

New Results on PWARX Model Identification Based on Clustering Approach

Zeineb Lassoued Kamel Abderrahim

Numerical Control of Industrial Processes, National School of Engineers of Gabes, University of Gabes,
St Omar Ibn-Khattab, 6029 Gabes, Tunisia

Abstract: This paper deals with the problem of piecewise auto regressive systems with exogenous input (PWARX) model identification based on clustering solution. This problem involves both the estimation of the parameters of the affine sub-models and the hyper planes defining the partitions of the state-input regression. The existing identification methods present three main drawbacks which limit its effectiveness. First, most of them may converge to local minima in the case of poor initializations because they are based on the optimization using nonlinear criteria. Second, they use simple and ineffective techniques to remove outliers. Third, most of them assume that the number of sub-models is known a priori. To overcome these drawbacks, we suggest the use of the density-based spatial clustering of applications with noise (DBSCAN) algorithm. The results presented in this paper illustrate the performance of our methods in comparison with the existing approach. An application of the developed approach to an olive oil esterification reactor is also proposed in order to validate the simulation results.

Keywords: Hybrid systems, piecewise autoregressive systems with exogenous input (PWARX) model, clustering, identification, density-based spatial clustering of applications with noise (DBSCAN) clustering technique, experimental validation.

1 Introduction

Hybrid systems are heterogeneous dynamical systems that arise out of the interaction of continuous and discrete dynamics. The continuous behavior is the fact of the natural evolution of the physical process whereas the discrete behavior can be due to the presence of switches, operating phases, transitions, computer program codes, etc. These hybrid dynamics characterize the behavior of a broad class of physical systems, for example, the real-time control systems where physical processes are controlled by embedded controllers. The notion of hybrid system can also be used to represent complex nonlinear continuous systems. In fact, the operating range of a nonlinear system can be decomposed into a group of operating point. For each operation points, we associate a simple sub-model (linear or affine) with it. Indeed, a complex system can be modeled as a hybrid system switching between simple sub-models. This paper addresses the problem of identification of hybrid systems represented by piecewise auto regressive systems with exogenous input (PWARX). This problem consists in building mathematical models of hybrid systems from observed input-output data. The PWARX models have attracted a considerable attention in recent years, since they provide an efficient solution for modeling a wide range of engineering applications. In addition, these models are able to approximate any nonlinear system with arbitrary accuracy^[1]. Moreover, the piecewise affine (PWA) model can be considered as a generic representation for other hybrid models such as jump linear (JL) models^[2], Markov jump linear (MJL) models^[3] mixed logic dynamical (MLD) models^[4, 5], max-min-plus-scaling systems (MMPS) models^[6], linear complementarity (LC) models^[7], extended linear complementarity (ELC) models^[8]. In fact, the transfer of the

results of PWARX models to other classes of hybrid systems is insured thanks to the properties of equivalence of PWARX models^[9]. The PWARX models are obtained by decomposing the regression domain into a finite number of non-overlapping convex polyhedral regions and by associating a simple linear model with each region. Consequently, two main problems must be considered for the identification of PWARX models: One is the estimation of the parameters of the sub-models, and another is the determination of the hyper planes defining the partitions of the state-input regression. Consequently, the identification of PWARX models is one of the most difficult problems that represent an area of research where considerable work has been done in the last decade^[10–13]. In fact, numerous solutions have been proposed in the literature for the identification of the PWARX models such as the algebraic solution^[14], the clustering-based solution^[12], the Bayesian solution^[15], the bounded-error solution^[16] and so on. Only the clustering one is considered in this paper. This solution consists of three main steps, which are data classification, parameter estimation and region reconstruction. It is easy to remark that the performance of this approach depends on the efficiency of the used classification algorithm. The early methods have favored the simplicity of implementation. For example, the method proposed in [12] uses a clustering technique based on a modified k -means algorithm^[17]. It assumes that the number of sub-models and their orders are known a priori. It consists in classifying the data into several clusters or local sets. The identified parameter vectors from these small local sets are then collected to the parameter vectors of the true sub-models. Therefore, the parameters of the sub-models can be obtained by clustering the local parameter vectors. This method is characterized by its simplicity of implementation. But the presence of outliers in the data and the poor initializations degrade its perfor-

mances. A modification of this approach was proposed in [18] to allow the number of sub-models to be estimated as the data are classified. This modified approach uses a hierarchical classification method called single-linkage. The idea of this method consists in defining a distance between two disjoint local sets. These sets are then fused if a given criterion is verified. It should be noted that the number of sub-models is critically dependent on a parameter supplied by the user. In fact, a small value of this parameter would produce a large number of sub-models. However, if the parameter has a too large value, it could give a single sub-linear model. Another method was introduced in [19]. It exploits a statistical classification approach based on a Gaussian mixture model which allows to completely separate the data without considering the parameter vectors. In fact, its performance degrades when the data of different sub-models do not form classes quite distant in the regression space. In addition, the equipartition of different models can favorite some models over others in decision making. It assumes that the number of sub-models and the orders are known a priori. Finally, the method proposed in [20, 21] is based on unsupervised classification algorithm which is combined with regression techniques in order to separate the data according to their respective sub-models and thus estimate their parameters. The classification uses the principle of the k -nearest neighbor's algorithm. In fact, each nearest neighbor of a given vector provides information on the allocation of this vector. This approach requires the knowledge of the orders of the system. In addition, a single adjustment parameter must be provided by the user. This study allows us to identify the main drawbacks of the main existing methods, which can be summarized as follows:

- 1) Most of them are based on the optimization of non-linear criteria. Consequently, they may converge to local minima in the case of poor initializations.
- 2) Their performances degrade in the case of the presence of outliers in the data to be classified.
- 3) Most of them assume that the number of sub-models is known a priori.

To overcome these problems, we recommend the use of density-based spatial clustering of applications with noise (DBSCAN) algorithm, which is a pioneer algorithm of density-based clustering^[22, 23]. This choice is justified by the fact that the algorithm automatically generates the number of models. In addition, it is characterized by its robustness to the classification of noisy measurements and the containing outliers.

This paper is organized as follows. Section 2 presents the model and its assumptions. In Section 3, we recall the main steps of the identification of PWARX systems approach based on clustering algorithm and its main drawbacks. Section 4 proposes a solution to overcome the main problems of the existing methods. This solution consists of using the DBSCAN algorithm for data classification. In Section 5, we present the simulation results in order to illustrate the performance of the proposed solution and to compare its efficiency with the modified k -means method. Section 6 proposes an application of the developed approach to an olive oil esterification reactor.

2 Model and assumptions

In this paper, we address the problem of identifying PWARX systems in the regression form described by

$$y(k) = f(\varphi(k)) + e(k) \quad (1)$$

where $y(k) \in \mathbf{R}$ is the system output, $e(k)$ is the noise, k is the current time index, $\varphi(k)$ is the vectors of regressors which belongs to a bounded polyhedron H in \mathbf{R}^d :

$$\varphi(k) = \begin{bmatrix} y(k-1), \dots, y(k-n_a), u(k-1), \dots, \\ u(k-n_b) \end{bmatrix}^T \quad (2)$$

where $u(k) \in \mathbf{R}^{n_u}$ 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

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

where $\bar{\varphi} = [\varphi^T \ 1]^T$, s is the number of sub-models, H_i are polyhedral partitions of the bounded domain H , and $\theta_i \in \mathbf{R}^{d+1}$ is the parameter vector.

The following assumptions are assumed to be verified:

Assumption 1. The number of sub-models s is unknown.

Assumption 2. The orders n_a and n_b are known.

Assumption 3. The noise $e(k)$ is an independent Gaussian sequence and identically distributed with zero mean and finite variance σ^2 .

Assumption 4. The regions $\{H_i\}_{i=1}^s$ are the polyhedral partitions of a bounded domain $H \subset \mathbf{R}^d$ such as

$$\begin{cases} \bigcup_{i=1}^s H_i = H \\ H_i \cap H_j = \emptyset, \quad \forall i \neq j. \end{cases} \quad (4)$$

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

3 The PWA regression algorithm

Based on the observed data, the identification problem amounts to determining the parameters of the ARX sub-models together with the regions of the regressor space where each model is valid. The main problem in the identification of PWARX models is the problem of data classification: Each data point must be assigned to one specific sub-model. When the data have been classified, the parameters of the sub-models can be determined, and the regions where each of the sub-models is valid can be estimated using techniques for pattern classification.

3.1 Data classification

For every pair of data $\{\varphi(k), y(k)\}_{k=1}^N$, we construct a local set C_k collecting $\{\varphi(k), y(k)\}$ and its $(n_\rho - 1)$ nearest

neighbors satisfying

$$\forall(\tilde{\varphi}, \tilde{y}) \in C_k, \|\varphi(k) - \tilde{\varphi}\|^2 \leq \|\varphi(k) - \hat{\varphi}\|^2, \quad \forall(\hat{\varphi}, \hat{y}) \in C_k. \quad (5)$$

We also define ρ_k such that $\rho_k = \{t_k^1, \dots, t_k^{n_\rho}\}$ containing in ascending order the indexes of the elements belonging to C_k . Among the obtained local sets (C_k), 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: The bigger this parameter is, the more the parameter estimation is improved and more effectively the noise is rejected. However, a large value of n_ρ increases the number of local mixed sets.

For each local set C_k , we can identify an affine model. To accomplish this task, we adopt the least square method to determine the local parameters θ_k as

$$\theta_k = (\phi_k^T \phi_k)^{-1} \phi_k^T Y_k \quad (6)$$

where $\phi_k = [\bar{\varphi}(t_k^1), \dots, \bar{\varphi}(t_k^{n_\rho})]^T$, and $Y_k = [y(t_k^1), \dots, y(t_k^{n_\rho})]^T$.

Our objective is to classify the vectors $\{\theta_k\}_{k=1}^N$ in s separated classes D_i using a suitable classification technique taking into account that the number of sub-models s is unknown.

The classification of data represents the main step toward the objective of PWARX system identification because the performance of the following steps (parameter estimation and regions estimation) depends essentially on the effectiveness of the used clustering algorithm. The early approaches use classical clustering algorithms for data classification such as k -means algorithms^[5, 12, 18]. These algorithms are interesting from a theoretical point of view because they allow to demonstrate the possibility of estimating PWARX model from measurements. In addition, they are defined by simple algorithms in terms of implementation and computational complexity. But, they are not recommended in engineering applications because they do not smooth out the effect of outliers. Moreover, they are based on the optimization of nonlinear cost functions which may converge to local minima in the case of bad initialization.

In Section 4, we propose an alternative to these methods based on the DBSCAN clustering algorithm. Thus, we improve the performances of the existing clustering methods for the identification of PWARX models. We describe the advantages of our approach: automatic generation of sub-models number, processing and deletion of outliers, and avoiding arbitrary initializations.

3.2 Parameter estimation

As the obtained data are classified, it is possible to determine the s auto regressive exogenous (ARX) sub-models. Since the application that attributes for every pair of data $(\varphi(k), y(k))$, one parameter vector is bijective, it is easy to obtain θ_i , $i = 1, \dots, s$ by appealing the least square method

again. More precisely, the i -th sub-model is obtained from data that belong to cluster D_i . Then, the points in each final cluster can be used for estimating the parameter vector of each sub-model according to the rule:

$$(\varphi(k), y(k)) \in D_i \Leftrightarrow \theta_k \in D_i. \quad (7)$$

3.3 Region estimation

The final step is to determine regions H_i . Methods of statistical learning such as the support vector machine (SVM) offer an interesting solution to this task^[24, 25]. SVM is a popular machine learning method for classification, regression and other learning tasks. This study is still an ongoing research issue^[26, 27].

In our case, it is matter of finding for every $i \neq j$ the hyperplane that separates points existing in H_i and H_j . Given two sets H_i and H_j , $i \neq j$, the linear separation problem is to find $w \in \mathbf{R}^d$ and $b \in \mathbf{R}$ such that

$$\begin{aligned} w^T \varphi_k + b &> 0, \quad \forall \varphi_k \in H_i \\ w^T \varphi_k + b &< 0, \quad \forall \varphi_k \in H_j. \end{aligned} \quad (8)$$

This problem can be easily rewritten as a feasibility problem with linear inequality constraints. The estimated hyperplane separating H_i from H_j is denoted by $M_{i,j} \varphi = m_{i,j}$, 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^[12]:

$$[M'_{i,1}, \dots, M'_{i,s}, M'] \bar{\varphi} \leq [m'_{i,1}, \dots, m'_{i,s}, m'] \quad (9)$$

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

4 DBSCAN clustering technique

4.1 Principle

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. The clustering algorithms can be classified into several categories such as: partitioning clustering, hierarchical clustering, density based clustering, grid based clustering, and model based clustering^[28, 29]. In this paper, we use the density-based spatial clustering of applications with noise (DBSCAN) algorithm which is a pioneer algorithm of density-based clustering^[22, 23]. The reason for us to choose this category is that it has significant advantages over partitioning and hierarchical clustering algorithms. It can find out clusters of arbitrary shapes and sizes from a large amount of data containing noise and outliers.

The key idea of the DBSCAN algorithm is that for each data object of a cluster, the neighborhood of a given radius (ϵ) has to contain at least a minimum number ($MinPts$) of objects. The main steps of the DBSCAN procedure can be summarized as follows:

- 1) DBSCAN starts with an arbitrary point. It counts all the neighbor points within distance ϵ of the starting point.
- 2) If the number of neighbors is at least $MinPts$, the point is marked as core.

3) If the number of neighbors is less than *MinPts* and the considered point is a neighborhood of a core point, the point is marked as border.

4) The remaining points are considered as noise points.

5) For the first core point, a cluster is formed and the neighbors are added to this cluster. The algorithm repeats the evaluation process for all the core points recursively: If the considered point is not previously associated to a cluster, a new cluster will be created, else the algorithm proceeds to iterate through the remaining unvisited points in the data set.

4.2 Algorithm

In the following, we present the basic version of DBSCAN algorithm that has two inputs (ϵ and *MinPts*) and generates the clusters and their centers:

- 1) Dispose of $\{\theta_i\}_{i=1}^N$ from a given data set (φ_i, y_i) .
- 2) Choose the input parameters: ϵ and *MinPts*.
- 3) Determine the core points as

$$D_i^2 = \|\theta_i - \theta_j\|, \quad i, j = 1, \dots, N. \quad (10)$$

Count the number Nb_{ng} of the neighbors of θ_i satisfying $D_i \leq \epsilon$

```

if  $Nb \geq MinPts$  then
     $\theta_i$  is a core point
else
    if ( $Nb < MinPts$ ) and ( $\theta_i$  is a core's neighbor)
    then
         $\theta_i$  is a border point
    else  $\theta_i$  is a noise point
    end if
end if

```

4) Cluster the core and the border points into cluster-label

```

for  $k = 1$  : number of core points do
    if cluster-label( $k$ ) = 0 then
        create new cluster-label
    end if
    for  $t = 1$  :  $Nb_{ng}(k)$  do
        if cluster-label ( $t$ ) = 0 then
            add the neighbor to the same cluster-label
        end if
    end for
end for

```

5) The number of clusters = size of (cluster-label)

6) The cluster center = mean(core points of the same cluster).

4.3 Properties

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

1) This method generates automatically the number of sub-models.

2) This method does not require the initialization of centers. Therefore, the problem of convergence towards local minima does not appear.

3) This method removes the misclassified points and consider them as noise. This operation improves the performance of the clustering step.

5 Simulation example

The purpose of these simulations is to illustrate the performance of the proposed method and to compare its efficiency with that of the modified *k*-means method^[12]. To achieve this objective, we consider the following quality measures^[30].

1) The maximum of relative error of parameter vectors is defined by

$$\Delta_\theta = \max_{i=1, \dots, s} \frac{\|\theta_i - \bar{\theta}_i\|_2}{\|\theta_i\|_2} \quad (11)$$

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

2) The averaged sum of the squared residuals is defined by

$$\sigma_e^2 = \frac{1}{s} \sum_{i=1}^s \frac{SSR_i}{|D_i|} \quad (12)$$

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 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.

3) The percentage of the output variation that is explained by the model is defined by

$$FIT = 100 \times \left(1 - \frac{\|\hat{y} - y\|_2}{\|y - \bar{y}\|_2} \right) \quad (13)$$

where \hat{y} and y are the estimated and the real outputs' vectors, respectively; and \bar{y} is the mean value of y .

The identified model is considered acceptable if *FIT* is close to 100.

Consider the following PWARX model^[20]:

$$y(k) = \begin{cases} [0.4 \ 0.5 \ 0.3] \bar{\varphi}(k) + e(k), & \text{if } \varphi(k) \in H_1 \\ [-0.7 \ 0.6 \ -0.5] \bar{\varphi}(k) + e(k), & \text{if } \varphi(k) \in H_2 \\ [0.4 \ -0.2 \ -0.2] \bar{\varphi}(k) + e(k), & \text{if } \varphi(k) \in H_3 \end{cases} \quad (14)$$

where $s = 3$, $n_a = 1$, $n_b = 1$, and $\varphi(k) = [y(k-1) \ u(k-1)]^T$ is the regressor vector.

$$\begin{aligned} H_1 &= \{ \varphi \in \mathbf{R}^2 : [1 \ 0.3 \ 0] \bar{\varphi} \geq 0 \text{ and } [0 \ 0.5 \ 0], \bar{\varphi} > 0 \} \\ H_2 &= \{ \varphi \in \mathbf{R}^2 : [1 \ 0.3 \ 0] \bar{\varphi} \leq 0 \text{ and } [1 \ -0.3 \ 0], \bar{\varphi} < 0 \} \\ H_3 &= \{ \varphi \in \mathbf{R}^2 : [1 \ -0.3 \ 0] \bar{\varphi} \geq 0 \text{ and } [0 \ 0.5 \ 0], \bar{\varphi} \leq 0 \}. \end{aligned} \quad (15)$$

System 15 is simulated using an input signal $u(k)$ and a noise signal $e(k)$ which are normal distributions with variances 0.5 and 0.05, respectively. The output $y(k)$ is presented in Fig. 1.

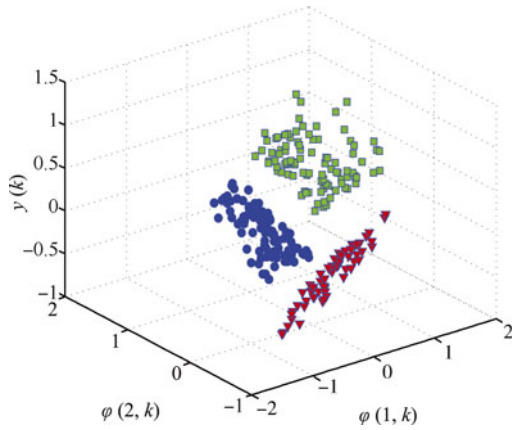


Fig. 1 The real output of the system (squares: output of sub-model 1; triangles: output of sub-model 2; dots: output of sub-model 3)

5.1 Data classification

Now, we determine the local parameter vectors using the least square method by taking a number of neighbors $n_p = 21$. The parameter vectors are presented in Fig. 2.

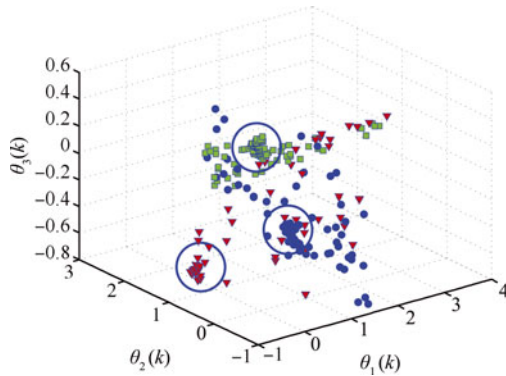


Fig. 2 Local parameters

The DBSCAN clustering technique allows not only to separate these local parameters but also to remove the misclassified ones. The noise points are eliminated from the data set before starting the computation of clusters.

Another important benefit of the proposed method is that it can determine automatically the number of sub-models s . This is shown in Fig. 3 that illustrates the local parameters separated into $s = 3$ sets, where the clusters' centers are depicted by the star symbols.

5.2 Parameter estimation

Table 1 presents the estimated parameter vectors obtained with the proposed method and the k -means one.

5.3 Region estimation

We use the SVM algorithm for the estimation of the regions H_i . For the proposed method they are defined by the following inequalities:

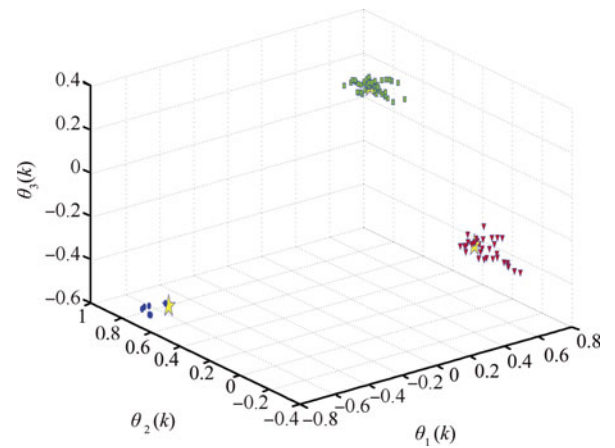


Fig. 3 Local parameters separated into 3 sets

Table 1 Estimated parameters

	True values	Proposed method	k -means
θ_1	$\begin{bmatrix} 0.4 \\ 0.5 \\ 0.3 \end{bmatrix}$	$\begin{bmatrix} 0.4054 \\ 0.4903 \\ 0.2992 \end{bmatrix}$	$\begin{bmatrix} 0.4064 \\ 0.5464 \\ 0.2598 \end{bmatrix}$
θ_2	$\begin{bmatrix} -0.7 \\ 0.6 \\ -0.5 \end{bmatrix}$	$\begin{bmatrix} -0.7369 \\ 0.6675 \\ -0.5239 \end{bmatrix}$	$\begin{bmatrix} -0.6955 \\ 0.5903 \\ -0.4939 \end{bmatrix}$
θ_3	$\begin{bmatrix} 0.4 \\ -0.2 \\ -0.2 \end{bmatrix}$	$\begin{bmatrix} 0.4679 \\ -0.1977 \\ -0.2298 \end{bmatrix}$	$\begin{bmatrix} 0.4792 \\ -0.2101 \\ -0.2406 \end{bmatrix}$

$$H_1 = \left\{ \varphi \in \mathbf{R}^2 : \begin{bmatrix} -0.9347 & 4.3010 & 0.6823 \\ 4.0579 & 0.7345 & 0.8633 \end{bmatrix}, \bar{\varphi} \geq 0 \right\} \quad (16)$$

$$H_2 = \left\{ \varphi \in \mathbf{R}^2 : \begin{bmatrix} 0.9342 & -4.2998 & -0.6817 \\ 3.5076 & -2.0378 & 0.2352 \end{bmatrix}, \bar{\varphi} \geq 0 \right\} \quad (17)$$

$$H_3 = \left\{ \varphi \in \mathbf{R}^2 : \begin{bmatrix} -4.0646 & -0.7318 & -0.8602 \\ -3.5083 & 2.0387 & -0.2350 \end{bmatrix}, \bar{\varphi} \geq 0 \right\}. \quad (18)$$

For the k -means method, the following inequalities are obtained

$$H_1 = \left\{ \varphi \in \mathbf{R}^2 : \begin{bmatrix} 0.3255 & 3.1594 & -0.1354 \\ 3.1094 & 1.4634 & -0.1090 \end{bmatrix}, \bar{\varphi} \geq 0 \right\} \quad (19)$$

$$H_2 = \left\{ \varphi \in \mathbf{R}^2 : \begin{bmatrix} -2.9235 & -1.4932 & 0.1469 \\ -2.4998 & 1.6640 & -0.1243 \end{bmatrix}, \bar{\varphi} \geq 0 \right\} \quad (20)$$

$$H_3 = \left\{ \varphi \in \mathbf{R}^2 : \begin{bmatrix} -0.3255 & -3.1594 & 0.1354 \\ 2.4998 & -1.6640 & 0.1243 \end{bmatrix}, \bar{\varphi} \geq 0 \right\}. \quad (21)$$

To obtain the output of the system with the two methods, we consider the input of Fig. 4.

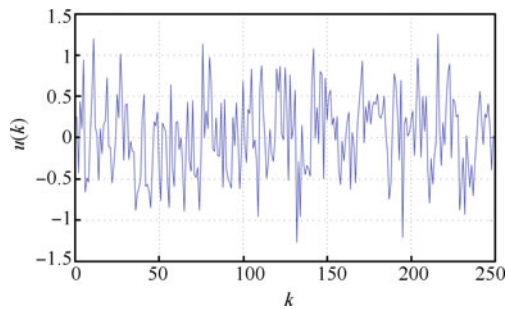


Fig. 4 The input used for the two methods

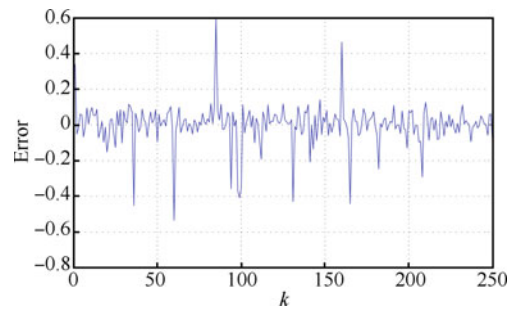


Fig. 8 Error for the k -means method

Figs. 5 and 6 illustrate the outputs and the error signals of the proposed method.

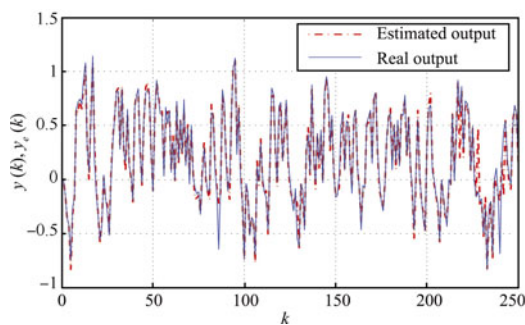


Fig. 5 Real and estimated outputs of the proposed method

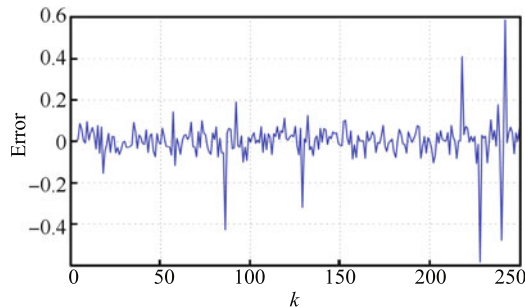


Fig. 6 Error for the proposed method

Figs. 7 and 8 illustrate the outputs and the error signals of the k -means method.

Table 2 presents the quality measures (11), (12) and (13) of the proposed method and the modified k -means method.

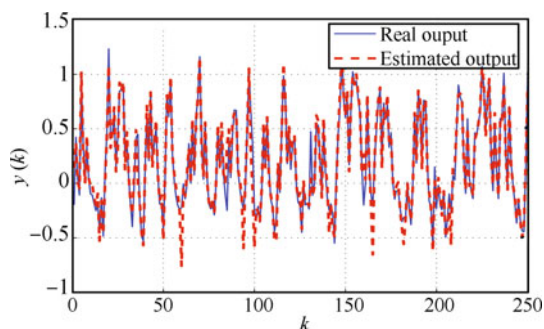


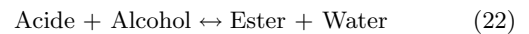
Fig. 7 Real and estimated outputs: modified k -means method

Quality measures	Proposed method	k -means method
Δ_θ	0.1514	0.1828
σ_e^2	0.0097	0.0109
FIT	79.0183	74.195

Based on the results presented in Table 2, we observe that the proposed method gives the best results compared with those obtained by the modified k -means method.

6 Experimental example: a semi-batch reactor

This section presents an application of the proposed method to an olive oil esterification reactor. This reactor carries out, by an alcohol, a chemical reaction of vegetable olive oil esterification. The reaction is given by the following scheme:



The obtained ester is a product with a very high added value. It is used in fine chemical industry. This reaction is highly exothermic and very difficult to operate manually^[31]. A block diagram of this reactor is given in Fig. 9.

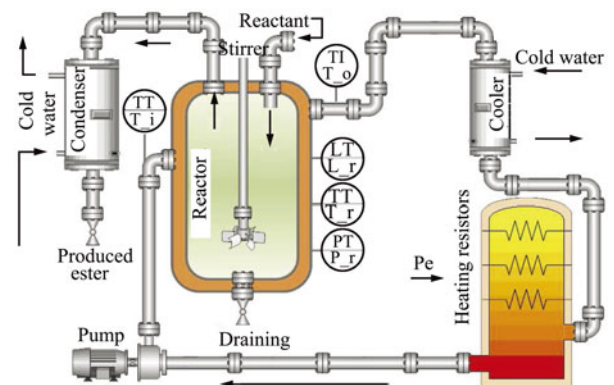


Fig. 9 Block diagram of the reactor

The core is a cylindrical stainless-steel reactor. Its temperature is regulated by means of a fluid circulating through a surrounding jacket, the fluid is either heated by a set of three resistors with electric power P_e , or it is cooled in the tabular cooler. A vertical stirrer, set along the reactor axis, is continuously rotated at a constant speed in order

to keep the medium as homogeneous as possible in terms of temperature and composition. Finally, a semi-batch recipe should be applied, a specific device that is located above the reactor allows to supply any additional reactant. A water-cooler condenser must be used to flow the solvent back to the reactor. We note three variables measured by sensors: inlet temperature T_i , outlet temperature T_o and reacting temperature T_r ^[31]. As shown in Fig. 10, the temperature must follow a specific trajectory.

1) Heating stage: The reactor's temperature T_r is increased to 105°C.

2) Reaction stage: The reactor's temperature T_r is maintained constant during the reaction (when no more water is dripped out of the condenser).

3) Cooling stage: The reactor's temperature is decreased.

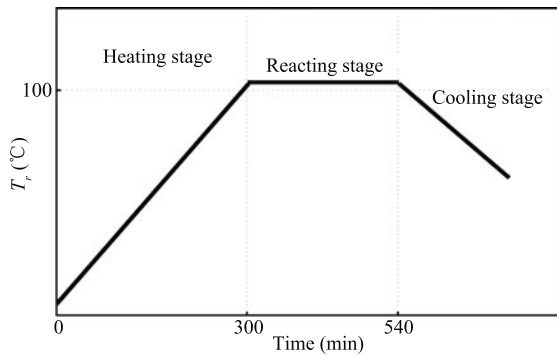


Fig. 10 Specific trajectory of the reactor temperature

An experimental study carried out on the reactor showed that the plant, having the heating power P_e as input and the reaction temperature T_r as output, is a nonlinear process. The variation of the quality of the reagent inside the reactor as well as the external effects can be regarded as random disturbances^[31]. So the classical modeling-method based on one global model can not lead to satisfactory results. That is why we are interested in decomposing the global model into different local models. Thus the alternative of considering a PWA map is very interesting because the characteristic of the system can be considered as piecewise linear in each operating phase: the heating phase, the reacting phase and the cooling phase.

The reactor is considered as a single-input-single-output system. The input $u(k)$ denotes the heating power P_e and the output $y(k)$ denotes the temperature T_r in the reactor.

Previous works have demonstrated that the adequate estimated orders n_a and n_b of each sub-model are equal to two^[32]. Thus, we can adopt the following structure

$$y(k) = \begin{cases} -a_{1,1}y(k-1) - a_{1,2}y(k-2) + b_{1,1}u(k-1) + \\ \quad b_{1,2}u(k-2), & \text{if } \varphi(k) \in H_1 \\ \vdots \\ a_{s,1}y(k-1) + a_{s,2}y(k-2) + b_{s,1}u(k-1) + \\ \quad b_{s,2}u(k-2), & \text{if } \varphi(k) \in H_s \end{cases} \quad (23)$$

where the regressor vector is defined by

$$\varphi(k) = [-y(k-1), -y(k-2), u(k-1), u(k-2)]^T$$

and the parameter vectors is denoted by

$$\theta_i(k) = [a_{i,1}, a_{i,2}, b_{i,1}, b_{i,2}], \quad i = 1, \dots, s.$$

We have taken some input-output measurements from the reactor in order to identify a model for this process. We have taken two measurement files, one for the identification having a length $N = 220$ and another one of length $N = 160$ for validation.

The measurement file used in this identification is presented in Fig. 11.

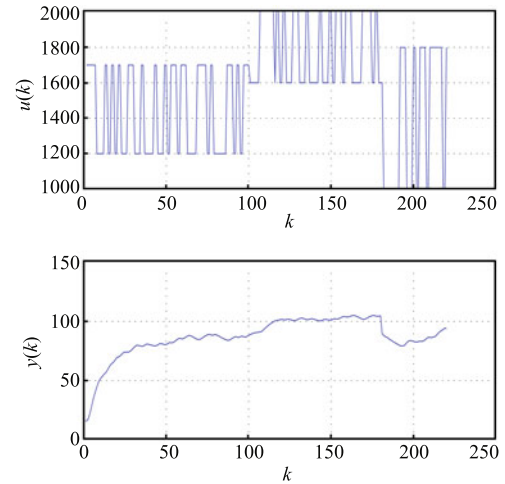


Fig. 11 The real input-output evolution

We apply the proposed identification procedure in order to represent the reactor by a PWARX model with the number of neighboring $n_\rho = 70$. Our purpose is to estimate the number of sub-models s , the parameter vectors $\theta_i(k)$, $i = 1, \dots, s$ and the hyper planes defining the partitions $\{H_i\}_{i=1}^s$. The number of sub-models and the parameter vectors are simultaneously estimated: We obtain $s = 3$, and the parameters are presented in Table 3.

Table 3 Estimated parameter vectors

Parameter vectors	Estimated parameters
θ_1	-1.4404
	0.4692
	0.0003
	0.0014
θ_2	-1.1144
	0.1772
	0.0003
	0.0032
θ_3	-1.0591
	0.1304
	0.0006
	0.0034

It remains only to determine the partitions $\{H_i\}_{i=1}^s$ that are defined by the following inequalities

$$H_i = \{\varphi \in \mathbf{R}^4 : M_i \bar{\varphi} \geq 0\} \quad (24)$$

$$M_1 = \begin{bmatrix} -0.0114 & 0.03124 & -7.8978 & -17.2775 & 0.0004 \\ 0.4138 & 0.4370 & -1.5539 & -8.5077 & 0.0005 \end{bmatrix} \quad (25)$$

$$M_2 = \begin{bmatrix} 0.0114 & -0.0312 & 7.8983 & 17.2783 & -0.0004 \\ 0.3168 & 0.3268 & 3.9548 & 4.4686 & 0.0001 \end{bmatrix} \quad (26)$$

$$M_3 = \begin{bmatrix} -0.4138 & -0.4370 & 1.5550 & 8.5083 & -0.0005 \\ -0.3169 & -0.3269 & -3.9556 & -4.4688 & -0.0001 \end{bmatrix}. \quad (27)$$

After estimating s , θ_i and H_i we obtain the output given by Fig. 12.

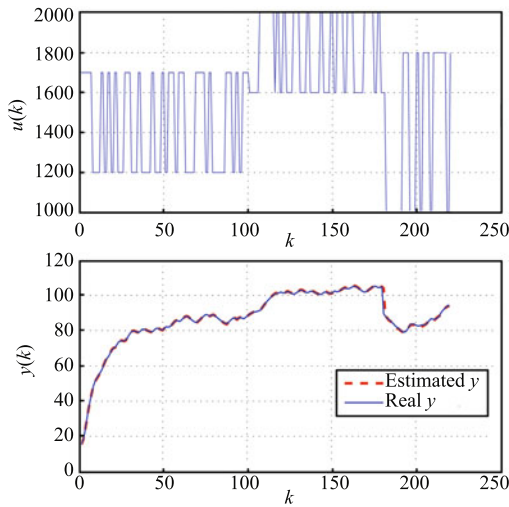


Fig. 12 The input, the real and the estimated outputs

To validate the obtained models, we have considered a new input-output measurement file having a length $N = 160$. Fig. 13 shows the real and the estimated outputs. Fig. 14 illustrates the error signal.

Based on Figs. 13 and 14, we can remark that the proposed method gives good results even in the experimental case, which proves the importance of using the DBSCAN clustering method.

7 Conclusions

In this paper, we have considered the clustering based procedure for identification of PWARX models. In fact, we have recalled the main steps of clustering based approach. This presentation shows that the used clustering algorithms require a good initial guess in order to converge to global minima. In addition, these algorithms are sensitive to the presence of outliers. To overcome this problem, we have suggested the DBSCAN algorithm for data classification. This algorithm presents several advantages. Firstly, this method does not require any initialization and the problem of convergence towards local minima is overcome. Secondly, this method is able to remove the outliers and consider them as noise points. Finally, our approach generates automatically the number of sub-models. Numerical simulation results are presented to demonstrate the performance of the proposed approach. Also, an experimental validation with

an olive oil reactor is presented to illustrate the efficiency of the developed method.

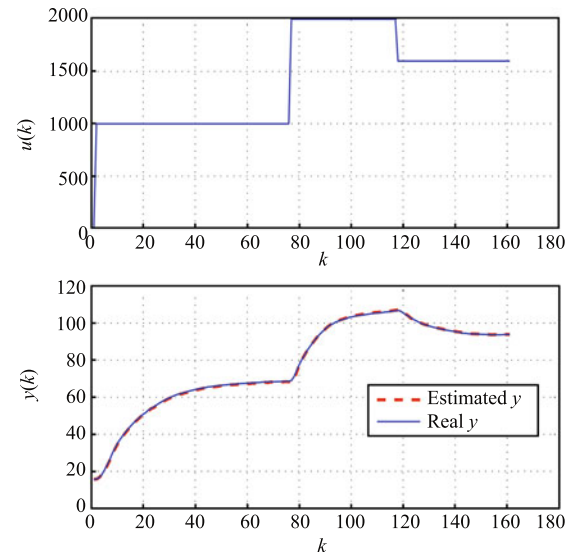


Fig. 13 The input, the real and the estimated validation outputs

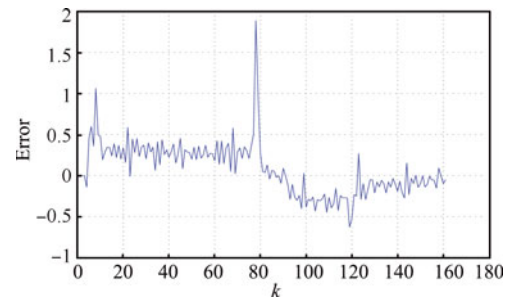


Fig. 14 Error for the validation measurements

References

- [1] J. N. Lin, R. Unbehauen. Canonical piecewise-linear approximations. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 39, no. 8, pp. 697–699, 1992.
- [2] X. Feng, K. A. Loparo, Y. Ji, H. J. Chizeck. Stochastic stability properties of jump linear systems. *IEEE Transactions on Automatic Control*, vol. 37, no. 1, pp. 38–53, 1992.
- [3] A. Doucet, N. J. Gordon, V. Krishnamurthy. Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, vol. 49, no. 3, pp. 613–624, 2001.
- [4] A. Bemporad, M. Morari. Control of systems integrating logic, dynamics, and constraints. *Automatica*, vol. 35, pp. 407–428, 1999.
- [5] A. Bemporad, G. Ferrari-Trecate, 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, T. J. J. 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, 2000.
- [7] A. van der Schaft, J. M. 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*, IEEE, Phoenix, AZ, vol. 4, USA, pp. 3978–3983, 1999.
- [9] W. P. M. H. Heemels, B. De Schutter, A. Bemporad. On the equivalence of classes of hybrid dynamical models. In *Proceedings of the 40th IEEE Conference on Decision and Control*, IEEE, Orlando, FL, USA, vol. 1, pp. 364–369, 2001.
- [10] L. Bako, K. Boukharouba, S. Lecoeuche. Recovering piecewise affine maps by sparse optimization. In *Proceedings of IFAC Conference on System Identification*, IFAC, Brussels, Belgium, vol. 16, no. 1, pp. 356–361, 2012.
- [11] L. Bako. Identification of switched linear systems via sparse optimization. *Automatica*, vol. 47, no. 4, pp. 668–677, 2011.
- [12] G. Ferrari-Trecate, M. Muselli, D. Liberati, M. Morari. A clustering technique for the identification of piecewise affine systems. *Automatica*, vol. 39, no. 2, pp. 205–217, 2003.
- [13] A. L. 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 *Proceedings of the 8th International Conference on Hybrid Systems: Computation and Control*, Springer-Verlag, Berlin Heidelberg, pp. 354–369, 2005.
- [14] Y. Tian, T. Floquet, L. Belkoura, 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.
- [15] A. Juloski, S. Weiland, W. P. M. H. Heemels. A bayesian approach to identification of hybrid systems. *IEEE Transactions on Automatic Control*, vol. 50, no. 10, pp. 1520–1533, 2005.
- [16] A. Bemporad, A. Garulli, S. Paoletti, A. Vicino. A bounded-error approach to piecewise affine system identification. *IEEE Transactions on Automatic Control*, vol. 50, no. 10, pp. 1567–1580, 2005.
- [17] G. Ferrari-Trecate, F. Cuzzola, D. Mignone, M. Morari. Analysis of discrete-time piecewise affine and hybrid systems. *Automatica*, vol. 38, no. 12, pp. 2139–2146, 2002.
- [18] G. Ferrari-Trecate, M. Muselli. Single-linkage clustering for optimal classification in piecewise affine regression. In *Proceedings of IFAC Conference on Analysis and Design of Hybrid Systems*, IFAC, pp. 33–38, 2003.
- [19] H. Nakada, K. Takaba, T. Katayama. Identification of piecewise affine systems based on statistical clustering technique. *Automatica*, vol. 41, no. 5, pp. 905–913, 2005.
- [20] K. Boukharouba. Modélisation et Classification de Comportements Dynamiques Des Systèmes Hybrides, Ph.D. dissertation, Université de Lille, France, 2011. (in French)
- [21] K. Boukharouba, L. Bako, S. Lecoeuche. Identification of piecewise affine systems based on dempster-shafer theory. In *Proceedings of IFAC Conference on System Identification*, IFAC, Saint-Malo, France, vol. 15, no. 1, pp. 1662–1667, 2009.
- [22] J. Sander, M. Ester, H. P. Kriegel, X. W. Xu. Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications. *Data Mining and Knowledge Discovery*, Springer, vol. 2, no. 2, pp. 169–194, 1998.
- [23] Chaudhari Chaitali G. Optimizing clustering technique based on partitioning DBSCAN and ant clustering algorithm. *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 2, no. 2, pp. 212–215, 2012.
- [24] S. Ingo, C. Andreas. *Support Vector Machines*, Berlin, Germany: Springer, 2008.
- [25] R. O. Duda, P. E. Hart, D. G. Stork. *Pattern Classification and Scene Analysis*, New York: Wiley.
- [26] C. Hsu, C. Lin. A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 415–425, 2002.
- [27] J. Weston, C. Watkins. Support vector machines for multiclass pattern recognition. In *Proceedings of the 7th European Symposium on Artificial Neural Networks*, IEEE, Bruges, Belgium, vol. 4, pp. 219–224, 1999.
- [28] M. T. Elbatta, R. M. Bolbol, W. M. Ashour. A vibration method for discovering density varied clusters. *ISRN Artificial Intelligence*, vol. 2012, Article ID 723516, 2011.
- [29] J. Erman, M. Arlitt, A. Mahanti. Traffic classification using clustering algorithms. In *Proceedings of the 2006 SIGCOMM Workshop on Mining Network Data*, ACM, New York, NY, USA, pp. 281–286, 2006.
- [30] A. L. Juloski, S. Paoletti, J. Roll. Recent techniques for the identification of piecewise affine and hybrid systems. *Current Trends in Nonlinear Systems and Control*, Birkhäuser Boston: Springer, pp. 79–99, 2006.
- [31] F. M’sahli, R. Abdennour, M. Ksouri. Experimental nonlinear model based predictive control for a class of semi-batch chemical reactors. *International Journal of Advanced Manufacturing Technology*, vol. 20, no. 6, pp. 459–463, 2002.
- [32] S. Talmoudi, K. Abderrahim, R. B. Abdennour, M. Ksouri. Multimodel approach using neural networks for complex systems modeling and identification. *Nonlinear Dynamics and Systems Theory*, vol. 8, no. 3, pp. 299–316, 2008.



Zeineb Lassoued received her engineer degree in electrical-automatic engineering and the M.Eng. degree in automatic and smart techniques from the National School of Engineers of Gabes (ENIG), Tunisia in 2010 and 2011, respectively. Currently, she is a Ph.D. candidate at Research Unit of Numerical Control of Industrial Processes (CONPRI), University of Gabes, Tunisia.

Her research interests include hybrid system identification and control.

E-mail: Zeineb.Lassoued1@gmail.com (Corresponding author)



Kamel Abderrahim received his engineer degree in electrical engineering from the National School of Engineers of Gabes (ENIG), Tunisia in 1992, and his M.Eng. degree in automatic control from Higher School of Sciences and Techniques of Tunis, Tunisia (ESSTT) in 1995, and his Ph.D. degree in electrical engineering from National School of Engineers of Tunis, Tunisia (ENIT) in 2000, and the Habilitation in

electrical engineering from the University of Gabes in 2009. He has been a member of Laboratory of Numerical Control of Industrial Processes (LACONPRI) at the ENIG since 1995. He joined the ENIG as an assistant professor in 2000, and now he works as a professor at the ENIG. From 2002 to 2005, he was the director of the Electrical Engineering Department at the ENIG. From 2005 to 2011, he was the director of Higher Institute of Industrial Systems, Tunisia (ISSIG).

His research interests include nonlinear process modeling, identification, and control.

E-mail: kamelabderrahim@yahoo.fr