A new clustering technique for the identification of PWARX hybrid models

Zeineb Lassoued
Numerical Control of Industrial Processes,
National School of Engineers of Gabes,
University of Gabes,
St Omar Ibn-Khattab, 6029 Gabes, Tunisia.
Email: zeineb.lassoued1@gmail.com

Kamel Abderrahim
Numerical Control of Industrial Processes,
National School of Engineers of Gabes,
University of Gabes,
St Omar Ibn-Khattab, 6029 Gabes, Tunisia.
Email: kamel.abderrahim@yahoo.fr

***Abstract*—This paper addresses the problem of clustering-based procedure for the identification of PWARX models. It consists in estimating both the parameter vector of each sub-model and the coefficients of each partition. It exploits three main techniques which are clustering, linear identification and pattern recognition. The performance of this approach depends on the used clustering technique. However, most of existing methods are based on classical approaches which are sensible to poor initialization and suffer from the presence of outliers. To overcome these problems, we propose to exploit the Chiu's clustering technique. Simulation results are presented to illustrate the performance of the proposed method.**

I. Introduction

Hybrid systems can be defined as systems where interact continuous and discrete-event phenomena. The PWA model has been the most popular since it can approximate any nonlinear system with arbitrary accuracy. Moreover, the parameters of each sub-model are linearly related to the output which allows the extension of some results of linear systems to nonlinear ones. On the other hand, the properties of equivalence [1] between PWA models and other representations of hybrid systems allow to transfer the results of PWA models to any subclasses of hybrid systems such as jump linear models (JL models)[2], Markov jump linear models (MJL models)[3], Mixed Logic Dynamical models (MLD models)[4], [5], Max-Min-Plus-Scaling systems (MMPS models)[6], Linear Complementarity models (LC models)[7], Extended Linear Complementarity models (ELC models)[8].
PWA systems are obtained by decomposing the state-input domain into a finite number of non-overlapping convex polyhedral regions, and by associating a simple linear or affine model to each region. This paper addresses the problem of identifying discrete-time hybrid systems in the Piecewise AutoRegressive forms with eXogenous input (PWARX) [9]. This problem consists in building mathematical models of hybrid systems from observed input-output data [5].

For the identification of PWARX systems, two main problems must be considered: one is the identification of the parameters of sub-models and two is the estimation of the coefficients of the hyperplanes which define the partition of the state-input domain. The existing methods for the identification of PWARX system can be classified in numerous categories of solutions such as the algebraic solution[10], the clustering-based solution[11], the Bayesian solution[12], the bounded-error solution[13], the Mixed-Integer Programming solution[14], etc. Only the clustering-based procedure is considered in this paper.
This approach use three main steps : data classification, parameter estimation and regions reconstruction. The early approaches use the k-means clustering algorithms for the data classification. However, these algorithms are blind to outliers and sensible to initialization. To overcome these problems, we propose a solution which consists in using Chiu's clustering technique [15].

This paper is organized as follows. Section 2 presents the model and its assumptions. Section 3 recalls the clustering-based procedure for the identification of PWARX systems. In section 4, we propose a new solution to overcome the problems of outliers and poor initialization. Results of simulations are illustrated in the last section.

II. Model and assumptions

In the following, we address the problem of identification of PieceWise AutoRegressive eXogenous (PWARX) systems described by:

$$y(t) = f(\varphi(t)) + e(t) \tag{1}$$

where
- $y(t) \in R$ is the system output,
- $e(t)$ is the noise,
- t is the now time index,
- $\varphi(t)$ is the vectors of regressors which belongs to a bounded polyhedron H in R^d:

$$\varphi(t) = [y(t-1), \ldots, y(t-n_a) \quad u(t-1), \ \ldots, u(t-n_b)]^T \tag{2}$$

where $u(t) \in R^{n_u}$ is the system inputs, n_a and n_b are the system orders and $d = n_a + n_u(n_b + 1)$.
- f is a piecewise affine function defined by:

$$f(\varphi) = \begin{cases} \theta_1^T \bar{\varphi} & if \quad \varphi \in \ H_1 \\ \vdots & \\ \theta_s^T \bar{\varphi} & if \quad \varphi \in \ H_s \end{cases} \tag{3}$$

where $\bar{\varphi} = \begin{bmatrix} \varphi^T & 1 \end{bmatrix}^T$, H_i are polyhedral partitions of the bounded domain H and $\theta_i \in R^{d+1}$ is the parameter vector.

978-1-4673-5769-2/13/$31.00 ©2013 IEEE

The following assumptions are assumed to be verified:

A1: The orders n_a and n_b and the number of sub-models s are known.

A2: The noise $e(t)$ is assumed to Gaussian sequence independent and identically distributed with zero mean and finite variance σ^2.

A3: The regions $\{H_i\}_{i=1}^{s}$ are the polyhedral partitions of a bounded domain $H \subset R^d$ such as:

$$\begin{cases} \bigcup_{i=1}^{s} H_i = H \\ H_i \bigcap H_j = \emptyset \forall i \neq j \end{cases} \tag{4}$$

Problem statement
Identify the partitions $\{H_i\}_{i=1}^{s}$ and the parameters vectors $\{\theta_i\}_{i=1}^{s}$ of the PWARX model using a data set $\{y(k), \varphi(k)\}_{k=1}^{N}$.

III. The k-means clustering method of identification

The identification of PWARX models involves the estimation of both the parameter vector of each sub-model and the coefficients of each partition. This is one of the most difficult problem because the estimation of the linear sub-models cannot be decoupled from the identification of the partitions. Several solutions have been proposed in the literature to solve this problem. Only the clustering-based approach is considered in this paper. The main steps of these methods can be summarized as follows: construct small data set from the initial data set, estimate a parameter vector for each small data set, classify the parameter vectors in s clusters, classify the initial data set, estimate s sub-models with their partitions.
It is easy to remark that this method is based on two main combined techniques which are data classification and linear regression. It is assumed that the number of sub-models and the orders of sub-models are a priori known. This section details the k-means clustering based method of identification [16], [11].

Step 1
For every paire of data $\{\varphi(t), y(t)\}_{t=1}^{N}$, we construct a local set $\rho_t = \left\{k_t^1, \cdots, k_t^{n_\rho}\right\}$ containing in ascending order the index t of $\{\varphi(t), y(t)\}$ and its $(n_\rho - 1)$ nearest neighbors satisfying:

$$\forall(\breve{\varphi}, \breve{y}) \in \rho_t, \left\|\varphi(t) - \breve{\varphi}\right\|^2 \leq \|\varphi(t) - \hat{\varphi}\|^2, \forall(\hat{\varphi}, \hat{y}) \notin \rho_t \tag{5}$$

Among the local sets obtained ρ_t, some may contain only data from the same model as they are called pure local sets, and others can collect data from multiple sub-models that are called mixed sets.
n_ρ is a parameter chosen randomly as $n_\rho > d+1$. It influences decisively on the performance of the algorithm. The optimal value of n_ρ is always a compromise between two phenomena: more this parameter is bigger more the parameter estimation is improved and the effect of noise is rejected. However, a large value of n_ρ increases the number of local mixed sets.

Step 2
Determine for each local set the local parameters θ_t using the least square method:

$$\theta_t = (\phi_t^T \phi_t)^{-1} \phi_t^T Y_t \tag{6}$$

where $\phi_t = \left[\bar{\varphi}(k_t^1) ... \bar{\varphi}(k_t^{n_\rho})\right]^T$

and $Y_t = \left[y(k_t^1) ... y(k_t^{n_\rho})\right]^T$

Step 3
Compute the average m_t of $\varphi(t)$ containing in ρ_t. Every data $(\varphi(t), y(t))$ is well matched to the characteristic vector ξ_t:

$$\xi_t = \left[\theta_t^T \quad m_t^T\right]^T \tag{7}$$

where $m_t = \frac{1}{n_\rho} \sum_{t \in \rho_t} \varphi(t)$.

The covariance matrix V_t of the local estimated parameter vector θ_t is given by:

$$V_t = \frac{SSR}{n_\rho - (d+1)} (\phi_t^T \phi_t)^{-1} \tag{8}$$

where $SSR = Y_t^T (I - \phi_t (\phi_t^T \phi_t)^{-1} \phi_t^T) Y_t$

and the variance Q_t of regressors containing in ρ_t is:

$$Q_t = \sum_{t \in \rho_t} (\varphi(t) - m_t)(\varphi(t) - m_t)^T \tag{9}$$

The characteristic vector $\xi_t \quad \forall t = 1, ..., N$ have the following variance R_t :

$$R_t = \begin{bmatrix} V_t & 0 \\ 0 & Q_t \end{bmatrix} \tag{10}$$

Classify the vectors $\{\xi_t\}_{t=1}^{N}$ in s separate classes D_j using a classification technique that minimize the suitable criterion. A modified version of the k-means algorithm is proposed in [11]. It allows to reduce the influence of the outliers and poor initializations and so it improves the performance of the classical k-means algorithm by using confidence measures on the points that have to be clustered. This solution is summarized by the following algorithm:

1) Initialise the centers $\mu_i, i = 1, ..., s$ and set a threshold $\varepsilon > 0$
2) Compute the clusters D_i minimizing the following criteria:

$$J = \sum_{i=1}^{s} \sum_{\xi_t \in D_i} \|\xi_t - \mu_i\|_{R_t^{-1}}^2 \tag{11}$$

the used norme is defined as:
$\|\xi_t - \mu_i\|_{R_t^{-1}} = \sqrt{(\xi_t - \mu_i)^T R_t^{-1} (\xi_t - \mu_i)}$
3) Update the centers according to the formula:

$$\tilde{\mu} = \frac{\sum_{t;\xi_t \in D_i} \xi_t w_t}{\sum_{t;\xi_t \in D_i} w_t} \tag{12}$$

with $w_t = \frac{1}{\sqrt{(2\pi)^{(2na+2nb+1)} \det(R_t)}}$

4) Set $\mu_i = \tilde{\mu}_i$ until having $\|\mu_i - \tilde{\mu}_i\| < \varepsilon$

Step 4

As the obtained data are now classified, it is possible to determine the s ARX sub-models. Since the application that attribute for every paire of data $(\varphi(t), y(t))$ one characteristic vector ξ_t is bijective, it is easy to obtain $\theta_i, \quad i = 1, ...s$ by appealing again the least square method.

Now, we have only to determine the regions H_i. Methods of statistical learning as the Support Vector Machine (SVM) offer an interesting solution to accomplish this task [17]. It's matter of finding for every $i \neq j$ the hyperplane that separate points existing in H_i and in H_j.

The estimated hyperplane separating H_i from H_j is denoted with $M_{i,j}\varphi = m_{i,j}$ where $M_{i,j}$ and $m_{i,j}$ are matrices of suitable dimensions. Moreover, we assume that the points in H_i belong to the half-space $M_{i,j}\varphi \leq m_{i,j}$.

The regions H_i are obtained by solving these linear inequalities. It is then enough to consider the bounded polyhedron [18]:

$$\left[M'_{i,1}...M'_{i,s}M'\right]\varphi \leq \left[m'_{i,1}...m'_{i,s}m'\right] \tag{13}$$

where $Mx \leq m$ are the linear inequalities describing H.

Remark

The classification is an important problem to achieve the objective of PWARX model identification because successful identification of both sub-models and partitions depends on the performance of the used clustering technique. In fact, this problem presents an area of research in which few results have been devoted in the past because most of existing methods for the identification of PWARX models are based on classical clustering algorithms such as k-means methods. However, the classical clustering methods even the modified k-means algorithm allows only to reduce the influence of outliers and poor initializations. Consequently, it still suffer from many drawbacks:

- It depends on the input signal which must be a persistent excitation to permit to the sub-models to have a balanced input [19].
- The parameter n_ρ must be a small number to not complicate the matricial calculation. However, this parameter must be larger to get better result.
- The k-means algorithm is not guaranteed to converge to the optimal clusters and may be trapped in local minima because of the randomly initialisation.

IV. Chiu's Classification method for identifying PWARX systems

Clustering of data forms the basis of many modeling and pattern classification algorithms. The purpose of clustering is to find natural groupings of data in a large data set, thus revealing patterns in the data that can provide a concise representation of the data behavior. Chiu proposed in [20] and [21] a simple and effective algorithm for data points clustering. Chiu's classification method consists in computing a potential value for each point from the data set based on its distances to the actual data points and consider each data point as a potential cluster center. The point having the highest potential value is chosen as the first cluster center. The key idea in this method is that once the first cluster center is chosen, the potential of all other points is reduced according to their distance from the cluster center. All points near the first cluster center will have greatly reduced potential. The next cluster center take then the highest remaining potential value. This procedure of acquiring new cluster center and reducing the potential of the surrounding points repeats until the potential of all points falls below a threshold or until reaching the number of clusters required.

This method works as follows:

Consider a collection of N data points $(\theta_i, i = 1, ..., N)$ picked out by applying step 1 and 2 of the procedure cited above. The potential of each θ_i is given by the following expression:

$$P_i = \sum_{j=1}^{N} e^{-\frac{4}{r_a^2}\|\theta_i - \theta_j\|^2} \tag{14}$$

where r_a is a positive constant.

The potential of each data point is a function of its distances to all other data point. Thus a data point with many neighboring data points will have the highest potential value. The constant r_a is the radius defining the neighborhood. Then, we select the data point with the highest potential value as the first cluster center. Let θ_1^* be this first center and $P_1{}^*$ be its potential. The potential is then updated by this formula :

$$P_i \Leftarrow P_i - P_1^* e^{-\frac{4}{r_b^2}\|\theta_i - \theta_1^*\|^2} \tag{15}$$

The data points near the first cluster center will have then a reduced potential and so they are unlikely to be selected as the next center. The parameter r_b is a positive constant that must be chosen larger than r_a to avoid obtaining closely spaced cluster centers.

In general after obtaining the *kth* cluster center, the potential of every data point is updated by the following formula :

$$P_i \Leftarrow P_i - P_k^* e^{-\frac{4}{r_b^2}\|\theta_i - \theta_k^*\|^2} \tag{16}$$

Where $P_k{}^*$ and $\theta_i{}^*$ are respectively the potential and the center of the *kth* data point.

This work is then repeated until obtaining s potential and s centers. Now, after obtaining s centers it is matter to search the elements belonging to each cluster. So, we calculate the distance between $\theta_i, i = 1, ..., N$ and $\theta_k{}^*, k = 1, ..., s$ and classify θ_i into the cluster whose distance is the minimum.

Properties

The new clustering technique has several interesting properties which can be summarized as follows:

- This method does not require the initialization of centers. Therefore the problem of convergence towards local minima is overcome.
- It is possible with this method to remove the misclassified points (θ_i) from the data set and repeat the overall identification procedure on the reduced set of data points. The outliers can be removed because of equation (14) that associate to these points a low potentials.

- The choice of the parameter n_ρ is more flexible. In fact, we can improve the performance with high value of n_ρ.

V. Numerical examples

We present now two simulation examples of the proposed classification method. Hence, the objective of the simulations is to compare the performance of the proposed method with that of the modified k-means approach.
The following quality measures are used to study the performance of each method [22].

- The maximum of relative error of parameter vectors is defined by:

$$\Delta_\theta = \max_{i=1,..,s} \frac{\left\|\theta_i - \bar{\theta}_i\right\|_2}{\left\|\bar{\theta}_i\right\|_2} \tag{17}$$

where $\bar{\theta}_i$ and θ_i are the true and the estimated parameter vectors for sub-model i. The identified model is deemed acceptable if Δ_θ is small or close to zero.

- The averaged sum of the squared residuals is defined by :

$$\sigma_e^2 = \frac{1}{s}\sum_{i=1}^{s} \frac{SSR_i}{|D_i|} \tag{18}$$

where $SSR_i = \sum_{(y(k),\varphi(k))\in D_i} (y(k) - [\varphi(k)'1]\theta_i)^2$

and $|D_i|$ is the cardinality of the cluster D_i.
The identified model is considered acceptable if σ_e^2 is small and/or close to the expected noise variance of the true system.

- The percentage of the output variation that is explained by the model is defined by:

$$FIT = 100.\left(1 - \frac{\|\hat{y} - y\|_2}{\|y - \bar{y}\|_2}\right) \tag{19}$$

where $\hat{y}$ and y are the estimated and the real outputs' vectors, $\bar{y}$ is the mean value of y.
The identified model is considered acceptable if FIT is close to 100.

A. Example 1

Consider the following PWARX system [16]:

$$y(k) = \begin{cases} [-1 \quad 0]\,[\mathrm{u}(k-1) \quad 1]' + e(k) & \text{if} \quad \varphi \in [-4 \quad 0[\\ [1 \quad 0]\,[\mathrm{u}(k-1) \quad 1]' + e(k) & \text{if} \quad \varphi \in [0 \quad 2[\\ [3 \quad -2]\,[\mathrm{u}(k-1) \quad 1]' + e(k) & \text{if} \quad \varphi \in [2 \quad 4] \end{cases} \tag{20}$$

where $s = 3$, $n_a = 0$, $n_b = 1$, $H = [-4 \quad 4]$, the input $u(k) \in R$ is generated randomly according to the uniform distribution on H, $e(k)$ is a white gaussian noise of variance $\sigma^2 = 0.05$ and $\varphi(k) = [u(k-1)]$ is the regressor vector.
The input $u(k)$ and the output $y(k)$ are presented in Fig.1.

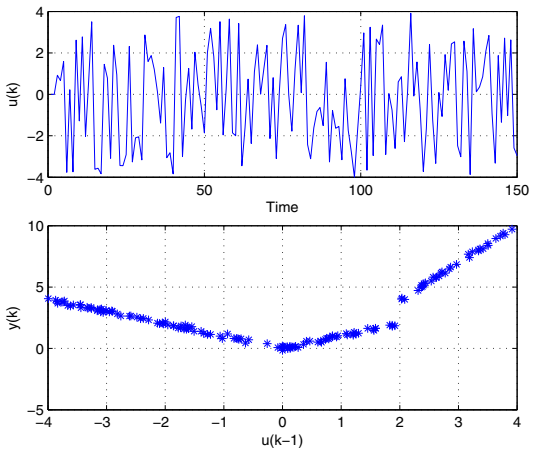

Fig. 1. Input and output signals of the system

This output is randomly generated so that we can't distinguish which sub-model has generated which output. Therefore as a first treatment, we will determine the local parameters by the least square method as described in step 2 taking a number of neighbors $n_\rho = 15$ for the Chiu's classification based algorithm. Fig.2 presents the local parameters.

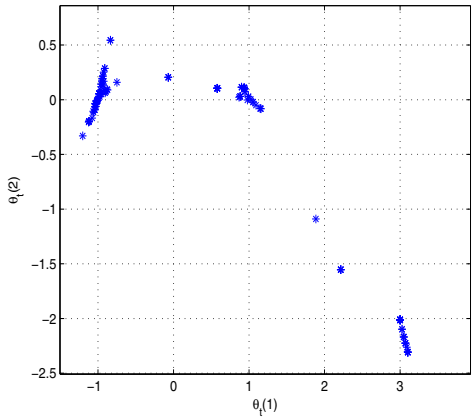

Fig. 2. The local parameters θ_t, $t = 1, ..., N$

After obtaining these local parameters, step 3 allows to classify these parameters into three clusters and to compute their centers using the proposed method for the classification based on Chiu's approach.

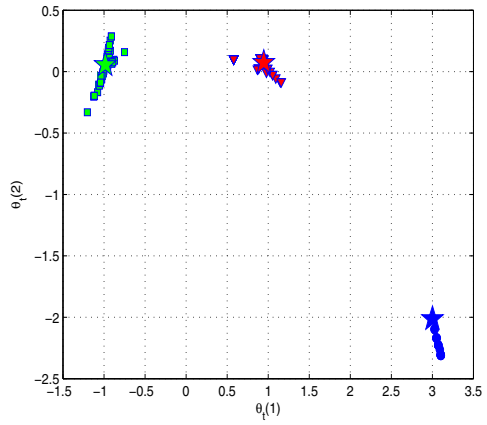

Fig. 3. Local parameters separated into 3 sets

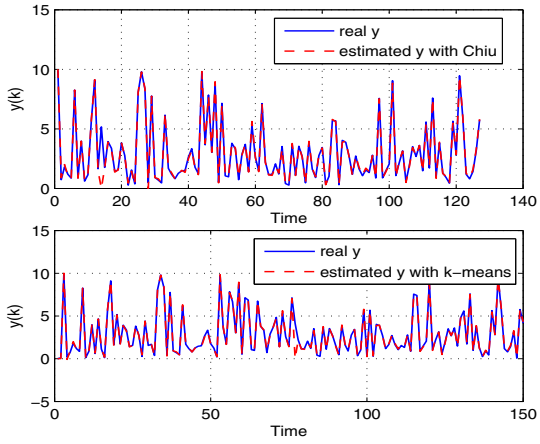

Fig. 4. Real and estimated outputs

The obtained results are illustrated in Fig.3 which shows that the outliers points are rejected and the set of data points are reduced. This is thanks to equation (14) that calculate the potential of these points. The centers of each set are depicted by the star symbols in Fig.3.

After this classification, the least square method is used to estimate the parameter vector of each sub-model. Then, by applying the SVM classification method, we obtain the s inequalities describing $\{H_i\}_{i=1}^{3}$.

A comparison is then done between this method and the modified k-means algorithm having a number of neighbors $n_\rho = 7$. Table I presents the obtained results.

TABLE I. ESTIMATED PARAMETERS WITH THE TWO METHODS

	True values	k-means	Chiu
θ_1	$\begin{matrix}-1\\0\end{matrix}$	$\begin{matrix}-0.9921\\0.0210\end{matrix}$	$\begin{matrix}-0.9970\\0.0080\end{matrix}$
θ_2	$\begin{matrix}1\\0\end{matrix}$	$\begin{matrix}1.0340\\-0.0511\end{matrix}$	$\begin{matrix}1.0385\\-0.0473\end{matrix}$
θ_3	$\begin{matrix}3\\-2\end{matrix}$	$\begin{matrix}3.0315\\-2.0810\end{matrix}$	$\begin{matrix}3.0033\\-1.9922\end{matrix}$
H_1	$\begin{matrix}-4\\0\end{matrix}$	$\begin{matrix}-4\\0.1082\end{matrix}$	$\begin{matrix}-4\\-0.06\end{matrix}$
H_2	$\begin{matrix}0\\2\end{matrix}$	$\begin{matrix}0.1722\\1.7457\end{matrix}$	$\begin{matrix}0.005\\1.93\end{matrix}$
H_3	$\begin{matrix}2\\4\end{matrix}$	$\begin{matrix}2.0572\\4\end{matrix}$	$\begin{matrix}2.19\\4\end{matrix}$

The real and the estimated outputs obtained by the two methods are depicted in Fig.4 :

Table II presents the quality measures (17), (18) and (19) of the proposed method and the modified k-means approach.

TABLE II. VALIDATION RESULTS (EXAMPLE1)

	k-means	Chiu
Δ_θ	0.0282	0.0161
σ_e^2	0.0092	0.0094
FIT	86.2390	93.1546

Based on the results presented in table I, table II and Fig.4, we observe that the proposed method gives better performances than the modified k-means method. The reason is that the proposed method reduces the influence of outlier and does not require any initialization.

B. *Example 2*

Consider the following PWARX system [23]:

$$y(k) = \begin{cases} [-0.4 \quad 1 \quad 1.5]\,\varphi(k) + e(k) & if \quad \varphi(k) \in H_1 \\ [0.5 \quad -1 \quad -0.5]\,\varphi(k) + e(k) & if \quad \varphi(k) \in H_2 \\ [-0.3 \quad 0.5 \quad -1.7]\,\varphi(k) + e(k) & if \quad \varphi(k) \in H_3 \end{cases} \tag{21}$$

with $H_1 = \{[4 - 1 \quad 10]\,\bar{\varphi}(k) < 0\}$,

$$H_2 = \left\{ \begin{bmatrix} -4 & 1-10 \\ -5 & 1-6 \end{bmatrix} \bar{\varphi}(k) \leq 0 \right\}$$

and $H_3 = \{[-5 - 1 \quad 6]\,\bar{\varphi}(k) < 0\}$

where $\varphi(k) = [y(k-1)u(k-1)]^T$, the input signal $u(k)$ and the noise signal $e(k)$ are uniformly distributed in $[-5, 5]$ and $[-0.1, 0.1]$.
The same procedure is followed.

TABLE III. ESTIMATED PARAMETERS WITH THE TWO METHODS

	True values	k-means($n_\rho = 7$)	Chiu ($n_\rho = 17$)
θ_1	$\begin{bmatrix} -0.4000 \\ 1.0000 \\ 1.5000 \end{bmatrix}$	$\begin{bmatrix} -0.4204 \\ 0.7637 \\ 1.4502 \end{bmatrix}$	$\begin{bmatrix} -0.4053 \\ 0.9895 \\ 1.4848 \end{bmatrix}$
θ_2	$\begin{bmatrix} 0.5000 \\ -1.0000 \\ -0.5000 \end{bmatrix}$	$\begin{bmatrix} 0.6744 \\ -1.0248 \\ -0.4301 \end{bmatrix}$	$\begin{bmatrix} 0.5034 \\ -0.9927 \\ -0.4580 \end{bmatrix}$
θ_3	$\begin{bmatrix} -0.3000 \\ 0.5000 \\ -1.7000 \end{bmatrix}$	$\begin{bmatrix} -0.2577 \\ 0.4315 \\ -1.8450 \end{bmatrix}$	$\begin{bmatrix} -0.2928 \\ 0.4998 \\ -1.7338 \end{bmatrix}$

TABLE IV. VALIDATION RESULTS (EXAMPLE2)

	k-means	Chiu
Δ_θ	0.1099	0.0234
σ_e^2	2.4150	0.0047
FIT	57.1020	84.0254

The obtained results from table III and table IV show that the proposed method gives the best estimation results.

VI. CONCLUSION

In this paper, a new clustering technique for the identification of PWARX models based on clustering approach is proposed. This technique allows to overcome the two problems of the existing methods. The first is the poor initializations which leads the algorithm to converge to local minima. The second is the presence of outliers which degrades the performance. Numerical simulations results are presented to illustrate the performance of the proposed approach.

ACKNOWLEDGMENT

This work was supported by the Ministry of the Higher Education and Scientific Research in Tunisia.

REFERENCES

[1] W. Heemels, B. De Schutter, and A. Bemporad, "On the equivalence of classes of hybrid dynamical models," in *Proceedings of the 40th IEEE Conference on Decision and Control, Vol. 1, pp. 364-369, 2001*.

[2] X. Feng, K. Loparo, Y. Ji, and H. Chizeck, "Stochastic stability properties of jump linear systems," *IEEE Transactions on Automatic Control*, vol. 37, pp. 38–53, 1992.

[3] A. Doucet, N. Gordon, and V. Krishnamurthy, "Particle filters for state estimation of jump markov linear systems," *IEEE Transactions on Signal Processing*, vol. 1, no. 2, pp. 613–624, 2001.

[4] A. Bemporad and M. Morari, "Control of systems integrating logic, dynamics, and constraints," *Automatica*, vol. 35, pp. 407–428, 1999.

[5] A. Bemporad, G. Ferrari-Trecate, and M. Morari, "Observability and controllability of piecewise affine and hybrid systems," *IEEE Transactions on Automatic Control*, vol. 45, no. 10, pp. 1864–1876, 2000.

[6] B. De Schutter and T. Van den Boom, "On model predictive control for max-min-plus-scaling discrete event systems," Technical Report bds 00-04: Control Systems Engineering, Faculty of Information Technology and Systems, Delft University of Technology, The Netherlands, Tech. Rep., 2000.

[7] A. van der Schaft and J. Schumacher, "Complementarity modeling of hybrid systems," *IEEE Transactions on Automatic Control*, vol. 43, no. 4, pp. 483–490, 1998.

[8] B. De Schutter, "Optimal control of a class of linear hybrid systems with saturation," in *Proceedings of the 38th IEEE Conference on Decision and Control, pp.3978-3983, 1999*.

[9] S. Paoletti, A. Juloski, G. Ferrari-Trecate, and R. Vidal, "Identification of hybrid systems: A tutorial," *European Journal of Control*, vol. 13, no. 2-3, pp. 242–260, 2007.

[10] Y. Tian, T. Floquet, L. Belkoura, and W. Perruquetti, "Algebraic switching time identification for a class of linear hybrid systems," *Nonlinear Analysis: Hybrid Systems*, vol. 5, no. 2, pp. 233–241, 2011.

[11] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari, "A clustering technique for the identification of piecewise affine systems," *Automatica*, vol. 39, no. 2, pp. 205–217, 2003.

[12] A. Juloski, S. Weiland, and W. Heemels, "A bayesian approach to identification of hybrid systems," *IEEE Transactions on Automatic Control*, vol. 50, no. 10, pp. 1520–1533, 2005.

[13] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino, "A bounded-error approach to piecewise affine system identification," *IEEE Transactions on Automatic Control*, vol. 50, no. 10, pp. 1567–1580, 2005.

[14] J. Roll, A. Bemporad, and L. Ljung, "Identification of piecewise affine systems via mixed-integer programming," *Automatica*, vol. 40, no. 1, pp. 37–50, 2004.

[15] S. Chiu, "Fuzzy model identification based on cluster estimation," *Journal of intelligent and Fuzzy systems*, vol. 2, no. 3, pp. 267–278, 1994.

[16] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari, "Identification of piecewise affine and hybrid systems," in *Proceedings of American Control Conference, Vol. 5, pp. 3521-3526, 2001*.

[17] C. Williams, "Support vector machines," School of Informatics, University of Edinburgh, 2008.

[18] G. Ferrari-Trecate, M. Muselli, D. Liberati, and M. Morari, "A clustering technique for the identification of piecewise affine systems," *Hybrid Systems: Computation and Control*, pp. 218–231, 2001.

[19] M. Petreczky and L. Bako, "On the notion of persistence of excitation for linear switched systems," in *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), pp. 1840-1847, 2011*.

[20] S. Chiu, "An efficient method for extracting fuzzy classification rules from high dimensional data," *Journal of Advanced Computational Intelligence*, vol. 1, no. 1, pp. 31–36, 1997.

[21] S.Chiu, "Extracting fuzzy rules from data for function approximation and pattern classification," *Chapter 9 in Fuzzy Information Engineering: A Guided Tour of Application. D. Dubois, H. Prade and R. Yager (eds), John Wiley Sons*, 1997.

[22] A. Juloski, S. Paoletti, and J. Roll, "Recent techniques for the identification of piecewise affine and hybrid systems," *Current trends in nonlinear systems and control*, pp. 79–99, Springer, 2006.

[23] A. Bemporad, A. Garulli, S. Paoletti, and A. Vicino, "A greedy approach to identification of piecewise affine models," *Hybrid Systems: Computation and Control*, pp. 97–112, Springer, 2003.