

# Localisation de défauts de capteur pour la validation des mesures hydrométriques dans réseaux d'assainissement

F. HAMIoud<sup>1,2</sup>, C. JOANNIS<sup>1</sup> et J. RAGOT<sup>2</sup>

<sup>1</sup>LCPC, Laboratoire Central des Ponts et Chaussées, Route de Bouaye BP 2941. 44341 Bouguenais cedex. France.

<sup>2</sup>Institut National Polytechnique de Lorraine, Centre de Recherche en Automatique de Nancy, UMR 7039 CNRS-UHP-INPL, 2, Avenue de la forêt de Haye. 54516 Vandoeuvre cedex. France.

Collaborateur : Groupement pour l'évaluation des mesures et des composants en eaux et assainissement, 149, rue Gabriel PERI. 54500 Vandoeuvre Les Nancy. France.

hamioud@lpc.fr; Joannis@lpc.fr; Jose.Ragot@ensem.inpl-nancy.fr

**Résumé :** Dans ce travail, nous proposons une approche de diagnostic hors ligne permettant d'identifier les capteurs en défauts parmi les capteurs installés sur le réseau de mesure du réseau d'assainissement. Cette approche est basée sur un raisonnement flou avec prise en compte du concept de persistance de défaut.

**Abstract—** In this study we represent an approach of diagnosis in differed times adapted of this kind of the sewer network to isolate the sensors in faults among the set of sensors on the measurement network. This approach is based on fuzzy concepts with taking into account the fault persistence parameter in our reasoning.

## 1. Introduction

Les mesures dans les réseaux d'assainissement sont difficiles et perturbées. Les résultats faux peuvent considérablement perturber un bilan. Il est donc nécessaire d'éliminer des analyses, les valeurs erronées. La validation des mesures comme traitée dans [1]-[6], permet d'analyser la représentativité de ces mesures et de contrôler leur consistance. Une méthode efficace pour valider ces mesures est de multiplier l'équipement de mesure. Cette redondance appelée matérielle introduit plus de complexité dans le système et peut être très coûteuse.

L'objectif de ce travail est de développer une approche automatique qui permet le diagnostic des défauts quand ceux ci se présentent simultanément sur un ou plusieurs capteurs dans le réseau de mesure. Cet objectif est atteint, en développant des modèles pour l'ensemble des données pour générer des relations de redondance de l'information (redondance analytique) [7]; et en appliquant par la suite, une approche de diagnostic pour isoler chaque défaut qui peut se présenter. Dans ce travail, la méthode de diagnostic à base de modèles est considérée. La première étape de cette méthode est la génération des résidus [8], [9]. Ils sont généralement utilisés dans différentes méthodes de diagnostic, spécialement dans celles basées sur les concepts du raisonnement flou [10]-[14]. Une approche de diagnostic basée sur le raisonnement flou est présentée dans ce papier. Elle est développée pour résoudre un problème de diagnostic concret appliqué au réseau de mesure défini dans la section suivante. Dans les autres sections, l'ensemble des modèles linéaires établi pour générer les résidus et la matrice des signatures des défauts expérimentale seront présentés. L'approche de diagnostic sera détaillée; elle consiste à la «fuzzification» des résidus et la définition du paramètre de persistance introduit dans le raisonnement flou. Et avant de conclure, les résultats expérimentaux obtenus seront présentés et discutés.

## 2. Réseau de mesure de Nantes

Le système de mesure installé sur le réseau d'assainissement de Nantes est représenté dans la figure 1.

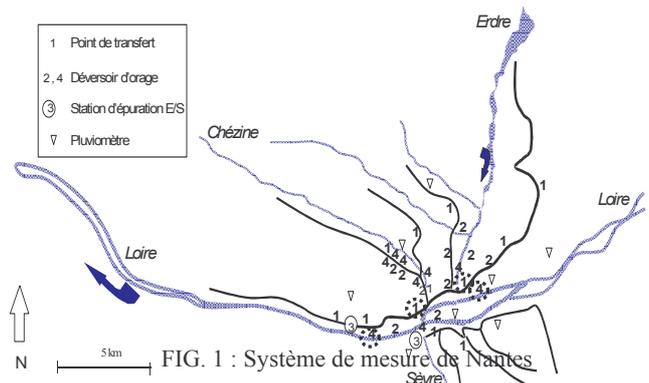


FIG. 1 : Système de mesure de Nantes. Il opère depuis 2001 et comporte 25 sites de mesure de débit et de 8 pluviomètres, pour un total de 53 capteurs utilisés. Chaque site de mesure est constitué d'un ou de plusieurs capteurs (mesure de hauteur et/ou mesure de vitesse d'eau dans le réseau).

Le présent travail est appliqué sur un ensemble de sites de mesure indiqués par des cercles en pointillés sur la figure 1. Ces sites comportent quatre variables de mesure de hauteur d'eau et une variable de mesure de vitesse d'eau. Ce travail peut être appliqué sur une partie du système ou sur tout le système. La formulation du problème traité reste similaire, seulement la taille des données à traiter change.

## 3. Modèles pour le réseau de mesure

Dans le cadre de cet article, un modèle représente une description mathématique des relations qui existent entre les différentes variables (mesures) sur le réseau de mesure. L'ensemble de modèles permet ainsi, de créer les relations de redondance analytique [7]. Celles ci permettent d'exprimer

une variable selon une ou plusieurs autres variables, qui sont en corrélation physique<sup>1</sup> et statistique avec cette variable.

La forme générale de ces modèles linéaires est donnée par :

$$Ax(k) = 0, \quad (1)$$

$A \in \mathfrak{R}^{l,n}$ ,  $x \in \mathfrak{R}^n$ , où  $A$  est la matrice des paramètres des modèles et  $x(k)$  est l'état général du système à l'instant  $k$ .

Dans le cas où toutes les variables sont mesurées, l'équation de mesure est donnée comme suit :

$$y(k) = x(k) + \varepsilon(k) + S\delta(k), \quad (2)$$

$y \in \mathfrak{R}^n$ ,  $\varepsilon \in \mathfrak{R}^n$ ,  $\delta \in \mathfrak{R}^q$ ,  $S \in \mathfrak{R}^{n,q}$ , où  $y(k)$  est le vecteur de mesure,  $\varepsilon(k)$  est le vecteur de bruit de mesures supposé aléatoire et suit une loi Gaussienne de moyenne nulle et de matrice de covariance  $V_\varepsilon \in \mathfrak{R}^{n,n}$ ,  $\delta(k)$  est le vecteur de défauts qui peuvent affecter quelques capteurs et  $S$  est la matrice qui exprime la direction des défauts.

La nature des variables d'entrées du modèle (les variables explicatives), est fixée par la nature de la variable à expliquer, Ainsi, on a des relations entre

- les hauteurs et les vitesses locales (même site de mesure);
- les hauteurs dans le site global (dans tout le réseau);
- les débits en aval et en amont traduits en hauteurs et vitesses.

Ces modèles sont établis à différents pas de temps, on a exploité ceux à pas de temps d'une minute.

Pour le réseau de mesure global, l'état  $x(k)$  est défini par :

$$x(k) = [h_1(k), \dots, h_{n_1}(k) \quad v_1(k), \dots, v_{n_2}(k)] \quad (3)$$

où  $n_1 + n_2 = n$ ,  $h_i(k)$  et  $v_j(k)$  sont respectivement, la hauteur et la vitesse d'eau dans le réseau,  $n_1$  et  $n_2$  sont les nombres de hauteurs et de vitesses d'eau respectivement.

## 4. Génération des résidus

Les résidus sont des signaux indicateurs de défauts [15]. La méthode utilisée pour générer ces signaux est basée sur l'exploitation des modèles établis. En éliminant les variables inconnues  $x(k)$  des deux équations (1) et (2), on obtient :

$$Ay(k) = A\varepsilon(k) + AS\delta(k). \quad (4)$$

À partir de cette équation, on peut obtenir les deux relations suivantes :

$$r(k) = Ay(k), \quad (5)$$

$$r(k) = A\varepsilon(k) + AS\delta(k), \quad (6)$$

où  $r(k) = [r_i(k)]_{i=1:l}$ .

La relation (5) permet de calculer le vecteur résidu à partir de mesures, et la relation (6) explique l'influence des erreurs de mesure et des défauts. En présence de(s) défaut(s) ( $\delta(k) \neq 0$ ),  $r(k)$  est non nul. Dans ce cas, on doit résoudre le problème inverse ; c. à. d. identifier à partir des valeurs des résidus les variables affectées par le(s) défaut(s). Le résidu doit être statistiquement nul pendant le fonctionnement normal du réseau et dévie de zéro dans le cas de présence de défaut. Ainsi, toute déviation affectant le vecteur résidu est un révélateur de présence d'un évènement anormal.

Les valeurs des résidus ne sont pas suffisantes pour isoler le(s) défaut(s), il est nécessaire aussi de connaître l'influence

de chaque variable sur les valeurs des résidus ; c. à. d. la connaissance de la relation entre les symptômes et leurs causes. La matrice des signatures des défauts théorique  $\Lambda$  est établie. Elle définit l'influence des défauts sur les valeurs des résidus. C'est une matrice ternaire (contient des valeurs de code ternaire) et est définie comme suit :

$$\begin{aligned} \Lambda_{ij} &= -1 \text{ if } A_{ij} < 0, \\ \Lambda_{ij} &= 0 \text{ if } A_{ij} = 0, \\ \Lambda_{ij} &= +1 \text{ if } A_{ij} > 0, \end{aligned} \quad (7)$$

où  $i = 1:l$ ,  $j = 1:n$  et  $\Lambda_j$  la matrice de signature de défaut affectant les mesures du  $j^{\text{ème}}$  élément de  $y(k)$ .

## 5. Isolation des défauts

### 5.1 Principe de la méthode.

Cette méthode est basée sur les concepts élémentaires reliés à la théorie des ensembles flous [10], [16], [17]. Un ensemble flou est défini par :

$$B = \{x, \mu_B(x) / x \in X\}, \quad (8)$$

où  $\mu_B(x)$  est la fonction d'appartenance de l'élément  $x$  à l'ensemble flou  $B$ , à partir de l'ensemble de définition  $X$  et prend des valeurs dans l'intervalle  $[0, 1]$  appelé  $\mu_B$ . On définit trois ensembles flous (modalités)  $r_i^-$ ,  $r_i^0$  and  $r_i^+$  correspondent à des résidus « négatif », « nul » et « positif » respectivement. Dans le cas général, on peut écrire  $r^g(k) = [r_i^g(k)]_{i=1:l}$ , avec  $g = 0, -$  ou  $+$ .

Les fonctions d'appartenance à ces modalités sont définies comme suit :

$$\mu_{r_i^0}(k) = 1 - \frac{(r_i(k)/\tau_i)^\rho}{1 + (r_i(k)/\tau_i)^\rho}, \quad (9)$$

$$\mu_{r_i^+}(k) = \begin{cases} 0 & \text{if } r_i(k) \leq 0 \\ 1 - \mu_{r_i^0}(k) & \text{if } r_i(k) > 0 \end{cases} \quad (10)$$

$$\mu_{r_i^-}(k) = \begin{cases} 1 - \mu_{r_i^0}(k) & \text{if } r_i(k) \leq 0 \\ 0 & \text{if } r_i(k) > 0 \end{cases}, \quad (11)$$

où  $\mu_{r_i^0}(k)$  représente la valeur de la fonction d'appartenance associée à l'ensemble flou  $r_i^0$  à l'instant  $k$ ,  $\rho$  est le facteur de forme qui doit être pair (il doit être pair pour obtenir des formes sigmoïdes) et  $\tau_i$  est la tolérance<sup>2</sup> associée au résidu  $r_i(k)$ . On définit à chaque instant  $k$ , le vecteur  $\mu_{r_i^g}(k) = [\mu_{r_i^g}(k)]_{i=1:l}$ , avec  $g_i = 0, +$  ou  $-$ .

La localisation des défauts est basée sur l'analyse conjointe des résidus et la matrice des signatures de défauts théorique. Elle permet de déduire à partir de la valeur des différents résidus, la variable affectée par les défauts. On a ainsi une base de règles [19] (règles d'inférence flou) entre symptômes (déviation des résidus) et causes (présence de défaut). Chaque règle comprend une partie prémisses donnée en

<sup>1</sup> Degré d'influence réel qui existe entre les variables (capteurs), il peut être plus ou moins significatif selon la dispersion des capteurs sur le réseau de mesure.

<sup>2</sup> Limite d'acceptance des valeurs du résidu (les valeurs acceptées sont celles supposées saines, non affectées par des erreurs). La tolérance est fixée en fonction de la variance du résidu

fonction des différents résidus et une partie liée à la décision. La règle est donc donnée par :

$$\text{si } r_1(k) \text{ is } r_1^{g_1} \text{ and } r_2(k) \text{ is } r_2^{g_2} \text{ and } \dots r_l(k) \text{ is } r_l^{g_l} \quad (12)$$

alors *decision*,

avec  $g_i = 0, +$  ou  $-$ ,  $i = 1 : l$ , (valeurs de  $g_i$  sont reliées aux valeurs de la signature du défaut  $\Lambda_j$ ), où

$$\text{decision} \equiv \begin{cases} \text{défaut(s) sur un ou plusieurs capteur(s)} \\ \text{pas de défaut sur le(s) capteur(s)} \end{cases}$$

L'inférence floue (proposition) est donc donnée par (12).

La quantification des différentes inférences floues est effectuée en fonction des valeurs de la fonction d'appartenance (appelé aussi degré d'appartenance). Le coefficient  $\mu_{r_i^{g_i}}(k)$  indique la quantification de la

proposition. Pour chaque instant et chaque règle, le degré de véracité noté  $\mu_f(k)$  d'une conséquence doit être évalué selon la validité des prémisses. Le « et » dans une proposition est un opérateur de conjonction ou d'agrégation, qui permet la combinaison des degrés d'appartenance des différentes propositions. Différents opérateurs de conjonction peuvent être utilisés [11]. Les opérateurs « minimum », « moyenne » et « Mandel », sont en cours d'utilisation [11], [15]. On a remarqué lors de l'utilisation de ces opérateurs que l'opérateur « Mandel » est équivalent à l'opérateur « moyenne » quand, pour un défaut spécifique on a  $g_j \in \{-, +\}$ . Par conséquent, un autre opérateur (13) est défini. Il augmente la robustesse de la méthode d'isolation appliquée sur les données du réseau de mesure

$$\mu_f(k) = \min \left( \frac{\sum_{i \in \{i / g_i \in \{0\}\}} \mu_{r_i^{g_i}}}{|\{i / g_i \in \{0\}\}|}, \min(\mu_{r_i^{g_i}} / g_i \in \{-, +\}) \right), \quad (13)$$

où  $|\{e_i\}|$  représente le cardinal de l'ensemble  $\{e_i\}$ ,  $f$  désigne le défaut affectant la variable  $y(k)$  ( $f_j^-(k)$ ,  $f_j^+(k)$  sont des défauts négatif et positif qui peuvent affectés chaque variable  $y_j(k)$ ) et  $g_i = 0, -$  ou  $+$  selon l'influence du défaut  $f$  sur  $r_i(k)$ . Les résultats d'isolation sont basés sur les degrés de véracité. Ainsi, le défaut le plus vraisemblable est celui décrit par le degré de véracité le plus élevé.

## 5.2 Persistance temporelle du défaut

La persistance temporelle des symptômes observés est un indicateur important qui permet de distinguer un vrai défaut d'une fausse alarme. Cela signifie qu'on ne doit pas considérer seulement l'amplitude de ces symptômes pour conclure sur la présence du défaut. En effet, quand seulement l'amplitude des défauts est considérée, un faible degré de véracité est systématiquement associé à ce défaut. Mais, la la considération de l'amplitude du défaut et de sa persistance, permet d'obtenir un degré de véracité élevé pour un défaut de faible amplitude, en condition que ce défaut soit persistant.

Le paramètre persistance intervient lors de l'évaluation des degrés d'appartenance des résidus aux différentes modalités. Pour chaque résidu,  $r_i(k)$  sont associés trois indicateurs de persistance de défaut  $p_i^-(k)$ ,  $p_i^0(k)$  et  $p_i^+(k)$ . Chaque

indicateur correspond au rapport nombre des valeurs du degré d'appartenance dépassant un seuil  $s_p$  sur  $L$  (largeur d'une fenêtre défini). Les degrés d'appartenance qui considèrent la persistance des défauts sont donnés par le calcul des moyennes entre les degrés d'appartenance (amplitudes des symptômes) et les indicateurs de persistance. Ces fonctions d'appartenance sont données comme suit :

$$\mu_{p_{r_i^-}}(k) = \frac{p_i^-(k) + \mu_{r_i^-}(k)}{2}, \quad (16)$$

$$\mu_{p_{r_i^0}}(k) = \frac{p_i^0(k) + \mu_{r_i^0}(k)}{2}, \quad (17)$$

$$\mu_{p_{r_i^+}}(k) = \frac{p_i^+(k) + \mu_{r_i^+}(k)}{2}. \quad (18)$$

Le poids donné pour chaque élément dans les relations (16), (17) et (18) est de 0.5. Il représente un choix arbitraire (spécifique à notre application). Dans ce cas, plus les symptômes du défaut observé sont persistants, plus les degrés d'appartenance sont élevés. Trois paramètres doivent être pris en considération pour une méthode d'isolation des défauts efficace : La tolérance  $\tau_i$  associé au résidu, la valeur de  $s_p$  et la largeur de la fenêtre  $L$  défini ci-dessus. Dans ce papier, on a considéré l'approche défini dans (5.1), en remplaçant les degrés d'appartenance  $\mu_{r_i^-}(k)$ ,  $\mu_{r_i^0}(k)$  et  $\mu_{r_i^+}(k)$  par les degrés d'appartenance  $\mu_{p_{r_i^-}}(k)$ ,  $\mu_{p_{r_i^0}}(k)$  et  $\mu_{p_{r_i^+}}(k)$  pour calculer les degrés de véracité correspondant à chaque défaut.

## 6. Résultats de simulation

Une application de l'approche d'isolation sur l'ensemble des variables de mesure  $y(k)$  est présentée dans cette section. Ces variables sont collectées des sites du réseau de mesure de Nantes montré dans la figure 1 (sites entourés par des cercles en pointillés). On défini de la relation (3), le vecteur  $x(k) = [h_1(k), h_4(k), h_{15}(k), h_{37}(k), v_5(k)]$ .

Dans une première étape, on établi l'ensemble des modèles linéaire pour ces variables, où leurs paramètres sont identifiés en appliquant la méthode des moindres carrés à l'ensemble  $y(k)$  dans le cas de fonctionnement normal du système (non présence de défauts). Cela est effectué en considérant la période du temps pendant laquelle les variables de mesure sont saines (pas de défauts sur les mesures). Les valeurs des résidus sont générées comme défini dans la relation (5) et sont données comme suit :

$$\begin{aligned} r_1(k) &= h_4(k) - 0.99h_1(k) + 0.014, \\ r_2(k) &= h_{37}(k) - 0.32h_1(k) - 0.68h_{15}(k) - 0.007, \\ r_3(k) &= v_5(k) - 0.44h_1(k) + 0.17h_4(k) - 0.28. \end{aligned} \quad (19)$$

On établi ensuite, la matrice des signatures des défauts (20) correspondante aux défauts envisagés.

$$\Lambda = \begin{array}{c|c|c|c|c} \hline -1 & +1 & 0 & 0 & 0 \\ \hline -1 & 0 & -1 & +1 & 0 \\ \hline -1 & +1 & 0 & 0 & +1 \\ \hline \end{array} \quad (20)$$

La matrice  $\Lambda$  est utilisée par la suite, pour générer une base de règles d'inférence. On procède dans une deuxième étape à la « fuzzification » des résidus. On calcule les valeurs des

fonctions d'appartenance  $\mu_{P_{r_i^-}}$ ,  $\mu_{P_{r_i^0}}$  et  $\mu_{P_{r_i^+}}$  pour chaque résidu. On a fixé le seuil  $s_p$  à 0.7 (pour minimiser le nombre de fausses alarmes) et la longueur de fenêtre  $L$  à 400 qui correspond approximativement à 7 heures (maximum de durée de défaut dans le cas de cette application). Les degrés de véracité de chaque hypothèse d'existence de défaut (correspondent à chaque variable) sont évalués par la relation (13) et quelques-uns sont représentés dans la figure 2

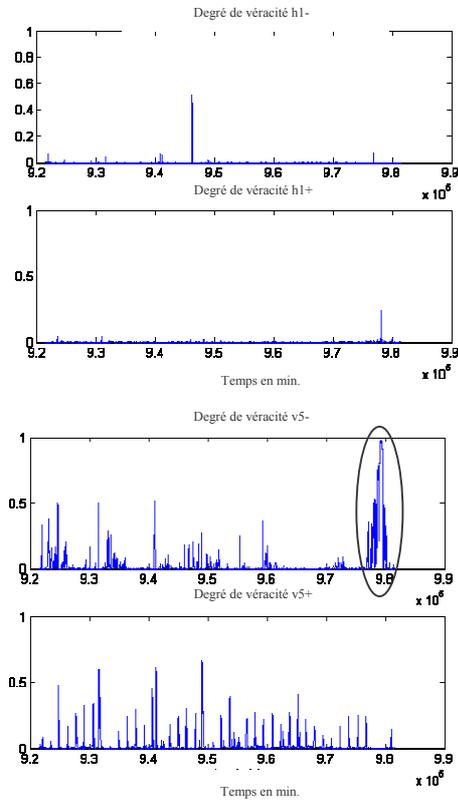


FIG. 2 : Degrés de véracité d'une anomalie + ou - affectant  $h_1$  et  $v_5$

L'analyse de ces degrés de véracité pour l'hypothèse correspondant à chaque variable montre que la défaillance (vrai défaut) est négative et affectant la variable  $v_5$  (où le degré de véracité est élevé (proche de 1) comme montré dans la figure 2). Ce vrai défaut est confirmé par le superviseur (validation par expertise). On a testé la procédure sur d'autres sites dans le même réseau de mesure et pour d'autres défauts réels. Les résultats obtenus confirment son efficacité.

## 7. Conclusion

Dans ce papier, on a utilisé un ensemble de modèles linéaires établi pour générer un ensemble de résidus et produire la matrice des signatures des défauts expérimentale. Par la suite, on a appliqué une méthode de diagnostic hors ligne basée sur les concepts du raisonnement flou qui est élaborée et implémentée pour la surveillance du réseau de mesure de la communauté urbaine de Nantes.

Dans le futur, on utilisera les modèles établis à pas de temps d'une heure et d'un jour pour améliorer l'ap'roche d'isolation ; et on complétera ce diagnostic en introduisant les règles de l'expertise et les variables de pluie pour donner les vrais causes de la présence de défauts sur les mesures.

Cela nous donnera la possibilité de prendre une décision concernant la solution à proposer pour que le système de surveillance continu sa mission correctement.

## Références

- [1] J. Ragot, M. Darouach, D. Maquin et G. Bloch, *Validation de données et diagnostic*, traité des nouvelles technologies, série diagnostic et maintenance, Hermès, Paris, 1990.
- [2] F. Berrada, S. Bennis et L. Gagnon, « Validation des données hydrométriques par des techniques univariées de filtrage », *Journal Canadien de Génie Civil*, vol. 23, pp. 872-892, 1996.
- [3] J. Ma, J. E. Amadi-Echendu et Y. Yan, « Non-redundancy based approach to validate a measuring device », *Proc. IEEE Instrum. and Meas. Technol. Conf.*, Ottawa, pp. 540-543, 1997.
- [4] M. Bourouis, L. P. Pascal F. S. Domenech et D. M. K. Al-Gobaisi, « Simulation and data validation in multistage flash desalination plants », *Desalination*, vol. 115, n°. 1, pp. 1-14, 1998.
- [5] S. Bennis et N. Kang, « Multivariate technique for validating historical hydrometric data with redundant measurement », *Nordic Hydrology*, vol. 31, n°. 2, pp. 107-126, 2000.
- [6] M. Mourad et J. L. B. Krajewski, « Procédure automatisée d'aide à la validation de données environnementales en hydrologie urbaine », *2<sup>ème</sup> Colloque Automatique et Environnement*, Saint-Etienne, 2001.
- [7] G. Olsson and B. Newel, *Modelling, Diagnostic and Control*, IWA Publishing, London, 1999.
- [8] D. Maquin, et J. Ragot, « Comparison of gross errors detection methods in process data », *Proc. of the 30<sup>th</sup> IEEE Conf. on Decision and Control*, Brighton, vol. 3, pp. 2253-2261, 1991.
- [9] C. Combastel, S. Gentil et J-P Rognon, « Génération de résidus adaptative pour le diagnostic de la machine asynchrone », *Conf. Internationale Francophone d'Automatique*, Nantes, 2002.
- [10] M. Ulieru, « A fuzzy logic based computer assisted fault diagnosis system », *European Journal of Diagnosis and Safety in Automation*, Hermès, vol. 4, n°. 3, pp. 405-441, 1994.
- [11] D. Mandel, D. Maquin et J. Ragot, « Process supervision based on fuzzy reasoning. Application to the urban water supply network », *3<sup>rd</sup> IFAC Workshop on Online Fault Detection and Supervision in the Chemical Processes Industries*, Solaize, 1998.
- [12] C. K. Mechefske, « Objective machinery fault diagnosis using fuzzy logic », *Mechanical Systems and Signal Processing*, vol. 12, n°. 6, pp. 855-862, 1998.
- [13] B. M. Izadi-Zamanabadi et S. Katebi, « Cheap diagnosis using structural modelling and fuzzy-logic-based detection », *Cont. Eng. Practice*, vol. 11, n°. 4, pp. 415-422, 2003.
- [14] H. D. Cheng, J. Wang et X. Shi, « Microcalcification detection using fuzzy logic and scale space approaches », *Pattern Recognition*, vol. 37, n°. 2, pp. 363-375, 2004.
- [15] D. Maquin et J. Ragot, *diagnostic des systèmes linéaires*, Hermès, Paris, 2000.
- [16] K. T. Atanassov, *Intuitionistic fuzzy sets*, Springer Phisica-Verlag, Heidelberg, 1999.
- [17] G. Chen, Q. Wie et E. Kerre, « Fuzzy data mining : discovery of fuzzy generalized association rules », *G. Bordogna, G. Pasi*, Springer Phisica-Verlag, pp. 45-66, 2000.