

New Clustering Algorithm for Identification of a Nonlinear Stochastic Model

TROUDI Ahmed, HOUCINE Lassad
Higher School of Sciences and Techniques of
Tunis (ESSTT)
Research unit (C3S)
Tunisia
(troudi.ahmed, houcinelassad@yahoo.fr

BOUZBIDA Mohamed, CHAARI Abdelkader
Higher School of Sciences and Techniques of
Tunis (ESSTT)
Research unit (C3S)
Tunisia
bouzbida_mohamed@hotmail.fr, assil.chaari@esstt.rnu.tn

Abstract— Many clustering algorithms have been proposed in literature to identify the premise and consequence parameters involved in the TS fuzzy model. In this paper these parameters are estimated at the same time and this from the minimization of four optimization criteria. The proposed algorithm constitutes an extension of the algorithm proposed by J.Q.Chen in 1998. However, in this paper we introduced some modification on the optimization criteria and especially the last two criteria, thus we replaced the Euclidean distance by another non-Euclidean distance when calculating the fuzzy partition matrix. The purpose of these modifications is to introduce more robustness with the algorithm especially for highly nonlinear systems and those operating in a stochastic environment. The efficiency of the algorithm is tested on an electro-hydraulic system.

Index Terms— Nonlinear system, fuzzy clustering, TS fuzzy model, stochastic environment, linguistic modeling, fuzzy identification, non-Euclidean distance.

II. INTRODUCTION

Most industrial processes are nonlinear. In this context, several techniques have been developed to carry out the modeling of this type of process. The neuro-fuzzy technique [7] and the fuzzy clustering technique [2] [3] [9] constitute one of the best approaches used for the representation of such process. Indeed, several researchers have noted that a nonlinear system can be approximated by the sum of several linear subsystems. In this case each subsystem is represented by a fuzzy rule. The number of rules (clusters) is fixed by an expert according to the type of application considered and the performances required by this last. Several clustering algorithms exist in literature allowing the identification of the premise and consequent parameters intervening in the TS fuzzy model, we can quote as an example the Fuzzy C-Mean algorithm (FCM) [4], the Gustafson Kessel

algorithm (GK) [1]. Moreover, these algorithms are sensitive to noises or outliers. To overcome these disadvantages Krishnapuram and Keller have proposed the Possibilistic C-Means algorithm (PCM) [6] by

abandoning the constraint of FCM and constructing a novel objective function. The PCM can deal with noisy data better than FCM, GK. However, FCM, GK, and PCM are all only allowed the identification of the premise parameters while the consequent parameters are estimated using the least squares methods. In 1998 J.Q. Chen proposed another clustering algorithm allowing the identification of the premise and consequent parameters at the same time and this by using an iterative optimization method and starting from the minimization of four optimization criterion. This algorithm has many drawbacks such as convergence to local optima, sensitivity to noise and the aberrant point, also the computation time is very slow. In this paper we propose a new clustering algorithm to overcome this problem. This algorithm consists to introduce some modification to the optimization criteria and more particularly the last two criteria. Inspired by Krishnapuram and Keller algorithm, we introduce two new objective functions into J.Q. Chen algorithm to replace the last two objective functions J3 and J4 in it. The algorithm which we have proposed overcomes the problems of sensitivity to noise and aberrant point better than FCM, GK, PCM and J.Q. Chen algorithms. However, FCM, GK, PCM and J.Q. Chen algorithm are all based on Euclidean distance in their objective function. In real world, the Euclidean distance is not complex enough to deal with more sophisticated problem. In order to introduce more robustness to the algorithm, Wu and Yang (2002) have proposed a non-Euclidean distance to replace the Euclidean distance in FCM algorithm. Inspired by Wu and Yang's algorithm, we introduce the new distance into J3 objective function to replace the Euclidean distance in it when calculating the fuzzy partition matrix. The introduction of this distance ensures convergence towards a global minimum in contrast to the other algorithms that have been proposed. The efficiency of this algorithm is tested on an electro-hydraulic system.

This paper is organized as follows:

Next section gives a brief overview of the TS fuzzy model. The criteria for fuzzy identification are presented in section 3. The proposed algorithm is introduced in

section 4. The simulation results are described in section 5. The validation model is presented in section 6. And finally section 7 concludes the paper.

II. TAKAGI-SUGENO FUZZY MODEL

The Takagi-Sugeno fuzzy model constitutes an appropriate model for approximating nonlinear systems [8]. It is constructed by a rule-based type If ... Then in which the consequent uses numeric variables rather than linguistic variables (case of Mamdani). The TS fuzzy models consist of linguistic if-then rules that can be expressed by the following form.

$$R_i : \text{If } x_k \text{ is } A_i \text{ then } y_i = a_i^T x_k + b_i \quad (1)$$

The "if" rule function defines the premise part, while the "then" rule function constitutes the consequent part of the TS fuzzy model. $A_i \in R^n$ is a multidimensional antecedent fuzzy set, defined by its membership function $\mu_{A_i}(x_k)$.

$x_k = [x_{k_1}, x_{k_2}, \dots, x_{k_n}] \in R^n$, is the input vector of the premise; $a_i \in R^n, b_i \in R$: are the polynomial coefficients that form the consequent parameters of the i^{th} rules, and $i=1, \dots, c$ (c : denotes the numbers of rules in the rule base). $y_i \in R$: Is the rule output variable.

The output of the general nonlinear system is calculated as the average of output corresponding to the rules multiplied by the degree of fulfillment of the antecedent γ_i of the form:

$$\hat{y} = \frac{\sum_{i=1}^c \gamma_i(x_k) y_i}{\sum_{i=1}^c \gamma_i(x_k)} \quad (2)$$

With:

$$\gamma_i = \mu_{i1}(x_{k_1}) \cdot \mu_{i2}(x_{k_2}) \cdot \dots \cdot \mu_{in}(x_{k_n}) \quad (3)$$

Introduce λ_i : the degree of achievement standard described by the following expression:

$$\lambda_i = \frac{\gamma_i(x_k)}{\sum_{i=1}^c \gamma_i(x_k)} \quad (4)$$

The estimated output of the Takagi-Sugeno fuzzy model can be expressed by:

$$\hat{y} = \sum_{i=1}^c \lambda_i(x_k) [a_i^T x_k + b_i] \quad (5)$$

III. CRITERIAS FOR FUZZY IDENTIFICATION

Unlike to the other clustering algorithms which have been proposed in the literature, named the Fuzzy C-Means algorithm (FCM), the GK algorithm and the Possibilistic C-Means algorithm (PCM), which only allow that the premise parameters identification intervening in the TS fuzzy model. J.Q.Chen proposes another algorithm that allows the identification of premise and consequent parameters simultaneously. It is

composed of fuzzy c-linear functions and Fuzzy C-Means clustering algorithm. Its obtaining requires the minimization of four optimization criteria. However, this algorithm has some disadvantages including: convergence to local optima, sensitivity with respect to noise these are due to the arbitrary choice of the third and fourth optimization criterion. To address this problem and in order to improve more robustness of the algorithm we tried to introduce modification to the optimization criteria which has been proposed, and more particularly on the last two criterions J_3 and J_4 . Moreover we replaced the Euclidean distance by another non-Euclidean distance in calculate of the J_3 criteria. This modification makes it possible to guarantee the robustness of the algorithm with respect to the noise and the aberrant points

III.1 OPTIMIZATION CRITERIAS

The minimization of the J_1 criterion allows the determination of the consequent parameters θ_i :

$$J_1 = \sum_{k=1}^N (y(k) - \sum_{i=1}^c \mu_{ik} y_i(k))^2 \quad (6)$$

The determination of the cluster centers v_i is obtained by minimizing the J_2 criterion:

$$J_2 = \sum_{k=1}^N (x_k - \sum_{i=1}^c \mu_{ik} v_i)^2 \quad (7)$$

The degree of membership μ_{ik} can be obtained by minimizing the following criterion:

$$J_3 = \sum_{i=1}^c (\mu_{ik})^{m_1} d_{ik}^2 + \sum_{i=1}^c \eta_i \sum_{k=1}^N (1 - \mu_{ik})^{m_1} \quad (8)$$

Where N is the total numbers of observations, $d_{ik} = \sqrt{1 - \exp(-\rho \|x_k - v_i\|^2)}$ is a non Euclidian distance from simple point x_k to the cluster center v_i , c is the number of clusters, $U = [\mu_{ik}]$ is a $c \times N$ matrix, denoted a fuzzy partition matrix. m_1 is a weighting exponent, $m_1 > 1$ and η_i is a suitable positive numbers. It should be noted that the first term demands that the distance between x_k to v_i be as low as possible, however the second term force μ_{ik} to be as large as possible.

The concluding truth degree f_{ik} can be obtained by minimizing the J_4 criterion

$$J_4 = \sum_{i=1}^c (f_{ik})^{m_2} (|y(k) - \bar{x}_k \theta_i|)^2 + \sum_{i=1}^c \beta_i \sum_{k=1}^N (1 - f_{ik})^{m_2} \quad (9)$$

$1 < m_2; i = 1, \dots, c; k = 1, \dots, N$

Where $\bar{x}_k = [x_k \ 1]$ and $\theta_i = [a_1^i \ a_2^i \ \dots \ a_n^i \ b_i^i]^T$ are coefficients of consequence linear matrix, m_2 : is a

weighting exponent for f_{ik} and β_i is a suitable positive numbers.

Thus the vector of the consequent parameters θ_i ($i = 1, \dots, c$) can be obtained by minimization of the following criterion:

$$J_1 = \sum_{k \in S_i} (\mu_{ik})^2 (f_{ik})^2 (y(k) - \bar{x}_k \theta_i)^2 \quad (10)$$

$$= \sum_{k=1}^N (\mu_{ik})^2 (f_{ik})^2 (y(k) - \bar{x}_k \theta_i)^2$$

μ_{ik} and f_{ik} are calculated by minimizing J_3 and J_4 criterion respectively

Similarly J_2 can be expressed by the following form:

$$J_2 = \sum_{k=1}^N (\mu_{ik})^2 (f_{ik})^2 (x_k - v_i)^2 \quad (11)$$

III.2. IDENTIFICATION ALGORITHM FOR PREMISE AND CONSEQUENCE PARAMETERS

The parameters identification intervening in the TS fuzzy model is obtained by minimization of four optimization criteria J_1, J_2, J_3 and J_4 expressed by “10”, “11”, “8” and “10” respectively.

Their minimization is complete in an iterative way and by using the following theorems:

Theorem 1: assume that v_i are fixed, then the minimization of the J_3 criteria gives:

$$\mu_{ik} = \frac{1}{1 + \left(\frac{d_{ik}^2}{\eta_i} \right)^{\frac{1}{m_1 - 1}}} \quad (12)$$

Theorem 2: Assume that θ_i ($i = 1, \dots, c$) are fixed. The minimization of the J_4 criteria gives:

$$f_{ik} = \frac{1}{1 + \left(\frac{|y(k) - \bar{x}_k \theta_i|}{\beta_i} \right)^{\frac{1}{m_2 - 1}}} \quad (13)$$

Theorem 3: Assume that μ_{ik} and f_{ik} are fixed, then coefficients θ_i are obtained from the minimization of the J_1 criterion:

$$\theta_i = (\bar{X}_k^T U_i^2 F_i^2 \bar{X}_k)^{-1} (\bar{X}_k^T U_i^2 F_i^2 Y_k); \quad (i = 1, \dots, c) \quad (14)$$

$$U_i = \text{diag}(\mu_{i1}, \mu_{i2}, \dots, \mu_{iN}), F_i = \text{diag}(f_{i1}, f_{i2}, \dots, f_{iN})$$

Theorem 4: assume that μ_{ik} and f_{ik} are fixed, then the cluster centers of the prototypes v_i are obtained from the minimization of the J_2 criterion:

$$v_i = \frac{\sum_{k=1}^N (\mu_{ik})^2 (f_{ik})^2 x_k}{\sum_{k=1}^N (\mu_{ik})^2 (f_{ik})^2}; \quad i = 1, \dots, c \quad (15)$$

Based on the optimization conditions “12”, “13”, “14” and “15”, the identification algorithm of the premise and consequent parameters is obtained from an iterative optimization algorithm described by the following steps:

IV. PROPOSED ALGORITHM

Given a data set (x_k, y_k) , the new clustering algorithm is given by the following steps.

Step 1: Initialization

- Choose the number of clusters
- Choose the weighting exponent m_1 and m_2
- Let err (0) be a large number

Step 2: initialize consequence parameters θ_i at random

- compute the fuzzy partition matrix U and the cluster centers v_i by the FCM algorithm

Step 3: compute η_i :

$$\eta_i = \frac{\sum_{k=1}^N \mu_{ik}^{m_1} D_{ik}^2}{\sum_{k=1}^N \mu_{ik}^{m_1}}$$

Where $D_{ik}^2 = \|x_k - v_i\|^2$

Compute β_i :

$$\beta_i = \frac{\sum_{k=1}^N f_{ik}^{m_2} (y(k) - \bar{x}_k \theta_i)^2}{\sum_{k=1}^N f_{ik}^{m_2}}$$

Step 4: compute the new clustering distance d_{ik} :

Step 5: Update the fuzzy partition matrix: “12”

Step 6: Update the matrix f_{ik} by: “13”

Step 7: compute the clusters center v_i “15”

Compute θ_i by: “14”

Step 8: compute the err = $\|\theta - \theta(0)\|$

Step 9: if err $< \varepsilon_1$ then turn to step10; else $\theta(0) = \theta$, turn to step 6.

Step 10: compute the estimated output and the sum of squared errors between the estimated output and the actual output by:

$$\hat{y}(k) = \sum_{i=1}^c \mu_{ik} y_i(k)$$

Step 11: if error $< \varepsilon_2$ or rate $< \varepsilon_3$ then stop; else $c = c + 1$, error(0) = error, and turn to step 2

$$\text{error} = \left(\frac{\sum_{k=1}^N (y(k) - \hat{y}(k))^2}{N} \right)$$

$$rate = \left| \frac{error - error(0)}{error} \right|$$

V. SIMULATION RESULTS

In this section, we will present two examples. All examples are nonlinear and difficult to be described by the ordinary method, so the fuzzy model presented in this paper is adopted.

Example 1: This system is described by the following equation:

$$y(k+1) = \frac{y(k)(y(k-1)+2)(y(k)+2.5)}{8.5+y^2(k)+y^2(k-1)} + u(k) + e(k)$$

Where $y(k)$, $u(k)$ are the output and the input of the system respectively?

$e(k)$ is a linear noise given by the recurrent equation.

$$e_1(k+1) = \cos(\beta)e_1(k) + \sin(\beta)e_2(k)$$

$$e_2(k+1) = -\sin(\beta)e_1(k) + \cos(\beta)e_2(k)$$

$$e(k) = 0.5e_1(k)$$

$$\beta = \frac{\pi}{6}$$

In this case, we present the simulation results concerning the identification of the algorithms we have introduced previously.

i. There exists the system by a random binary signal given in Fig. 1.

ii. For another input, the simulation results given by the proposed algorithm is given in Fig. 2.

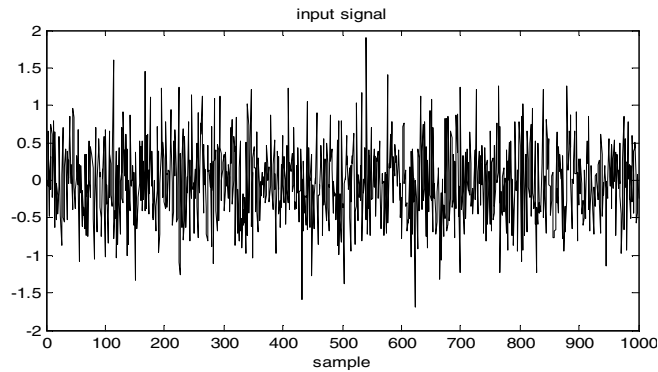


Figure 1. Sequences of input-output

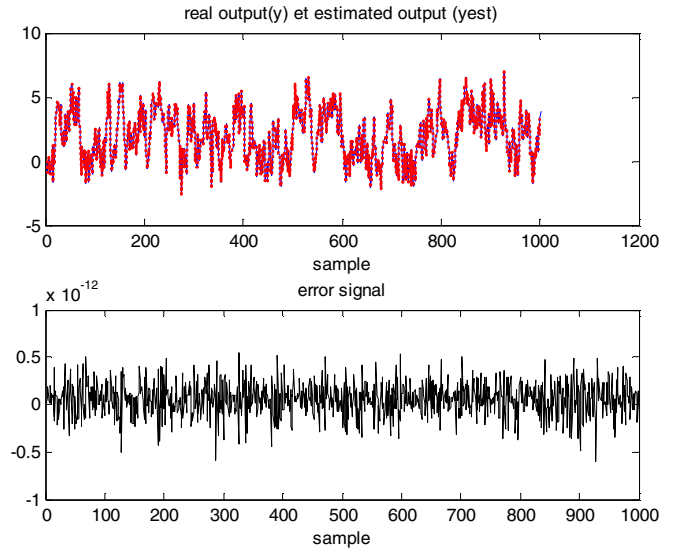


Figure 2. Identification results for the proposed algorithm

Example 2: APPLICATION TO AN ELECTRO-HYDRAULIC SYSTEM

The effectiveness of the identification algorithm we proposed in this paper is tested on an electro-hydraulic system described by the schematic diagram in Fig. 3.

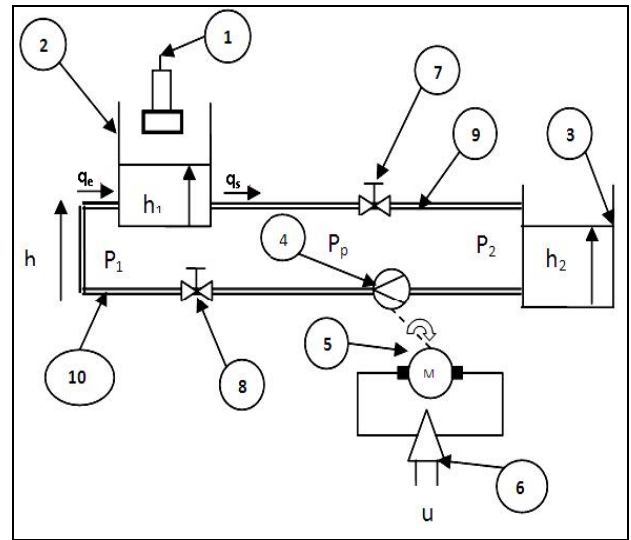


Figure 3. Bloc diagram

- 1 : ultrasonic level sensor
- 2 : Tank 1
- 3 : Tank 2
- 4 : centrifugal pump
- 5 : DC motor
- 6 : Variable speed
- 7 : manual valve v1
- 8 : manual valve v2
- 9 : Pipe 1
- 10 : Pipe 2

IDENTIFICATION OF SYSTEM PARAMETERS

To identify the parameters of this system, we applied a proposed clustering algorithm. The set of observations we have taken is illustrated in Fig. 4.

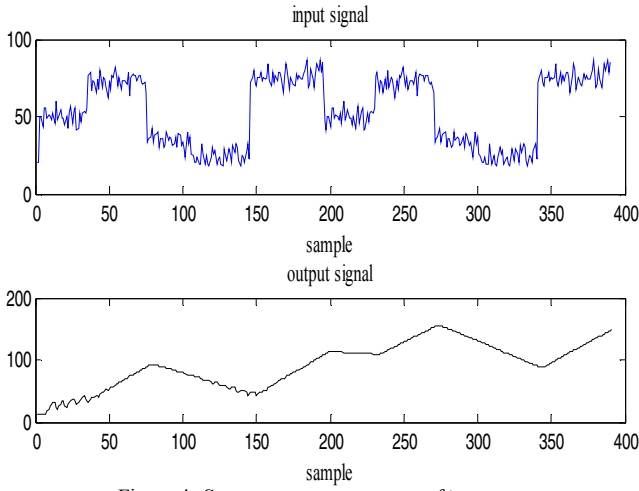


Figure 4. Sequences of input-output

For another sequences of input-output, the simulation result given by the proposed algorithm is given as follows:

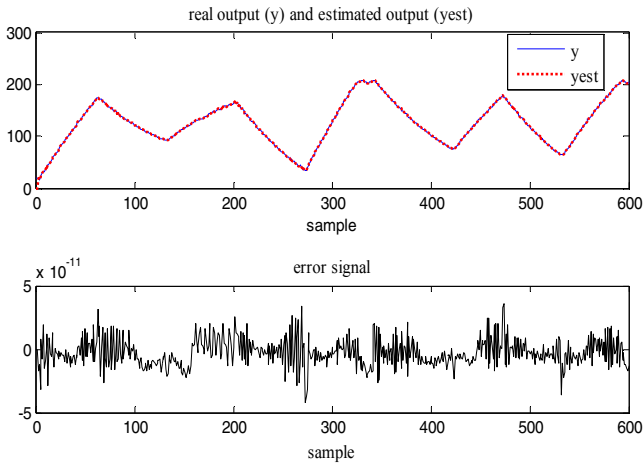


Figure 5. Identification results for the proposed algorithm

VI. VALIDATION MODEL

Therefore, to ensure that the model obtained from the estimation it is compatible with other forms of inputs to represent correctly system operating to identify it. It we present, in this paragraph, statistical tests to validate a prediction model based on the RMSE test and the VAF test.

VI.1. RMSE (Root Mean Square Error)

This test calculates the mean squared error between the measured output and model output.

$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (y(k) - \hat{y}(k))^2} \quad (16)$$

When the model output and actual output are very near, the test tends to zero.

VI.2. VAF (Variance Accounting For)

This test calculates the percentage standard deviation of the variance between the measured output and model output. It is defined this way:

$$VAF = 100\% \left[1 - \frac{\text{var}(y(k) - \hat{y}(k))}{\text{var}(y(k))} \right] \quad (17)$$

TABLE I. VALIDATION RESULTS (EXAMPLE 1)

	RMSE	VAF (%)
Proposed algorithm	1.831410^{-13}	100
J.Q. Chen algorithm	$3.62 \cdot 10^{-7}$	98.97

The simulation results (table 1) show that the proposed algorithm can effectively solve the problem of the J.Q.Chen algorithm. The validation tests used have shown good performance of these algorithms (Figure 2).

TABLE II. VALIDATION RESULTS (EXAMPLE 2)

	RMSE	VAF
Proposed algorithm	$0.1478 \cdot 10^{-9}$	99.9999
J.Q. Chen algorithm	$8.22 \cdot 10^{-4}$	98.97

The validation results (RMSE and VAF test) show well the effectiveness of the proposed algorithm compared to the algorithm proposed by J.Q. Chen

VII. CONCLUSION

In this paper another algorithm for the identification of nonlinear stochastic systems is used. The proposed algorithm can estimate simultaneously the premise and consequence parameters by using an iterative optimization method it is starting from the minimization of four optimization criteria. In fact this algorithm is an extension of the algorithm proposed in J.Q.Chen 1998. In this paper to introduce some modifications in the last two criteria were further replaced the Euclidean distance by another non Euclidean distance. The simulation results illustrate the effectiveness of this algorithm.

REFERENCES

- [1] D.E. Gustafson, and W.C. Kessel. "Fuzzy Clustering with a fuzzy covariance matrix." In Proc. IEEE CDC, San Diego, CA, USA, pp. 761-766, 1979.
- [2] I. Gath and A.B. Geva. "Unsupervised optimal fuzzy clustering." IEEE Transactions on Pattern Analysis and Machine Intelligence, 7, pp. 773-781, 1989.
- [3] J.Adonyi, R. Babuska and F. Szeifert, "Modified Gath Geva Fuzzy Clustering for Identification of Takagi-Sugeno Fuzzy Models" IEEE Trans. Syst, Man, Cybern.Part B: vol.32, no 5, pp. 612-621, oct.2002.
- [4] J. C. Dunn. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters, J. Cybern., vol. 3, no. 3, pp. 32-57, 1974.
- [5] Jian-Qin Chen, Yu-Geng Xi and Zhong-Jun Zhang. "A clustering algorithm for fuzzy model identification." Fuzzy Sets and Systems 98, pp. 319-329, 1998.

- [6] R. Krishnapuram and J. Keller. The Possibilistic c-Means Algorithm: Insights and Recommendations, IEEE Trans. Fuzzy Systems, vol. 4. No. 3, pp. 385-393, 1996.
- [7] R. Babuska and H. Verbruggen. "Neuro-fuzzy methods for nonlinear system identification." Annual Reviews in control 27, pp. 73-85, 2003.
- [8] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its application to modelling and control. IEEE Trans. On systems, Man and Cybernetics, vol. 15, pp. 116-132, 1985.
- [9] T. Ahmed, C. Abdelkader G. Moncef. "Algorithmes de clustering pour l'identification d'un modèle flou. JTEA, pp. 1085-1094, 2008.
- [10] T. Ahmed, H. Lassad and C. Abdelkader G. "New Extended Possibilistic C-Means algorithm for identification of an Electro-hydraulic system". 12th. Inter. Conf. STA. ACS. 1695, pp. 1-10, 2011.