ONLINE CLASSIFICATION OF SWITCHING MODELS BASED ON SUBSPACE FRAMEWORK

K. M. Pekpe * **S. Lecœuche** **

* *LAGIS - UMR CNRS 8146*
Polytech'Lille - IMA 2 Bd Langevin, 59 655 Villeneuve d'Ascq
Phone: +33(0)320434565, Fax : +33(0)320337189
Email : Midzodzi.Pekpe@polytech-lille.fr
** *LAGIS - UMR CNRS 8146*
DØpartement G.I.P., Mines de Douai, 59508 Douai Cedex
Phone: +33(0)327712445, Fax : +33(0)327712980
Email: lecoeuche@ensm-douai.fr

Abstract: The paper deals with the modelling of switching systems and focuses on the characterization of the local functioning modes using online clustering approach. The considered system is represented as a weighted sum of local linear models where each model could have its own structure. That implies that the parameters and the order of the switching system could change when the system switches. The presented method consists in two steps. First, an online estimation method of the Markov parameters matrix of the local linear models is established. Secondly, the labelling of theses parameters is done using a dynamical decision space worked out with learning techniques, each local model being represented by a cluster. The paper ends with an example, in view to illustrate the method performances.

1. INTRODUCTION

An online classification method for switching systems is proposed in this paper. Switching systems are a particular class of hybrid systems which can be considered as the weighted sum of the linear local models (or local models) with normalized weights. In this modelling, only one model is active at each time (binary weights). On the one hand, several methods are proposed for offline clustering of the hybrid systems (Breiman, 1993), (Bemporad *et al.*, 2001), (Ferrari-Trecate *et al.*, 2003), and (Pekpe *et al.*, 2004). A clustering method based on hyperplane determination is proposed in (Breiman, 1993), but this method consists of the estimation of two hyperplanes continuously joined together. The method proposed in (Bemporad *et al.*, 2001) uses a mixed integer programming which is NP-hard in the worst case, then this is practically applicable only when the number of the data is very small (see (Ferrari-Trecate *et al.*, 2003)). Moreover, the order of the system cannot change in the two above methods. In 2004, (Vidal, 2004) proposes an identification of pwarx hybrid models with unknown and different orders. Also, a recent offline method is proposed in (Pekpe *et al.*, 2003) for data classification and parameter identification of switching state-space models with variable structure. This method uses a change detection technique to estimate the switches and a subspace method to estimate the Markov parameters (state representation) of the local models.

On the other hand, online data classification does not seem to have a great attention in hybrid systems community. Recently, (Lecœuche *et al.*, 2006) presents an approach based on the combination of a recursive identification technique and a classifier set for non stationary environment. But, this approach is limited to non-stationnary systems with a fixed structure.

The method proposed in this paper supposes that the number and the order of the local linear models are unknown/ Moreover, the order could change for each local model. In fact, no *a priori* knowledge, even the number of models, is required, the knowledge on the switching model is gained using online estimation and continuous learning. The approach developped here is based on a similar way than (Lecœuche *et al.*, 2006) : one stage for the estimation and one stage for the classification.

The paper structure is as follow. After presenting the overall formulation of the problem in the second section, the third and fourth sections are respectively dedicated to the presentation of the online estimation method and to the online clustering technique. The last section presents first results based on simulated switching system.

2. PROBLEM FORMULATION

The output of the switching system is represented as a weighted sum of the outputs of h local models:

$$y_k = \sum_{s=1}^{h} p_{s,k} . y_{s,k} \tag{1}$$

where the scalar k represents the time index, the vector $y_k \in R^{\ell}$ the output of the system, the vector $y_{s,k} \in R^{\ell}$ the output of local model s and the scalar $p_{s,k}$ the weight associated to $y_{s,k}$. For each time k, the weights verify the following condition :

$$p_{s,k} \in \{0,1\} \text{ and } \sum_{s=1}^{h} p_{s,k} = 1, \ \forall k \tag{2}$$

Each local model is supposed to be linear and described by the state space model of order n_s :

$$\begin{aligned} x_{s,k+1} &= \mathbf{A_s} x_{s,k} + \mathbf{B_s} u_k + v_{s,k} \\ y_{s,k} &= \mathbf{C_s} x_{s,k} + \mathbf{D_s} u_k + w_{s,k} \end{aligned} \tag{3}$$

where the process noises $v_{s,k} \in R^{n_s}$ and the measurement noises $w_{s,k} \in R^{\ell}$ of the s^{th} local model are zero mean white noises which are uncorrelated with the inputs $u_k \in R^m$. The vector $x_{s,k} \in R^{n_s}$ represents the state vector of model s.

The h local models are assumed to be stable and being active during a minimal time τ (Hespanha and Morse, 1999), this parameter is called the dwell time.

From available measurements of inputs u_k and outputs y_k, our goal is to characterize the current functioning mode using online classification of the regressors vectors and, in other words, to estimate the weight $p_{s,k}$.

This task is done in two stages described hereafter :

- *Online local model Markov parameters estimation*
First, the online estimation of the Markov parameters matrix of the local models is fulfilled. This black-box estimation uses the state space representation, which is particularly adapted to MIMO processes and does not require any canonical parameterization. Moreover, the use of the Markov parameters suits for the estimation of the system parameters with variable structure or order. The presented method is based on the FIR modelling of the local models. It is shown that the regressors of a FIR model lie in the hyperplane which orthogonal matrix is the Markov parameters matrix. Markov parameters of each local model are estimated by least squares method from data derived from a sliding window.

- *Determination of the current functioning mode*
This classification is based on a dynamical decision space obtained by online learning techniques. The novelty of the proposed approach consists in exploiting a specific clustering technique making possible the continuous modelling of the functioning modes.These ones are modelled by an online neural network technique (Lecœuche and Lurette, 2003). When a new observation is presented at this algorithm (new estimation of Markov parameters matrix), the decision space is updated according to the information brought by this observation and the current functioning mode is determined (the closest linear local model). Then, the characterization of the current functioning mode is given in term of membership degree of the identified parameters vector to updated classes representing the actual local linear modes.

These two stages will be more precisely presented in the following two sections.

3. LOCAL MODELS PARAMETERS ESTIMATION

The online estimation of the local models Markov parameters matrix is discussed in this section. First, it establishes that the regressors of one local model belong to the same hyperplane and an orthogonal matrix of this hyperplane is the Markov parameters matrix. Then, the online estimation of the Markov parameters matrix by least squares method is proposed.

3.1 *Local model hyperplane equation*

The aim of this paragraph is to perform the online estimation of the Markov parameters of the local models. To reach this goal, it is established first that all the regressors built from the data resulting from the same local model belong to the same hyperplane. This could be proved from the expression of the local model output according to the state and the inputs multiplied by the Markov parameters. This expression being developed, the state influence is deleted by weighting it by a high power of the local state matrix ($\mathbf{A}_s$) which is supposed to be steady. This implies that the local model output is equal to the inputs multiplied by the Markov parameters, similar to the approximation by the FIR model. This equality can be rewritten as the orthogonality of the regressors (which contains the local model output and the inputs) and the Markov parameters matrix.

In the following, it is supposed that the system stays in each mode during a minimal time (or dwell time) τ.

The s^{th} local model being active, its output can be expressed (see relation (3)) as:

$$\begin{aligned} y_{s,k} = \mathbf{C}_s \mathbf{A}_s^{i-2} \mathbf{B}_s u_{k-i+1} + \ldots + \mathbf{C}_s \mathbf{B}_s u_{k-1} \\ + \mathbf{D}_s u_k + \mathbf{C}_s \mathbf{A}_s^{i-2} v_{s,k-i+1} + \ldots + \mathbf{C}_s v_{s,k-1} + w_{s,k} \\ + \mathbf{C}_s \mathbf{A}_s^{i-1} x_{s,k-i+1} \end{aligned} \tag{4}$$

As the local model is assumed to be stable, the term $\mathbf{C}_s\mathbf{A}_s^{i-1}x_{s,k-i}$ could be neglected for high values of "i". This term is considered negligible if:

$$\left\|\mathbf{C}_s\mathbf{A}_s^{i-1}x_{s,k-i+1}\right\| \leq \sqrt{variance(w_{s,k})} \tag{5}$$

Thus, (4) becomes:

$$y_{s,k} = \mathbf{C}_s\mathbf{A}_s^{i-2}\mathbf{B}_s u_{k-i+1} + \ldots + \mathbf{C}_s\mathbf{B}_s u_{k-1} + \mathbf{D}_s u_k + \beta_{s,k} \tag{6}$$

with:

$$\beta_{s,k} = \mathbf{C}_s\mathbf{A}_s^{i-1}x_{s,k-i+1} + \mathbf{H}_{s,i}^{v}\bar{v}_{s,k-1} + w_{s,k} \tag{7}$$

where

$$\begin{array}{l} \mathbf{H}_{s,i}^{v} = \left(\mathbf{C}_s\mathbf{A}_s^{i-2} \;\; \mathbf{C}_s\mathbf{A}_s^{i-3} \;\; \ldots \;\; \mathbf{C}_s\mathbf{A} \;\; \mathbf{C}_s \right) \in R^{\ell\times m(i-1)} \\ \bar{v}_{s,k-1} = \left(v_{s,k-i+1}^T \;\; \ldots \;\; v_{s,k-2}^T \;\; v_{s,k-1}^T \right)^T \in R^{\ell(i-1)\times n_s} \end{array} \tag{8}$$

the vector $\beta_{s,k}$ is the perturbation due to the noises and the approximation of the local model by a FIR model. The previous equation can be written as:

$$y_{s,k} = \mathbf{H}_{s,i}^{v}\left(u_{k-i+1}^T \;\; \ldots \;\; u_{k-1}^T \;\; u_k^T \right)^T + \beta_k \tag{9}$$

Indeed, equation (9) gives the FIR approximation of the local model, it can be rewritten as :

$$\mathbf{M}_s z_{s,k} = \beta_k \tag{10}$$

where $z_{s,k}$ is the regressors vector of the local model:

$$z_{s,k} = \left(u_{k-i+1}^T \;\; \ldots \;\; u_{k-1}^T \;\; u_k^T \;\; y_{s,k}^T \right)^T \in R^{(mi+\ell)} \tag{11}$$

and $\mathbf{M}_s$ is a orthogonal matrix of the regressors, this matrix is called also as the augmented Markov parameters matrix:

$$\mathbf{M}_s = \left(\mathbf{H}_{s,i} \;\; -\mathbf{I}_\ell \right) \in R^{n_s\times(mi+\ell)} \tag{12}$$

where $H_{s,i} \in R^{\ell\times mi}$ is the Markov parameter matrix and defined as:

$$\mathbf{H_{s,i}} = \left(\mathbf{C}_s\mathbf{A}_s^{i-2}B_s \;\; \mathbf{C}_s\mathbf{A}_s^{i-3}\mathbf{B}_s \;\; \ldots \;\; \mathbf{C}_s\mathbf{B}_s \;\; \mathbf{D}_s \right) \tag{13}$$

The sub-script s denotes the index of the active local model, this sub-script disappears if the local model index is not indicated. In the deterministic case, all the regressors $z_{s,k}$ lie in the hyperplane:

$$\mathbf{M}_s z_{s,k} = 0 \tag{14}$$

But this relation changes in the presence of the noises, that makes the modelling problem more difficult.

3.2 *Online estimation of the local models Markov parameters*

The local model hyperplane equation has been established in the previous paragraph, least squares estimation of a orthogonal matrix of the hyperplane is proposed now. From the hyperplane equation (10), the orthogonal matrix ($\mathbf{M}_s$) is estimated. This estimation can be done if there are mi independent regressors $z_{s,k}$. These regressors are independent if the inputs are persistently excited of order mi. Consider the regressors matrices $\underline{z}_{s,k}$:

$$\underline{z}_{s,k} = \left(z_{s,k-\rho+1} \;\; \ldots \;\; z_{s,k+1} \;\; z_{s,k} \right) \tag{15}$$

$$\underline{z}_{s,k} = \begin{pmatrix} \mathbf{U}_k \\ \underline{y}_{s,k} \end{pmatrix} \tag{16}$$

with

$$\underline{y}_{s,k} = \left(y_{s,k-\rho+1} \;\; \ldots \;\; y_{s,k-1} \;\; y_{s,k} \right) \in \mathbb{R}^{\ell\times\rho} \tag{17}$$

$$\mathbf{U}_k = \begin{pmatrix} u_{k-\rho-i+2} & \cdots & u_{k-i} & u_{k-i+1} \\ u_{k-\rho-i+3} & \cdots & u_{k-i+1} & u_{k-i+2} \\ \vdots & \ldots & \vdots & \vdots \\ u_{k-\rho+1} & \cdots & u_{k-1} & u_k \end{pmatrix} \in R^{mi\times\rho} \tag{18}$$

and ρ is an integer equals or greater than the number of the row (mi) of the matrix $\mathbf{U}_k$. The unique matrix which is orthogonal (and defined by equation (12)) to this regressors matrix is determined. To estimate the matrix $\mathbf{M}_s$, it is enough to determine the Markov parameters matrix $\mathbf{H}_{s,i}$ (12). The following theorem gives the least squares estimation of this matrix.

Theorem 3.1. Under the following conditions :

- the matrix $\mathbf{U}_k \in R^{mi\times\rho}$ is full row rank,
- the local model s is stable,

the Markov parameters matrix is given by :

$$\underline{y}_{s,k}\mathbf{U}_k^T(\mathbf{U}_k\mathbf{U}_k^T)^{-1} = \mathbf{H}_{s,i} + \underline{\mathbf{B}}_{s,k}\mathbf{U}_k^T(\mathbf{U}_k\mathbf{U}_k^T)^{-1} \tag{19}$$

where the perturbation matrix $\underline{\mathbf{B}}_{s,k}$ is defined as:

$$\underline{\mathbf{B}}_{s,k} = \left(\beta_{s,k-\rho+1} \;\; \ldots \;\; \beta_{s,k-1} \;\; \beta_{s,k} \right) \in \mathbb{R}^{\ell\times\rho} \tag{20}$$

The mathematic expectation of the matrix $\underline{y}_{s,k}\mathbf{U}_k^T(\mathbf{U}_k\mathbf{U}_k^T)^{-1}$ is:

$$\begin{aligned} \mathbf{E}[\underline{y}_{s,k}\mathbf{U}_k^T(\mathbf{U}_k\mathbf{U}_k^T)^{-1}] &= \mathbf{E}[\mathbf{H}_{s,i}] + \mathbf{E}[\underline{\mathbf{B}}_{s,k}\mathbf{U}_k^T(\mathbf{U}_k\mathbf{U}_k^T)^{-1}] \\ &= \mathbf{H}_{s,i} + \mathbf{E}[\underline{\mathbf{B}}_{s,k}]\mathbf{U}_k^T(\mathbf{U}_k\mathbf{U}_k^T)^{-1} \end{aligned}$$

whereas (see equation (7)):

$$\mathbf{E}[\underline{\mathbf{B}}_{s,k}] = \mathbf{C}_s\mathbf{A}_s^{i-1}\left(x_{s,k-\rho-i+2} \;\; \ldots . \;\; x_{s,k-\rho-i+2} \;\; x_{s,k-\rho-i+1} \right)$$

which can be neglected if integer i is great enough, or be considered as a deterministic perturbation if the term $\left\|\mathbf{C}_s\mathbf{A}_s^{i-1}x_{s,k-i+1}\right\|$ is not negligible.

If the consecutive regressors

$$z_k = \left(u_{k-i+1}^T \;\; \ldots \;\; u_{k-1}^T \;\; u_k^T \;\; y_k^T \right)^T,\; k = 1,\ldots\rho,\; \rho \geqslant mi \tag{21}$$

are generated by the same local model "s" (i.e. $y_k = y_{s,k}$) then the matrices $\underline{y}_{s,k}$ (resp. $\underline{z}_{s,k}$) and $\underline{y}_k$ (resp. $\underline{z}_k$) are equal. That suppose the dwell time τ should be greater than $i+\rho$ (see figure (1)). In this case the local output matrix (for local model s) $\underline{y}_{s,k}$ can be replace in the theorem by the global output matrix $\underline{y}_{s,k}$

$$\underline{y}_k = \left(y_{k-\rho+1} \;\; \ldots \;\; y_{k-1} \;\; y_k \right) \in \mathbb{R}^{\ell\times\rho} \tag{22}$$

If the matrix $\underline{y}_k$ is built with a data from two local models then the system is in transient mode. In the next section, the clustering method used to determine the class of the estimated matrix is given.

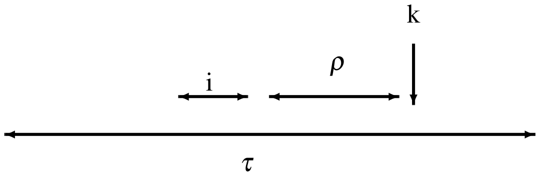

Fig. 1. Illustration of the dwell time

4. CLUSTERING METHOD

The modes modelling tool treats the data extracted from the identified Markov parameters matrices as they arrive. The new information is incorporated continuously in order to redefine the structure of the functioning modes (also named linear local models) and thus to model continuously the decision space. The taking into account of the various situations related to non stationarity environment requires the setting of specific adaptation rules through a continuous learning process. In the area of machine learning, some techniques exist with architectures exploiting incremental learning (Deng and Kasabov, 2003), (Eltoft and deFigueiredo, 1998), (Mouchaweh, 2004). Most of these algorithms present some disadvantages related to a coarse classes modelling and/or limited adaptation capacities in non stationary environment. In order to fill these gaps, two neural algorithms for the dynamic classification of the evolutionary data have been previously developed (Lecœuche and Lurette, 2003; Amadou-Boubacar and Lecœuche, 2005). These algorithms use a multi-prototype approach making possible to accurately model the structure of complex classes. In this paper, the AUDyC network based on a Gaussian modelling is used. Each functioning mode corresponds to a label of a complex class Ω which could be defined by an assembly of Gaussian prototypes Φ. The activation function of each hidden neuron determines the membership degree μ_j^t of the observation X_t to the prototype Φ_j. In order to obtain a fine classes representation, this one is based on the Mahanalobis distance

$$\mu_j^t = \exp\left(\frac{1}{2}\left(X_t - \overline{X}_j\right)^T * \Sigma_j^{-1} * \left(X_t - \overline{X}_j\right)\right) \quad (23)$$

where $\overline{X}_j$ and Σ_j are respectively the center and the covariance matrix of the prototype. The use of the membership function allows the implementation of the learning rules. With the first acquisition X_1, the network is initialized: creation of the first prototype Φ_1 constituting the first class Ω_1 (first functioning mode). The prototype is parameterized by its center $\overline{X}_1 = X_1$ and an initial covariance matrix Σ_{ini} beforehand selected. Then, according to new acquisitions, various situations can arise by comparison of the membership degree with two fixed thresholds μ_{min} and μ_{max} (resp. limit of prototype and class membership). Each case leads to a specific procedure (see Table 1).

Then, the AUDyC learning process is established in three principal phases:

Table 1. Classifier Adaptation rules.

	If	Then
1	$\mu_j^t < \mu_{min} \forall j \in \{1..J\}$	Creation $\Phi_{new} \in \Omega_{new}$
2	$\exists \Phi_j \in \Omega_i, \mu_{min} < \mu_j^t < \mu_{max}$	Creation $\Phi_{new} \in \Omega_i$
3	$\exists \Phi_j, \mu_{max} < \mu_j^t$	Adaptation $\widehat{\Phi}_j = \Phi_j(X_t)$
4	$\exists \Phi_l \in \Omega_p \cup$ $\exists \Phi_m \in \Omega_q, \mu_{min} < \mu_{l,m}^T < \mu_{max}$	Ambiguity $X_t \in \chi_{amb}$

4.1 *First phase: classification*

The classification stage corresponds to the creation and adaptation of prototypes and classes. In cases 1 and 2 of table 1, the observation X_t is not close to any existing prototype. These cases are similar to a distance rejection which could be used to detect the novelty in the multiclass environment. If the observation is not sufficiently close to any class (case 1), it leads to the creation of a new prototype and a new class corresponding to a new system mode. In the case 2, a new prototype is created and affected to the nearest class in order to contribute to a better definition of the mode model. In situation 3, the observation is rather close to a prototype to take part in its definition. The functioning mode adaptation is then carried out by using the following recursive equations

$$\overline{X}_j^t = \overline{X}_j^{old} + \frac{1}{N_p}\left(X_t - X_{t-N_p+1}\right) \quad (24)$$

$$\Sigma_j^t = \Sigma_j^{old} + \Delta X \begin{pmatrix} \frac{1}{N_p} & \frac{1}{N_p(N_p-1)} \\ \frac{1}{N_p(N_p-1)} & \frac{-(N_p+1)}{N_p(N_p-1)} \end{pmatrix} \Delta X^T \quad (25)$$

with $\Delta X = \left[X_t - \overline{X}_j^{old} \quad X_{i-N+1} - \overline{X}_j^{old}\right]$, N_p : prototype size.

4.2 *Second phase: fusion*

The case 4 of table 1 depicts the case of the rejection in ambiguity when an observation is sufficiently close to two or several prototypes (e.g. l, m) to contribute to their structure. The fusion procedure consists in evaluating the similarity of two densities by using an acceptance criterion based on the Kullback-Leibler distance (Zhou and Chellappa, 2004). When this criterion is higher than a threshold, the different classes (e.g. p, q) merge onto a unique new functioning mode.

4.3 *Third phase: evaluation*

The evaluation phase is significant to eliminate the parasite prototypes and classes possibly created by the noise influence. To detect not-representative modes, this phase is based on the cardinality of the models.

For more details on the AUDyC network, the reader can consult (Lecœuche and Lurette, 2003), (Lurette, 2003) and (Amadou-Boubacar *et al.*, 2005) which give theoretical and practical analysis of the AUDyC.

5. ILLUSTRATION EXAMPLE

Consider a switching system which is the sum of three models. The state-space model is used and each matrix ($\mathbf{A}_s, \mathbf{B}_s, \mathbf{C}_s$ and $\mathbf{D}_s$) is specific to each local model. The matrices of the local models are defined below :

$$\mathbf{A}_1 = \begin{pmatrix} 0{,}4 & 0{,}1 & 0 \\ 0{,}8 & 0{,}4 & 0 \\ 0 & 0 & 0{,}8 \end{pmatrix}, \mathbf{B}_1 = \begin{pmatrix} 1{,}5 & 0{,}9 \\ 1 & -1 \\ -1{,}5 & 2{,}3 \end{pmatrix}$$
$$\mathbf{C}_1 = \begin{pmatrix} 0{,}8 & 1{,}1 & 2 \\ -1{,}3 & 0{,}7 & 1{,}7 \\ 1{,}5 & 0{,}7 & -0{,}9 \end{pmatrix}, \mathbf{D}_1 = \begin{pmatrix} 0{,}4 & 0{,}8 \\ -0{,}6 & 1{,}4 \\ 1{,}3 & -0{,}75 \end{pmatrix}$$

$$\mathbf{A}_2 = \begin{pmatrix} 0{,}4 & 0{,}6 \\ 0{,}5 & 0{,}1 \end{pmatrix}, \mathbf{B}_2 = \begin{pmatrix} 1{,}5 & 0{,}9 \\ 1 & -1 \end{pmatrix}$$
$$\mathbf{C}_2 = \begin{pmatrix} 0{,}8 & 1{,}1 \\ -1{,}3 & 0{,}7 \\ 1{,}5 & 0{,}7 \end{pmatrix}, \mathbf{D}_2 = \begin{pmatrix} 0{,}4 & 0{,}8 \\ -0{,}6 & 1{,}4 \\ 1{,}3 & -0{,}75 \end{pmatrix}$$

$$\mathbf{A}_3 = \begin{pmatrix} 0{,}3 & 0{,}2 & 0 \\ 0{,}8 & 0{,}2 & 0 \\ 0 & 0 & -0{,}75 \end{pmatrix}, \mathbf{B}_3 = \mathbf{B}_1$$
$$\mathbf{C}_3 = \mathbf{C}_1, \mathbf{D}_3 = \mathbf{D}_1$$

The changes of the system dynamics are summarized as follows: $s = 1$ over the intervals $[1, 099]$, $[300, 399]$ and $[600, 699]$, $s = 2$ over the intervals $[100, 199]$, $[400, 499]$ and $[700, 800]$ and $s = 3$ over the intervals $[200, 299]$ and $[500, 599]$. The figure 2 gives the index of the active local model according to the time.

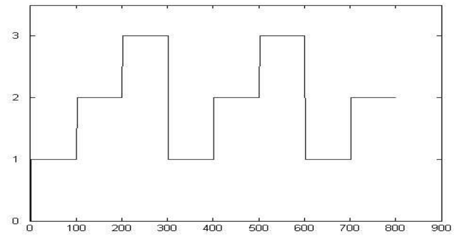

Fig. 2. The index of the active local model

The inputs u_k are Pseudo-Random Binary Sequences (PRBS) with variable amplitudes. The signal to noise ratio of the outputs with respect to the measurement noises is 25db. The process noises v_k are a Gaussian white noises and have a covariance $var(v_k) \simeq 7 \times 10^{-4} I_3$; the covariance matrix of the inputs is $var(u_k) \simeq 7 \times 10^{-2} I_2$. During the simulation, 900 input-output data have been processed on the system.

From, these IO data, the estimation of the Markov parameters matrix on slidding window is achieved by theorem *3.1*. The size of the sliding window ρ is chosen equal 35 and the integer i is fixed equal 13.

These identified parameters are sent to the clustering tool. The values of the identified matrix $\hat{\mathbf{H}}_{\mathbf{s},\mathbf{i}}$ constitute the X_t observation vector. In fact, this vector consists of all the monitored parameters. According to the complexity of the application, its dimension could be reduced (Markov parameters selection or reduction) or increased by adding, for example, complementary physical information. In this paper, in order to present tight and comprehensible results, we have applied a space reduction in order to obtain a 3D representation.

From the X_t information, the dynamical classifier updates the decision space (functioning modes models) and determines the current functioning mode. For the whole of this study, the AUDyC parameters are fixed as follows: $\Sigma_{ini} = 2.5, \mu_{min} = 0.1, \mu_{max} = 0.3, NP_{min} = 40, N_P = 500$ and $N_{amb} = 5$. For more information about the choice of the parameters, the reader can refer to (Amadou-Boubacar *et al.*, 2005). Figure 3.a illustrates the final representation space of the X_t raw data (the circle are non-classified data) and the figure 3.b gives the final classes locations.

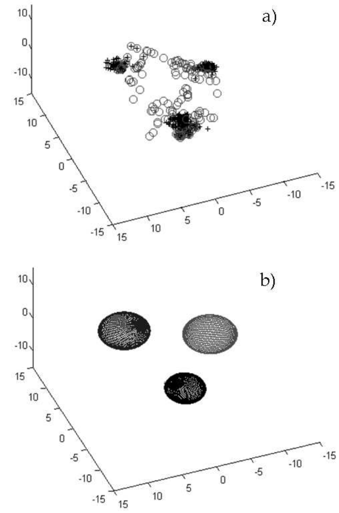

Fig. 3. Classified data and Final decision space

In fact, the data are online classified and the decision space is built and updated in a recursive way. Any *a priori* knowledge is required. From the first similar data, the classifier creates the first class corresponding to the first local model. When a switch occurs, according to the choice of the membership threshold, the data generated after the switch are not recognized as a known class. After a delay corresponding to a stabilization of these data (when the number of stable observations exceeds the NP_{min} threshold) in a particular area of the decision, a new class that characterizes the second functioning mode is created. All the new classes are created in this way (cf. Table 1).

At each time, the current functioning mode is determined by using the membership degree of the observation (membership ratio rule according to a threshold θ). On figure 4 where θ equals 0.25, it can be noticed that the observations located between modes are non classified and the class creation (3 first situations) is effective after an extra delay corresponding to NP_{min}. When the mode is already known, the decision is done quicker (e.q way through mode 3 to mode 1). In this case, the delay of recognition is mainly due to the convergence of the Markov parameters matrices estimation and to the noise influence.

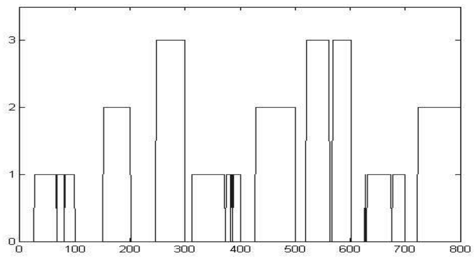

Fig. 4. Classification of the Markov Parameters Matrices; Membership threshold = 0.25

The figure 5 illustrates the behaviour of this method for a small value of θ. In this case, the membership is forced to known classes.

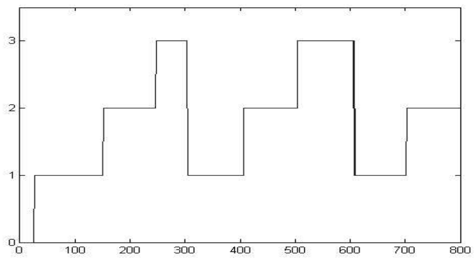

Fig. 5. Classification of the Markov Parameters Matrices; Membership threshold = ε

One can notice, even if this is not presented on the figures, that the classes parameters are continuously adapted. This allows a better accurate definition of the decision space and make possible the modelling of non-stationary local linear models (Lecœuche *et al.*, 2006). Of course, from this point, expert knowledge should be introduced in order to define if this evolution is "normal" (running in periods, chemical transition...) or if a progressive failure appears to the process.

6. CONCLUSION

In this paper, a new approach has been presented for the online determination of the functioning mode of a switching system. This approach is based on a recursive estimation tool coupled with a dynamic classification algorithm. The method is dedicated to switching systems where each linear local model could have its own structure (parameters and order). The interest of this approach is also to take into account commutations and evolutions of modes without *a priori* knowledge thanks to the use of a dynamical classifier.

7. REFERENCES

Amadou-Boubacar, H. and S. Lecœuche (2005). A new kernel-based algorithm for online clustering. In: *7th International Conference on Artificial Neural Networks*. Warsaw, Poland. pp. 350–358.

Amadou-Boubacar, H., S. Lecœuche and S. Maouche (2005). Audyc neural network using a new gaussian densities merge mechanism. In: *7th International Conference on Adaptive and Natural Computing Algorithms*. Coimbra, Portugal. pp. 155–158.

Bemporad, A., J. Roll and L. Ljung (2001). Identification of hybrid systems via mixed-integer programming. In: *13th International Conference on Artificial Neural Networks*. Orlando, FL. pp. 786–792.

Breiman, L. (1993). Hinging hyperplanes for regression, classification, and function approximation. *IEEE Transaction on Information Theory* **39**(3), 999–1013.

Deng, D. and N. Kasabov (2003). On-line pattern analysis by evolving self-organizing maps. *Neurocomputing* **51**, 87–103.

Eltoft, T. and R. deFigueiredo (1998). A new neural network for cluster-detection-and-labeling. *IEEE Transactions on Neural Networks* **9**, 1021–1035.

Ferrari-Trecate, G., M. Muselli, D. Liberati and M. Morari (2003). A clustering technique for the identification of piecewise affine systems. *Automatica* **39**(2), 205–217.

Hespanha, J. P. and A. S. Morse (1999). Stability of switched systems with average dwell-time. In: *38th CDC*. Phoenix, USA. pp. 2655–2660.

Lecœuche, S. and C. Lurette (2003). Auto-adaptive and dynamical clustering neural network. In: *13th International Conference on Artificial Neural Networks*. Istanbul, Turkey. pp. 350–358.

Lecœuche, S., G. Merc'ere and H. Amadou-Boubacar (2006). Modelling of non stattionary systems based on a dynamical decision space. In: *14th IFAC symposium on System Identification*. Newcastle, Australia.

Lurette, Christophe (2003). Développement d'une technique neuronale auto-adaptative pour la classification de données non-stationnaires.. PhD thesis. Lille University.

Mouchaweh, M. (2004). Diagnostic in real time for evolutionary processes in using pattern recognition and possibility theory. *International Journal of Computational Cognition* **2**(1), 79–112.

Pekpe, K. M., K. Gasso, G. Mourot and J. Ragot (2003). Subspace identification of switching model. 13th IFAC Symposium on System Identification. Rotterdam, The Netherlands.

Pekpe, K. M., K. Gasso, G. Mourot and J. Ragot (2004). Identification of switching systems using change detection technique in the subspace framework. 43rd IEEE Conference on Decision and Control. Atlantis, Bahamas.

Vidal, René (2004). Identification of pwarx hybrid models with unknown and possibly different orders. In: *IEEE American Conference on Control*. Boston, MA, USA.

Zhou, S. K. and R. Chellappa (2004). Kullback-leibler distance between two gaussian densities in reproducing kernel hilbert space. In: *International Symposium on Information Theory*. Chicago, USA.