DECISION METHOD FOR STATES VALIDATION IN DRINKING WATER PLANT MONITORING

C. Isaza[1+], E. Diez-Lledo[1], H. Hernandez de Leon[1], J. Aguilar-Martin[1], M.V. Le Lann[1,2]

[1] *Université de Toulouse, LAAS-CNRS, 7, Avenue du Colonel Roche, 31077 Toulouse, France*
[2] *INSA, 135, Avenue de Rangueil, 31077 Toulouse, France*
{ cisaza, ediezll, hhernand, aguilar ,mvlelann }@laas.fr

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. The classification algorithm provides the membership degrees of the instant system data (individual) to all possible identified functional states (classes) of the water plant. Usually, the higher membership degree determines the recognized class. However, this criterion leads to uncertainty levels when decision is based on a bad conditioned fuzzy set due to similar membership degrees. The proposed method validates the recognized functional state in presence of uncertainty.

Keywords: Fuzzy Entropy Measures, Decision Index, Transition Validation, Water Plant Monitoring.

1. INTRODUCTION

Water industry is facing the problem of producing quality water at lower costs. In drinking water to ensure the water supply to the population, in terms of quality and quantity, while protecting the environment.

The drinking water plant object of this study is the "SMAPA" (SMAPA,2005) of Tuxtla city in Mexico. For this station we found worthy to include an automatic monitoring system in order to program a proper maintenance due to the irregular performance observed in the plant with the actual one. The objective is to establish a preventive and flexible maintenance according to the current system functional state instead of nowadays maintenance based on a fixed program.

Finding a general mathematical model for biotechnological process is somehow complex. Moreover the model should be adequate for monitoring, which is otherwise evident. On the other hand, the existence of large quantity of historical data of that plant offers the possibility to apply fuzzy classification methods for the monitoring system. Using this approach, the several functional states (classes) of the SMAPA plant are identified. Nevertheless; to eliminate false alarms and spurious classes, the transition validation method proposed in this document is applied in order to improve the monitoring performance.

The dynamic process monitoring based on data classification consists in the identification of the current class associated to each functional system state. When the system is monitored on-line the aim is to recognize the class for the new individual (set of data) at the present time. The fuzzy classification methods provide the adequacy degree of the individual to each class under the form of a fuzzy set.

[+] Professor at the Universidad de Antioquia – Medellin, Colombia

Usually, the logic applied when selecting a class among all possible classes is to take the one which the individual shows the maximum membership degree. Making a decision among all possible classes is particularly critical at transitions times between classes.

In biological process monitoring uncertainties may be caused by the inaccuracy in measures, real process disturbances or the inherent algorithm variances. In such situations, the fuzzy set representing the individual adequacy to each class could be badly conditioned in order to make a decision accurately. Therefore, we found necessary to introduce a criterion that may help to improve the state transition validation reliability, as a contribution to health monitoring.

Previous studies were carried out, focused in identifying the process states and their correspondent transitions (Kempowsky et al., 2006). However, those approaches have not taken in account the problem of uncertainty in transitions between classes.

The purpose of that study is so to apply a methodology for decision validation in presence of uncertainty. To reach that aim, this approach is focused in the entropy theory since the objective is to make a reliable decision based on the present information provided by the classifier.

The entropy theory was introduced by Shannon (1948) as a measure of information in the probabilistic domain. DeLuca and Termini (De Luca and Termini,1972) proposed an extension of Shannon (Shannon,1948) theory to the non probabilistic domain of the fuzzy sets. This approach is largely used in literature as a proved information index.

The fuzzy entropy is a measure of information of a fuzzy set compared to a crisp set, taking in account that the crisp set is the most informative one. Though, the decision theory problem studied in this paper is not represented by the comparison to a crisp set but to a singleton, since the singleton is the most informative set in decision making. As a result, the validation method is based on fuzzy entropy (Diez-Lledo and Aguilar-Martin, 2006) which is defined in the context of the classical fuzzy entropy. This approach provides an instantaneous information measure which allows making a reliable decision in presence of uncertainty when validating a state transition.

In section 2 a brief plant description and the initial monitoring system are presented. The proposed transition validation method is explained in section 3. The results of the applied method to the SMAPA plant are showed in section 4. Finally the conclusions and perspectives are presented.

2. DRINKING WATER PLANT MONITORING

2.1 Brief Drinking water Plant Description

The “SMAPA” water treatment plant (Tuxtla city, Mexico), provides water to more than 800,000 habitants and has a nominal capacity to process 800 l/s of water per day. The complete usual chain comprises the 5 great following units: pre-treatment, pre-oxidation, clarification, disinfection, and refining. Raw water is collected at the river “Grijalva” and pumped to the treatment plant. Water treatment plant includes two main process units, clarification and filtration. The behavior plant depends of the turbidity parameters and the chemical reagents consummation. There are two principal types de maintenance:

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 rain 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.

Filters maintenance: 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.

2.2 Monitoring Using Classification Methods General Theory

The use of the classification algorithms into the systems monitoring allows obtaining interpretable results and offers useful information for decision making in state system dynamics.

At first step (learning), the objective is to find, through the analysis of historical data, the characteristics of the behaviour which will permit to identify the system states (Fig.1). At a second step, the data recognition allows us to identify on line the current system state.

The data pre-treatment is necessary and includes traditional operations of Control Theory and Signal Processing. In all cases, a set of variables that describe the system ('descriptors', often called 'attributes') summarizes the accessible information provided to monitoring. The current set of descriptors is called individual. Monitoring algorithms that include fuzzy classification methods provide the membership degree of the individual to each system state (class).

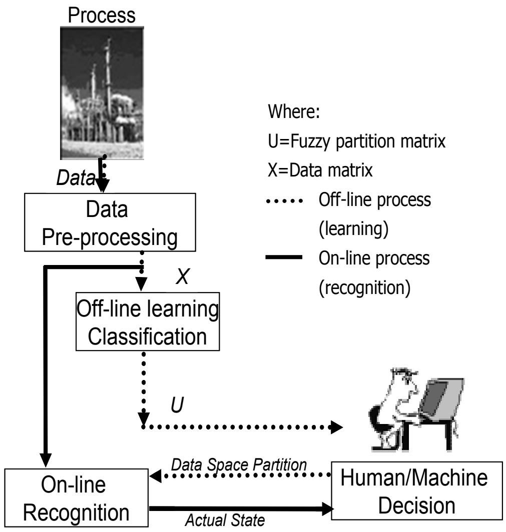

Fig. 1. General monitoring scheme

2.3 Drinking Water Plant Monitoring

The LAMDA (Learning Algorithm for Multivariate Data Analysis) (Aguilar and Lopez,1982; Isaza et al.,2006) method was selected for data fuzzy classification. This method has a comparable performance with other very recognized classification methods (e.g. neural networks, gk-means). Moreover, LAMDA offers the advantage of making a not supervised classification without giving the a priori number of classes. The algorithm shows a short training time and the parameters selection is intuitive relatively simple (Isaza et al.,2006). These advantages give the possibility to obtain a model for monitoring without the need of a precise knowledge on the system dynamics. Indeed, this is interesting for the SMAPA monitoring station since there is no a priori knowledge of the system states.

Training data base is building up by 105 samples (from November 2000 to mid-February 2001). Four variables of the plant are chosen to identify the state: turbidity at the decanters input (TurbE), filter backwashing frequency (Retrolavag) and the coagulant dose added (Dose). The fourth descriptor is calculated as the difference between the descriptor TurbE and the output turbidity value (TurbEAF).

By using the LAMDA classification method with not supervised training (the number of classes is not established a priori), the functional states and alarms of the system are identified. After that, the plant expert associates the classes to functional states and alarms (Hernandez,2006). The classification results are presented in section 4.

The alarm class indicates the need for filter maintenance so as to prevent faulty operation states (high sludge, and critical mode). This alarm suggests preventive decanters maintenance long before the actual programmed maintenance.
Previously, during the faulty states described before (high sludge and critical mode), the filter backwashing frequency and the coagulant dose should be increased in order to guarantee the quality of drinking water provided. Therefore, this implies additional costs and filter damaging. Since the alarm is taken in account the preventive maintenance avoids these critical situations.

However, the classification algorithm finds out a spurious class. Moreover, there are false alarms occurring during the normal state. For these reasons is necessary to include a transition validation method as we propose in the next section.

3. TRANSITIONS VALIDATION METHOD

3.1 Algorithmic structure

This paper proposes to analyse the quantity information of the membership degrees vector obtained as result of the fuzzy classification algorithm. When the monitoring system recognizes a functional state, the validation approach estimates the information decision degree (Eq.4). If this value of information is smaller than the uncertainty level (Eq.5) the transition is not validated and the decision is postponed.

The validation of a transition applies whenever at some instant $t+R$ a higher information value that the uncertainty level at that time is reached. The general flow chart of the transition state validation method proposed is presented in Fig. 2

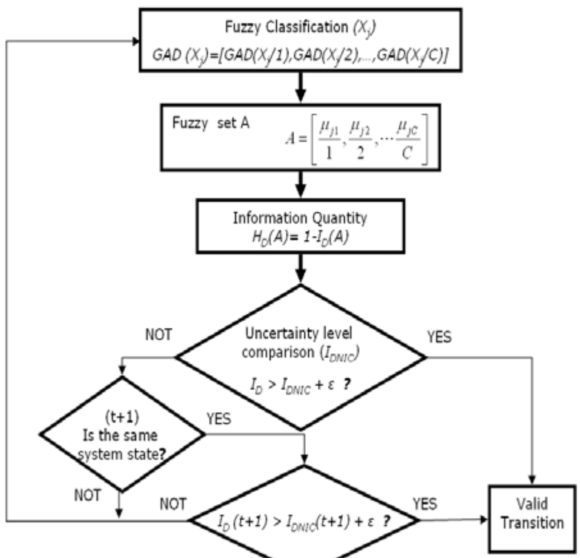

Fig. 2. Validation state scheme

Delay (R) permits to eliminate the noises or disturbances effects present during the transitions between functional states. In certain cases, due to the inaccuracy in measures, the monitoring system may oscillate between two classes. This generates false alarms; hence these oscillations must be eliminated in an automatic way.

3.2 Fuzzy Decision Validation Index

Non probabilistic entropy functions indicate the fuzziness degree of a given discrete finite set that can be defined by the vector $\mu = [\mu_1, ..., \mu_i, \mu_c]^T$ where $0 \le \mu_i \le 1$.

DeLuca & Termini proposed the following set of axioms for non probabilistic entropies so that a function $H(\mu)$ can be considered as an entropy measure if it satisfies the following axioms. (DeLuca and Termini,1972;Trillas and Alsina,1979;Kosko ,1986; Pal and Bezdek,1993)

$$
\begin{aligned}
&\text{P1}: \; H(\mu) = 0 \Leftrightarrow \mu(x_i) \in \{0,1\} \\
&\text{P2}: \; \max H(\mu) \;\Leftrightarrow\; \forall i \;\; \mu(x_i) = 1/2 \\
&\text{P3}: \; H(\eta) \le H(\mu) \;\Leftrightarrow\; \eta \ge_S \mu
\end{aligned}
\tag{1}
$$

The order relation $\ge_S$ is called "*sharpness".* A fuzzy set η is considered sharper than the μ fuzzy set if:

$\mu(x) \le 0.5 \;\Rightarrow\; \eta(x) \le \mu(x)$

and $\mu(x) \ge 0.5 \;\Rightarrow\; \eta(x) \ge \mu(x)$

They proposed as fuzzy entropy the function (DeLuca and Termini,1972):

$$H(\mu) = K \cdot \sum_{i}^{C} S(\mu_i) \tag{2}$$

Where

$S(\mu_i) = \mu_i \cdot \ln \mu_i + (1 - \mu_i) \cdot \ln(1 - \mu_i)$.

This entropy has proven to be useful in many situations to quantify the uncertainty of fuzzy sets (fuzziness measure).

To make a decision based on the membership functions that define a fuzzy set, the degree of fuzziness must be considered. As a result the fuzzy entropy shall be used to propose a decision validation index, as a measure of specificity (Garmendia et al.,2003). According to this remark, the information to make a decision is the complement of fuzzy entropy.

Generally the choice that leads to the decision corresponds to the maximum membership $\mu_M = \max[\mu(x_i)]$. Then the difference between the membership of the element that is chosen by the decision process and the membership of the other elements gives rise to a fuzzy set $\delta(\mu) = \{\mu_M - \mu_i\}_{i \ne M}$ representing the adequacy of each element to the actual decision associated to $\mu_M = \mathbf{max}[\mu_i]$.

Some modifications to the De Luca and Termini axiomatic definition have been proposed (Diez-Lledo, Aguilar-Martin, 2006). Essentially they consist in replacing the "sharpness" relation by $\eta \ge_{\mathbf{F}} \mu \Leftrightarrow \mathbf{REL}(\eta) \ge \mathbf{REL}(\mu)$, and to take as reference the singleton instead of the corresponding crisp set. The REL relation, or reliability of a fuzzy set, $\mu = [\mu_1, ..., \mu_i, \mu_c]^T$, is given by $\mathbf{REL}(\mu) = \mu_M + card[\delta(\mu)]$. As a result this fuzzy entropy is defined by:

$$\mathbf{H}(\mu) = 1 - \frac{\sum_{\mathbf{i}} \delta_{\mathbf{i}} \cdot \mathbf{e}^{(\delta_{\mathbf{i}})}}{\mathbf{C} \cdot \mu_{\mathbf{M}} \cdot \mathbf{e}^{\mu_{\mathbf{M}}}} \tag{3}$$

Finally the information index proposed is the following. where *C* is the number of classes:

$$I_D(\mu) = \frac{\sum_i \delta_i \cdot e^{(\delta_i)}}{C \cdot \mu_M \cdot e^{\mu_M}} \tag{4}$$

3.3 Uncertainty level

The memberships of an individual to all the classes of the partition are a finite discrete fuzzy set $\Gamma = \{\mu_i\}_{i=1}^{C}$. Whenever a threshold ζ of membership has been introduced, it can be viewed as the membership to a non informative class (NIC) such that if $\forall i; \;\; \mu_i < \mu_{NIC} = \zeta$, the corresponding element is assigned to that NIC class, meaning that the information is not sufficient to assign it to any class.

By choosing as fuzzy classification method LAMDA (Aguilar-Martin and Lopez, 2001; Isaza et al. 2006), the value of the threshold that defines the NIC class is automatic. Therefore, the information, or uncertainty level, related to the NIC class is

$$I_{DNIC} = \frac{\left(\sum_{i \ne M} (\delta_i) \cdot e^{\delta_i} + (\delta_{NIC}) \cdot e^{\delta_{NIC}}\right)}{C \cdot \mu_M \cdot e^{\mu_M}} \tag{5}$$

where C includes the NIC class.

As stated in the flow chart of figure 1, in a sequential classification, whenever a change of class occurs, it will only be considered meaningful if the information of the corresponding fuzzy set is greater than the minimum, i.e. $I_D > I_{DNIC}$ (for security reasons a margin ε is considered $I_{DNIC} > I_D + \varepsilon$)

4. RESULTS APPLYING THE TRANSITION VALIDATION MEHOD

Initially, the states identification of the SMAPA plant is obtained using the plant descriptors mentioned in Section 2. The classification results without the transitions validation are presented in Fig. 3 corresponding to the training phase of the fuzzy classification algorithm. The training is not supervised, i.e. that the number of classes is not established a priori.

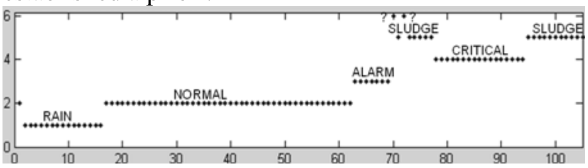

Fig.3. LAMDA results in training phase without states validation (2001-2002)

As a result, 6 classes were identified: 2 classes correspond to normal operation states (end of rain period and normal), one corresponds to the alarm and two classes were associated to non desirable states (high sludge and critical state). Class 6 was considered a spurious or badly conditioned class. The result of the on-line process monitoring using the system states recognition suggests the decanters maintenance 87 days before that the actual programmed maintenance.

To validate the classes and to retire the badly conditioned decisions, the transition validation method proposed was applied. At every moment, the minimum information level to validate a transition was calculated by means of the uncertainty degree, as described in Section 3. To apply this method, it is appropriate to choose a security value (ε). ε = -0.0018 for the training data set permits not to validate the class 6. As it will be shown later, this value is representative for the system regular operation and not only of the training data set. All the other transitions are correctly validated. The transitions validation results obtained for the training data set are shows on Figure 4.

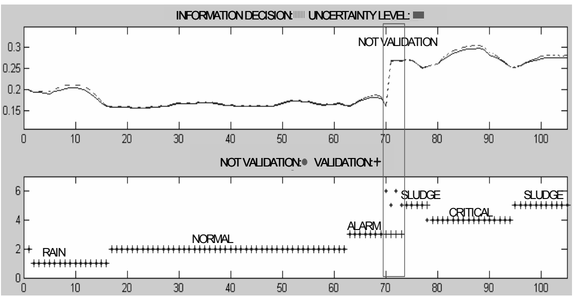

Fig.4. States validation results for training data (2001-2002)

In the recognition phase, two data sets corresponding to years 2001-2002 and 2003-2004 were treated. The data were analyzed in a sequential way in order to simulate on line operation. The Figure 5.a. and Figure 5.b. present the variables evolution and the classes for the low water period in 2001-2002. This figure represents the recognized states and the alarm. The Figure 5.c. shows the correspondent transition validations. Class 6 was automatically invalidated.

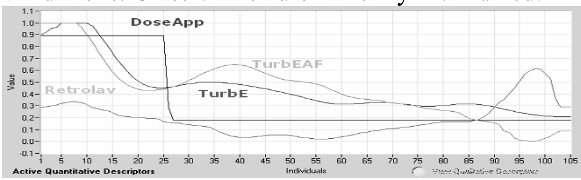

Fig.5. a. System variables (2001-2002)

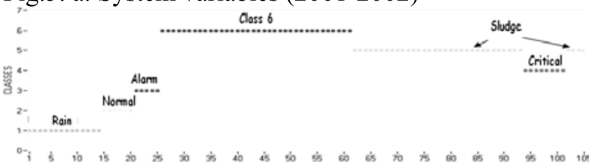

Fig.5.b. Recognized states (without the transition validation) (2001-2002)

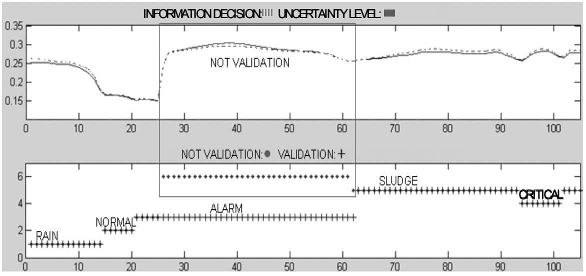

Fig. 5.c. States validation results: data test 2001-2002

The states recognition to period 2003- 2004 was carried out in the same way as period 2001-2002. The system variables evolution is showed in Figure 6.a. The Figure 6.b. shows fuzzy classification results. The proposed validation method results for this period are presented in Figure 6.c. The class 6 was automatically invalidated in all cases, so proving that class 6 is non informative. The maintenance alarm is also presented before to the actual programmed maintenance.

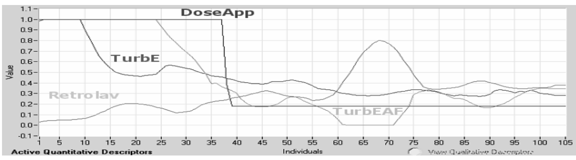

Fig.6.a. Analyzed variables (2003-2004)

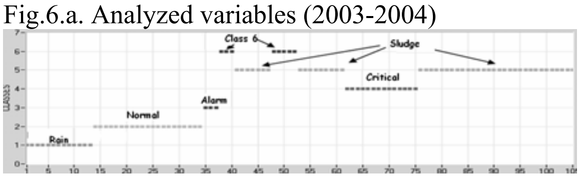

Fig.6.b. Results from LAMDA: phase of test without validation of states (2003-2004)

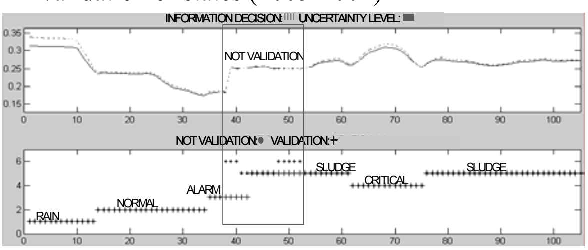

Fig.6.c. States validation results: data test (2003-2004)

The validation method has been also applied to a data set where the actually applied coagulant dose has been replaced by the calculated value using a neural network (Hernandez and LeLann,2006), in order to be placed in the same conditions as the on line operation. In this case, false alarms are also removed by the transition validation method (Figure 7). The systems states identification and validation is carried out without changing the ε parameter value.

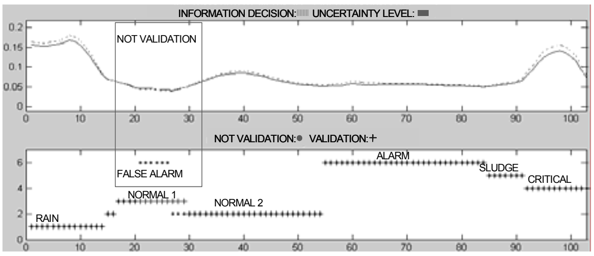

Fig.7. States validation results: data test 2001-2002 (Dose calculated)

In this experiment, three classes have been associated with normal operation (end of rain, normal 1 and normal 2). Class 6 is now an alarm. In the example a false alarm occurs if no (normal 1 state). This problem is eliminated as previously by the method here proposed.

5. CONCLUSIONS AND PERSPECTIVES

A new method for transition validation is introduced. 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 instantaneously 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.

ACKNOWLEDGMENTS

The authors wish to thanks Mr. Saavedra, Technical Manager of SMAPA who provided us help and experience to recover historical data.

REFERENCES

Aguilar-Martin J. and R. Lopez de Mantaras, (1982). The process of classification and learning the meaning of linguistic descriptors of concepts *Approximate Reasoning in Decision Analysis* p. 165-175. North Holland.

Al-Sharhan S., F. Karray, W. Gueaieb, O. Basir, (2001). Fuzzy Entropy: a Brief Survey, *IEEE International Fuzzy Systems Conference*.

DeLuca A. and S. Termini (1972). A Definition of Non Probabilistic Entropy in the Setting of Fuzzy Set Theory. *Information and Control*, Vol.20:301-312

Diez-Lledo E. and J. Aguilar-Martin; (2006). Proposition of Non-probabilistic Entropy as Reliability Index for Decision Making, *CCIA'2006*

Garmendia L., R. Yager, E. Trillas, A. Salvador, (2003). On T-Norms based specificity measures. *Fuzzy Sets and Systems* Vol.133-2, p.p. 237-248.

Hernandez de Leon H. and M.V. Le Lann (2006), Development of a neural sensor for the on-line prediction of coagulant dosage in a potable water treatment plant in the way of its diagnosis, *IBERAMIA* 23-27 October 2006

Hernandez de Leon H., (2006). Supervision et diagnostic des procédés de production de l'eau potable, *Thèse de Doctorat* à l'INSA de Toulouse.

Isaza C., T. Kempowsky, J. Aguilar-Martin, M.V. LeLann, A. Gauthier (2006). Etude comparative de la méthodologie LAMDA et autres techniques de classification basées sur la fouille de données dans le cadre du diagnostic. *Rapport LAAS No.06540.*.

Kempowsky T., A. Subias, J. Aguilar-Martin; (2006) Process situation assessment: From a fuzzy partition to a finite state machine, *Engineering Applications of Artificial Intelligence*,Vol.19, N°5, pp.461-477.

Kosko B., (1986). Fuzzy entropy and conditioning. *Information Sciences*,Vol. 40, pp.165-174,

Pal N. R. and J. C. Bezdek, (1993). Several new classes of measures of fuzziness, *Proc. IEEE Int. Conf. on Fuzzy Syst*., 928-933.

SAMPA (2005). Sistema Municipal de Agua Potable y Alcantarillado de Tuxtla*, Manual de Procedimientos*, Tuxtla Gtz., Chiapas, México,.

Shannon C., (1948). A mathematical theory of communication*. Bell Syst., Tech*., 27:379-423,.

Trilla E. S. and C. Alsina, (1979). Sur les mesures du dégrée du flou*. Stochastica*,Vol. III pp. 81-84,