24th ABCM International Congress of Mechanical Engineering
December 3-8, 2017, Curitiba, PR, Brazil

COBEM-2017-2900
RECONCILIATION OF PATTERNS IN THE CLUSTERING OF TIME SERIES

Izete Celestina dos Santos Silva
Graduate Program in Industrail Engineering - PEI, UFBA, Rua Aristides Novis n°2, 6° andar – Federação, 40210-630
izabox@gmail.com

Pedro Moreira Arruti Aragão
Polytechnic School, UFBA, Rua Aristides Novis n°2, 6° andar – Federação, 40210-630
pedrom.aragao@outlook.com

Cristiano Hora de Oliveira Fontes
Graduate Program in Industrail Engineering - PEI, UFBA, Rua Aristides Novis n°2, 6° andar –Federação, 40210-630
cfontes@ufba.br

Raony Maia Fontes
Graduate Program in Industrail Engineering - PEI, UFBA, Rua Aristides Novis n°2, 6° andar – Federação, 40210-630
raonyfontes@gmail.com

Marcelo Embiruçu de Souza
Graduate Program in Industrail Engineering - PEI, UFBA, Rua Aristides Novis n°2, 6° andar – Federação, 40210-630
embirucu@ufba.br

Abstract. *The significant increase in the relevance of issues related to reliability and safety in the production processes of the various segments of the production network has led organizations to seek efficient methods to diagnose and detect possible failures in their processes. Although there are many works related to Fault Detection and Diagnosis (FDD), few of them are based on clustering and pattern recognition in time series, especially the multivariate ones. In addition, there are no works related to the pattern recognition in time series that consider the process model as a constraint. This paper proposes a new method for the pattern recognition in uni and multivariate time series, based on Fuzzy C-Means (FCM), which directly considers the process dynamics in the clustering problem in order to ensure the feasibility of the recognized patterns. The proposed method is applied in a case study that comprises the clustering and pattern recognition of abnormal (failures) operation of a nonisothermal Continuous Stirred Tank Reactor (CSTR), a well-known benchmark system used to compare various monitoring solutions and also used for the assessment of FDD techniques. The results show that the proposed method (FCM coupled with a process model) is able to recognize patterns consistent with the behavior of the process without worsening the quality of clustering and classification.*

Keywords: *clustering, multivariate time series, fault diagnosis, process model*

1. INTRODUCTION

Due to the complexity of the production processes, companies from various segments of the production network are looking for technological solutions to monitor, store and manipulate data in order to acquire knowledge of the process. One of the applications of knowledge Discovery in Data bases (KDD) comprises the Fault Detection and Diagnosis (FDD) (Fontes and Pereira, 2016).

Data clustering and pattern recognition in time series have been investigated through the use of traditional techniques, especially in the case of univariate series (Liao, 2005; Keogh and Kasetty, 2002; Trebuňa and Halčinová, 2013). The Fuzzy C-Means (FCM) clustering algorithm represents a robust and consolidated alternative of non-hierarchical clustering capable of recognizing patterns in uni or multivariate time series associated with process variables (Liao, 2005; Izakian and Pedrycz, 2015; Trebuňa and Halčinová, 2013; D'urso and Maharaj, 2012). The clustering of multivariate time series is less investigated and comprises a more complex problem due to its intrinsic features such as the choice of the similarity metric and clustering validation. One of the widely used similarity metrics for the comparison between multivariate series comprises the PCA-based similarity metrics (SPCA) in its different versions (Singhal and Seborg, 2006; Khediri, Limam and Weihs, 2011; Deng, Tian and Chen, 2013).

However, traditional SPCA and the classic FCM algorithm are not able to ensure that the recognized patterns are consistent with the dynamic behavior of the process, which, depending on the quality of information in the data, can lead to obtaining patterns distant from the dominant dynamics or even not achievable. So far, no work involving the reconciliation of patterns in optimization-based clustering problems has been verified. This paper reveals for the first time possible inconsistencies in the pattern recognition in multivariate time series using an optimization strategy (FCM) and presents two approaches (sequential and simultaneous) associated with two different types of problems involving the reconciliation of these patterns through the dynamic process model. As in the traditional practice of data reconciliation (Shuanghua, McLean and Thibault, 2007), pattern reconciliation means modifying patterns obtained from historical data making them consistent with the reality of the process while preserving the quality of the clustering/classification results.

2. PRELIMINARIES

2.1 Multivariate time series and the PCA-based Similarity metrics (SPCA)

The univariate time series (UST) and multivariate time series (MST) are considered as objects in problems involving clustering and pattern recognition (Maharaj and D'urso, 2012; Li and Wen, 2014). Time series comprise a series of observations over time (time series) associated with a specific process variable $\boldsymbol{z}_j(t)$ $(j=1,\ldots,p\ ;\ t=1,\ldots,m)$ (p is the number of variables, m is the number of observations and t indexes the measurements made at each time instant). A MTS object comprises the case in which $p \geq 2$ and can be represented by the following $m \times p$ matrix (Fontes and Budman, 2017):

$$\boldsymbol{Z}_i = \begin{bmatrix} z_{i1}(1) & \cdots & z_{ip}(1) \\ \vdots & \ddots & \vdots \\ z_{i1}(m) & \cdots & z_{ip}(m) \end{bmatrix} \tag{1}$$

where $\boldsymbol{Z}_i$ is the object, $z_{ij}(t)$ is the measurement of variable j (j=1, …, p) at time instant t (t=1, …, m) in the object $\boldsymbol{Z}_i$ (i=1,…,n objects). The column j contains the time series related to the variable j.

The Similarity metrics (SPCA) is based on the Principal Component Analysis (PCA) and quantifies the level of similarity between two different objects (MTS represented by two matrices $m \times p$, respectively) by comparing the directions associated with the main components of each object (Yang and Shahabi, 2004; Singhal and Seborg, 2006). The matrices to be compared can have the same number of columns (p), but not necessarily the same number lines or observations (m) (MTS). The SPCA is restricted to the interval [0; 1] and values close to the unit indicate high similarity.

$$SPCA_{\lambda}\ (\boldsymbol{A},\boldsymbol{B}) = \frac{1}{\sum_{i=1}^{k_0}(\lambda_i^A . \lambda_i^B)} . \sum_{i=1}^{k_0}\sum_{j=1}^{k_0}(\lambda_i^A\ .\ \lambda_j^B)\cos^2\theta_{ij} \tag{2}$$

where A and B are the matrices (multivariate time series), θ_{ij} is the angle between the ith principal component of A and the jth principal component of B, k_1 and k_2 are the quantities of principal components capable of describing more than 95% of the variability of matrices A and B, respectively, and k_0 is the largest of k_1 and k_2. where $\boldsymbol{\lambda}^{\boldsymbol{A}}$ and $\boldsymbol{\lambda}^{\boldsymbol{B}}$ are vectors with the eingenvalues of $\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}$ and $\boldsymbol{B}^{\boldsymbol{T}}\boldsymbol{B}$, respectively and "." indicates scalar product.

Equation (2) presents the modified SPCA$_{\lambda}$, in which each major component is weighted by the square root of its corresponding eigenvalue (Singhal and Seborg, 2006), thus considering that each component describes different levels of variability.

2.2 Method Fuzzy C-Means (FCM)

Cluster analysis or simply clustering aims at partitioning a given set (sample) of objects into homogeneus and well separated clusters (subsets) according to criteria of similarity and dissimilarity (Dao et al., 2017, Döring and Lesot, 2006; Hair, Black, Babin, Anderson and Tathan, 2009). The fuzzy c-means algorithm (FCM) is a non-hierarchical clustering method that requires the pre-specification of the number of groups (Bezdek, 1981). The FCM is based on an optimization problem Eq. (3) and (4) whose decision variables are the centers (v_i, i = 1, …, c, prototypes or patterns) of each one of the c clusters ($c \geq 2$) and the membership degree of each object to each cluster (Liao, 2005; Bezdek, Keller and Krisnapuram, 2005).

The term *fuzzy* derives simply from the inherent characteristic of the FCM in assigning levels of intermediate pertinence (in the interval [0, 1]) for each object in relation to each one of the groups (Berget, Mevik and Neas, 2008, Hoppner, Klawonn, Kruse and Runkler, 1999).

$$\min_{(U,V)} \boldsymbol{J}_{\varepsilon}(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n}\left\{u_{ik}^{\varepsilon}\left\|x_k - v_i\right\|^2\right\} \tag{3}$$

where c is the number of clusters, n is the number of objects, u_{ik} is the membership degree of the k^{th} object to the i^{th} cluster, $\boldsymbol{U}$ is the partition matrix ($c{\times}n$ matrix). The parameter ε (fuzzification coefficient) is related to the inherent uncertainty of the partition problem and was the recommended value in the literature ($\varepsilon = 2$). $\boldsymbol{V}$ is the set of prototype vectors $\{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_c\}$. Two additional constraints must be considered:

$$U \in \Re^{cxn} : u_{ik} \in [0,1] \quad \sum_{k=1}^{n} u_{ik} > 0 \quad \forall i \quad \text{and} \quad \sum_{i=1}^{c} u_{ik} = 1 \tag{4}$$

The analytical solution of the optimization problem represented by equations (5) and (6) comprises the following expressions for the centers (patterns) of the clusters and levels of pertinence:

$$\boldsymbol{v}_i = \frac{\sum_{k=1}^{n}(u_{ik})^{\varepsilon}.\ x_k}{\sum_{k=1}^{n}(u_{ik})^{\varepsilon}}, \quad i = 1,...,c \tag{5}$$

and

$$u_{ik} = \frac{\left(\frac{1}{\left\|x_k - \boldsymbol{v}_i\right\|^2}\right)^{\frac{1}{\varepsilon-1}}}{\left(\frac{1}{\left\|x_k - \boldsymbol{v}_j\right\|^2}\right)^{\frac{1}{\varepsilon-1}}} \quad k = 1,...,n \quad i = 1,...,c \tag{6}$$

The classical FCM algorithm comprises an iterative procedure through equations (5) and (6) and the metric adopted is Euclidian (Berget, Mevik and Neas, 2008). In problems involving multivariate series, in which Euclidean distance does not apply, equation (6) can not be used, and a classical optimization algorithm must be used to solve the problem described by Eq. (3) and (4) (Fontes and Pereira, 2016).

In this work the modified $SPCA_{\lambda}$, Eq. (2) was adopted as a similarity metric. In addition, a bi-criterion constrained clustering was adopted in order to maximize the split between clusters (patterns) avoiding the coalescence of clusters.

$$\min_{(U,V)} J_{\varepsilon}(U,V) = \left\{\sum_{i=1}^{c}\sum_{j=1}^{n} u_{ik}^{\varepsilon}\left[SPCA_c(x_i,v_j)\right]^2 + \beta\cdot\sum_{i=1}^{c}\sum_{\substack{j=1\\ j>i}}^{c}\frac{1}{\left[SPCA_c(v_i,v_j)\right]^2}\right\} \tag{7}$$

where β is an adjustable parameter. It was verified that even values close to zero are able to avoid coalescence between groups. $SPCA_c$ is the complement of $SPCA_{\lambda}$ $\left(SPCA_c\left(x_k,v_i\right) = 1 - SPCA_{\lambda}\left(x_k,v_i\right)\right)$.

2.3 Reconciliation of Patterns

Although the reconciliation of measured process data has been extensively analyzed over the past few years (Shuanghua, McLean and Thibault, 2007), there are no studies to date suggesting the need for reconciliation of patterns involving clustering and classification of time series.

In this work, two types of problems are presented with the respective approaches of sequential and simultaneous resolution aiming at the reconciliation of patterns in time series.

2.3.1. MTS with output variables

The first type of problem comprises the situation in which the time series present in each object comprise only process output variables. In this case, the reconciliation of the patterns obtained consists in determining the temporal trajectory of each output variable considered to be closer to the standard recognized by the clustering algorithm, but at the same time, it attends the process dynamics. The resolution of this problem comprises the determination of a time sequence of inputs consistent with reality (satisfies pre-specified process constraints) and, as a result, the respective trajectory of the output variable predicted by the process model (reconciled pattern).

Once the centers (patterns) are recognized through the modified FCM algorithm (v_i, i = 1, …, c, [Eq. 7 and 4]), the reconciliation comprises the following resolution of the following optimization problem, output variables present in each object are affected by n and input variables, we have:

$$\min_{w_1(t),\ldots,w_{ne}(t)} J_m = \sum_{i=1}^{c} \left\| y_i - y_i^r \right\|^2 \tag{8}$$

Subject to:

$$L_{hi} \le w_i(t) \le S_{hi}, \Delta L_{hi} \le \Delta w_i(t) \le \Delta S_{hi} \qquad i = 1,\ldots,n_e \tag{9}$$

$$y_i^r = f(w_i(t),\ldots,w_{n_e}(t),t) \;\rightarrow\; \text{Process model} \tag{10}$$

where w_i ($i=1,\ldots, n_e$) are the input variables. The Eq. (9) constraints comprise maximum and minimum limits on the absolute values and respective variations of each input, at each instant of time. y_i e y_i^r $(i = 1,\ldots,c)$ are the trajectories of the output variables referring to the patterns recognized by the clustering algorithm and the reconciled trajectories, respectively. The metric distance used for each output variable is the Euclidean distance in order to reconcile the dynamics of the recognized trajectory to the dominant dynamics of the process (described through the process model).

2.3.2. MTS with input and output variables

The second type of problem comprises multivariate time series involving output variables and input variables. In this case, the reconciliation consists in relating the time series referring to the output variables, in each recognized standard, with the time series referring to the inputs in the same standard, through the process model.

The reconciliation strategy is carried out in this case through a simultaneous approach. A *bi-criterion constrained* clustering comprises an objective function whose the first criterion is related to the fuzzy clustering itself Eq. (7) and the second criterion sets a bond between the time series of each pattern based on the process model.

Consider a process with n_u input variables of which only n_1 entries ($n_1 < n_u$) are part of each object, with n_2 remaining entries ($n_2 + n_1 = n_u$). The first ones are called *Prototype Inputs* (PI) and the others are called *Free Inputs* (FI). Therefore, the set of input variables is divided into two subsets, namely, $w_j^{PI},\ (j = 1,\ldots,n_1)$ and $w_j^{FI},\ (j = 1,\ldots,n_2)$.

Each recognized standard (and each of the objects) v_i ($i=1, \ldots, c$) is composed of time series referring to n_y output variables and n_1 input variables (PI). Each pattern v_i includes an array of n_y+n_1 columns, where v_{ij}^y $(i = 1,\ldots,c$ and $j = 1,\ldots,n_y)$ refers to the time series of the output variable j in the pattern associated with group i and v_{ij}^u $(i = 1,\ldots,c$ and $j = 1,\ldots,n_1)$ refers to the time series of the input variable w_j^{PI} in the pattern associated with group i.

$$v_i = \left[\underbrace{v_{i1}^y \quad \cdots \quad v_{in_y}^y}_{n_y\ outputs} \ \underbrace{v_{i1}^u \quad \cdots \quad v_{in_1}^u}_{n_1\ prototype\ inputs} \right] \tag{11}$$

where $v_{ij}^u \equiv w_j^{PI}(t),\ j = 1,\ldots,n_1$.

The optimization problem is structured as follows:

$$\min_{(U,V)} J_{\varepsilon}(U,V) = \alpha \left\{ \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ik}^{\varepsilon}\left[SPCA_c(x_i,v_j)\right]^2 + \beta \cdot \sum_{i=1}^{c}\sum_{\substack{j=1\\ j>i}}^{c} \frac{1}{\left[SPCA_c(v_i,v_j)\right]^2} \right\} + (1-\alpha)\sum_{i=1}^{c}\sum_{y=1}^{n_y}\left\| v_{iy} - v_{iy}^{r} \right\|^2 \quad (12)$$

Subject to:

$$\sum_{k=1}^{n} u_{ik} > 0 \quad \forall_{i,} \quad \sum_{i=1}^{c} u_{ik} = 1 \quad \forall_{k} \quad (13)$$

$$L_i \le \Delta w_i(t) \le S_i \quad i = 1, \ldots, n_u \quad (14)$$

$$\Delta L_i \le \Delta w_i(t) \le \Delta S_i \quad i = 1, \ldots, n_u \quad (15)$$

$$v_{iy}^{r} = f\left(p, \underbrace{v_{i1}^{u}, \ldots, v_{in_1}^{u}(t)}_{prototype\ inputs}, \underbrace{w_1^{FI}(t), \ldots, w_{n_2}^{FI}(t)}_{free\ inputs}\right) \rightarrow \text{Process model} \quad (16)$$

The *free inputs* can be pre-fixed or not, depending on the features of the problem or process. If fixed, these are not decision variables. The Euclidian distance is used as metrics in the part of the objective function (Eq. 12). α $\left(\alpha \in [0,1]\right)$ is a tuning parameter (trade-off parameter). v_{iy} is the time series for output y (y = 1, …, ny), associated with the standard of group i (i = 1, …, c). The constraints (14) and (15) are analogous to (9).

3. CASE STUDY

The work comprised two steps: obtaining the database (time series); and recognizing patterns for different kinds of fault (fault diagnosis). The nonisothermal Continuous Stirred Tank Reactor (CSTR) with cooling jacket dynamics and a variable liquid level was simulated in order to generate historical data (Singhal and Seborg, 2002; Vaidyanathan and Venkatasubramanian, 1992). This virtual plant (Fig. 1) is a benchmark process used to compare and analyze FDD approaches.

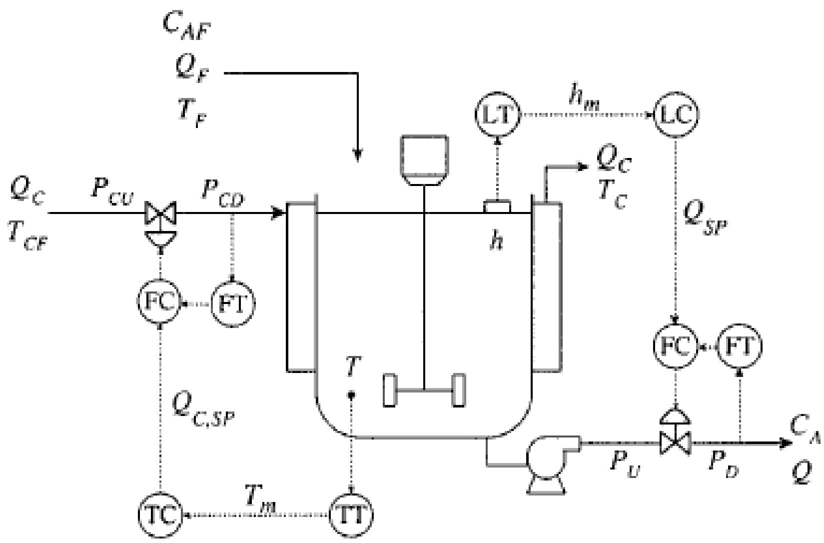

Figure 1: Nonisothermal continuous stirred tank reactor (Singhal and Seborg, 2002)

The objects are represented by time series (Fig. 2). Two types of failures were analyzed, both related to the disturbances in reactor charge flow (Q_F): 30 fault objects were related to the step changes (with different amplitudes); and 30 other objects were related to the oscillations of damped and sustained types. Figure 2 presents the respective time series (time window of 5 min) associated with two state variables, i.e., reactant concentration in the reactor (C_A), and reactor temperature (T).

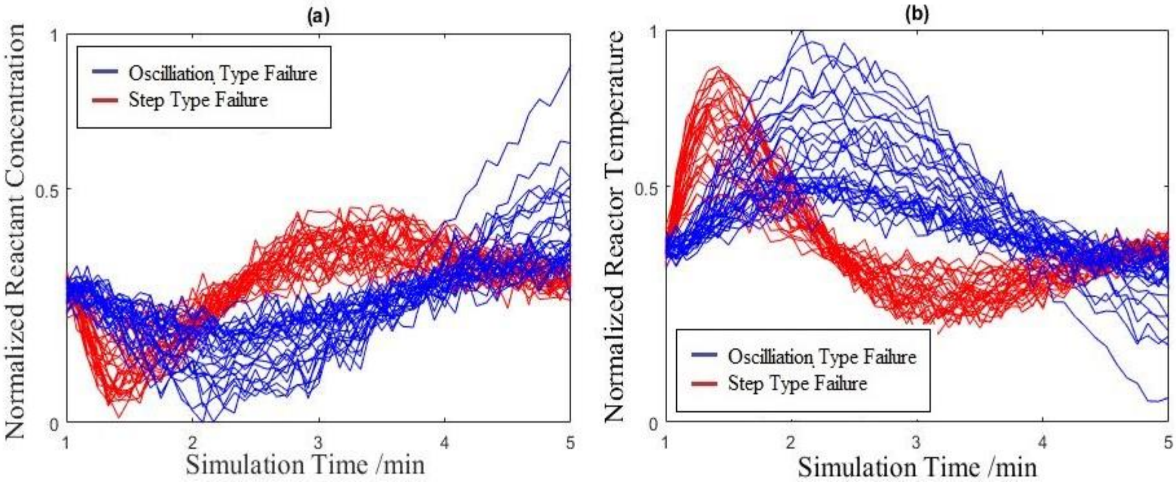

Figure 2: Time series. (a) reactant concentration; (b) reactor temperature

4. RESULTS

The first test comprised objects composed of output variables (C_A and T). Figure 3 shows the patterns of each one of the 2 clusters considered using FCM. The clustering results using the classical FCM algorithm (Eq. 3 and 4) presented quite satisfactory results and no case of misclassification occurred. SPCA was able to recognize the complete distinction between the failure objects (step disturbance and oscillatory disturbance in charge flow) and no misclassification was verified. Although in both type of failures the temperature patterns (with and without reconciliation) are dynamically quite similar, the same is not true for concentration. The pattern of the reactant concentration (C_A) recognized for the oscillatory disturbance without reconciliation (just FCM) shows inconsistent dynamic behavior. In both type of failures, the initial increase in the reactor temperature should cause an initial decrease in the reactant concentration which is really verified in the reconciled pattern. The heterogeneity among the objects associated to this type of failure (oscillatory disturbance in the feed flow) suggests that the SPCA metric and the classic algorithm FCM do not ensure the adherence to the process once there is an increase in the complexity of the clustering problem associated with the absence of uniform behavior between the objects of the same label (oscillatory disturbance). This is a typical and frequent situation in data extracted from real industrial systems subject to noises and unknown disturbances. In addition, the reconciliation approach (Eq. 8, 9 and 10) did not change the quality of the clustering and misclassification results have not been verified.

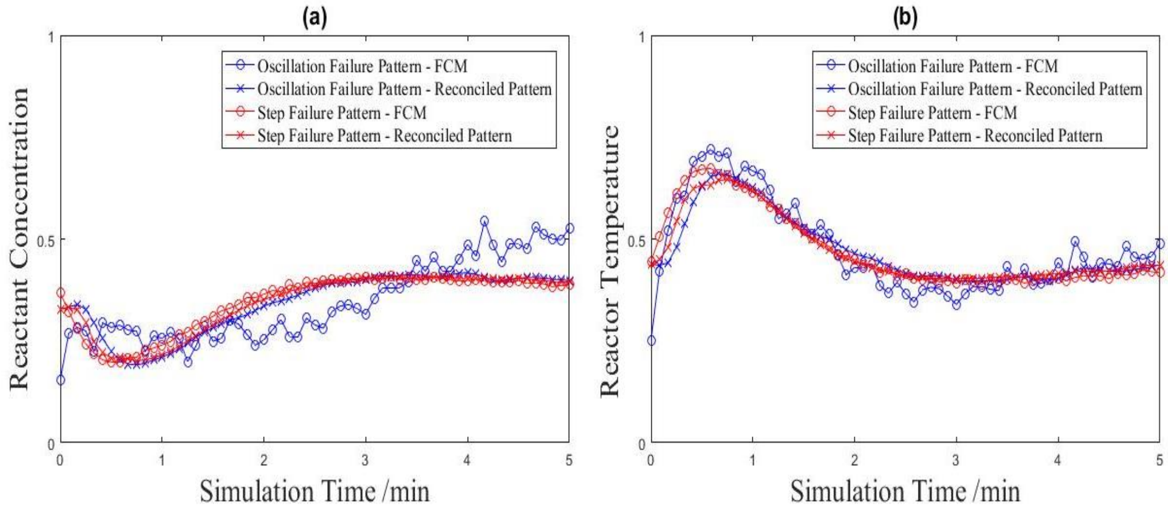

Figura 3: Centers or Patterns (sequential reconciliation approach).
(a) reactant concentration; (b) reactor temperature

A second test comprised objects with the two output variables (C_A e T) as well as the feed rate (Q_F) and the simultaneous reconciliation approach was applied. The results obtained are shown in Figure 4. The reconciled patterns also maintained the same quality of the classification results obtained with classic FCM (without reconciliation).

Figure 4 shows that reconciled patterns present dynamic behavior different from the original patterns (without reconciliation). In some cases, such as in the patterns related to the feed rate (Q_F), the results obtained without reconciliation are not consistent with the reality of the process and shows an initial decrease in this variable, as opposed to the expected type of failure.

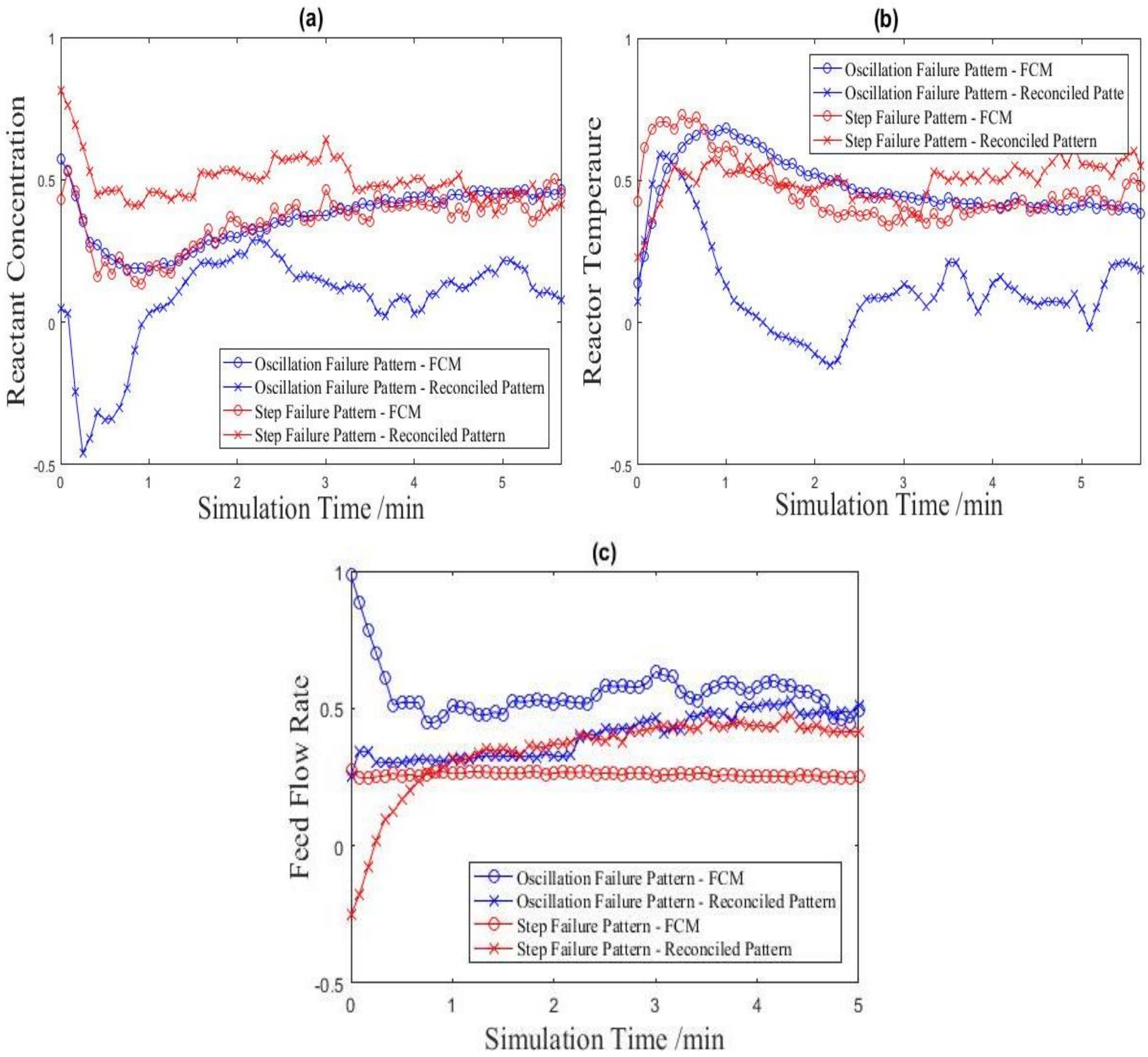

Figure 4: Centers or Pattern (simultaneous reconciliation approach).
(a) reactant concentration; (b) reactor temperature; (c) feed flow rate.

5. CONCLUSION

This work presented the clustering and pattern recognition of abnormal (failures) operation for a nonisothermal CSTR, which is a benchmark process used to compare and analyze FDD approaches. The results obtained show that FCM algorithm coupled with SPCA are not capable of recognizing patterns consistent with the process dynamics even with e achievement of classification results.

Two reconciliation approaches are proposed and apply according to the nature of the variables (inputs and/or outputs) present in each sample object. The two approaches were able to obtain reconciled pattern without compromising the quality of the clustering and classification results, i.e., without impairing the quality of fault diagnosis.

6. ACKNOWLEDGEMENTS

This work was supported by Brazilian Federal Agencies for Supporting Research and Graduate Education: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); and Conselho Nacional de Pesquisa e Desenvolvimento (CNPq).

7. REFERENCES

Berget, I., Mevik, B. and Neas, T., 2007. "New modifications and applications of fuzzy C-means methodology". *Computational Statistics & Data Analysis*, Vol. 52, p 2403-2418.

Bezdek, J. C., 1981. "Pattern Recognition with Fuzzy Objective Function Algorithms". New York: Plenum Press.

Bezdek, J. C., Keller, J., Krisnapuram, R. and Pal, N., 2005. "Fuzzy Models and Algorithms for Pattern Recognition and Image Processing". Springer Science Business Media, New York, 1° Edition.

Brockwell, P. J. and Davis, R. A., 1987. "Time Series: Theory and Methods". Springer-Verlag, Colorado-USA, 1° Edition

D'urso, P. and Maharaj, E. A., 2012 "Wavelets-based clustering of multivariate time series," *Fuzzy Sets Syst.*, vol. 193, p. 33–61.

Dao, T., Duong, K. and Vrain, C., 2017. "Constrained clustering by constraint programming", *Artificial Intelligence*, Vol. 244, pp. 70-94.

Deng, X. Tian, X. and Chen, S., 2013. "Modified kernel principal component analysis based on local structure analysis and its application to nonlinear process fault diagnosis". *Chemom. Intell. Lab. Syst*. Vol. 127, p. 195-209.

Döring, C. and Lesot, M., 2006. "Data analysis with fuzzy clustering methods". *Computational Statistics & Data Analysis*, Vol. 51, p. 192-214.

Fontes, C. H. O. and Budman, H., 2017. "A Hybrid clustering approach for multivariate time series – a case study applied to failure analysis in a gas turbine", *ISA Transactions*.

Fontes, C.H.O and Pereira, O., 2016. "Pattern recognition in multivariate time series – A case study applied to fault detection in a gas turbine". *Enginnering applications of artificial intelligence*, Vol 49, p. 10-18.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. and Tathan, R., 2009. "Análise Multivariada de Dados", Porto Alegre-Brasil, Editon Bookman, 6° edition.

Hoppner, F., Klawonn, F., Kruse, R. and Runkler, T., 1999. "Fuzzy cluster analysis - Methods for classifican, data analysis and image recognition", *New York, Editora Jonh Wiley & Son*, LTD.

Izakian, H., Pedrycz, W. and Jamal, I., 2015. "Fuzzy clustering of time series data using dynamic time warping distance", *Eng. Appl. Artif. Intell*, Vol. 39, p. 235-244.

Keogh, E. J. and Kasetty, S. 2002. "On the need for time series data mining benchmarks: a survey and empirical demonstration", *in Proceedings of the International Conference on Knowledge Discovery and Data Mining (ACM SIGKDD)*, p. 23-26.

Khediri, B., Limam, M. and Weihs, C. 2011. "Variable window adaptive Kernel Principal Component Analysis for nonlinear nonstationary process monitoring". *Comput. Ind. Eng.*, Vol. 61, p. 437-446.

Li, S. and Wen, J., 2014. "Application of pattern matching method for detecting faults in air handling unit system," *Autom. Constr.*, Vol. 43, p. 49–58.

Liao, T. W., 2016. "Clustering of time series data - a survey". *Pattern Recognit*, Vol 38, p. 1857- 1874.

Maquin, D., Adrot, O. and Ragot, J., 2000. "Data reconciliation with uncertain models". *ISA Transactions*, Vol. 39, p. 35-45.

Shuanghua, B., McLean, D. D. and Thibault, J., 2007. "Impact of model structure on the performance of dinamic data reconciliation. *Computers and Chemical Engineering*, Vol. 31, p. 127-135.

Singhal, A. and Seborg, D. E., 2002. "Pattern matching in multivariate time series databases using a moving-window approach". *Ind. Eng. Chem. Res.* Vol. 41, p. 3822-3838.

Singhal, A. and Seborg, D. E., 2006. "Evaluation of a pattern matching method for the Tennessee Eastman challenge process". *J. Process Control*, Vol. 16, p. 601-613.

Trebuňa, P. and Halčinová, J., 2013. "Mathematical Tools of Cluster Analysis". *Applied Mathematics*, Vol. 4, p. 814-816.

Vaidyanathan, R. and Venkatasubramanian, V., 1992. "Representing and diagnosing dynamic process data using neural networks". *Engineering Applications of Artificial Intelligence*, Vol. 5 p. 11-21.

Yang, K. and Shahabi, C., 2004. "A PCA-based Similarity Measure for Multivariate Time Series," in p*roceedings of the International Workshop on Multimedia Databases, ACM-MMDB, Washington DC, USA*, pp. 1–10.

8. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.