Applied Soft Computing 17 (2014) 67–78

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Hybrid-fuzzy modeling and identification

Alfredo Núñez[a,*], Bart De Schutter[b], Doris Sáez[c], Igor Škrjanc[d]

[a] Section of Road and Railway Engineering, Delft University of Technology, Stevinweg 1, 2628 CN Delft, The Netherlands
[b] Delft Center for Systems and Control, Delft University of Technology, Mekelweg 2, 2628 CD Delft, The Netherlands
[c] Electrical Engineering Department, University of Chile, Av. Tupper, 2007 Santiago, Chile
[d] Faculty of Electrical and Computer Engineering, University of Ljubljana, Tržaška 25, 1000 Ljubljana, Slovenia

ARTICLE INFO

Article history:
Received 26 May 2012
Received in revised form
18 November 2013
Accepted 22 December 2013
Available online 8 January 2014

Keywords:
Fuzzy identification
Hybrid systems
Fuzzy clustering
Principal component analysis
Switching regression models

ABSTRACT

In this paper a class of hybrid-fuzzy models is presented, where binary membership functions are used to capture the hybrid behavior. We describe a hybrid-fuzzy identification methodology for non-linear hybrid systems with mixed continuous and discrete states that uses fuzzy clustering and principal component analysis. The method first determines the hybrid characteristic of the system inspired by an inverse form of the merge method for clusters, which makes it possible to identify the unknown switching points of a process based on just input–output (I/O) data. Next, using the detected switching points, a hard partition of the I/O space is obtained. Finally, TS fuzzy models are identified as submodels for each partition. Two illustrative examples, a hybrid-tank system and a traffic model for highways, are presented to show the benefits of the proposed approach.

1. Introduction and backgrounds

Hybrid systems represent a class of dynamical systems that contain continuous and discrete/integer variables. Different types of models can be used to represent hybrid systems [1,2], for example mixed logical dynamic (MLD) models, complementarity systems, piece-wise affine (PWA) models, max–min plus scaling systems, timed or hybrid Petri-nets, differential automata, switched systems, hybrid inclusions, and real-time temporal logics, among others. Each sub-class has its own advantages over the others. For example, control techniques have been developed for MLD hybrid models, stability criteria for PWA systems, and conditions of existence and uniqueness of solution trajectories for linear complementarity systems (see [3,4] and the references within).

For non-linear systems, a broad family of identification methodologies are available, for fuzzy, neural networks, neuro-fuzzy models [5–9]. However, few methodologies consider non-linear models with continuous and discrete variables, i.e. identification of hybrid systems. In general, the identification of hybrid systems requires to solve two issues: to classify the different modes of operation (discrete behavior) and to estimate the parameters for each mode. Assuming prior knowledge about the discrete modes is not the interest of this paper, because the identification of the model parameters for each mode can be performed straightforwardly using conventional identification techniques. In the literature, the identification methods for hybrid systems mainly focus on Piecewise AutoRegressive eXogenous (PWARX) systems. In [10] an extensive comparison between some of those methods and their drawbacks is presented, and in [11] a recent and complete review of identification methods for hybrid systems (including among other topics like system description, state estimation, control, etc.) can be found. Next, some of those procedures are briefly described.

1.1. Identification methods for hybrid systems

Ferrari-Trecate et al. [12] propose a methodology for the identification of discrete-time hybrid systems in the PWA form, formulated as a discontinuous PWA map. The algorithm is based on clustering, linear identification, and pattern recognition techniques. An algebraic identification procedure to cope with the identification problem of Switched AutoRegressive eXogenous (SARX) systems was proposed by Ma and Vidal [13]. Multiple ARX models are encoded in a single polynomial expression that decouples the determination of parameters from the switching mechanism. The Bayesian procedure for the identification of Piecewise AutoRegressive eXogenous (PWARX) systems proposed by Juloski et al. [14], exploits some prior knowledge about the discrete states and parameters of the submodels. The parameters of

* Corresponding author. Tel.: +31 152789355; fax: +31 152783443.
E-mail addresses: a.a.nunezvicencio@tudelft.nl (A. Núñez), b.deschutter@tudelft.nl (B. De Schutter), dsaez@ing.uchile.cl (D. Sáez), igor.skrjanc@fe.uni-lj.si (I. Škrjanc).

1568-4946/$ – see front matter
http://dx.doi.org/10.1016/j.asoc.2013.12.011

submodels are treated as random variables, and described through their probability density functions. A bounded-error procedure was proposed by Bemporad et al. [15] in order to identify PWARX systems. The main feature of the method is to ensure that the identification error is bounded for all data points. Nakada et al. [16] address the problem of identifying PWARX systems by using statistical clustering. The method consists of clustering the measured data, while estimating the boundary hyperplane and the parameters. Gegundez et al. [17] present an identification method for PWA systems based on fuzzy clustering and competitive learning. The method estimates the number of submodels of the system, the parameters corresponding to each submodel, and the regions in the regression space. Lauer et al. [18] propose a nonlinear hybrid system identification method based on kernel functions in order to estimate arbitrary nonlinearities without prior knowledge.

1.2. Fuzzy identification

Many advances in fuzzy systems identification and applications are available in the literature [19–24], including observers [25,26] and control methods [27,28]. Nefti et al. [29] present a method for merging fuzzy sets based on clustering in the parameter space. The fuzzy sets are replaced by the most compatible prototypical fuzzy set, which is determined from a inclusion-based clustering algorithm. Hadjili and Wertz [30] propose an identification method for Takagi–Sugeno (TS) models, incorporating the selection of optimal rules and input variables. The subtractive clustering algorithm, based on compactness and the separation of clusters, is performed in order to determine the number of rules. Roubos and Setnes [31] propose a complexity-reduction algorithm based on genetic-algorithm optimization procedures to find redundancy among the rules with a criterion based on maximum accuracy and maximum set similarity. In addition, Kim et al. [32] present a combined identification method, based on the TS and Sugeno–Yasukawa models. The approach implements fuzzy regression clustering for initial tuning of the parameters and a gradient-descent method to adjust them accurately. In Abonyi et al. [33] a modified Gath–Geva fuzzy clustering algorithm for the identification of TS models is proposed to directly obtain the parameters of the membership functions. A linear transformation of the input variables makes it possible to recover accurately the fuzzy partition of the antecedents. Li et al. [34] propose a new fuzzy c-regression model clustering algorithm where the clustering prototype in fuzzy space partition is a hyperplane. The new clustering algorithm is used in the identification of TS fuzzy model, obtaining good results in the identification of the premise parameters of the model.

1.3. Hybrid-fuzzy identification

For hybrid-fuzzy models, stability analysis and control designs have been proposed in the literature [35,36]. Regarding the identification of hybrid-fuzzy systems, although most of the developments have been made in conventional fuzzy systems, a few hybrid-fuzzy identification methods have been proposed. Palm and Driankov [37] present a hierarchical identification approach for fuzzy switched systems. The proposed method considers a black-box fuzzy identification approach by using fuzzy clustering and measurable discrete states in order to obtain the hybrid-fuzzy model. Although good performance is obtained, prior knowledge about the discrete modes is required. Next, Girimonte and Babuška [38] describe two structure-selecting methods for non-linear models with mixed discrete and continuous inputs. The first method, based on fuzzy clustering, uses fuzzy sets to obtain the relevant inputs. The second approach involves an induction algorithm included in a search method. The results show that fuzzy clustering is faster in terms of computation time. Zeng et al. [39] propose a new representation theorem for hierarchical systems when a discrete input space is considered. The theorem states that one-to-one mappings for low-level functions are required to obtain a flexible hierarchical representation. Moreover, they demonstrate that flexible hierarchical fuzzy systems satisfy the universal approximation property, which allows us to estimate any hierarchical function to any degree of accuracy. A new hierarchical structure of hybrid systems integrating modeling and control is presented by Cheng et al. [40], where the fuzzy controller is synthesized based on the identification of continuous and discrete components. The authors of [40] assume that measurements of the discrete components are available, which allows the use of fuzzy adaptive identification techniques or other ways to directly learn a TS model by clustering or by identifying a neuro-fuzzy model for each of the separate regions. In our paper, direct measurements of the discrete component are not available, and as a consequence it is not possible to do an experimental contrast within those other hybrid-fuzzy identification frameworks.

In this paper, a new identification method is proposed for nonlinear hybrid systems that identifies first the discrete transitions and then all other non-linearities through fuzzy models only using input–output data of the process, where the main difference with the literature is that prior knowledge of the discrete modes is not required. Next the hybrid-fuzzy models and the identification problem are presented.

2. Problem statement

For the modeling of hybrid systems two of the most popular model types used in the literature are piecewise affine (PWA) systems and mixed logical and dynamical (MLD) systems [11]. In this paper the use of another type of model called hybrid-fuzzy systems is proposed, which combine the characteristics of fuzzy models to represent nonlinearities, and of hybrid systems to include quantized variables.

A hybrid discrete-time nonlinear dynamic system is considered with input $\mathbf{u}(t) \in \mathbb{R}^m$, and to explain the identification method a single output $y(t) \in \mathbb{R}$ is assumed (the method is easily extendible for multiple outputs). Let $\mathbf{u}^{t-1} = [\mathbf{u}^T(t-1), \ldots, \mathbf{u}^T(t-n_b)]^T \in \mathbb{R}^{m \cdot n_b}$ be past inputs, and $\mathbf{y}^{t-1} = [y(t-1), \ldots, y(t-n_a)]^T \in \chi \subset \mathbb{R}^{n_a}$ be past outputs, up to time $t-1$, where n_a and n_b are the model orders (given a priori). The class of hybrid systems considered is described as:

$$\begin{aligned} y(t) &= \sum_{i=1}^{s} f_i(\mathbf{y}^{t-1}, \mathbf{u}^{t-1}) \varrho_i(\mathbf{y}^{t-1}), \\ \varrho_i(\mathbf{y}^{t-1}) &= \begin{cases} 1, & \text{if} \quad \mathbf{y}^{t-1} \in \chi_i \\ 0, & \text{otherwise} \end{cases}, \end{aligned} \tag{1}$$

where s is the number of discrete modes (submodels). The local behavior of the system is described by the functions $f_i(\cdot)$ and the discrete mode $\varrho_i(\mathbf{y}^{t-1})$ is a binary variable that equals 1 if $\mathbf{y}^{t-1}$ belongs to the region of $\chi_i \subset \mathbb{R}^{n_a}$, and 0 otherwise. The regions χ_i form a complete partition of the regressor set χ, i.e. $\bigcup_{i=1}^{s} \chi_i = \chi$ and $\chi_i \cap \chi_j = \emptyset$, $\forall i \neq j$. Note that discrete dynamics (transitions) of the system are assumed to occur when $\mathbf{y}^{t-1}$ satisfies some conditions, and they will not depend on the inputs. The aim in this work is to present a systematic method for determining the functions $f_i(\cdot)$ and the regions χ_i given only the input–output data of the process. The functions $f_i(\cdot)$ could be any non-linear function that will be identified by the TS models and the regions χ_i are assumed to be convex polyhedra, described by

$$\chi_i = \{\mathbf{y}^{t-1} \in R^{n_a} : H_i \mathbf{y}^{t-1} \preceq h_i\} \tag{2}$$

where $H_i \in R^{q_i \times n_a}$, $h_i \in R^{q_i}$ $i = 1, \ldots, s$, and $\preceq$ denotes componentwise inequality, where some inequalities are strict to prevent the boundaries of the regions from overlapping. Note that q_i is the number of linear inequalities defining the ith polyhedral region. In this paper, as a consequence of the algorithm, the resulting H_i, $i = 1, \ldots, s$, are diagonal matrices, so the partition will be a box (hyperrectangle). To assume the partition is a box will be sufficient to model a limited class of hybrid systems; however, when the model does not fit this characteristic (like in more general classes of hybrid systems [44]), the partition can be can approximate arbitrarily close at the cost of increased complexity. Optimization-based and other techniques could be applied to solve this issue, to obtain more general partitions, and then to model a larger class of hybrid systems [45,46].

The system given by (1) can be represented by a two-level fuzzy model, which was described by Tanaka et al. [47]. The corresponding two levels are the local fuzzy level and the discrete/quantized level. The local fuzzy level is a set of $\hat{s}$ TS fuzzy models with local validity in one region of an estimated partition $\hat{\chi}_i$, $i = 1, \ldots, \hat{s}$. The discrete/quantized level is given by a set of crisp functions $\delta_i(\mathbf{y}^{t-1})$, which activate the ith local TS model if $\mathbf{y}^{t-1}$ is in $\hat{\chi}_i$.

Assume that input–output data is available: ($y(t)$, $\mathbf{y}^{t-1}$, $\mathbf{u}^{t-1}$), $t = 1, \ldots, N$. The structure of a hybrid-fuzzy model to be identified for the variable $y(t)$ is described as:

$$y(t) = \sum_{i=1}^{\hat{s}} f_i^{\mathrm{TS}}(\mathbf{z}^{t-1}, \mathbf{y}^{t-1}, \mathbf{u}^{t-1})\delta_i(\mathbf{y}^{t-1}),$$
$$\delta_i(\mathbf{y}^{t-1}) = \begin{cases} 1, & \text{if} \mathbf{y}^{t-1} \in \hat{\chi}_i \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

$$f_i^{\mathrm{TS}}(\mathbf{z}^{t-1}, \mathbf{y}^{t-1}, \mathbf{u}^{t-1}) = \sum_{j=1}^{R_i} \beta_{ij}(\mathbf{z}^{t-1}) y_{ij}(\mathbf{y}^{t-1}, \mathbf{u}^{t-1}),$$
$$y_{ij}(\mathbf{y}^{t-1}, \mathbf{u}^{t-1}) = (\mathbf{a}_{ij})^T \mathbf{y}^{t-1} + (\mathbf{b}_{ij})^T \mathbf{u}^{t-1} + r_{ij}, \tag{4}$$
$$\beta_{ij}(\mathbf{z}^{t-1}) = \frac{\prod_{r=1}^{p} A_{ij,r}(z_r(t-1))}{\sum_{j=1}^{R_i} \prod_{r=1}^{p} A_{ij,r}(z_r(t-1))},$$

where the vector of the premises is $\mathbf{z}^{t-1} = [z_1(t-1), \ldots, z_p(t-1)]^T$ and p is the number of inputs at the premises. The premises variables are permitted to be inputs or outputs, and in this paper we will assume $\mathbf{z}^{t-1} = [(\mathbf{y}^{t-1})^T, (\mathbf{u}^{t-1})^T]^T$, so $p = n_a + m \cdot n_b$. Moreover, $(\mathbf{a}_{ij})^T$, $(\mathbf{b}_{ij})^T$, r_{ij} are the parameters of the fuzzy model $f_i^{\mathrm{TS}}(\cdot)$ for the region i in rule j, R_i is the number of rules of the fuzzy model at the ith region, $A_{ij,r}(z_r(t-1))$ is the membership degree for the input $z_r(t-1)$ in premise r at the ith region and rule j, and $\beta_{ij}(\mathbf{z}(t-1))$ is the activation degree of the jth rule that belongs to the fuzzy model of the ith region.

The hybrid-fuzzy model can be seen as a multimodel [48], where the validity indices (relevance degree of each model) are the on/off conditions that define whether a point belongs to a region of the complete partition of the system; as a consequence of the fusion principle, only one model (the most dominant) represent the dynamics in that region. Fig. 1 shows a simple scheme of the hybrid-fuzzy model structure (3).

Remarks:

- Note that the model given by (3) and (4) is a Takagi–Sugeno fuzzy model, with $\hat{s} \cdot R_i$ rules and activation degree $\beta_{ij}(\mathbf{z}^{t-1})\delta_i(\mathbf{y}^{t-1})$. One of the most important parts of the hybrid-fuzzy model is the fuzzy rule base. The rule i, j is the following: R_{ij} : ***if*** $\mathbf{y}^{t-1} \in \hat{\chi}_i$ ***and*** $z_1(t-1)$ *is* $A_{ij,1}$ ***and*** $z_2(t-1)$ *is* $A_{ij,2}$ ***and***. . . ***and*** $z_p(t-1)$ *is* $A_{ij,p}$ ***then*** $y_{ij}(t) = (\mathbf{a}_{ij})^T \mathbf{y}^{t-1} + (\mathbf{b}_{ij})^T \mathbf{u}^{t-1} + r_{ij}$, $\quad j = 1, .., R_i, \quad i = 1, \ldots, \hat{s}$.

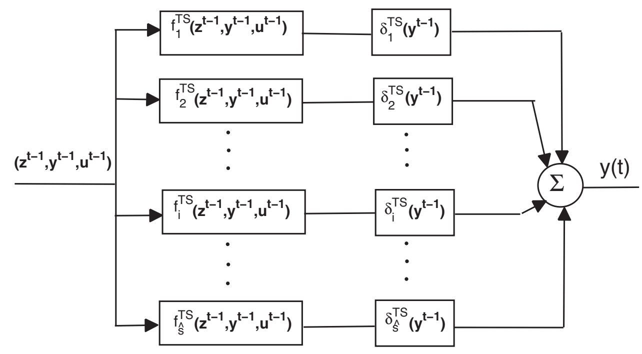

Fig. 1. Fusion principle for hybrid-fuzzy models.

- The first component of the fuzzy rule ($\mathbf{y}^{t-1} \in \hat{\chi}_i$) evaluates the binary membership function $\delta_i(\mathbf{y}^{t-1})$. This component explicitly incorporates the discrete transitions of the system.

By only using a finite input–output data set of the process ($y(t)$, $\mathbf{y}^{t-1}$, $\mathbf{u}^{t-1}$), $t = 1, \ldots, N$, the identification problem of a hybrid-fuzzy model given by (3) and (4) consists of estimating the following parameters: the number of regions $\hat{s}$, the partition $\hat{\chi}_i$, $i = 1, \ldots, \hat{s}$, and for each TS model, the number of rules R_i, the parameters of the membership functions $A_{ij,r}(\cdot)$, and the parameters $(\mathbf{a}_{ij})^T$, $(\mathbf{b}_{ij})^T$, r_{ij}. Usually an identification procedure is carried out by minimizing a cost function with respect to the unknown parameters [15]:

$$V_N = \frac{1}{N} \sum_{t=1}^{N} V\left(y(t) - \sum_{i=1}^{\hat{s}} f_i^{\mathrm{TS}}(\mathbf{z}^{t-1}, \mathbf{y}^{t-1}, \mathbf{u}^{t-1})\delta_i(\mathbf{y}^{t-1}) \right) \tag{5}$$

where V is a penalty function for the error, typically a quadratic function. The optimization problem should also include additional terms to avoid overfit (to limit for example $\hat{s}$ or R_i), or the premises may be obtained by using a specific clustering method. The minimization of (5) is in general a non-convex non-linear mixed-integer optimization (MINLP) problem. Therefore, in the next section, an identification procedure for hybrid-fuzzy systems based on well-known principles (avoiding to solve the MINLP problem) is described.

3. A hybrid-fuzzy identification method

As a motivation example, the hybrid tank system in Fig. 2 is considered, which is a modified version of the one used in [17] for PWA identification. In the figure, A_1, A_2 and A_3 are the cross-section of the tanks, S_1, S_2 and S_3 are the cross-section of the outlet holes, g is the acceleration due to gravity, $Q(t)$ is the input flow at time t, and $h(t)$ is the level of the tank. The hybrid tank system is divided into three regions because the cross-section of the tank is larger when the level is higher than h_1 or h_2.

Thus, for a fixed input flow, it will take more time to increase the level $h(t)$ when it is higher than h_1 or h_2, because the cross-section is larger. This means that the level values h_1 and h_2 are switching points in the sense that those levels are the boundaries of the three different operating regions, the dynamics of which are different. In order to avoid complicated optimization-based methods, we propose to detect the switching point by analyzing the principal component of the clusters' variance matrices, as provided by the Gustafson–Kessel (G–K) algorithm [41]. The G–K algorithm is typically used in the identification of TS models [43], so no extra information is required to perform the algorithm. The cluster slopes are analyzed by using the principal components, to be able to identify switching points of systems whose discrete behavior is reflected in the data by an abrupt variation of the spatial orientation

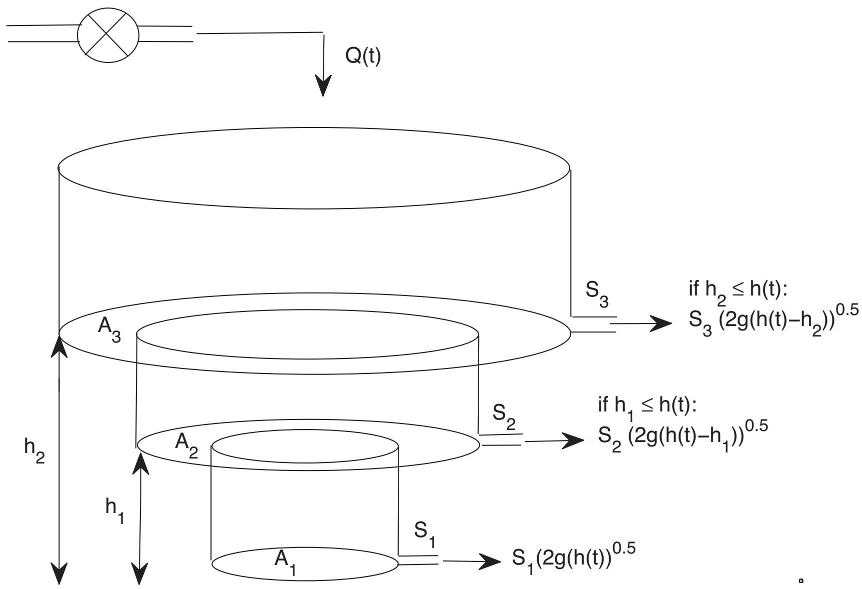

Fig. 2. Hybrid tank system.

of clusters. The method presented is interpretable as an "inverse" form of the merge method of clusters presented in [42,49], where, instead of merging similar clusters, clusters that differ a lot are used to detect switching points and to define a complete partition $\hat{\chi}_i$, $i = 1, \ldots, \hat{s}$, over the regressor space. The slope changes concept has been used recently for the statistical monitoring of nonlinear profiles in [50].

Once the partition is known, the local TS models $f_i^{\mathrm{TS}}(\cdot)$, $i = 1, \ldots, \hat{s}$, are identified by classifying the data according to the rule: the datum $(y(t), \mathbf{y}^{t-1}, \mathbf{u}^{t-1})$ is used for the identification of $f_i^{\mathrm{TS}}(\cdot)$ if $\mathbf{y}^{t-1} \in \hat{\chi}_i$.

3.1. Hybrid-fuzzy model identification procedure

Throughout this paper, we assume that a sequence of N input/output data have been collected:

$$\Phi = \begin{bmatrix} y(1) & (\mathbf{y}^0)^T & (\mathbf{u}^0)^T \\ y(2) & (\mathbf{y}^1)^T & (\mathbf{u}^1)^T \\ \vdots & \vdots & \vdots \\ y(N) & (\mathbf{y}^{N-1})^T & (\mathbf{u}^{N-1})^T \end{bmatrix}_{N, n_a+m\cdot n_b+1}, \tag{6}$$

where N denotes the number of data samples.

The identification procedure consists of seven steps. In the *Step 1*, based on the G–K algorithm, the information of the clusters (center, covariance matrices, eigenvectors and eigenvalues) is obtained. Then, in *Step 2* the principal components are extracted. Next, with the projections on the regressors space obtained in *Step 3* and in *Step 4*, the cluster slopes are calculated. With the cluster slopes, in *Step 5* the switching points are determined. In *Step 6* the partition is generated, and finally in the *Step 7* the local fuzzy models are identified. The details of each step are explained next.

Step1: Clustering. Determine the fuzzy clusters over the data Φ, using the G–K algorithm [41]. This algorithm includes fuzzy covariances in an n-dimensional space, and closely resembles maximum likelihood estimation on mixture densities. It is suitable for the identification of hybrid-fuzzy models because the consequents of hybrid-fuzzy models are hyperplanes in the premise-consequent product space. The algorithm will cluster the data given a specified number of clusters c, the parameters for the cluster fuzziness, and the stopping criterion. The G–K algorithm provides the centers of the clusters $\mathbf{v}_l = [v_l^1, \ldots, v_l^{n_a+m\cdot n_b+1}]^T$, a covariance matrix for each fuzzy cluster l, with $n_a + m \cdot n_b + 1$ eigenvectors $\varphi_{1,l}, \ldots, \varphi_{n_a+m\cdot n_b+1,l}$, and with the corresponding eigenvalues $\lambda_{1,l}, \ldots, \lambda_{n_a+m\cdot n_b+1,l}$.

It is well known that the G–K algorithm does not give an indication of the correct number of clusters c needed. A large number of clusters will result in a complicated rule-base model, while a small number of clusters will result in a poor model. So to obtain the optimum number of clusters, the use of a compatible cluster merging method is proposed, just like it is suggested for the identification of TS models in [42,49]. This method works as follows: let the center of two clusters be $\mathbf{v}_{l_1}$ and $\mathbf{v}_{l_2}$, with φ_{1,l_1} and φ_{1,l_2} the eigenvectors associated with the minimum eigenvalue λ_{1,l_1} and λ_{1,l_2} respectively. The criteria to merge the clusters proposed in [49] consider that the nearly parallel major axes of consecutive clusters should be merged ($|\varphi_{1,l_1} \cdot \varphi_{1,l_2}| \geq k_1$, with k_1 close to 1) and also the cluster centers should be sufficiently close for merging ($\| \mathbf{v}_{l_1} - \mathbf{v}_{l_2} \|_2 \leq k_2$, with k_2 close to 0). In this step we could also use a set of cluster validity metrics like fuzzy entropy, fuzzy hypervolume, etc.

Step 2: Principal components. First select the eigenvector $\varphi_l^* = [\phi_l^1, \phi_l^2, \ldots, \phi_{l_1}^{n_a+m\cdot n_b+1}]^T$ associated with the maximum eigenvalue λ_l^* for each cluster $l = 1, \ldots, c$:

$$\lambda_l^* = \max\{\lambda_{1,l}, \lambda_{2,l}, \ldots, \lambda_{n_a+m\cdot n_b+1,l}\}. \tag{7}$$

The switching points are obtained by analyzing the most important eigenvectors (the principal vectors or the principal components), in the direction of which the most information is given. Inspired by the merge method for clusters [49], instead of merging clusters with nearly-parallel major axes, we will split the output-regressor space when those consecutive clusters are very different (i.e. the angle between the hyperplanes is big). It is assumed that the switching points are defined by the outputs, so the analysis is done for each component of the output-regressor space $y(t-k)$, $k = 1, \ldots, n_a$.

Step 3: Projections. For every cluster $l = 1, \ldots, c$ and every component of the output regressor space $y(t-k)$, $k = 1, \ldots, n_a$, calculate the vector $\hat{\pi}_{lk}$, which represents the projection of the eigenvector

φ_l^* on the subspace given by the inputs and the output $y(t-k)$, and which is given by:

$$\hat{\pi}_{lk} = \frac{\Phi_k \varphi_l^*}{\| \Phi_k \varphi_l^* \|_2}, \quad \forall l = 1, \ldots, c, \quad \forall k = 1, \ldots, n_a, \tag{8}$$

where φ_l^* is the eigenvector chosen in *Step* 2 and Φ_k is the matrix of dimension $(n_a + m \cdot n_b + 1) \times (n_a + m \cdot n_b + 1)$, the elements of which are defined as:

$$(\Phi_k)_{\ell,\wp} = \begin{cases} 1 & \text{if} \ell = \wp = k+1, \\ 1 & \text{if} \ell = \wp \text{ and } \ell > n_a + 1, \\ 0 & \text{if} \quad \text{otherwise}. \end{cases} \tag{9}$$

Note that the vector is normalized, so $\| \hat{\pi}_{lk} \|_2 = 1$. Note also that (7) is just the projection of the eigenvector on the sub-space generated by the inputs and the output $y(t-k)$.

For every vector $\hat{\pi}_{lk}$ determine $\hat{\pi}_{lk}^u$ which represents the projection of $\hat{\pi}_{lk}$ in the subspace generated by the inputs, and which is obtained in the following way:

$$\hat{\pi}_{lk}^u = \frac{\Phi_u \hat{\pi}_{lk}}{\| \Phi_u \hat{\pi}_{lk} \|_2}, \quad \forall l = 1, \ldots, c, \quad \forall k = 1, \ldots, n_a, \tag{10}$$

where $\hat{\pi}_{il}$ is the vector obtained in *Step 3*, and Φ_u is the matrix of dimension $(n_a + m \cdot n_b + 1) \times (n_a + m \cdot n_b + 1)$, the elements of which are defined as:

$$(\Phi_u)_{\ell,\wp} = \begin{cases} 1 & \text{if} \quad \ell = \wp \quad \text{and} \quad \ell > n_a + 1, \\ 0 & \text{if} \quad \text{otherwise}. \end{cases} \tag{11}$$

Note that the vector is normalized, so $\| \hat{\pi}_{lk}^u \| = 1$. Note also that in (10) we are just projecting (7) on the subspace generated by the inputs. This vector is used to define the angle we need to estimate the switching point.

Step 4: Cluster slope. Let $\hat{\gamma}_{lk}$ be the estimation of the angle between $\hat{\pi}_{lk}$ and $\hat{\pi}_{lk}^u$. It is possible to obtain this angle by calculating $\arccos(\hat{\pi}_{lk}^T \hat{\pi}_{lk}^u)$. Then, for each cluster l and every output variable $y(t-k)$, compute the cluster slope $\Gamma_{lk} = \tan(\hat{\gamma}_{lk})$ given by:

$$\Gamma_{lk} = \tan(\arccos\left(\hat{\pi}_{lk}^T \hat{\pi}_{lk}^u\right)) = \frac{|\hat{\pi}_{lk}^T \hat{\pi}_{lk}^u|}{\hat{\pi}_{lk}^T \hat{\pi}_{lk}^u} \sqrt{\frac{1}{(\hat{\pi}_{lk}^T \hat{\pi}_{lk}^u)^2} - 1},$$
$$\forall l = 1, \ldots, c, \quad \forall k = 1, \ldots, n_a, \tag{12}$$

As an example consider Fig. 3 with $\mathbf{y}^{t-1} = y(t-1)$, $\mathbf{u}^{t-1} = u(t-1)$. Fig. 3a shows the data with the corresponding clusters, and the lines inside the cluster represent the vectors φ_l^* associated with the maximum variance for each cluster. Fig. 3b shows the projections of the vectors φ_l^* and the angles $\hat{\gamma}_{lk}$.

Step5: Switching points. In this step the switching points candidates are determined for each variable $y(t-k)$, $k = 1, \ldots, n_a$. Consider two consecutive clusters l_1 and l_2, with centers $\mathbf{v}_{l_1}$ and $\mathbf{v}_{l_2}$ in descending order for the component k of the vector related with the variable $y(t-k)$, $(v_{l_1}^k < v_{l_2}^k)$.

The candidate switching point should be in between the coordinates $v_{l_1}^k$ and $v_{l_2}^k$. A good estimator of the switching point could be the coordinate $v_{l_1}^k + \sqrt{\lambda_{l_1}^*} \phi_{l_1}^k$, which is interpretable as the coordinate obtained when running from the center of the ellipsoid cluster l_1, given by $v_{l_1}^k$, through the kth axis, up to the edge of the ellipsoid. The value $\lambda_{l_1}^*$ is the eigenvalue obtained in *Step* 2 corresponding to cluster l_1, and $\phi_{l_1}^k$ the kth coordinate of the corresponding eigenvector. For the same reason, if the cluster l_2 is used, which is on the other side of the switching point, the value $v_{l_2}^k - \sqrt{\lambda_{l_2}^*} \phi_{l_2}^k$ can

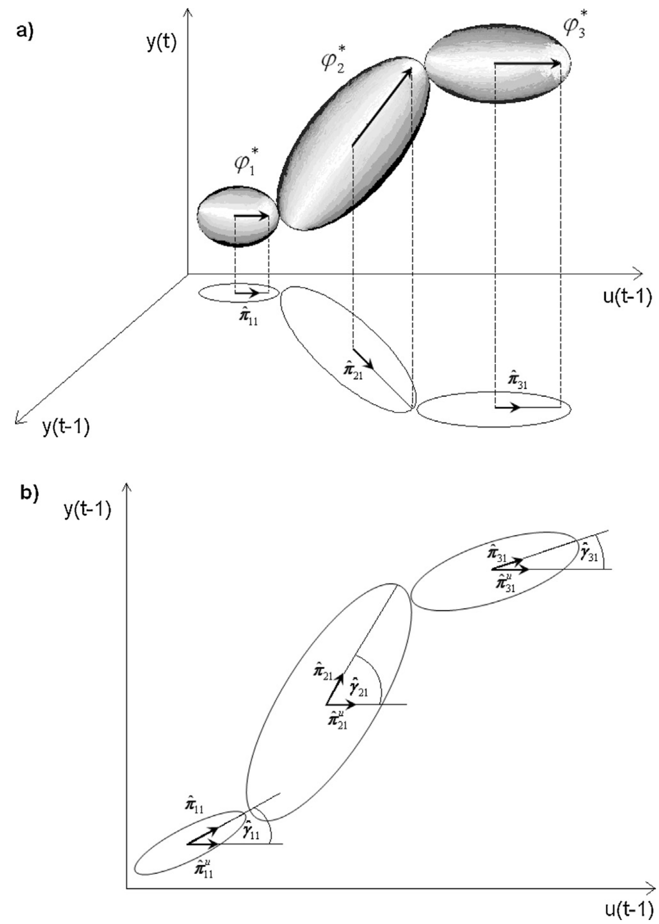

Fig. 3. (a) Representation of projections obtained in *Step 3* and (b) angles obtained in *Step 4*.

be chosen as a candidate switching point. Then, for estimating the location of the switching point $V_k^{l_1 l_2}$, a weighted sum of those coordinates is proposed:

$$V_k^{l_1 l_2} = \frac{((v_{l_1}^k + \sqrt{\lambda_{l_1}^*} \phi_{l_1}^k)/\lambda_{l_1}^*) + ((v_{l_2}^k - \sqrt{\lambda_{l_2}^*} \phi_{l_2}^k)/\lambda_{l_2}^*)}{(1/\lambda_{l_1}^*) + (1/\lambda_{l_2}^*)}, \tag{13}$$

where $\phi_{l_1}^k$ and $\phi_{l_2}^k$ are the kth coordinates of the corresponding eigenvectors. The weighted sum represents the fact that from experiments we found out that the switching points are usually closer to the cluster with the smaller eigenvalue. Better methods to estimate the switching point location can be further investigated.

The next step is to choose the switching point candidates $V_k^{l_1 l_2}$ the rate $\Delta\Gamma_{l_1 l_2 k}$ of which satisfies a criterion. The rate $\Delta\Gamma_{l_1 l_2 k}$ is given by:

$$\Delta\Gamma_{l_1 l_2 k} = \left| \Gamma_{l_1 k} - \Gamma_{l_2 k} \right|. \tag{14}$$

One criterion could be to select rates $\Delta\Gamma_{l_1 l_2 k}$ that are larger than a given threshold. For example, a threshold could be the mean value of $\Delta\Gamma_{l_1 l_2 k}$ (denoted as $\Delta\bar{\Gamma}_k$), plus two times its standard deviation $(\sigma_{\Delta\Gamma_k})$. So, if all the rates are similar, this means that there is not a switching point in the variable $y(t-k)$. Otherwise, just the clusters with a larger variation will be considered. The main advantage of this method is the chance to obtain a good estimation of all the switching points in a row, without any further analysis. The main drawback in this criterion is the possibility to miss switching points (or include not switching points) if the threshold is not appropriate.

As an alternative to the threshold criterion, a sensitivity analysis could be performed to evaluate if the inclusion of a switching point improves the performance of the prediction model while keeping the complexity of the model reduced. If so, one extra switching point will be included, the corresponding hybrid-fuzzy model identified, and then *Step 5* is analyzed again, to determine the inclusion of another switching point. The process will finish once a the performance of the hybrid-fuzzy model does not improve significantly (within some thresholds) by the inclusion of new switching points. So, let us assume we have generated a partition $\{\hat{\chi}_i\}_{i=1}^{\hat{s}-1}$. We now analyze the inclusion of a new switching point in the model, by splitting the region $\hat{\chi}_i$ into two new regions defined by the new switching point. So, let consider the switching point candidate, with the maximum rate, given by:

$$V_{\hat{s}} = \{V_{\bar{k}}^{\bar{l}_1\bar{l}_2} : (\bar{l}_1, \bar{l}_2, \bar{k}) = \mathrm{argmax}\{\Delta\Gamma_{l_1 l_2 k}\}\}. \quad (15)$$

Step6: Generation of the partition. The region $\hat{\chi}_i$ is divided into two new regions. Recall that the region $\hat{\chi}_i$ is defined as follows:

$$\hat{\chi}_i = \{\mathbf{y}^{t-1} : H_i\mathbf{y}^{t-1} \preceq h_i\}, \quad (16)$$

where $H_i \in R^{q_i \times n_a}$ is diagonal, $h_i \in R^{q_i}$ $i = 1, \ldots, s$, the symbol $\preceq$ denotes componentwise inequality, where some inequalities are strict to avoid the boundaries of the regions to have multiple values. Given the new switching point $V_{\hat{s}}$ in the variable $y(t-k)$, the two new regions are defined as follows:

$$\hat{\chi}_1^i = \{\mathbf{y}^{t-1} : H_i\mathbf{y}^{t-1} \preceq h_i \quad \wedge \quad y(t-k) \leq V_{\hat{s}}\}. \quad (17)$$

$$\hat{\chi}_2^i = \left\{\mathbf{y}^{t-1} : H_i\mathbf{y}^{t-1} \preceq h_i \quad \wedge \quad -y(t-k) < -V_{\hat{s}}\right\}. \quad (18)$$

Step7: Fuzzy identification. For the sub-regions $\hat{\chi}_1^i$ and $\hat{\chi}_2^i$, a local TS model is identified. First, the data belonging to the region $\hat{\chi}_i$ is split into the two new regions, by the rule: if $y(t-k) \leq V_{\hat{s}}$ then $(y(t), (\mathbf{y}^{t-1})^T, (\mathbf{u}^{t-1})^T) \in \hat{\chi}_1^i$, else $(y(t), (\mathbf{y}^{t-1})^T, (\mathbf{u}^{t-1})^T) \in \hat{\chi}_2^i$, $t = 1, \ldots, N$. Then, for each new partition, just considering data that belongs to the subregion, the number of rules R_i and the membership functions $A_{ij,r}(\cdot)$ are obtained with a clustering method (G–K). Each TS model is optimized for the number of fuzzy clusters and their regressor structure is obtained by a sensitivity analysis (see [30,29,51]).

The next step is to identify the consequent parameters of each rule of the TS model (see [52–54,34]). Let us write the consequent parameters for the fuzzy rule j in the region i as follows:

$$\Theta_{ij} = \begin{bmatrix} \mathbf{a}_{ij} \\ \mathbf{b}_{ij} \\ r_{ij} \end{bmatrix}, \quad (19)$$

An identification procedure is carried out by minimizing the following cost function with respect to the unknown parameters Θ_{ij}:

$$V_{N_{ij}} = \frac{1}{N_{ij}} \sum_{t=1}^{N_{ij}} (\beta_{ij}(\mathbf{z}^{t-1}))^2 (y(t) - [(\mathbf{y}^{t-1})^T \quad (\mathbf{u}^{t-1})^T \quad 1]\Theta_{ij})^2 \quad (20)$$

where N_{ij} is the number of input–output data pairs corresponding to the rule j of the region i considering only the data that belongs to the region i and for which $\beta_{ij}(\mathbf{z}(t-1)) \geq \delta$, with δ a small positive number.

The model parameters for the rule j of region i can be obtained using the least squares identification method as follows:

$$\Theta_{ij} = (\Psi_{ij}^T \Psi_{ij})^{-1} \Psi_{ij}^T Y_{ij} \quad (21)$$

where the matrices Ψ_{ij} and Y_{ij} are as follows:

$$\Psi_{ij} = \begin{bmatrix} \beta_{ij}(\mathbf{z}^0)[(\mathbf{y}^0)^T \quad (\mathbf{u}^0)^T \quad 1] \\ \beta_{ij}(\mathbf{z}^1)[(\mathbf{y}^1)^T \quad (\mathbf{u}^1)^T \quad 1] \\ \vdots \\ \beta_{ij}(\mathbf{z}^{N_{ij}-1})[(\mathbf{y}^{N_{ij}-1})^T \quad (\mathbf{u}^{N_{ij}-1})^T \quad 1] \end{bmatrix}, \quad (22)$$

$$Y_{ij} = \begin{bmatrix} \beta_{ij}(\mathbf{z}^0)y(1) \\ \beta_{ij}(\mathbf{z}^1)y(2) \\ \vdots \\ \beta_{ij}(\mathbf{z}^{N_{ij}-1})y(N_{ij}) \end{bmatrix}, \quad (23)$$

By the identification of each rule (not the overall model), and also by weighting the data for the corresponding activation degree of each rule β_{ij}, a better conditioning of the matrices is obtained, compared to the conditioning of the whole data matrix. This approach leads to a better estimation of the hybrid-fuzzy model parameters as the data close to the center of the cluster will be more important to minimize than the data at the borders. This fact can help with the problem of misclassification of data, where due to errors in the estimation of the switching points or due to the noise of the system, a data point belonging to one region is considered to belong to another.

Remarks:

- The proposed generation of the partition (2) is quite limited as it is defined by diagonal matrices H_i, $i = 1, \ldots, s$. Systems defined by another kind of partitions like polyhedral ($H_i\mathbf{y}^{t-1} - h_i \preceq 0$ with H_i non diagonal), or a non-linear partition ($H_i(\mathbf{y}^{t-1}) \preceq 0$ with H_i a non-linear function), will not fit into this configuration. The proposed method improves the monitoring and prediction capabilities in some processes, but the analysis and algorithms for a more general case is topic for the further research.
- Note that, as we first perform the identification of the discrete transitions, our method can be adapted for use as a key element in other modeling methods for hybrid systems. For instance, other modeling tools can be used instead of a fuzzy system like neural networks, PWA systems, etc., to create subsystems models after the successful identification of the discrete modes.

4. Simulation results

Next, an illustrative experiment on a hybrid-tank system is described. Then, as an empirical validation of the method, based on real-life data measured on a part of a highway in the Netherlands, the results of the hybrid-fuzzy identification of a first-order traffic model is presented.

4.1. Hybrid tank system

Let us consider the hybrid tank system shown in Fig. 2. The following non-linear equations describe the dynamics of the tank system:

$$\frac{dh}{dt} = \begin{cases} \frac{1}{A_1}(Q(t) - Q_1(h(t))) & \text{if} \quad h(t) \leq h_1 \\ \frac{1}{A_2}(Q(t) - Q_2(h(t)))) & \text{if} \quad h_1 < h(t) \leq h_2 \\ \frac{1}{A_3}(Q(t) - Q_3(h(t))) & \text{if} \quad h_2 < h(t) \end{cases}, \quad (24)$$

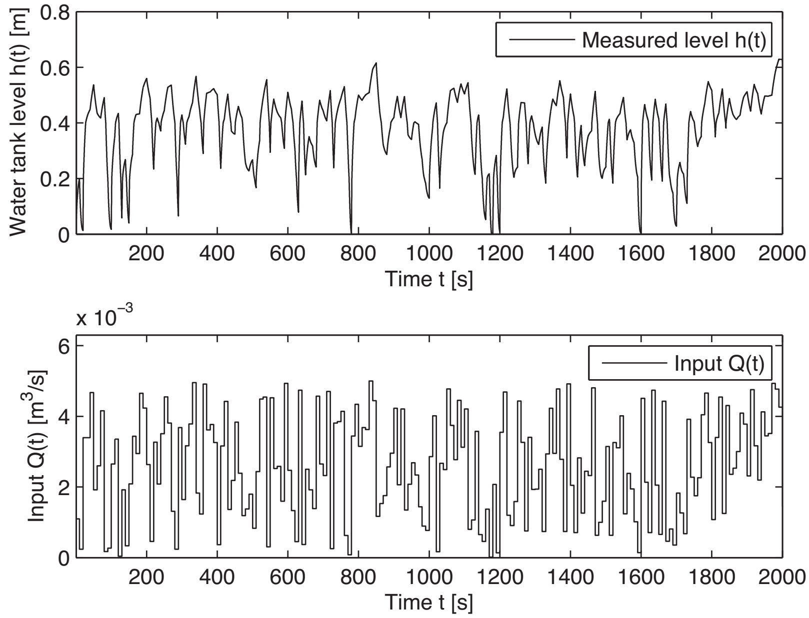

Fig. 4. Input/output data.

where $h(t)$ $[m]$ is the level of the tank at time t, $u(t) = Q(t)$ $[m^3/s]$ is the input flow, $Q_1(h(t)) = S_1\sqrt{2gh(t)}$ is the outflow of the first part of the tank, the outflow of the second part is $Q_2(h(t)) = S_2\sqrt{2g(h(t)-h_1)} + Q_1(h(t)$, $Q_3(h(t)) = S_3\sqrt{2g(h(t)-h_2)} + Q_2(h(t)) + Q_1(h(t))$ is the outflow of the third part, $A_1 = 0.0154$ $[m^2]$ is the cross-section of the first part of the tank, the cross-sections of the second and third parts are given by $A_2 = 3A_1$, $A_3 = 9A_1$, $S_1 = S_2 = S_3 = 0.0005\,\mathrm{m}^2$ are the cross-sections of the outlet holes, and $g = 9.81\,\mathrm{m}^2/\mathrm{s}$ is the acceleration due to gravity, $h_1 = 0.2$ m and $h_2 = 0.4$ m. We will assume that just input–output data shown in Fig. 4 are available for identification and validation. For the input $Q(t)$ a uniformly distributed random signal with minimum value 0, maximum value 0.005, and sample time 0.1 s was used. The experiment was designed in a way to just have a good number of h_1 and h_2 crossings. A total of 4000 samples were used as identification set, and 4000 as the validation set. The results presented next were obtained with signals sampled with $T_s = 0.1$ s.

The identification problem is to find the relation between $h(t)$ and $Q(t)$ considering the input/output data. The main goal is to find the number of switching regions and the switching points (in this case $h_1 = 0.2$ m and $h_2 = 0.4$ m), that define the partition. After an optimization of the order of the models, first-order model is considered good enough to predict the dynamics of the system, so the input/output data vectors considered are $\mathbf{y}^{t-1} = h(t-1)$ as the output and $\mathbf{u}^{t-1} = Q(t-1)$ as the input. In order to evaluate the performance of the hybrid-fuzzy model (with one and two switching points detected) and a TS model (with no switching point included), the Root Mean Squared (RMS) error is used.

4.1.1. TS modeling results

The G–K algorithm was used to obtain the clusters. The premises were obtained by denormalization of the cluster parameters given by the G–K algorithm with normalized data. The consequent parameters were obtained using the method explained in *Step 7*. The number of clusters obtained from a sensitivity analysis was six. The number of clusters can be increased if a smaller RMS is needed; however, the complexity of the model will then also increase. Just to be fair in the comparison with the hybrid-fuzzy model, the number of rules will be the same as the TS model.

The TS model is given by:

R_j : ***if*** $h(t-1)$ *is* $A_{j,1}$ ***and*** $Q(t-1)$ *is* $A_{j,2}$ ***then*** $h(t) = a_{j1}h(t-1) + b_{j1}Q(t-1) + r_j$, $j = 1, \ldots, 6$. where $A_{j,l}(z_l(t-1)) = e^{-0.5(c_{1,j,l}(z_l(t-1)-c_{2,j,l}))^2}$.

The parameters of the premises and the consequents of the TS model are listed in Table 1.

4.1.2. Hybrid-fuzzy model results

The same procedure based on the G–K algorithm used for TS model is performed. Using the covariance matrices, given by the G–K algorithm from *Step 1*, the eigenvalues and eigenvectors associated with each cluster are determined. Each cluster has 3 eigenvalues and 3 eigenvectors. Then, the eigenvector (φ_l^*) associated with the largest eigenvalue for each cluster is considered. The projections of the eigenvectors obtained from *Step 2* are determined and the slopes of the projected eigenvectors with respect to the coordinate $h(t-1)$ were computed. The rates of consecutive clusters are calculated. Each rate is associated with the estimated switching point in the coordinate $h(t-1)$, by using (13). A switching point was estimated to be in $h(t-1) = 0.2037$ m (the real value is 0.2). After splitting the data into the new regions $h(t-1) \geq 0.2037$ and $h(t-1) < 0.2037$, the rates between consecutive clusters belonging to each region are calculated again, the second switching point being estimated to be in $h(t-1) = 0.4018$ m (the real value is 0.4).

Two hybrid-fuzzy models are identified. The first one (hybrid-fuzzy 1) considering one estimated switching point $h(t-1) = 0.2037$. The second model (hybrid-fuzzy 2) considers switching points $h(t-1) = 0.2037$ and $h(t-1) = 0.4018$. For the model hybrid-fuzzy 1 there are two subregions ($\hat{s} = 2$): the first

Table 1
Parameters of TS model.

Rules	$c_{1,j,1}$	$c_{2,j,1}$	$c_{1,j,2}$	$c_{2,j,2}$	a_{j1}	b_{j1}	r_j
$j = 1$	0.0093	0.2393	2.3866	0.0015	0.9886	1.5045	0.0005
$j = 2$	0.6257	0.1013	0.0935	0.0006	0.9884	1.5274	0.0005
$j = 3$	3.2164	0.4244	0.0086	0.0028	0.9899	1.2738	0.0006
$j = 4$	0.0211	0.3417	2.0847	0.0020	0.9886	1.5045	0.0005
$j = 5$	2.4465	0.3778	0.0156	0.0031	0.9891	1.4083	0.0005
$j = 6$	2.1503	0.5101	0.0261	0.0035	0.9898	1.2978	0.0006

one is $\hat{\chi}_{11}$, where $h(t-1) \geq 0.2037$. The second is $\hat{\chi}_{12}$, where $h(t-1) < 0.2037$. For the model hybrid-fuzzy 2 There are three subregions ($\hat{s} = 3$): The first one is $\hat{\chi}_{21}$, where $h(t-1) < 0.2037$. The second is $\hat{\chi}_{22}$, where $h(t-1) < 0.4018$ and $h(t-1) \geq 0.2037$. The third is $\hat{\chi}_{23}$, where $h(t-1) \geq 0.4018$. Finally, using the proposed identification method, local TS models for the corresponding switching regions are computed, optimizing the number of clusters per region. In total six rules are used in each hybrid-fuzzy model, so the results will be comparable with the 6 rules of the TS model obtained before. Finally, the structure of hybrid-fuzzy model with two switches (three subregions) is given by:

R_{1j}: ***if*** $h(t-1) \in \hat{\chi}_{21}$ ***and*** $h(t-1)$ *is* $A_{1,j,1}$ ***and*** $Q(t-1)$ *is* $A_{1,j,2}$, ***then*** $h(t) = a_{1j1}h(t-1) + b_{1j1}Q(t-1) + r_{1j}$, $j = 1, 2$.

R_{2j}: ***if*** $h(t-1) \in \hat{\chi}_{22}$ ***and*** $h(t-1)$ *is* $A_{2,j,1}$ ***and*** $Q(t-1)$ *is* $A_{2,j,2}$, ***then*** $h(t) = a_{2j1}h(t-1) + b_{2j1}Q(t-1) + r_{2j}$, $j = 1, 2$

R_{3j}: ***if*** $h(t-1) \in \hat{\chi}_{23}$ ***and*** $h(t-1)$ *is* $A_{3,j,1}$ ***and*** $Q(t-1)$ *is* $A_{3,j,2}$, ***then*** $h(t) = a_{3j1}h(t-1) + b_{3j1}Q(t-1) + r_{3j}$, $j = 1, 2$ where $A_{ij,r}(z_r(t-1)) = e^{-0.5(c_{1,ij,r}(z_r(t-1)-c_{2,ij,r}))^2}$.

The parameters for hybrid-fuzzy models are given in Tables 2(hybrid-fuzzy 1) and 3(hybrid-fuzzy 2).

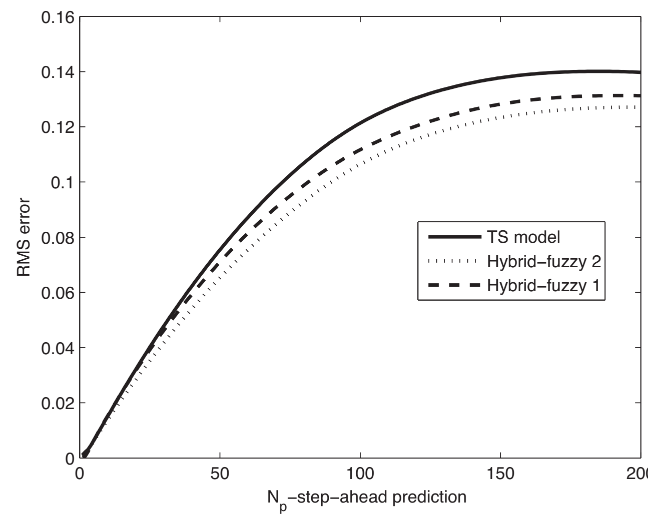

Fig. 5. N-step-ahead prediction error.

4.1.3. Comparative analysis

Fig. 5 contains the RMS for the TS model, hybrid-fuzzy model with one switch (hybrid-fuzzy 1) and hybrid-fuzzy model with two switches (hybrid-fuzzy 2), considering the validation data set for N_p step-ahead predictions. Fig. 6 shows the measured output and the predicted modes using hybrid-fuzzy model 2, in the case of 200 step-ahead-prediction.

The switching points detected at $h(t-1) = 0.4018$ and $h(t-1) = 0.2037$ are a fairly good estimation. From Figs. 5 and 6, the main advantage of hybrid-fuzzy modeling are its fuzzy rules,

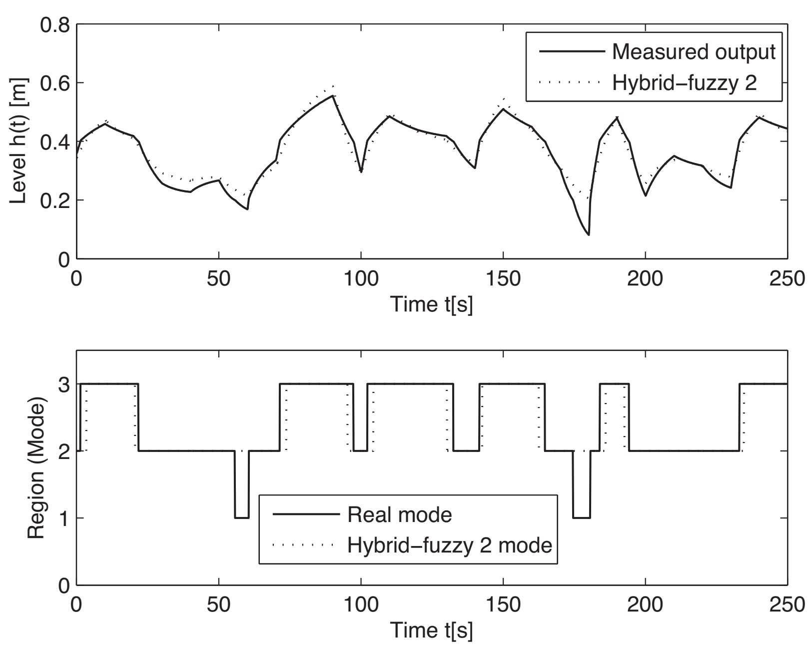

Fig. 6. Measured and 200-step-ahead predicted: (a) output and (b) modes using hybrid-fuzzy model 2.

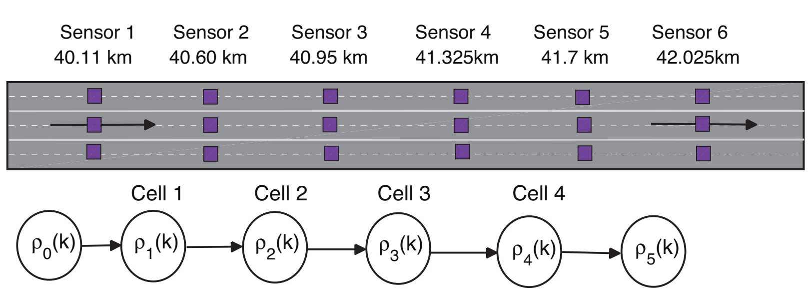

Fig. 7. Schematic sensor positions and cells of the part of the A12 freeway in the Netherlands.

Table 2
Parameters of the model hybrid-fuzzy 1 (2 subregions).

$\hat{\chi}_{12}$	$c_{1,1j,1}$	$c_{2,1j,1}$	$c_{1,1j,2}$	$c_{2,1j,2}$	a_{1j1}	b_{1j1}	r_{1j}
$j=1$	4.5845	179.2749	0.5278	0.0026	−0.1549	0.7641	79.5937
$j=2$	7.0266	116.9673	0.4418	0.0022	−0.1711	0.7818	82.4465
$j=3$	6.5462	125.5516	0.4694	0.0024	−0.1668	0.7787	81.3518
$j=4$	4.3358	189.5602	0.4832	0.0024	−0.1594	0.7684	80.6628
$j=5$	3.7419	219.6468	0.4455	0.0022	−0.1619	0.7709	81.4176
$\hat{\chi}_{11}$	$c_{1,2j,1}$	$c_{2,2j,1}$	$c_{1,2j,2}$	$c_{2,2j,2}$	a_{2j1}	b_{2j1} 0	r_{2j}
$j=1$	2.9461	167.8887	0.2734	0.0014	0.0391	0.1361	100.4664
$j=2$	4.5124	109.614	0.0563	0.0003	0.047	0.0849	100.8289
$j=3$	3.8836	127.362	0.2698	0.0013	0.0375	0.144	99.7747
$j=4$	2.0879	236.8959	0.23	0.0012	0.037	0.1288	103.0967
$j=5$	3.2343	152.9293	0.3015	0.0015	0.0383	0.1424	99.9024

Table 3
Parameters of the model hybrid-fuzzy 2 (3 subregions).

$\hat{\chi}_{23}$	$c_{1,1j,1}$	$c_{2,1j,1}$	$c_{1,1j,2}$	$c_{2,1j,2}$	a_{1j1}	b_{1j1}	r_{1j}
$j=1$	4.4844	183.2774	0.4689	0.0023	−0.1617	0.7711	81.0272
$j=2$	5.3606	153.3211	0.4403	0.0022	−0.1675	0.7776	81.9565
$j=3$	5.1186	160.5671	0.5224	0.0026	−0.1572	0.7679	79.6844
$\hat{\chi}_{22}$	$c_{1,2j,1}$	$c_{2,2j,1}$	$c_{1,2j,2}$	$c_{2,2j,2}$	a_{2j1}	b_{2j1}	r_{2j}
$j=1$	7.8234	129.1493	0.3513	0.0018	0.0213	0.2117	98.2485
$j=2$	4.7129	214.3854	0.2462	0.0012	0.0119	0.224	101.4305
$j=3$	4.4403	227.5493	0.2857	0.0014	0.0137	0.2217	101.0792
$\hat{\chi}_{21}$	$c_{1,3j,1}$	$c_{2,3j,1}$	$c_{1,3j,2}$	$c_{2,3j,2}$	a_{3j1}	b_{3j1}	r_{3j}
$j=1$	6.679	146.8651	0.0933	0.0005	0.0539	−0.0366	102.6413
$j=2$	5.555	176.5839	0.0277	0.0001	0.0531	−0.0437	103.5835
$j=3$	5.6231	174.4456	0.1562	0.0008	0.0511	−0.0331	103.9583
$j=4$	9.1069	107.7121	0.1488	0.0007	0.0515	−0.0251	102.7786

which can be used directly to detect the modes of the system as shown in Fig. 6.

The three-level tank system is a simple example to show the benefits of the new identification method. In a real-world implementation of this system some effects like turbulence and boundary-layer effects in the transition between the tanks will be found. Next, in order to get an empirical validation using real-life data, the method is applied for the hybrid-fuzzy identification of the density in a highway.

4.2. Hybrid-fuzzy identification for density traffic monitoring

A 1.915 km long stretch of the A12 freeway, in the Netherlands has been used as test field to validate the identification method. The stretch used is in the segment that crosses the Dutch province of South Holland. In Fig. 7 the scheme of the stretch is depicted. In this paper, the identification procedure will be explained with a single-link case study, with no on-ramps or off-ramps.

A period of 6 h (5:00–11:00) representative of typical working Monday will be modeled. The data from two representative days were used for identification, and for validation another day was used. The highway is divided in 4 sections or cells (see Fig. 7). The density in the section i is denoted by $\rho_i(t)$, and the density in the next section of the highway is denoted by $\rho_{i+1}(t)$, and so on. The model structure of a typical first-order in time traffic model is considered. Then, for example, to model the density $\rho_i(t)$ the regressors $\rho_{i-1}(t-1)$, $\rho_i(t-1)$, and $\rho_{i+1}(t-1)$ are used. Note that those are the same regressors used by the cell transmission model [55].

Thus, the hybrid-fuzzy model to identify for each section $i=1, \ldots, 4$ is:

$$\rho_i(k+1) = \sum_{p=1}^{s_i} f_{i,p}^{\mathrm{TS}}(\rho_i(k), \rho_{i-1}(k), \rho_{i+1}(k))\delta_{i,p}(\rho_i(k)) \tag{25}$$

where the output is $y(k)=\rho_i(k)$, the densities at the neighboring links are the inputs $u(k)=[\rho_{i-1}(k), \rho_{i+1}(k)]^T$ and the number of

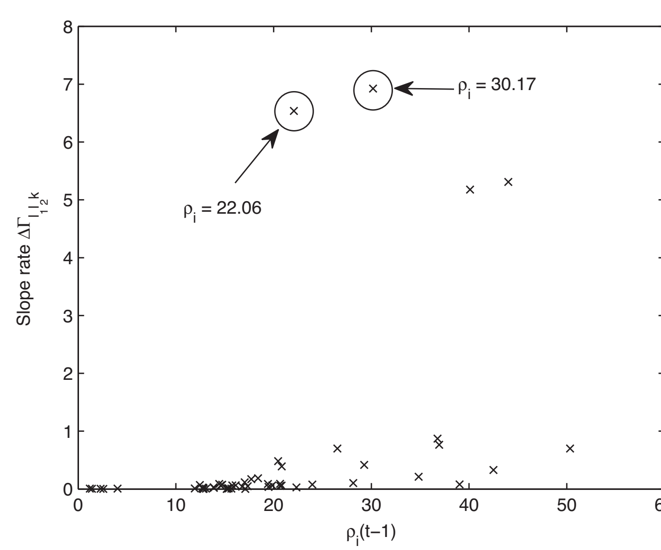

Fig. 8. Slope rates.

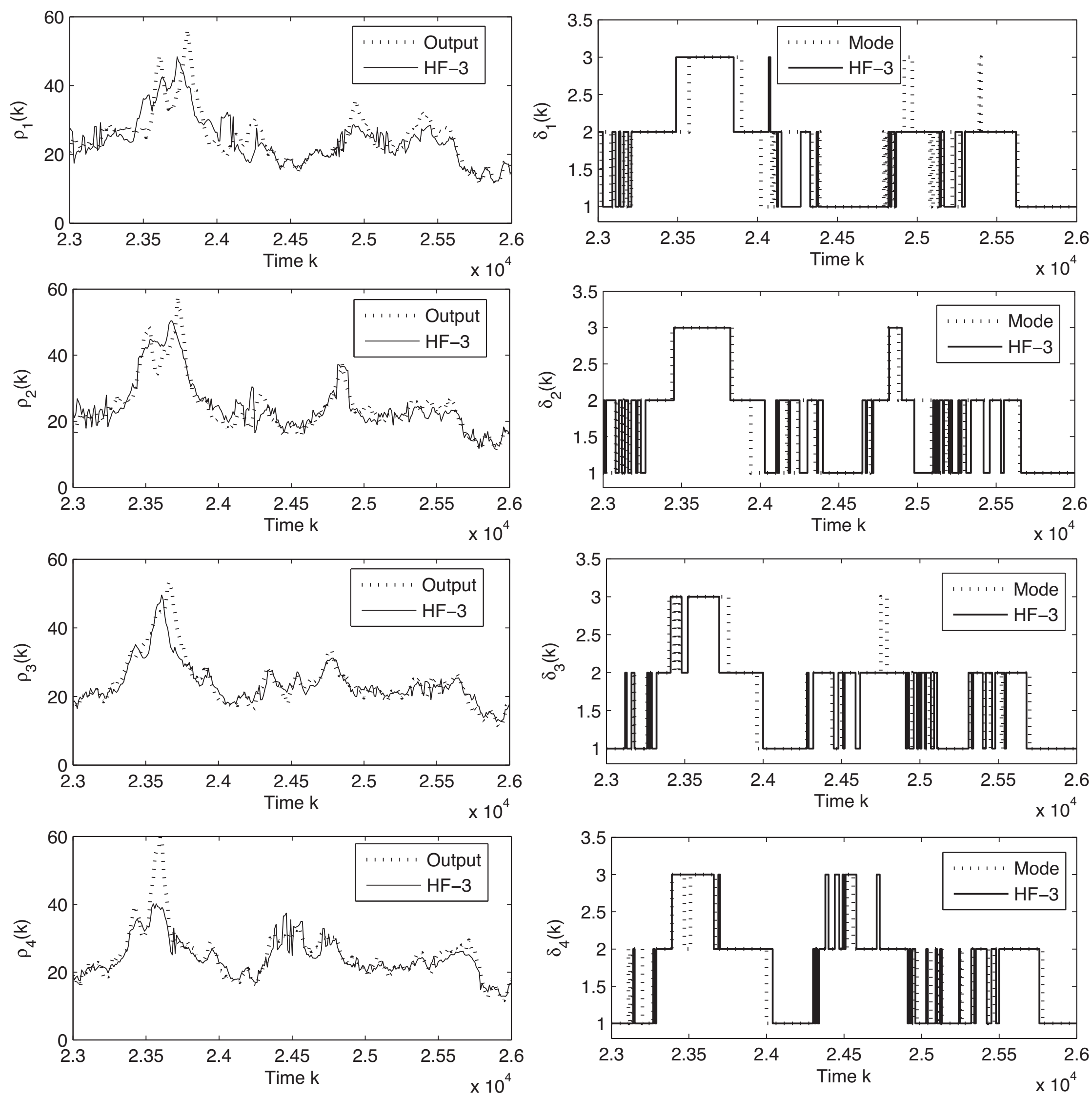

Fig. 9. Predicted densities and modes.

switching points is s_i. To evaluate the model for the four segments, the densities at the boundaries $\rho_0(k)$ and $\rho_5(k)$ are considered as inputs. For simplicity, the four hybrid-fuzzy models of the freeway will have the same discrete modes representation, thus $\delta_{i,p}(\rho_i(k)) = \delta_p(\rho_i(k))$. Next, results for the case $s_i = 3$ are presented.

In Fig. 8 the slope rates obtained for the 4 densities of the highway. Only the two higher slope rates are used as criterion to split the data.

From Fig. 8, the switches (sorted from the most relevant) are $\rho_i(k) = 30.17$ and $\rho_i(k) = 22.06$. In Fig. 9 the predicted density for each segment and the estimated modes $\delta_i(k) = \sum_{s=1}^{s_i} \delta_{i,s}(k)$ are shown. The most relevant part of the hybrid-fuzzy models is the estimation of the modes, not only because the method suggests a natural classification of the operational modes of the system, (in this case three modes, congested if $\rho_i(k) > 30.17$, free flow if $\rho_i(k) < 22.06$, otherwise transition mode), but also because the prediction model is good enough to estimate the densities in current mode.

5. Conclusions

In this paper a new identification method for non-linear hybrid systems that identifies discrete transitions by using only input–output data has been presented. A hybrid-fuzzy model was identified that consists of a local fuzzy level and a discrete/quantized level. Thus, the hybrid-fuzzy model incorporates explicitly the hybrid behavior of the process. Moreover, the method was implemented and applied to a hybrid tank system and a first-order traffic model. The use of principal component analysis demonstrated to be very useful in the detection of switching points. In the simulation results we obtain a small prediction error using the hybrid-fuzzy model. However, we must point out that the main advantage of hybrid-fuzzy modeling are the fuzzy rules with explicit information about the modes of the plant, which can be used directly to detect the discrete transitions of the system.

In further research, new approaches of hybrid-fuzzy identification will be analyzed such as a fuzzy clustering that generates the partitions (for the fuzzy rules together with the hybrid behavior). With the fuzzy clustering approach, the switching point candidates are located close to the border of the ellipsoid clusters. Therefore, further research should focus on different methods to characterize this phenomenon. Also, after the successful identification of the discrete modes, other models instead of fuzzy systems could be used to create the subsystem models (neural networks, linear models, PWA, among others). State-space model identification and estimation is also an interesting topic for this class of non-linear systems.

Online clustering, or learning methods could be also applied in a further stage.

Acknowledgments

This research has been supported by the European 7th Framework Network of Excellence "Highly complex and networked control systems (HYCON2)" Grant 257462, by the Millennium Institute "Complex Engineering Systems" (ICM: P-05-004-F, CONICYT: 522 FBO16), Fondecyt Chile Grant 1110047, CONICYT/FONDAP/15110019, and by the Ministry of Science, Higher Education and Technology of the Republic of Slovenia.

References

[1] B. De Schutter, W.P.M.H. Heemels, J. Lunze, C. Prieur, Survey of modeling, analysis, and control of hybrid systems, in: J. Lunze, F. Lamnabhi-Lagarrigue (Eds.), Handbook of Hybrid Systems Control – Theory, Tools, Applications, Cambridge University Press, Cambridge, UK, 2009, pp. 31–55 (Chapter 2).
[2] W.P.M.H. Heemels, B. De Schutter, J. Lunze, M. Lazar, Stability analysis and controller synthesis for hybrid dynamical systems, Philosophical Transactions of the Royal Society A 368 (1930) (2010) 4937–4960.
[3] A. Bemporad, M. Morari, Control of systems integrating logic, dynamics and constraints, Automatica 35 (3) (1999) 407–427.
[4] W.P.M.H. Heemels, B. De Schutter, A. Bemporad, Equivalence of hybrid dynamical models, Automatica 37 (7) (2001) 1085–1091.
[5] R. Babuška, H. Verbruggen, Neuro-fuzzy methods for nonlinear system identification, Annual Reviews in Control 27 (1) (2003) 73–85.
[6] S. Srivastava, M. Singh, M. Hanmandlu, A.N. Jha, New fuzzy wavelet neural networks for system identification and control, Applied Soft Computing 6 (1) (2005) 1–17.
[7] H. Du, N. Zhang, Application of evolving Takagi–Sugeno fuzzy model to nonlinear system identification, Applied Soft Computing 8 (1) (2008) 676–686.
[8] S. Beyhan, M. Alci, Fuzzy functions based ARX model and new fuzzy basis function models for nonlinear system identification, Applied Soft Computing 10 (2) (2010) 439–444.
[9] R.H. Abiyev, O. Kaynak, T. Alshanableh, F. Mamedov, A type-2 neuro-fuzzy system based on clustering and gradient techniques applied to system identification and channel equalization, Applied Soft Computing 11 (1) (2011) 1396–1406.
[10] A. Juloski, W.P.M.H. Heemels, G. Ferrari-Trecate, R. Vidal, S. Paoletti, J.H.G. Niessen, Comparison of Four Procedures for the Identification of Hybrid Systems, in: M. Morari, L. Thiele (Eds.), Hybrid Systems: Computation and Control, Lecture Notes in Computer Science, 3414, Springer, Berlin Heidelberg, 2005, ISBN 978-3-540-25108-8, pp. 354–369, http://dx.doi.org/10.1007/978-3-540-31954-2_23.
[11] E.F. Camacho, D.R. Ramirez, D. Limon, D. Muñoz de la Peña, T. Alamo, Model predictive control techniques for hybrid systems, Annual Reviews in Control 34 (1) (2010) 21–31.
[12] G. Ferrari-Trecate, M. Muselli, D. Liberat, M. Morari, A clustering technique for the identification of piecewise affine systems, Automatica 39 (2) (2003) 205–217.
[13] Y. Ma, R. Vidal, Identification of Deterministic Switched ARX Systems via Identification of Algebraic Varieties, in: M. Morari, L. Thiele (Eds.), Hybrid Systems: Computation and Control, Lecture Notes in Computer Science, 3414, Springer, Berlin Heidelberg, 2005, ISBN 978-3-540-25108-8, pp. 449–465, http://dx.doi.org/10.1007/978-3-540-31954-2_29.
[14] A. Juloski, S. Weiland, W. Heemels, Bayesian approach to identification of hybrid systems, IEEE Transactions on Automatic Control 50 (10) (2005) 1520–1533.
[15] A. Bemporad, A. Garulli, S. Paoletti, A. Vicino, A bounded-error approach to piecewise affine system identification, IEEE Transactions on Automatic Control 50 (10) (2005) 1567–1580.
[16] H. Nakada, K. Takaba, T. Katayama, Identification of piecewise affine systems based on statistical clustering technique, Automatica 41 (5) (2005) 905–913.
[17] M.E. Gegundez, J. Aroba, J.M. Bravo, Identification of piecewise affine systems by means of fuzzy clustering and competitive learning, Engineering Applications of Artificial Intelligence 21 (8) (2008) 1321–1329.
[18] F. Lauer, G. Bloch, R. Vidal, Nonlinear hybrid system identification with kernel models, in: Proceedings of the 49th IEEE Conference on Decision and Control (CDC), Atlanta, USA, 2010, pp. 696–701.
[19] R.J. Hathaway, J.C. Bezdek, Switching regression models and fuzzy clustering, IEEE Transactions on Fuzzy Systems 1 (3) (1993) 195–204.
[20] A. Trabelsi, F. Lafont, M. Kamoun, G. Enea, Fuzzy identification of a greenhouse, Applied Soft Computing 7 (3) (2007) 1092–1101.
[21] A. Tewari, M. Macdonald, Knowledge-based parameter identification of TSK fuzzy models, Applied Soft Computing 10 (2) (2010) 481–489.
[22] A. Celikyilmaz, I. Burhan, Enhanced fuzzy system models with improved fuzzy clustering algorithm, IEEE Transactions on Fuzzy Systems 16 (3) (2008) 779–794.
[23] S. Blažič, I. Škrjanc, S. Gerkšič, G. Dolanc, S. Strmčnik, M. Hadkiski, A. Stathaki, Online fuzzy identification for an intelligent controller based on a simple platform, Engineering Applications of Artificial Intelligence 22 (4) (2009) 628–638.
[24] R.E. Precup, H. Hellendoorn, A survey on industrial applications of fuzzy control, Computers in Industry 62 (3) (2011) 213–226.
[25] S. Tong, H.Z. Li, W. Wang, Observer-based adaptive fuzzy control for SISO nonlinear systems, Fuzzy Sets and Systems 148 (3) (2004) 355–376.
[26] Zs. Lendek, R. Babuška, B. De Schutter, Sequential stability analysis and observer design for distributed TS fuzzy systems, Fuzzy Sets and Systems 174 (1) (2011) 1–30.
[27] S. Tong, H.Z. Li, G. Chen, Adaptive fuzzy control for decentralized control for a class of large-scale nonlinear systems, IEEE Transactions on Systems, Man and Cybernetics, Part B 34 (1) (2004) 770–775.
[28] S. Tong, H.Z. Li, Fuzzy adaptive sliding-mode control for MIMO nonlinear systems, IEEE Transactions on Fuzzy Systems 11 (3) (2003) 354–360.
[29] S. Nefti, M. Oussalah, U. Kaymak, A new fuzzy set merging technique using inclusion-based fuzzy clustering, IEEE Transactions on Fuzzy Systems 16 (1) (2008) 145–161.
[30] M. Hadjili, V. Wertz, Takagi–Sugeno fuzzy modeling incorporating input variables selection, IEEE Transactions on Fuzzy Systems 10 (6) (2002) 728–742.
[31] H. Roubos, M. Setnes, Compact and transparent fuzzy models and classifiers through iterative complexity reduction, IEEE Transactions on Fuzzy Systems 9 (4) (2001) 516–524.
[32] E. Kim, M. Park, S. Ji, M. Park, A new approach to fuzzy modeling, IEEE Transactions on Fuzzy Systems 5 (3) (1997) 328–337.
[33] J. Abonyi, R. Babuška, F. Szeifert, Modified Gath–Geva fuzzy clustering for identification of Takagi–Sugeno fuzzy models, IEEE Transactions on Systems, Man, and Cybernetics, Part B 32 (5) (2002) 612–621.
[34] C. Li, J. Zhou, X. Xiang, Q. Li, X. An, T–S fuzzy model identification based on a novel fuzzy c-regression model clustering algorithm, Engineering Applications of Artificial Intelligence 22 (4–5) (2009) 646–653.
[35] W.J. Wang, Y.J. Chen, C.H. Sun, Relaxed stabilization criteria for discrete–time T–S fuzzy control systems based on a switching fuzzy model and piecewise Lyapunov function, IEEE Transactions on Systems, Man, Cybernetics – Part B 37 (3) (2007) 551–559.
[36] A. Núñez, D. Sáez, S. Oblak, I. Škrjanc, Fuzzy-model-based hybrid predictive control, ISA Transactions 48 (1) (2009) 24–31.
[37] R. Palm, D. Driankov, Fuzzy switched hybrid systems – modeling and identification, in: Proceedings of the 1998 IEEE ISCI/CIRA/SAS Joint Conference, Gaithersburg, MD, USA, 1998, pp. 130–135.
[38] D. Girimonte, R. Babuška, Structure for nonlinear models with mixed discrete and continuous inputs: a comparative study, in: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, The Hague, The Netherlands, 2004, pp. 2392–2397.
[39] X. Zeng, J. Goulermas, P. Liatsis, D. Wang, J. Keane, Hierarchical fuzzy systems for function approximation on discrete input spaces with application, IEEE Transactions on Fuzzy Systems 16 (5) (2008) 1197–1215.
[40] S. Cheng, R. Dong, W. Pedrycz, A framework of fuzzy hybrid systems for modelling and control, International Journal of General Systems 39 (2) (2010) 165–176.
[41] D.E. Gustafson, W.C. Kessel, Fuzzy clustering with a fuzzy covariance matrix, in: Proceedings of the IEEE Conference on Decision and Control including the 17th Symposium on Adaptive Processes, San Diego, CA, USA, 1978, p. 761 L-766.
[42] R. Babuška, Fuzzy Modelling for Control, Kluwer Academic Publishers, Boston, 1998.
[43] T. Takagi, M. Sugeno, Fuzzy identification of systems and its applications to modeling and control, IEEE Transactions on Systems, Man, Cybernetics 15 (1) (1985) 116–132.
[44] H. Witsenhausen, A class of hybrid-state continuous time dynamic systems, IEEE Transactions on Automatic Control 11 (2) (1966) 161–167.
[45] S. Kozak, J. Stevek, Improved piecewise linear approximation of nonlinear functions in hybrid control, in: Proceedings of the 18th IFAC World Congress, Milano, Italy, 2011, pp. 14982–14987.
[46] A. Kroll, S. Soldan, On data-driven Takagi–Sugeno modeling of heterogeneous systems with multidimensional membership functions, in: Proceedings of the 18th IFAC World Congress, Milano, Italy, 2011, pp. 14994–14999.
[47] K. Tanaka, M. Iwasaki, H. Wang, Switching control of an R/C hovercraft: stabilization and smooth switching, IEEE Transactions on Systems, Man, Cybernetics 31 (6) (2001) 853–863.
[48] N. Elfelly, J.Y. Dieulot, M. Benrejeb, P. Borne, A new approach for multimodel identification of complex systems based on both neural and fuzzy clustering algorithms, Engineering Applications of Artificial Intelligence 23 (7) (2010) 1064–1071.
[49] U. Kaymak, R. Babuška, Compatible cluster merging for fuzzy modeling, in: Proceedings of the FUZZ-IEEE/IFES95, Yokohama, Japan, 1995, pp. 897–904.
[50] S.K. Fan, N.C. Yao, Y.J. Chang, C.H. Jen, Statistical monitoring of nonlinear profiles by using piecewise linear approximation, Journal of Process Control 21 (8) (2011) 1217–1229.
[51] D. Saez, A. Cipriano, A new method for structure identification of fuzzy models and its application to a combined cycle power plant, Engineering Intelligent Systems for Electrical Engineering and Communications 9 (2) (2001) 101–107.

[52] T.A. Johansen, R. Babuška, Multiobjective identification of Takagi Sugeno fuzzy models, IEEE Transactions on Fuzzy Systems 11 (6) (2003) 847–860.
[53] G. Karer, G. Music, I. Škrjanc, B. Zupancic, Hybrid fuzzy model-based predictive control of temperature in a batch reactor, Computers & Chemical Engineering 31 (12) (2007) 1552–1564.
[54] S. Jakubek, C. Hametner, N. Keuth, Total least squares in fuzzy system identification: an application to an industrial engine, Engineering Applications of Artificial Intelligence 21 (8) (2008) 1277–1288.
[55] C.F. Daganzo, The cell transmission model. Part II: Network traffic, Transportation Research Part B: Methodological 29 (2) (1995) 79–93.