15th Triennial World Congress, Barcelona, Spain

FUZZY CLUSTERING ALGORITHM FOR LOCAL MODEL CONTROL

José Luis Díez, Antonio Sala, José Luis Navarro [1]

Departamento de Ingeniería de Sistemas y Automática
Universidad Politécnica de Valencia; Camino de Vera, 14;
46022 Valencia, Spain. Email:{jldiez,asala,joseluis}@isa.upv.es

Abstract:
Fuzzy modelling has interpretability of the obtained models as a fundamental goal. In this paper a control-oriented local-model fuzzy clustering algorithm will try that local models approximate the linearized plant model on their validity zones. A family of clustering algorithms is presented so that it incorporates some desirable characteristics regarding convexity and smoothness of the final identified clusters, with advantages regarding other methodologies such as Gustaffson-Kessel. The algorithm simultaneously provides local linear models and input clustering, being suitable for Takagi-Sugeno models and local linear models decomposition of complex systems. *Copyright ©2002 IFAC*

Keywords: fuzzy systems, identification, fuzzy clustering, local model control, neurofuzzy systems

1. INTRODUCTION

A suitable modelling and identification of a system is essential for a successful subsequent control design, and the same applies to supervision and fault detection systems (Aström and Wittenmark, 1997). However, classic modelling techniques are not applicable to complex systems such as those in biochemical and chemical engineering, social models, aerospace, etc. Many of these systems are characterised by noisy data and significant nonlinearities.

Artificial intelligence (AI) and, in particular, fuzzy logic (FL) are widely used techniques to deal with complex systems because of its universal function approximation (UFA) capabilities (Wang, 1997) and the parallelism to human reasoning processes. *Fuzzy models* try to incorporate additional qualitative or imprecise information that engineers or operators have about systems.

The main problem in identifying fuzzy models from data appears when *structure identification* is performed (Sugeno and Yasukawa, 1993). Subsequently, *parameter identification* must be carried out. This second identification step can be easily done, for example, by least mean squares (Babuska, 1996) if the system is linear in parameters or by gradient or genetic techniques in other cases (Wang, 1997).

Some *rule extraction* tasks can be given or provided by an expert but, in general, the more tasks to perform in structure identification, the more difficult the identification becomes. There exist a number of methods for solving rule extraction problem (Wang, 1997) based on genetic algorithms, neural networks, templates or clustering techniques. In this paper, emphasis will be made on control-oriented rule extraction using **fuzzy clustering** (Sugeno and Yasukawa, 1993; Emami *et al.*, 1998; Babuska, 1996). The objective of clustering is to partition a data set into a reduced number of clusters. Results are given by assigning a membership function value for each cluster to each data point.

[1] Partially supported by project GV00-100-14, regional government and the authors' institution.

The structure of the contribution is as follows. An identification approach for fuzzy identification for local models control is chosen in section 2, after highlighting some misunderstandings in literature. Section 3 analyzes fuzzy clustering for local-models identification, providing the main characteristics this clustering algorithm should have in control context. Local model control oriented fuzzy clustering algorithms are shown in section 4 and a more suitable family of clustering algorithms is presented in section 5. Its capabilities are tested in two identification examples in section 6.

2. MODELLING CRITERIA FOR LOCAL-MODEL CONTROL

Let us have a nonlinear discrete-time dynamic system:

$$x_{k+1} = F(x_k, u_k) + e(k) \tag{1}$$

where F arguments are measurable and $e(k)$ represents noise and unmodelled dynamics.

The general system (1) has a form $Y = F(z) + e(t)$, where z aggregates state and input variables, e(t) is a noise term and Y is the predicted value (usually x_{k+1}). In the spirit of fuzzy models, a state representation with *physical sense* would be preferred instead of a collection of past input-output values.

An *approximation function* $f()$ to the real system, is an approximation of $F()$ calculated by identification techniques from a set of n-dimensional input-output data with N elements $Z = \{w_k = (z_k, y_k), \quad k = 1, \ldots N\}$.

Two different concepts (steps) appear when using fuzzy models as $f()$: *locality* (to divide the complexity of the system into a set of local behaviours) and *interpolation* (integration of these local models). The ideas of modelling for control have recently arisen in the linear system field (**?**), and those approaches should be extended to fuzzy identification techniques because the use of fuzzy models to design control systems may modify the criteria used in identification for prediction. Two main fuzzy identification approaches can be thought of. The first one consist of a minimization of a "global" prediction error (difference between y(t) and f(t)) (**?**). This **global models** approach to modelling complex systems presents an objective function to minimize defined by:

$$J_g = \sum_{k=1}^{N} (y_k - f(z_k))^2 \tag{2}$$

where $f(\cdot)$ is any UFA, such as Takagi-Sugeno (Takagi and Sugeno, 1985) fuzzy models, radial basis networks, etc. Many of them can be expressed as

$$f(z_k) = \sum_{i=1}^{c} \mu_i(z_k) f_i(z_k, \beta_i) \tag{3}$$

where c represents the number of regions in which the operating space has been divided, β_i are parameter vectors of "local" model f_i, and μ_i is a validity (membership) function.

On the following, the data $\mu_i(z_k)$ will be arranged in an array μ_{ik} of dimensions $c \times N$ and they will be forced to verify the conditions for a *fuzzy partition*:

$$\mu_{ik} \in [0,1], \quad \sum_{i=1}^{c} (\mu_{ik}) = 1 \tag{4}$$

Global objective function system identification has the advantage of easy training of the model (if f_i functions are linear in parameters, least squares can be used) and good accuracy. Regarding control design, techniques such as LMI (Tanaka and Sugeno, 1992), gain-scheduled control (Hunt and Johansen, 1997), generalised minimum variance controllers networks (Díez and Previdi, 2001), etc. can be used.

However, the widespread in literature (Hunt and Johansen, 1997) name of "local model" to f_i in (3) may have no particular meaning: the identified f_i may not correspond to any local behaviour, *i.e.*, the global behaviour is explained by the convex aggregation of the f_i but its meaning is unrelated to local models (such as linearized ones) or other distinctive characteristics of the underlying system unless fuzzy set overlapping is very small. Another issue on those models regards conditioning when calculating memberships μ_i and f_i: if sufficient freedom were available to the μ_i, they would be able to approximate a global model for arbitrarily fixed f_i.

The alternate option for modelling complex systems is to partition the operating regime space into a number of sets (either used-defined or generated in the identification process), and minimize the "local" error of a number of **local models** that represent the system in a region of its operating space (Babuska, 1996) and approximate the plant behaviour on the given set. An objective function to minimize in this case is:

$$J_i^l = \sum_{k=1}^{N} \mu_{ik} (y_k - f_i(z_k))^2, \quad i = 1, ..., c \tag{5}$$

where μ_{ik} verify conditions (4). In this case, $f_i(z_k)$ has to be adjusted to the data, weighting the adjustment proportionally to membership values. With narrow membership functions, the f_i minimizing the index in (5) will truly approach the local models around a set of prototype points (z_i^*, y_i^*) which $\mu_{ik}^* = 1$. To get global models from (5), alternative convex interpolation must be used (Babuska *et al.*, 1996). The main disadvantage of this kind of systems is that local designs can not ensure global stability with identified interpolation functions μ_{ik}, unless appropriate gain-scheduling techniques (Rugh and Shamma, 2000) are used for control design.

To summarize, the *global* approach (3) may achieve better accuracy between real and estimated output of the process, while the *local* one (5) improves readability of the identified model. In identification with AI techniques, this trade-off between readability and accuracy must be taken into account, and the use of the concepts (fuzzy sets) and the models originated depends on which of the two approaches is taken. If the objective is the application to local linear control design or the models are designed for human interface, user (control engineer) interpretability is fundamental and local modelling criteria might be more suitable.

3. IDENTIFICATION OF FUZZY MODELS WITH FUZZY CLUSTERING

As previuosly outlined, the objective of clustering is to partition a finite data set $Z = \{z_1, z_2, \ldots, z_N\}$ into c clusters, giving as a results a $c \times N$ matrix $U = [\mu_{ik}]$. To express those results as fuzzy rules, antecedent calculations must be done. The easiest solution is to project the obtained membership functions to one-dimensional clusters and adjust the shape to a proper predefined membership function. Another option is to keep the obtained membership function in the n-dimensional data space. In any case, the lower the modelling error usually leads to a lower degree of rule interpretability. This section reviews the most common fuzzy clustering methodologies.

3.1 *Fixed distance algorithms*

Fuzzy C-Means (FCM) is the mainly used algorithm in this class. These *distance-based* algorithms minimize a *c-means* objective function (Bezdek, 1987) or any of its modifications. The FCM functional is usally formulated as:

$$J(Z;U,C) = \sum_{i=1}^{c} \sum_{k=1}^{N} (\mu_{ik})^m D_{ik}^2 \qquad (6)$$

where $C = [c_1, c_2, \ldots, c_c]$, $c_i \in R^n$ are the cluster centers (prototypes) to be determined,

$$D_{ik}^2 = (w_k - c_i)^T B (w_k - c_i) \qquad (7)$$

is a distance (norm) defined by B, and $m \in (1, \infty)$ determines the fuzziness of the resulting clusters.

Minimization of (6) is usually carried out in two independent steps:

- minimize for U, with fixed C by constrained minimization (must add one),
- minimize for C, with fixed U by unconstrained minimization.

Combination of both conditions makes (6) descend towards a (maybe local) minimum (Bezdek, 1987). FCM drawbacks (all identified clusters have similar size and orientation, with hyperellipsoid shape) can be avoided by using extensions to the algorithm, such as those presented in next sections, to make its results better suited to control-relevant problems.

3.2 *Adaptive distance algorithms*

The Gustafson-Kessel (GK) algorithm (Gustafson and Kessel, 1979) is the FCM extension most used in identification, and allows a different norm B_i for each cluster:

$$D_{ikB_i}^2 = (w_k - c_i)^T B_i (w_k - c_i) \qquad (8)$$

The procedure adds a third minimization step for B_i with constant U, C under the constraint of constant cluster volume ($\det(B_i) = \rho_i$, being ρ_i user-defined parameters).

GK algorithm detects quasi-linear behaviours of existing operating regimes quite correctly. However, the hyperellipsoidal clusters have to be adjusted to linear structures eliminating the least significant eigenvalue and eigenvector.

GK has interesting properties that make it an appropriate algorithm for identification (Babuska, 1996) but in the usual case of lack of information about clusters volumes, all initial values are the same and final detected clusters cannot have very big differences in size. Additionally, when only a small number of data are available, noise free data are presented or when data are linearly dependent, numeric problems could appear because the cluster covariance matrix becomes almost singular.

Finally, it must be emphasized that a common drawback of all methods for identification of fuzzy models via clustering is that they forget their final goal (the fuzzy model) while grouping data. Derivation of rules is always a step to be done after clustering. Additionally, in order to determine rule consequents, detected clusters have to be adjusted to linear structures if a TS fuzzy model is needed (for example, by least mean squares). All these problems suggest some improvements in clustering techniques:

- Include modelling error information in the iterative clustering algorithm.
- Enhance rule interpretability, for example by favouring convex cluster shapes.

The objective of this paper is to find a clustering algorithm in which local linear models are obtained in the clustering process and a balance criteria between interpretability and modelling error can be done.

4. LOCAL MODEL CONTROL ORIENTED FUZZY CLUSTERING ALGORITHMS

4.1 *Linear prototypes algorithms*

FCM limitation of shape and dimensions of the clusters is solved in this case by keeping a constant norm but defining r-dimensional prototypes $(0 \le r \le N)$, linear or non-linear, in the input data subspace of dimension N. Algorithms of our interest and based on this approach are those referred to linear spaces: FCV fuzzy c-varieties algorithm (Bezdek, 1987) and FCRM fuzzy c-regression models algorithm (Hathaway and Bezdek, 1993). The most interesting linear prototypes algorithm, in our case, is FCRM, because objective function to minimize is based on substitution of D_{ik} of (6) by $E_{ik}(\beta_i) = (y_k - f_i(z_k;\beta_i))^2$, and the resulting index to minimize is the same presented in (5).

Although its philosophy is good (obtains local linear models and includes modelling error in clustering process), FCRM algorithm sometimes does not achieve the desired objectives due to spurious local minima originated by the excessive degrees of freedom in the membership functions and local model parameters, resulting in error minimized by different models ponderation instead of local model closeness to data.

4.2 *Mixed prototypes algorithms*

The problem of FCM limitation of cluster shape and dimension can be handled in a third way apart from variable norms and linear prototypes, by mixing both approaches. These algorithms are fuzzy c-elliptotypes (FCE) and adaptive fuzzy c-regression models (AFCR).

AFCR is based on (**?**), where it is shown that if D_{ik}^2 in FCM is substituted by a convex combination of distances in a generic criteria $D_{a\,ik}$, then a Û strict local minimum of J can be calculated. AFCR objective function (9) can be calculated by D_{ik} substitution in equation (6) for a new distance (10) defined as a combination of FCRM (E_{ik}) and FCM (D_{ik}) distances with $\alpha \in [0,1]$.

$$J_{a\,ik} = \sum_{i=1}^{c}\sum_{k=1}^{N}(\mu_{ik})^m D_{a\,ik} \quad (9)$$

$$D_{a\,ik} = \alpha_k E_{ik}(\beta_i) + (1-\alpha_k)\eta D_{ik}^2 \quad (10)$$

First term of (10) provides the same criterion as FCRM and second term increases partition capability in input space by taking into account data distance from clusters prototypes. α selection is dynamically determined by the algorithm and is as closer to 1 as cluster structure becomes more linear. Calculation of this parameter is based on:

$$\alpha_k = 1 - \frac{\min_l\{\lambda_{kl}\}}{\max_l\{\lambda_{kl}\}},\ k = 1,2\ldots,c \quad (11)$$

where λ_{kl} are eigenvalues of cluster k covariance (used, for example, in GK). In this way, a single algorithm includes advantages of FCRM and GK. Parameter η is used for balance between criterion terms when mean size is very different and there is no studies about its determination. In general (AFCR and GK), results are not very good if attention is focused on membership functions, because interpretability is lost in exchange of a lower modelling error.

5. AN ALGORITHM IMPROVING READABILITY

In this paper, a modification of the objective function of AFCR linear clustering algorithm including appropriate criteria in it for μ convexity assurance will be presented, with improvements over original behaviour towards a better performance and membership function interpretability. Of course those additions are at the expense of a maybe higher modelling error if number of clusters is preserved.

Two new additional terms are going to be added to performance index defined in equation (9): a first term to penalise high membership of points far from a prototype ($J_{far\,ik}$), and a second term to penalise low membership of points near to a prototype ($J_{near\,ik}$):

$$J_{far\,ik} = \sum_{i=1}^{c}\sum_{k=1}^{N}\mu_{ik}^m[1 - exp(-\frac{L_{ik}^2}{2\sigma_1^2})] \quad (12)$$

$$J_{near\,ik} = \sum_{i=1}^{c}\sum_{k=1}^{N}(1-\mu_{ik})^m[exp(-\frac{L_{ik}^2}{2\sigma_2^2})] \quad (13)$$

$$L_{ik}^2 = (z_k - p_i)^T B(z_k - p_i) \quad (14)$$

where distance L_{ik} and prototypes p_i are measured on *input* space. In both cases, Gaussian decay functions are used although other monotonic functions could be though of. Then, minimisation of new performance index:

$$J_{C\,ik} = J_{a\,ik} + \gamma_1 J_{far\,ik} + \gamma_2 J_{near\,ik} \quad (15)$$

leads to a solution biased towards smoother and more convex membership function shapes.

Minimisation of $J_{C\,ik}$ can be done following the original FCM-AFCR procedures with modified update expressions for prototypes and memberships at each iteration step.

New parameters γ_1 and γ_2 are included in $J_{far\,ik}$ and $J_{near\,ik}$, respectively, with the purpose of balancing the diverse cost index terms.

This new algorithm is called **AFCRC** (Adaptive Fuzzy C-Regression models with Convexity enhancement). Obtained membership functions are more interpretable and clusters are more similar to desired general structure.

5.1 *Parameters setting*

The use of normalized data, so that all variables are scaled to zero mean and unit variance, helps in preserving parameters values for different sets of data. However, to facilitate parameters tuning, its meanings have to be clarified:

- σ_1 represents approximate maximum allowed cluster size,
- σ_2 corresponds to approximate minimum cluster size,
- η provides a balance between cluster size and modelling error at startup,
- γ_1 and γ_2 weight the new index terms against AFCR ones

As γ_1 and γ_2 represent relative importance of interpretability versus modelling error at the final iteration steps of the algorithm, determination of them should be user supervised. A possible procedure is to run AFCR algorithm to determine the lowest feasible modelling error and increase γ_1 and γ_2 until an affordable rise in modelling error is achieved (for example, 10%) enhancing interpretability. The noisier the data the bigger those parameters should be for acceptable membership functions shape.

6. CASE STUDIES

6.1 *Nonlinear static function identification*

Suggested clustering algorithms have been tested in the identification from data of growth rate kinetics of a simulated bioreactor. Results are shown in the following figures for GK (figure 1) and AFCRC (figure 2).

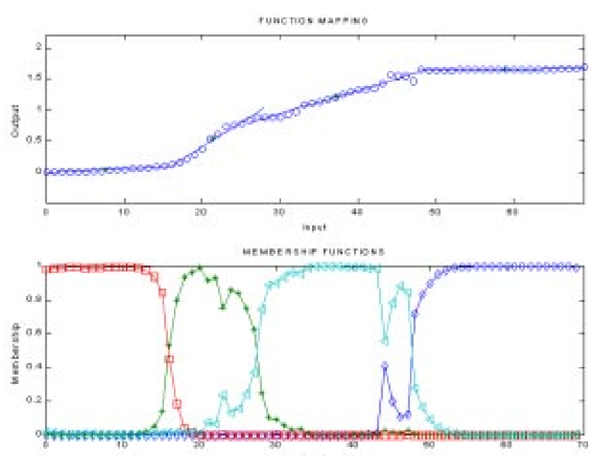

Fig. 1. GK detection of real data structure (circle: experimental data, line: identified local model).

Results shown in the last figure overcomes other tested approaches in the membership functions convexity sense, while local linear models detection is maintained and its parameters are calculated in the clustering process taking into account modelling error.

6.2 *Nonlinear dynamic function identification*

A simulator of a nonlinear dynamic system has been built, based on the longitudinal vehicle dynamics presented in (Hunt and Johansen, 1997), in order to

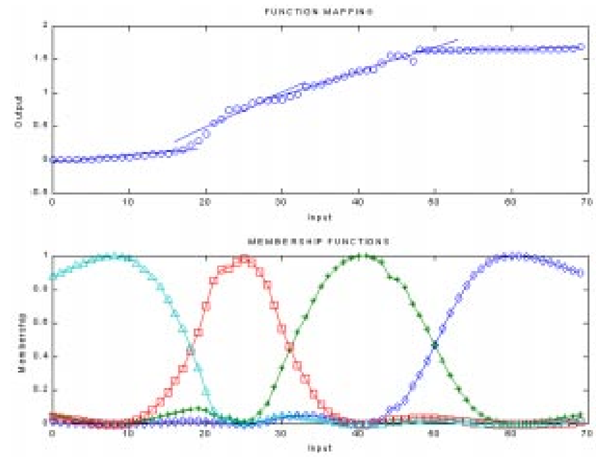

Fig. 2. AFCRC detection of real data structure (circle: experimental data, line: identified local model).

test AFCRC algorithm identification capabilities for nonlinear dynamic systems. Simplified vehicle model consists of a single control signal u combining throttle angle and brake pressure, and vehicle longitudinal speed y as model output.

The model consists of three operating regimes, each corresponding to a certain range around an operating point (u^*, y^*). Defined local models are presented in table 1.

Table 1. Local models.

Regime	Op. point	Transfer function	Offset
1	(0.1,3.5)	$\frac{0.3z^{-1}}{1-0.98z^{-1}}$	0.04
2	(0.5,15)	$\frac{1.8z^{-1}}{1-0.94z^{-1}}$	0.0
3	(0.9,26.5)	$\frac{0.6z^{-1}}{1-0.96z^{-1}}$	0.52

Relatively small variations in parameters between two models makes operating mode separation harder but improves subsequent switching between designed controllers. The surface to identify corresponds to the nonlinear discrete dynamic system, with trapezoidal validity functions f_1, f_2 and f_3, is (16):

$$\begin{aligned} y_{k+1} = {} & f_1(0.98y_k + 0.3u_k + 0.04) + \\ & + f_2(0.94y_k + 1.8u_k + 0.0) + \\ & + f_3(0.96y_k + 0.6u_k + 0.52) \end{aligned} \quad (16)$$

AFCRC algorithm leads, following the performance index evolution shown in figure 3, to identified local models presented in table 2.

Table 2. AFCRC identified local models.

Regime	Op. point	Transfer function	Offset
1	(0.1004,3.1084)	$\frac{0.2583z^{-1}}{1-0.9852z^{-1}}$	0.0291
2	(0.5035,14.7769)	$\frac{1.7313z^{-1}}{1-0.9420z^{-1}}$	0.0044
3	(0.9030,26.2319)	$\frac{0.5473z^{-1}}{1-0.9647z^{-1}}$	0.4436

AFCRC results are better than those obtained by GK algorithm (convergence with GK required many runs

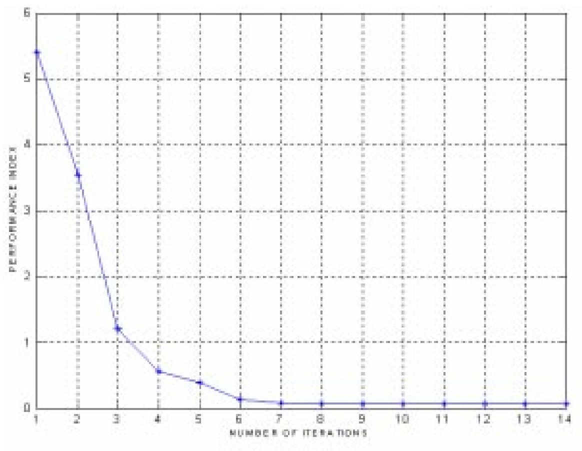

Fig. 3. Evolution of performance index in AFCRC algortihm.

with different initilization parameters) and presented for comparison in table 3.

Table 3. GK identified local models.

Regime	Op. point	Transfer function	Offset
1	(0.0998,3.0824)	$\frac{0.4402z^{-1}}{1-0.9739z^{-1}}$	0.0433
2	(0.4993,14.6181)	$\frac{2.4549z^{-1}}{1-0.9174z^{-1}}$	0.0038
3	(0.8993,26.0735)	$\frac{0.4781z^{-1}}{1-0.9623z^{-1}}$	0.5723

7. CONCLUSIONS

In this paper, after defining proper criteria for fuzzy identification, fuzzy clustering approaches for local-model identification have been analysed, providing the main characteristics a clustering algorithm for identifying local linearized fuzzy models should have in the control context. A new family of fuzzy clustering algorithms that overcomes some problems present in current ones has been presented, enhancing the convexity and smoothness (*i.e.*, the readability) of the obtained clusters.

8. REFERENCES

Aström, K.J. and B. Wittenmark (1997). *Computer controlled systems : theory and design*. Prentice-Hall, Information and System Sciences series. New Jersey.

Babuska, R. (1996). *Fuzzy Modeling and Identification*. PhD dissertation, Delft University of Technology. Delf, The Netherlands.

Babuska, R., C. Fantuzzi and H. B. Verbruggen (1996). Improved inference for takagi-sugeno models. *Proc. IEEE Conf. on Fuzzy Systems*.

Bezdek, J.C. (1987). *Pattern recognition with Fuzzy Objective Function Algorithms*. Plenum Press. New York.

Díez, J. L. and F. Previdi (2001). Guaranteed closed loop precision in multiple model based control. *Proc. IFAC Workshop on Adaptation and Learning in Control and Signal Processing* pp. 139–144.

Emami, M. R., I.B. Türksen and A.A. Goldenberg (1998). Development of a systematic methodology of fuzzy logic modeling. *Transactions on Fuzzy Systems* **6**, 346–366.

Gustafson, E.E. and W.C. Kessel (1979). Fuzzy clustering with a fuzzy covariance matrix. *Proc. of IEEE CDC, San Diego, California* pp. 761–766.

Hathaway, R. J. and J.C. Bezdek (1993). Switching regression models and fuzzy clustering. *Transactions on Fuzzy Systems* **1**, 195–204.

Hunt, K.J. and T.A. Johansen (1997). Design and analysis of gain-scheduled control using local controller networks. *International Journal of Control* **66**, 619–651.

Rugh, W. J. and J. S. Shamma (2000). Research on gain scheduling. *Automatica* **36**, 1401–1425.

Sugeno, M. and T. Yasukawa (1993). A fuzzy-logic-based approach to qualitative modeling. *Transactions on Fuzzy Systems* **1**, 7–31.

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

Tanaka, K. and M. Sugeno (1992). Stability analysis and design of fuzzy control systems. *Fuzzy Sets and Systems* **45**, 135–156.

Wang, L.-X. (1997). *A Course in Fuzzy Systems and Control*. Prentice-Hall. New Jersey.