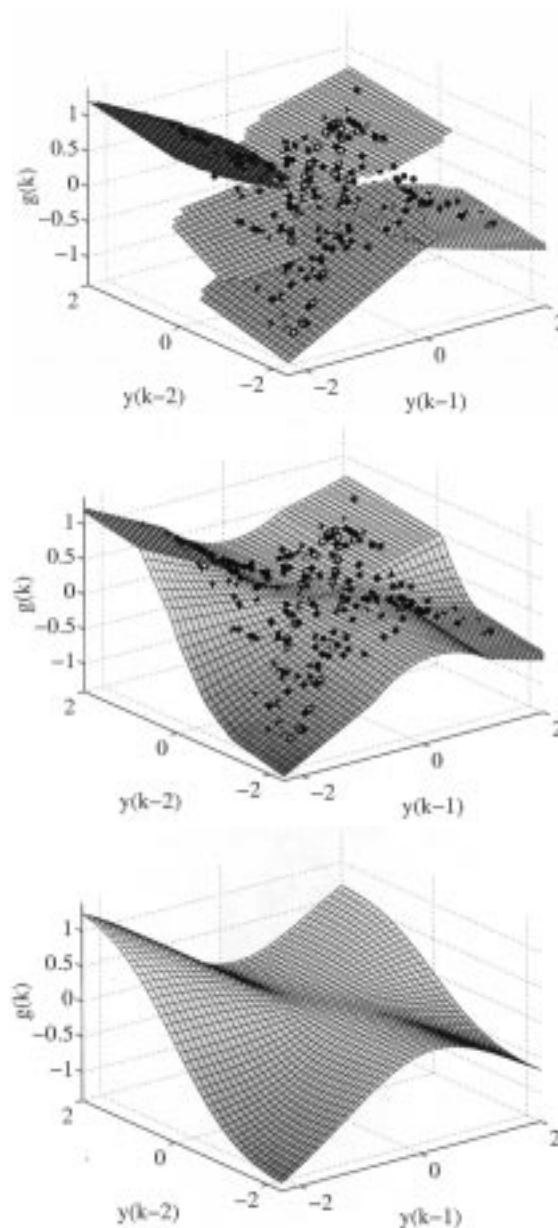


Fig. 8. Local linear models (top), TS model (middle), and real-surface (bottom).

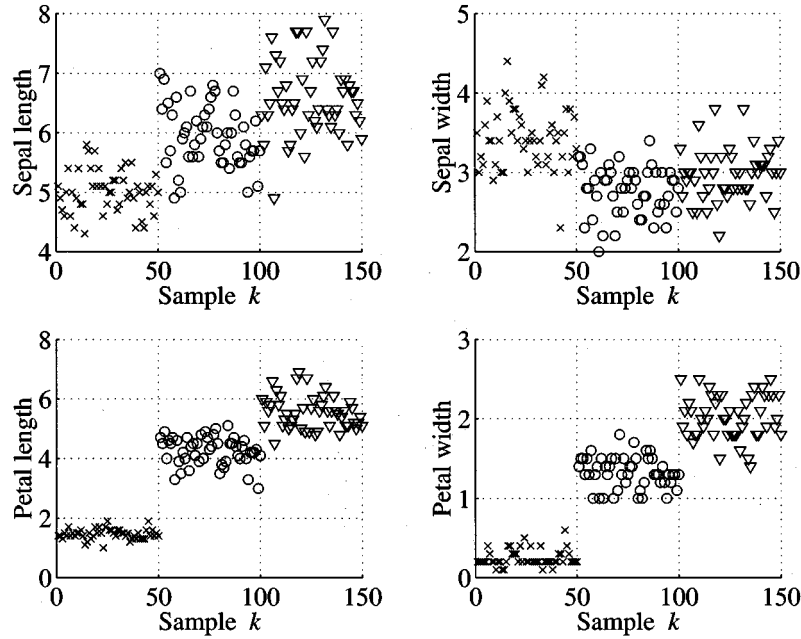


Fig. 9. Iris data: (1) *Iris setosa* ×, (2) *Iris versicolor* o, and (3) *Iris virginica* ∇.

One of the compact and accurate fuzzy models is given by Shi et al. [21] who used a GA with integer coding to learn a Mamdani type fuzzy model. Starting with three fuzzy sets associated with each feature, the membership function shapes and types, and the fuzzy rule set, including the number of rules, were evolved using a GA. Furthermore, a fuzzy expert system was used to adapt the GA's learning parameters. After several trials with varying learning options, a four-rule model was obtained, which gave three errors in learning the data. Recently, Russo [11] applied a hybrid GA+NN approach to learn fuzzy models. He present a five-rule fuzzy model with 18 fuzzy sets and 0 misclassifications. In [1] we presented a two-rule (8 fuzzy sets) and a three-rule fuzzy model (12 fuzzy sets), both having one misclassification only.

We apply the proposed modeling schemes (Fig. 4). The fuzzy  $c$ -means clustering was applied to obtain an initial TS model with singleton consequents. In order to perform classification, the output  $y_k$  of the TS model was used with the following classification rule:

$$c_k = \begin{cases} 1, & \text{if } 0.5 < y_k \leq 1.5 \\ 2, & \text{if } 1.5 < y_k \leq 2.5 \\ 3, & \text{if } 2.5 < y_k \leq 3.5. \end{cases} \quad (14)$$

The classes are for transparency reasons limited to a certain model-output range. An initial model with three rules was constructed from clustering where each rule described a class (singleton consequents). The classification accuracy of the initial model was rather discouraging, giving 33 misclassifications on the training data. The rule antecedents sets are shown in Fig. 10 and the estimated rule consequents were  $\{1.00, 2.10, 2.95\}$ , which is close to the class labels as expected. These are changed for transparency reasons into  $\{1, 2, 3\}$ , before further optimization.

The GA was applied with the same parameter as in Section IV-A. The weight  $\sigma$  in the objective (9) was 1 and the threshold  $\lambda$  for the redundancy searches is 0 for scheme 1 and

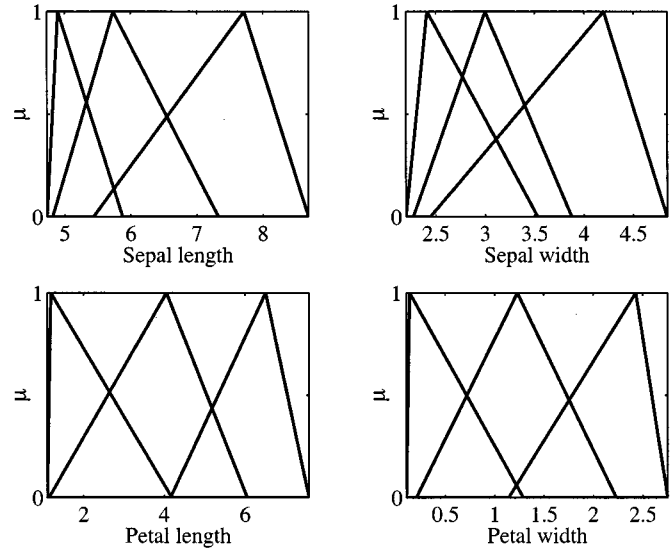


Fig. 10. Initial fuzzy rule based model with three rules and 33 misclassifications.

0.5 for scheme 2, and  $\lambda = -0.5$  for the final optimization step. The search-space parameters  $\alpha_1$  and  $\alpha_2$  are varied and given in Table II. Note, that if  $\alpha_2$  is 0 then the consequent values are fixed.

First, scheme 1 (Fig. 4) was applied. We consider the complete data set, i.e., there is not a separate test and evaluation data set. This is mentioned in the manuscript. It is done for reasons of simplicity and comparison with other methods as this is also the approach in some of the compared approaches. The result is expected to be similar to the leave-one-out or resubstitution error, which needs many repetitions for an accurate average result and highly depends on the chosen samples. Generally, for more complex problems, or problems with a bad data distribution, one should apply bootstrap methods to avoid overtraining.

TABLE II

FUZZY RULE-BASED CLASSIFIERS FOR THE IRIS DATA DERIVED BY MEANS OF SCHEME 1 (A, B, C) AND SCHEME 2 (D, E, F). RULES GIVES THE NUMBER OF SETS FOR EACH INPUT AND MISCLASSIFICATIONS THE PERFORMANCE, AFTER EACH RULE-BASE SIMPLIFICATION STEP

No.	$\alpha_1$	$\alpha_2$	$\lambda$	Rules	Misclassifications
A	0.25	0.25	0/-0.5	{3333}	{4}
B	0.25	0	0/-0.5	{3333}	{3}
C	0.5	0	0/-0.5	{3333}	{2}
D	0.25	0.25	0.5/-0.5	{2133}, {2032}, {0032}	{4}, {3}, {4}
E	0.25	0	0.5/-0.5	{2233}, {2032}, {0032}	{3}, {3}, {4}
F	0.5	0	0.5/-0.5	{2032}, {0032}	{1}, {3}, {4}

The results for three typical runs with different parameters are presented in Table II (A, B, C). The number of misclassification is quickly reduced to 3 or 4. The obtained model is accurate and is suitable for interpretation since the rules consequents are the same or close to the actual class labels, such that each rule can be taken to describe a class. The fuzzy sets of the optimized model B are shown in Fig. 11. The corresponding rules are

- $R_1$ : **If**  $x_1$  is *short* **and**  $x_2$  is *wide* **and**  $x_3$  is *short* **and**  $x_4$  is *narrow*, **then** class is 1
- $R_2$ : **If**  $x_1$  is *medium* **and**  $x_2$  is *narrow* **and**  $x_3$  is *medium* **and**  $x_4$  is *medium*, **then** class is 2
- $R_3$ : **If**  $x_1$  is *long* **and**  $x_2$  is *medium* **and**  $x_3$  is *long* **and**  $x_4$  is *wide*, **then** class is 3. (15)

Second, scheme 2 was applied. The results for three typical runs with different parameters are presented in Table II (D, E, F). The number of misclassification is quickly reduced to 1, 3 or 4. One intermediate model had 1 misclassification only but was not very transparent due to overlapping sets, however, resulted in a perfect rule-interpolation which shows the good optimization property of the GA. The rule base reduction is one or two times applied and subsequently, the model is optimized for transparency. The resulting models are highly reduced, while the misclassification error is not much increased. The fuzzy sets of the optimized model E are shown in Fig. 12. The corresponding rules are

- $R_1$ : **If**  $x_3$  is *short* **and**  $x_4$  is *narrow*, **then** class is 1
- $R_2$ : **If**  $x_3$  is *medium* **and**  $x_4$  is *narrow*, **then** class is 2
- $R_3$ : **If**  $x_3$  is *long* **and**  $x_4$  is *wide*, **then** class is 3. (16)

The proposed iterative reduction scheme removed 7 sets from the three-rule model, thereby, removing two inputs. By comparing the reduced fuzzy model with the data in Fig. 9, one observes that the inputs with the highest information content are maintained. In addition, it is noted that most models from research show three or more misclassifications, indicating that there are a few samples that are difficult to classify correctly. These are often close to the decision curve; e.g., in [1] we showed that our 1 misclassification had the typical properties of another class and may be an outlier in the data. A good method to describe the plausibility of the classification results is to use a certainty factor that describes the expectation that a data-pair belongs to the different classes [25]. This certainty factor can

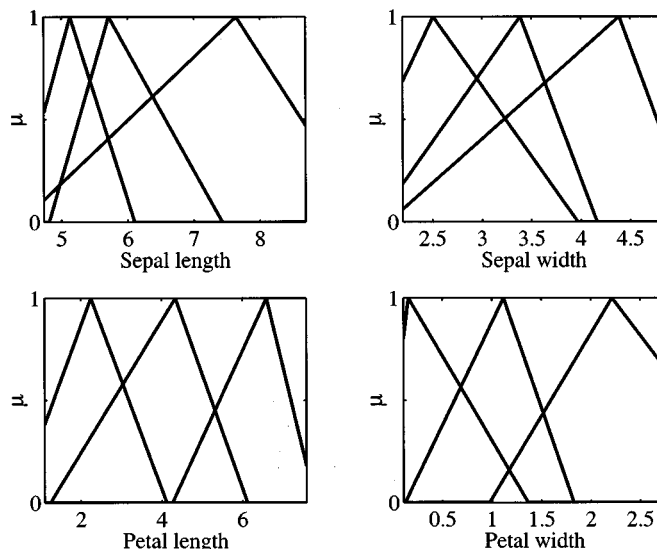


Fig. 11. Optimized fuzzy rule based model with three rules and three misclassifications (Table II-B).

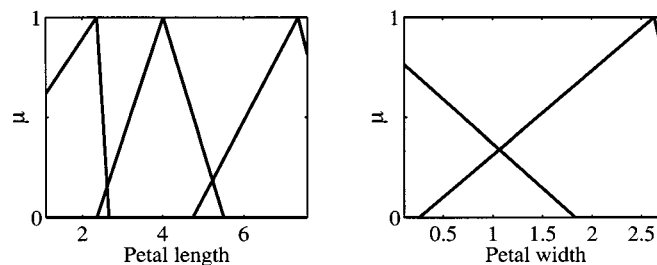


Fig. 12. Optimized and reduced fuzzy rule based model with three rules and four misclassifications (Table II-E).

also be aggregated in the multi-objective model optimization to improve the fuzzy classifier.

The results obtained with the proposed modeling approach for the Iris data case illustrate the power of the GA for optimizing fuzzy rule based classifiers. By simultaneously optimizing the antecedent and/or consequent parts of the rules, according scheme 1, the GA found an optimum for the model parameters in the neighborhood of the initializations, which gave drastic improvements in the classification performance. Moreover, compact fuzzy models with a low amount of inputs and fuzzy sets were obtained by the proposed model reduction scheme 2.

V. CONCLUSION

We have presented an approach to construct compact and transparent, yet accurate fuzzy rule based models and classifiers from measured input-output data based on iterative complexity reduction. Methods for modeling, complexity reduction, and optimization are combined in the approach. Fuzzy clustering is first used to obtain an initial rule based model. Similarity based simplification and multi-objective GA-based optimization are then used in an iterative manner to decrease the complexity of the model while maintaining high accuracy. The proposed algorithm was successfully applied to two problems known from the literature. The accuracy of the obtained models was comparable



to the results reported in the literature, however, the obtained models use fewer rules and less fuzzy sets than other models reported in the literature. Compact and transparent, yet accurate fuzzy TS models are the results.

#### ACKNOWLEDGMENT

The authors are grateful to Dr. Liang Wang, (co)author of references [20], [19], [18], for unconditionally sharing his data for this research.

#### REFERENCES

- [1] M. Setnes and J. A. Roubos, "GA-fuzzy modeling and classification: Complexity and performance," *IEEE Trans. Fuzzy Syst.*, vol. 8, pp. 509–522, 2000.
- [2] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.
- [3] R. Babuška, *Fuzzy Modeling for Control*. Norwell, MA: Kluwer, 1998.
- [4] R. Isermann, "On fuzzy logic applications for automatic control, supervision, and fault diagnosis," *IEEE Trans. Syst., Man, Cybern. A*, vol. 28, pp. 221–235, 1998.
- [5] B. Kosko, "Fuzzy systems as universal approximators," *IEEE Trans. Comput.*, vol. 28, no. 1, pp. 1329–1333, 1994.
- [6] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," *IEEE Trans. Fuzzy Syst.*, vol. 1, pp. 7–31, 1993.
- [7] M. Setnes, R. Babuška, and H. B. Verbruggen, "Rule-based modeling: Precision and transparency," *IEEE Trans. Syst., Man, Cybern. C*, vol. 28, pp. 165–169, 1998.
- [8] J. Valente de Oliveira, "Semantic constraints for membership function optimization," *IEEE Trans. Fuzzy Syst.*, vol. 19, pp. 128–138, 1999.
- [9] O. Cordon and F. Herrera, "A proposal for improving the accuracy of linguistic modeling," *IEEE Trans. Fuzzy Syst.*, vol. 8, pp. 335–344, 2000.
- [10] J. G. Martin-Blázquez, "From approximative to descriptive models," in *9th IEEE Int. Conf. of Fuzzy Systems*, San Antonio, Texas, USA, May 7–10, 2000, pp. 829–834. IEEE.
- [11] M. Russo, "Genetic fuzzy learning," *IEEE Trans. on Evolutionary Computation*, vol. 4, pp. 259–273, 2000.
- [12] M. Setnes, R. Babuška, U. Kaymak, and H. R. van Nauta Lemke, "Similarity measures in fuzzy rule base simplification," *IEEE Trans. Syst., Man, Cybern. B*, vol. 28, no. 3, pp. 376–386, 1998.
- [13] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modeling and control," *IEEE Trans. Syst., Man, Cybern.*, vol. 15, pp. 116–132, 1985.
- [14] J. Yen, L. Wang, and C. W. Gillespie, "Improving the interpretability of TSK fuzzy models by combining global learning with local learning," *IEEE Trans. Fuzzy Syst.*, vol. 6, pp. 530–537, 1998.
- [15] M. Setnes, R. Babuška, and H. B. Verbruggen, "Rule-based modeling: Precision and transparency," *IEEE Trans. Syst., Man, Cybern. C*, vol. 28, no. 1, pp. 165–169, 1998.
- [16] J. A. Roubos and M. Setnes, "Compact fuzzy models through complexity reduction and evolutionary optimization," in *9th IEEE Int. Conf. on Fuzzy Systems*, San Antonio, Texas, USA, May 7–10, 2000, pp. 762–767.
- [17] Z. Michalewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, 2nd ed. New York: Springer Verlag, 1994.
- [18] J. Yen and L. Wang, "Application of statistical information criteria for optimal fuzzy model construction," *IEEE Trans. Fuzzy Syst.*, vol. 6, no. 3, pp. 362–371, 1998.
- [19] L. Wang and J. Yen, "Extracting fuzzy rules for system modeling using a hybrid of genetic algorithms and Kalman filter," *Fuzzy Sets Syst.*, vol. 101, pp. 353–362, 1999.
- [20] J. Yen and L. Wang, "Simplifying fuzzy rule-based models using orthogonal transformation methods," *IEEE Trans. Syst., Man, Cybern. B*, vol. 29, no. 1, pp. 13–24, 1999.
- [21] Y. Shi, R. Eberhart, and Y. Chen, "Implementation of evolutionary fuzzy systems," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 2, pp. 109–119, 1999.
- [22] H. Ishibuchi and T. Nakashima, "Voting in fuzzy rule-based systems for pattern classification problems," *Fuzzy Sets Syst.*, vol. 103, pp. 223–238, 1999.
- [23] E. Anderson, "The Irises of the Gaspé peninsula," *Bulletin American Iris Society*, vol. 59, pp. 2–5, 1935.
- [24] J. C. Bezdek, J. M. Keller, R. Krishnapuram, L. I. Kuncheva, and N. R. Pal, "Will the real Iris data please stand up?," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 3, pp. 368–369, 1999.
- [25] J. A. Roubos, M. Setnes, and J. Abonyi, "Learning fuzzy classification rules from labeled data," in *International Journal of Information Sciences*, 2001, to be published.



**Hans Roubos** (S'97) was born in 1973, Sprang-Capelle, The Netherlands. He received the M.Sc. degree in bioprocess engineering from Wageningen University, The Netherlands, in 1997 and is currently working on the Ph.D. degree.

He was with the European FAMIMO project at the Systems Control Engineering Group, Delft University of Technology, The Netherlands, working on MIMO fuzzy model-based predictive control for eight months. In December, 1997, he started his current Ph.D. work on model-based process development of fed-batch fermentation processes. His research interests include genetic algorithms and fuzzy systems for modeling, identification and control, hybrid models and biotechnological processes.



**Magne Setnes** (S'94–M'95) was born in 1970, Bergen, Norway. He received the B.Sc. degree in robotics from the Kongsberg College of Engineering, Norway, in 1992. He received the M.Sc. degree in electrical engineering from the Delft University of Technology, The Netherlands, the Degree of Chartered Designer in Information Technology, and the Ph.D. degree from the Systems and Control Engineering Group in 1995, 1997 and 2001, respectively.

Currently, he is with Heineken Technical Services, Research and Development, the Netherlands. His interests include fuzzy systems, computational intelligence techniques for modeling, control and decision making.