Intelligent Fault Diagnosis and Prognosis using state validations in a drinking water plant

Hector Hernandez[1], Jorge Camas[1], Nicolás Juárez[1], Madaín Pérez[1], Rafael Mota[1] and Claudia Isaza[2]

[1]Instituto Tecnológico de Tuxtla, ITTG
Carretera Panam. Km. 1080, 29050-Tuxtla Gutiérrez, Chiapas, México
(hhernandezd,jcamas,njuarez,mperez,rmota)@ittg.edu.mx
http://www.ittg.edu.mx

[2]Facultad de Ingeniería, Departamento de Ingeniería Electrónica
Calle 67 Número 53-108, Medellín, Colombia
cisaza@udea.edu.co
http://www.udea.edu.co

Abstract. This paper proposes a transitions validation method between system functional states in drinking water plant monitoring. The method is based in fuzzy entropy measure. The water plant is monitored by means of fuzzy classification method. Diagnosis of the system is the main objective of the work, nevertheless as a complement, prognosis is also proposed whenever maintenance takes an important place of the water plant. In this process, periodic maintenance is fundamental, and its schedule is commonly applied. Then, the objective of the study is coordinate intelligent fault detection with prognosis, in order to propose an adaptive and preventive maintenance schedule to achieve in integrated drinking water plant monitoring. This knowledge is then organized as a partition of the data set into classes representing the functional states of the process (normal or faulty operations). The proposed method validates the recognized functional state in presence of uncertainty using diagnosis and prognosis technique.

1 Introduction

Water industry is facing increased pressure to produce higher quality treated water at a lower cost. In drinking water industrial production processes, the correct operation and maintenance of the complex processes have a crucial role to ensure the supply of the quantity of adequate water to the population and the safeguarding of the environment. Monitoring principle of a dynamic process from a method of classification consists in determining at every moment, the current class which was associated beforehand a functional state of the system.

It's because that in the production phase (stage of recognition), it's a question of deciding sampling at every moment which is the operating condition. This decision is particularly delicate to take at the time of the transitions, i.e. when there is a change in the class to which the whole of measurements (individuals to be classified) is allotted.

© M. Martínez, A. Alarcón (Eds.)
Advances in Computer Science and Engineering.
Research Computing Science 45, 2010, pp. 179-190

Received 30/04/10
Final version 19/05/10
Accepted 14/05/10

The result of fuzzy classification techniques provides the individual adequacy degrees analyzed with each class. In the majority of the algorithms, the decision of classification is obtained by the research of the class to which the individual presents the maximum of membership, or adequacy. In the presence of uncertainties caused by the inaccuracy in measurements, or by the possible not very significant disturbances of the process operation, the transition from a state to another, it can to have a little real justification. In this case, we say that there is a "bad conditioning" to make the decision of change of state. This is why the introduction of a criterion of validation of the transitions was regarded as a significant contribution to the effective process monitoring.

This article approaches the validation of change of state to avoid false transitions or transitions in badly conditioned states. The validation is made in the stage of recognition. This approach is based on the information which east calculates from degrees of adequacy obtained by a fuzzy classification algorithm. For this approach, the essential criterion associated a decision was regards as the evaluation of the information which was necessary to take it.

Generally, the quantity of information is associated the entropy of the data. In the case of fuzzy sets, the entropy nonprobabilistic formula suggested by Luca and Termini [2] is largely used. For transition validations, it is necessary to use the individual instantaneous information that produced the change of state. Consequently, the analysis is based on the unit (vector) degrees of adequacy of this individual to each class. The adequacy vector degrees can be regarded as a fuzzy unit. Transitions validation based on the quantity of information uses a traditional measurement of fuzzy entropy [2][3][4], the decision is considered well conditioned when the quantity of information is significant. However, a decision which was made with low values of the adequacy degrees gives a high value of information. For this, the paper propose to use a reliability index which was inspires by the measurement of fuzzy information (fuzzy entropy of Luca and Termini). This index allows measuring the instantaneous information which caused the change of state and allows holding in account the relationship between the small adequacy degrees and uncertainty on the decision. To validate the approach in experimental form, the method was applied to the process monitoring of drinking water plant of Tuxtla-Mexico. The practical significance of results will be in order to improve the water treatment process.

In section 2, this paper presents a brief description of water treatment process [5]. The transition validation method is presented in section 3, in section 4 the method is evaluated on a real case. Finally the conclusions are presented.

2 Water treatment process

2.1 Brief Drinking Water Plant description

The water is the most abundant compound on the surface of the world. Water treatment involves physical, chemical and biological processes that transform raw water into drinking water which satisfies a whole of standards of quality at a reasonable price for the consumer.

The “SMAPA” water treatment plant (Tuxtla city, Mexico)[10], which was used as an application site for this study, provides water to more than 800,000 inhabitants and has a nominal capacity to process 800 l/s of water per day. The figure 1 presents a schematic overview of the various operations necessary to treat the water.

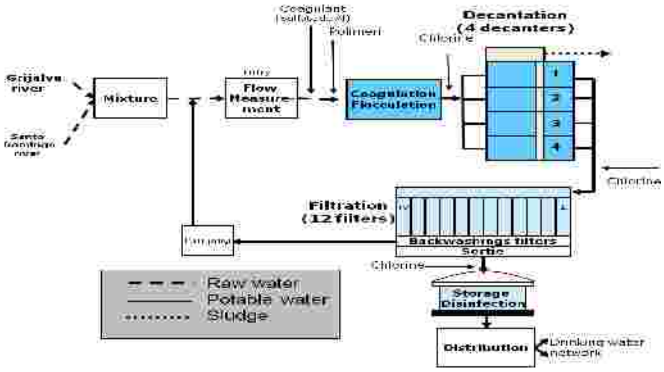

Fig. 1. SMAPA Potable Water Plant

The complete usual chain comprises the following units: clarification (coagulation-flocculation + decantation), disinfection, and filtration. The present work concerns essentially the clarification process which represents the main step in the treatment process. Raw water is collected at the rivers (Grijalva and Santo Domingo), and pumped to the treatment plants.

For decanters maintenance, the critical variable is the water turbidity at the input and the exit of the decanters (before filtering). The sludge accumulation in decanters is considered a non desired state of the plant. During the rainy season eventually there is a solids accumulation which blocks the sludge removal system. Consequently, a part of the solids is transferred to the filtering phase instead of being evacuated properly. Nowadays decanters maintenance is predetermined in February-March. The filters maintenance is programmed before the rain season so as to ensure the filter proper conditions at that period. As well, each filter has a backwashing task which is executed in function of the inlet and outlet filter pressure. A monitoring system of the process state would allow a flexible maintenance. Coagulation process is brought about by adding a highly ionic salt (aluminum sulphate) to the water. A bulky precipitate is formed which electrochemically attracts solids and colloidal particles. The solid precipitate is removed by allowing it to settle to the bottom of the tank and then periodically removing it as sludge. Water is stored in a tank and ready to be transported through the water supply network.

2.2 Monitoring Using Classification Methods General Theory

Process monitoring using classification method consists in determining at each sample time, the current class which was associated beforehand to a functional state of the process. There are two principal phases: the training and the recognition. In the first step (training), the objective is to find the process behavior characteristics which will allow differentiating the process states (each one being associated to a class). The initial algorithm parameters are selected by the process expert who validates the obtained behavioral model. In a posterior step, the data recognition allows to identify one line the current process state. At each sampling time, a vector collects the accessible information (raw data or pre-treated data such as filtered, FFT) which is provided for monitoring, and the class recognition procedure yields the operator what is the current functional state of the process. In order to optimize the obtained partition we propose to include into the training phase a step to automatically validate and adjust the clusters. The proposed new approach automatically improves, in terms of compactness and class separation, a non optimal initial partition helping therefore the discrimination between classes i.e. between operation modes.

3 Transitions validation method

3.1 Fuzzy degree index (Decision index)

Entropy nonprobabilistic functions represent the fuzzy degree of a fuzzy discrete unit whole (μ) respect to the elements which make it up [2][6][7]. The analysis is made according to the adequacy degrees $\mu(x_i)$ of each element (x_i). According to the approach suggested by Luca and Termini, the entropy fuzzy functions type $H(\mu)$ must respond the following axioms [2][6][7][8]:

$$\begin{aligned} &\text{P1 :}\quad H(m) = 0 \Leftrightarrow m(x_i) \in \{0,1\} \\ &\text{P2 :}\quad \max H(m) \quad \Leftrightarrow \quad \forall i \quad m(x_i) = \tfrac{1}{2} \\ &\text{P3 :}\quad H(h) \le H(m) \quad \Leftrightarrow \quad h \ge_S m \end{aligned} \tag{1}$$

The order relation =*s* is a comparison operator called "*sharpness*". A fuzzy unit *h* is regarded as plus "acute" (*sharp*) that the fuzzy unit *m* if:

$$\begin{aligned} \forall x \in E \qquad &\text{if } m(x) \le 0.5 \text{ then } h(x) \le m(x) \\ &\text{if } m(x) \ge 0.5 \text{ then } h(x) \ge m(x) \end{aligned} \tag{2}$$

Functions which obey these axioms can be expressed by the general formula:

$$H(m) = h\left(\sum_{i}^{C} w_i \cdot T(m(x_i))\right) \tag{3}$$

C corresponds to the individual numbers in the dialogue universe (E) where is defined the fuzzy unit μ. According to [2][9]:

$$\begin{aligned} &\text{(i)}\quad w_i \in \Re^+ \\ &\text{(ii)}\quad T(0) = T(1) = 0 \\ &\text{(iii)}\quad T(m(x_i)) : [0,1] \longrightarrow \Re^+ \quad \text{it has a maximum value in} \end{aligned} \tag{4}$$

$m(x_i) = \frac{1}{2}$, and is monotonous for $m(x_i) < \frac{1}{2}$ and for $m(x_i) > \frac{1}{2}$

(iv) the function $h : \Re^{+} \longrightarrow \Re^{+}$ is monotonous increasing

De Luca and Termini [2] proposed to use as function $T(.)$ the Shannon probabilistic entropy [4] applied to the pair formed by the element and its complement to the fuzzy unit (equation 5):

$$T(m(x_i)) = m(x_i) \cdot \ln m(x_i) + (1 - m(x_i)) \cdot \ln(1 - m(x_i)) \tag{5}$$

Then the Luca and Termini entropy expression is [2]:

$$H_{DLT}(m) = K \cdot \sum_{i}^{C} S(m(x_i)) \tag{6}$$

where $K \in \Re^{+}$ is a normalization constant.

Many studies showed the validity of this expression like fuzzy information measures [1]. However, other families of functions with the same base form that the Luca and Termini can be used especially in the field of the decision-making and classification [11][12]. They are always based on the axioms presented in (1) and in the equation form of (3), where $w_i = K$ and the function T is given by:

$$T(m(x_i)) = f(m(x_i)) + f(1 - m(x_i)) \tag{7}$$

3.2 Validation index for states transition

Individual adequacy degree expresses his adequacy to a class. On the other hand, in the case of a decision-making between several groups or classes, the adequacy degrees express the adequacy of an individual to several classes. In the case of systems dynamic monitoring, the individual is represented by the whole of the variables which define the current state of the system and the classes are the possible states. The state with the highest adequacy degree is considered as the system that evolves at this time there. The reliability of the choice of the state at the moment present is directly proportional to the capacity of election among the adequacy degrees. On the contrary of fuzzy degree indices (e.g. fuzzy entropy), the problem in this case is not any more the analysis of adequacy of several individuals to a class, but the choice between several classes (states) to which an individual can belong. For this, this paper define a new fuzzy unit where the dialogue universe E is defined by the number of classes C and adequacy degrees correspond to the values of adequacy of each individual to each class. Plus one unit is ordered, plus informative is and then the entropy is lower. The unit that the paper consider most informative in the case of a choice is that which assigns the individual in a class with the maximum of adequacy while the adequacy degree of the other classes is null. On the other hand, the unit decision entropy becomes maximum if all adequacy degrees are equal. Consequently, the information provided by the unit would be then null, with respect to the reliability of the decision. In this context, as validation index this paper use the complement of the entropy of the decision suggested in [3] which is based on the Luca and Termini fuzzy entropy.

Considering that the choice corresponds to the maximum of adequacy, thus μ_M = *max* [$\mu(x_i)$]. Then, *the fuzzy decision indexes* are defined like the difference between this maximum value and each unit adequacy degree (case of state validations that corresponds to the individual adequacy degrees to each class):

$$d_i = m_M - m(x_i) \quad \forall i \neq M \tag{8}$$

The decision entropy for the fuzzy unit μ is the dual of total information (eq. (9)):

$$H_D(m) = 1 - I_D(m) \tag{9}$$

The useful information provided by the fuzzy unit whole for the decision-making $I_D(\mu)$ corresponds to:

$$I_D(m) = K \cdot \sum_i d_i \cdot e^{(d_i)} \quad \text{where} \quad K = \frac{1}{C \cdot m_M \cdot e^{m_M}} \tag{10}$$

This entropy is based on axioms which follow the philosophy of those proposed by Luca and Termini using the fuzzy entropy like fuzzy degree index. These axioms are:

$$\begin{aligned} &\text{R1:} \quad H(m) = 0 \Leftrightarrow \forall i \neq M \ ; \quad m(x_i) = 0 \\ &\text{R2:} \quad \max H(m) \Leftrightarrow \quad \forall i, j \quad ; \quad m(x_i) = m(x_j) \\ &\text{R3:} \quad H(\eta) \leq H(\mu) \Leftrightarrow \eta \geq_F \mu \end{aligned} \tag{11}$$

The relation $=_F$ is proposed like a comparator of reliability to unit. This operator is defined in a way similar to the operator "*sharpeness* " ($=_S$) proposed by Luca and Termini. A decision based on a nonprobabilistic fuzzy unit *h* is considered more reliable than another based on the fuzzy unit *m* if the reliability index of *h* is larger than that of *m*[15].

$$h \geq_F m \quad \Leftrightarrow \quad FIA(h) \geq FIA(m) \tag{12}$$

Operator $FIA(\mu)$ is defined like:

$$FIA(m) = m_M + card[d(m)] \tag{13}$$

3.3 Transitions validations algorithm diagram

When a transition between the moment *t-1* and the moment *t* is proposed by the recognition algorithm, this paper propose to analyze the information index (equation 10) of the new adequacy vector individual *x(t)*, if this one exceeds a certain uncertainty level the transition is validated. In the contrary case, the transition is put on standby, as long as the algorithm of recognition continues to propose the same class, it will be validated if at one posterior moment $t + r$ the quantity of information is considered reliable. The validation method diagram suggested is presented in Figure 2.

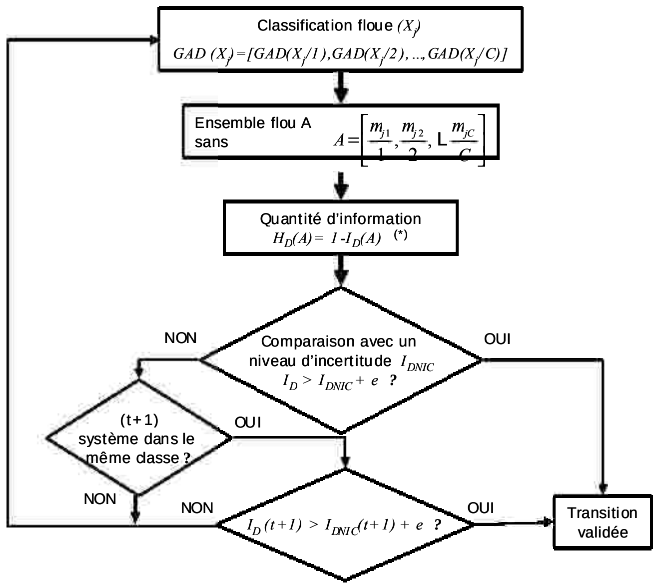

Fig. 2. Transitions validation algorithm diagram proposed

The information measure is the information index *ID(μ)* given by equation 10. It can arrive, either that a transition is not never validated or remains been unaware of system, or that it is validated with a delay *R* compared to the moment when the algorithm of recognition had detected it. The introduction of this delay makes it possible to eliminate the effects of noises or disturbances which provide an appearance of transition. In certain cases, there are oscillations between two classes which are only caused by the inaccuracies in measurements and these oscillations will then be eliminated in an automatic way, as well as false alarm which theirs is generally associated.

3.4 Uncertainty level

It is necessary to define a value minimum of information, which makes it possible to regard the decision as sufficiently reliable validating the change of state. This value can be established by analyzing the quantity of training information, in such manner according to the criterion of the system expert, the transitions considered as non valid will not be held in account during the stage from recognition. In this case there the uncertainty level is constant for all the time of supervision. However, it is more interesting automatically to obtain a value which corresponds to instantaneous minimal information to validate a transition at every moment from sampling. To obtain this value, on the LAMDA fuzzy classification method (Learning Method for Multivariate Data Analysis) [1][13], the information brought by attribution to one of the informative classes, (different from the class from Not Information, NIC), is equal to the uncertainty of this class NIC, since the union of the informative classes is exactly the complement of class NIC. The adequacy value of any individual to class

NIC is a constant provided by the algorithm which plays the minimum role of adequacy value, it is thus natural to use as uncertainty level (IDNIC) the formula of total information by including the adequacy degree of the class NIC, given by the equation 14.

$$I_{DNIC} = K \cdot (\sum_i (d_i) \cdot e^{d_i} + (d_{NIC}) \cdot e^{d_{NIC}}) \quad \forall i \neq M \,, \quad K = \frac{1}{C \cdot m_M \cdot e^{m_M}} \tag{14}$$

By analyzing the equation 14, if the decision change of class was made with an adequacy degree significantly higher than the minimal adequacy degree (μNIC), the relationship between the quantities of information corresponds to equation 15:

$$I_{DNIC} > I_D \Leftrightarrow m_M \geq m_{NIC} \tag{15}$$

To make the decision of validation this paper added a margin of decision *e*. This value makes the possibility to guarantee the information for the change of state ID and the minimal uncertainty level in such manner that the transition is valid if:

$$I_{DNIC} > I_D + e \tag{16}$$

If the decision margin is big, the transition will be validated and better conditioned.

With this approach, this paper validate only the transitions which have a sufficient information degree compared to total information, by including the class of minimal adequacy, which in the case of the method of classification LAMDA corresponds to the NIC.

4 Application to method to SMAPA potable water plant

The treatment plant concerned is the drinking water drinking water station "SMAPA" [10] of Tuxtla city in Mexico. For this station is significant to include an automatic monitoring system which makes it possible to have an alarm to avoid the states of bad operation. The objective is to establish a preventive maintenance according to the current state of the system and not on fixed dates like was made at the moment. For the monitoring system an approach of fuzzy classification was chosen, in such manner that according to the historical data a model station which allows, in the stage of recognition and in line can be established, to identify the states and to obtain an alarm to avoid the situations of bad operation.

LAMDA method was selected for data classification. The training time is very short and the parameters election method is easy to make for an expert in the procedure which does not have necessarily a strong knowledge on the classification methods [13]. To have a better performance and to retire the false alarms and conditioned bad classes, the transition validation method proposed in this document, was applied.

Training data base consists of 105 samples (from November 2000 to mid-February 2001). For each sample, the variable values which constitute a whole of 4 descriptors of the raw water quality are: turbidity input, a filter retrowashing number, plus the coagulant dose added. The fourth descriptor used is the difference between the value of turbidity input and the measured value after the decanters of the station. By using the LAMDA classification method, the reference model is obtained made up of the functional states and alarms. After identifying the various classes, the system expert associates the classes to functional states and class was associates like alarm. The classification results (without the transition validations) which correspond at training stage are presented in Figure 3. The training is of the type not supervised, i.e. that the number of classes is not established a priori.

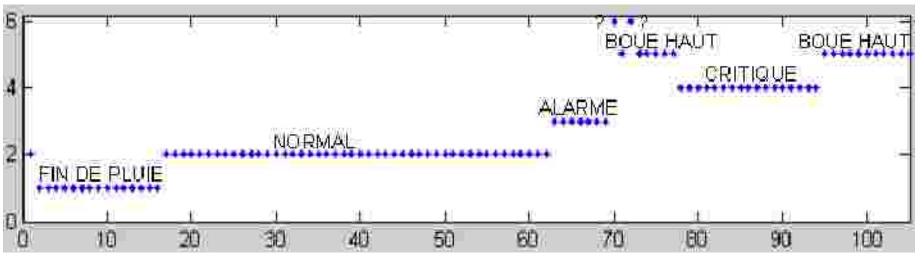

Fig. 3. LAMDA results: training phase without states validation (2000-2001)

Six classes were obtained. There are 2 classes which correspond to the normal operation, one of alarm which indicates the need for maintenance of the filters and two classes of faulty operation (high sludge, and critical stage). Class 6 is not a priori associated with a state with the system. This alarm which indicates a need for maintenance of the decanters intervenes very before the date established for maintenance. Indeed, for this whole of data the date of maintenance of the decanters (according to alarm) is proposed 87 days before the date plans. It's very significant to include this alarm to avoid high or critical sludge states (classes 4 and 5). During these states of bad operation, to give a quality of adequate water, it is necessary to increase the dose of coagulant and filter retrowashing number. If alarm is included, it is possible to carry out a preventive maintenance which will allow avoiding these nondesirable situations.

To validate the classes and to retire the badly conditioned transitions, the transition validation method between states was applied. At every moment, the minimum level of information to validate a transition was calculated by holding of account the adequacy degree of each data with the class of noninformation NIC [13]. To utilize this method, it is appropriate to choose a value of requirement (e). With the whole of the training data, the value of requirement $e = 0.0018$ makes it possible to invalidate class 6. As it will be shown later, this value is representative of the system and not a priori of the training base. All the other transitions are validated perfectly. The result of the validation of transitions for this unit from data corresponds on Figure 4.

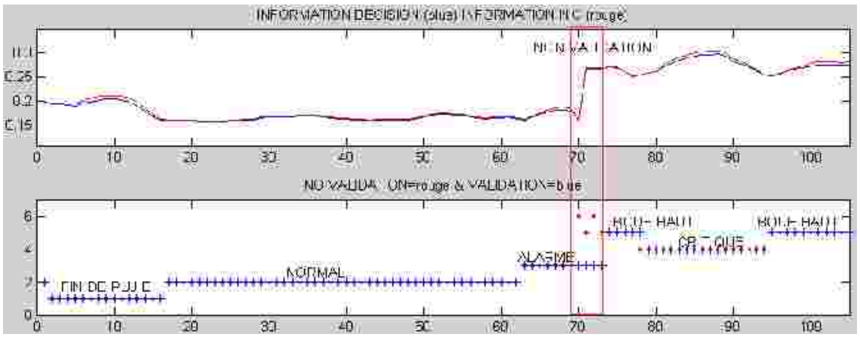

Fig. 4. States validation results: training data (2000-2001)

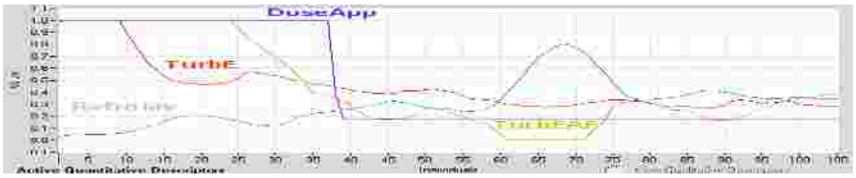

Fig. 5.a Analyzed variables, data test (2003-2004)

The variables which was analyzed (period 2003-2004) are presented in the figure 5.a. Classes identification results with the method LAMDA are watch in figure 5.b. and 5.c. shows the results applying this transition validation technique proposed. Class 6 was invalidated in all the cases, then it is regarded as state badly conditioned, alarm for the maintenance of the station was also identifies with new the data.

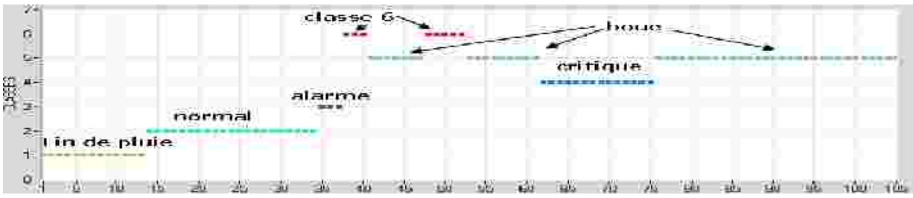

Fig. 5.b. Results from LAMDA: phase of test without validation of states (2003-2004)

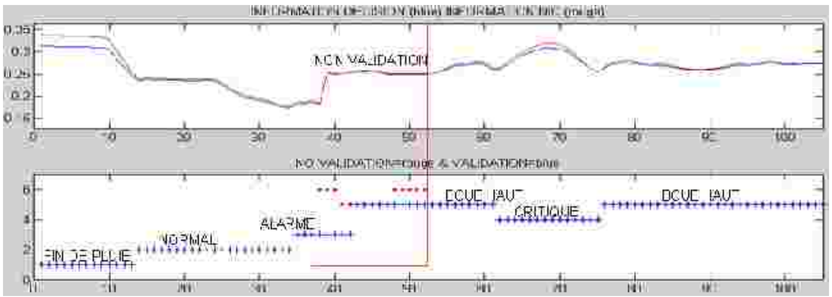

Fig. 5.c States validation results: data test 2003-2004

The method of validation suggested was also applied to a unit of data by replacing the dose of coagulant applied by the dose of coagulant calculated with neural networks [14]. In this case, there are false alarms which are removed with the transition validation method (Figure 6). The identification and validation of the system states for all the periods were made, without changing the value of the parameter *e* of validation method.

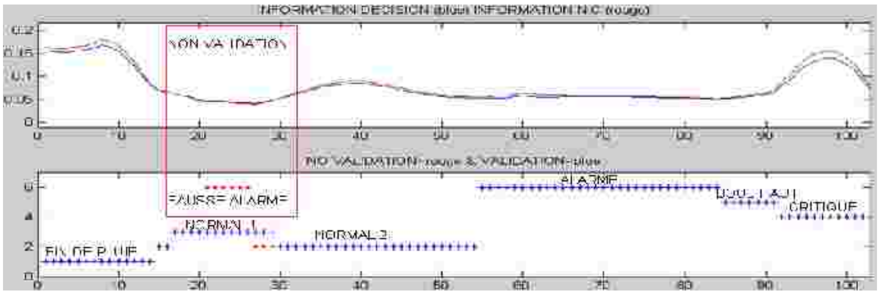

Figure 6. States validation results: data test 2001-2002 (Dose calculated)

In this case, there are three classes associated with normal operation (end of rains, normal 1 and normal 2). Class 6 corresponds to alarm but there is false alarm whereas one is in the normal state, this false alarm was eliminated by the method suggested.

5 Conclusions

The new methodology for transition validation based in fuzzy entropy measure is introduced. This strategy addresses the needs for diagnosis and prognosis in order to provide adequate preventive and predictive maintenance to potable water plants, which can lead to problems of dependability as well as significant economic losses. This approach provides a criterion for decision making when associating a class to an individual in presence of uncertainty or bad conditioned individuals. As a result, false alarms are eliminated.

Moreover, the effect of disturbances has been minimized when eventually they lead to non reliable transitions. In consequence, the system monitoring becomes more robust since apparent transitions due to inaccuracy measures are not validated. One of the advantages of the method is also that the transition may be validated since the method uses the output of the fuzzy classifier, which is not a big amount of data, and the computing time is as well reduced. On the other hand, some further studies are looking forward to introducing historical individuals memberships so that transitions could be validated based on finite time sliding window observations analysis.

References

1. T. KEMPOWSKY, A. SUBIAS, J. AGUILAR-MARTIN. Process situation assessment: From a fuzzy partition to a finite state machine, Engineering Applications of Artificial Intelligence, In Press, Corrected Proof, Available online 11 April 2006
2. D. LUCA and S. TERMINI. A Definition of Non Probabilistic Entropy in the Setting of Fuzzy Set Theory. Information and Control, 20:301-312,1972
3. B. KOSKO. Fuzzy entropy and conditioning, Info Sciences, vol. 40, pp.165-174, 1986.
4. C. SHANNON. A mathematical theory of communication. Bell SystTech,pp379-423, 1948.
5. H. HERNANDEZ DE LEON. Supervision et diagnostic des procédés de production de l'eau potable. Thèse de Doctorat à l'INSA de Toulouse 2006. 04160 (2006).
6. E. TRILLAS, C. ALSINA. Sur les mesures du dégrée du flou. Stoch, vol.III pp. 81-84, 1979
7. E. TRILLAS, T. RIERA. Entropies in finite fuzzy sets. Info Sciences 15,pp.159-168, 1978
8. E. TRILLAS, C. SANCHIS. On entropies of fuzzy sets deduced from metrics. Estadistica Española 82-83, pp. 17-25, 1979.
9. N. R. PAL and J. C. BEZDEK. Several new classes of measures of fuzziness. Proc. IEEE Int. Conf. on Fuzzy Syst., 928-933, Mar. 1993.
10. SMAPA Sistema Municipal de Agua Potable y Alcantarillado de Tuxtla. Manual de Procedimientos, Tuxtla Gtz., Chiapas, México, 2007.
11. S. AL-SHARHAN et al. Fuzzy Entropy: a Brief Survey, IEEE Int Fuzzy Syst Conf 2001.

A. DELUCA, S. TERMINI. Entropy of L-fuzzy sets. Info and control, 24, pp.55-73 1974.

B. ISAZA, T. KEMPOWSKY, J. AGUILAR-MARTIN, M.V. LELANN, A GAUTHIE; « Etude comparative de la méthodologie LAMDA pour le diagnostic » ed. Cepadues , 2006.

12. H. HERNANDEZ DE LEON, M.V. LE LANN. Development of a neural sensor for the on-line prediction of coagulant dosage in a potable water treatment plant. IBERAMIA 2006.
13. E. DIEZ, J. AGUILAR-MARTIN; Proposition of Non-probabilistic Entropy as Reliability Index for Decision Making, CCIA'2006.