Preprints of the 2012 20th Mediterranean Conference on Control & Automation (MED), Barcelona, Spain, July 3-6, 2012

Leakage Localization in Water Networks using Fuzzy Logic

Gerard Sanz, Ramon Pérez and Antoni Escobet

Abstract— **This paper presents a methodology for leakage localization using FIR (Fuzzy Inductive Reasoning). A real water network situated in Barcelona has been subdivided in zones which could contain a leakage. Two sensors measure pressures on two separated points of the network. A faulty fuzzy model for each zone and sensor is generated. Test data have been used for classification of leakages in order to evaluate how this methodology helps in leakage localization. Results are compared with another isolation methodology. All the work has been done using simulations carried out by EPANET connected with Matlab. FIR applications used are programmed in Matlab too.**

I. INTRODUCTION

LEAKAGES in water networks are a major issue when an efficient management is intended for such huge infrastructure. Leakages appear inevitably sooner or later due to the ageing, high pressures or chemical reactions. Water loss due to leakages can be evaluated [1]. When the efficiency of the network falls leakages should be located and repaired.

There are different ways of facing the leakage detection and localization [2]. The research group of the authors has long experience in locating leakages using hydraulic simulation models [3] combined with sensors that provide the on-line information. A new approach is presented in this paper, where the models are based on fuzzy logic and generated with hydraulic simulations [4]. In the on-line process the simulations are not necessary. Of course the sensors are the providers of the information coming from the network.

The case study used for illustrating the methodology has been gently provided by Water Company of Barcelona AGBAR. It is a DMA (District Metered Area) called Nova Icària. The company has been working hard improving the efficiency of the network and there are pressure sensors installed for pilot test of leakage localization methodologies [5].

In section II the fuzzy logic methodology used for modeling leakages is briefly described. The application to the water networks is explained in section III. Results and comparison are presented in section IV. Finally, some conclusions and future works are suggested in section V.

This work was supported in part by the project DPI2009-13744 (WATMAN) of the Science and innovation ministry of Spain. Models and data of the real network were provided by AGBAR (Water Distribution Company in Barcelona)

G. Sanz, R. Pérez and A. Escobet are with the Department of Automatic Control, Universitat Politècnica de Catalunya, Rambla Sant Nebridi 10, 08222 Terrassa, Spain gerard.sanz@upc.edu

II. METHODOLOGY

A. Fuzzy Inductive Reasoning

Fuzzy Inductive Reasoning is a modeling and simulation methodology which finds the relations between qualitative and causal variables of the system from the observation of these variables behaviours during a certain time. Prediction of future behaviours is done using these relations.

The methodology is divided in four processes, shown in Fig. 1. Initially, the fuzzification process converts data into qualitative triples (class, membership degree and side). Next, the qualitative modeling process generates the optimal mask. This mask indicates which input variables are better for estimating the output. Moving this mask all over the dynamic identification data generates static patterns which indicate the input-output relations. Third, the qualitative simulation is performed. A qualitative triple is predicted by means of the optimal mask and the static patterns generated in the previous process. The k-nearest neighbour algorithm is used in order to increase the prediction consistency. Finally, a defuzzification process is performed to convert the predicted qualitative triple in a quantitative value. A more detailed description of the FIR methodology is presented in [6].

It is important to keep in mind that the FIR methodology is based on the system's behaviour rather than on its structure, and therefore, the amount and richness of the data available from the system are crucial in order to assure the identification of an accurate and reliable model representing it.

This whole process to find patterns of behaviour of each sensor is done by means of Visual-FIR. A description of this tool can be found in [7].

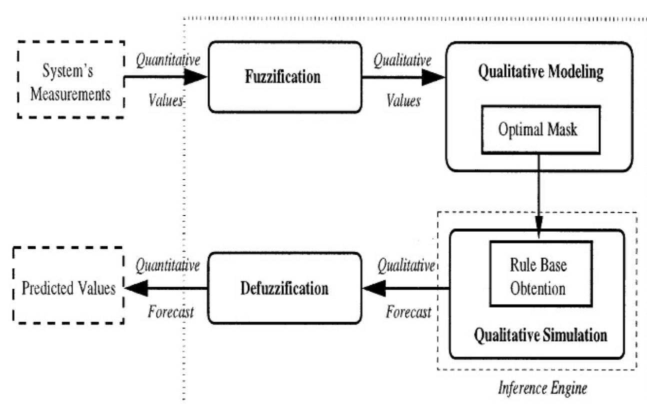

Fig. 1. FIR methodology structure

B. Fault Detection, Isolation and Identification

The methodology of leakage localization used in this paper is mainly based on standard theory of model-based diagnosis [8]. Model-based diagnosis can be divided in two subtasks: fault detection and fault isolation. The principle of model-based fault detection is to check the consistency of observed behaviour while fault isolation tries to isolate the component that is in fault.

An FDS (Fault Diagnosis System) is a monitoring system that is used to detect faults and diagnose their location and significance in a system. The fault detection process is described in Fig. 2. The grey boxes represent FIR processes, whereas the white boxes constitute the fault detection procedure.

As said in previous section, data measured from the system are converted into qualitative triples by means of the FIR fuzzification process. Subsequently, the fuzzy forecasting process predicts the next output value, a qualitative triple, from the qualitative data using the model (mask and pattern rule base) that represents the current behaviour of the system. The fuzzy forecasting process computes also the enveloping interval that drives the detection process. The enveloping concept is based on the 5 nearest neighbours that are computed inside the FIR inference engine by means of the k-nearest neighbour rule. A distance measure is computed between the input pattern, from which the output prediction should be obtained, and all patterns stored in the pattern rule base that match with that input pattern. The 5 patterns with shortest distance are selected as the 5 nearest neighbours.

The enveloping is composed by an upper bound (maximum value) and a lower bound (minimum value) delimiting the space where the real output signal should be present. If the real value falls outside the envelope, an instantaneous error occurs, meaning that the model used in the prediction does not correctly represent the system in that specific point. The instantaneous errors occurred inside a predetermined time window are accumulated. When the cumulative errors within the window are greater than the threshold specified by the modeler, an alarm is issued, and it is then necessary to identify the fault that has occurred.

The fault identification process is presented in Fig. 3. The grey boxes represent FIR processes whereas the white boxes constitute the fault identification procedure. This process runs along a diagnostic time window when the alarm triggers due to the detection of an anomalous behaviour. The size of this time window defines the number of prediction values that will be used in order to identify the fault that has been produced.

Therefore, the time window guides the prediction during the identification process. A small size of the time window is desired because it implies fast model identification. The prediction errors produced during each of the forecasting processes are accumulated, thus each fault model stored in the library has an associated cumulative error. This error is used to compute the model acceptability measure. The acceptability measure is a relative index ranking the models in terms of their ability to predict the new behaviour of the system. This measure allows, in a reliable way, to identify the fault that has occurred. It also offers guidance when the identification process faces additional problems, e.g. when the produced fault is not a foreseen fault and therefore is not available in the fault model library, or when an observed fault can be classified in two different models.

The model with the largest acceptability measure is selected as the one that best represents the new behaviour of the system, so the detected fault has been identified.

Software VisualBlock-FIR [9] performs these tasks by means of a user-friendly framework that runs under the Simulink platform. Several modules are available to the user in order to build a graphical FIR model of each fault. These models generate as output the prediction signal and the upper and lower bounds of the envelope.

Once each FDS based on a measurement has signalled the fault, this information has to be integrated. This paper presents a possible approach while other ideas are being developed meanwhile. This approach uses a voting process based on the FIR diagnosis. An advantage is that if new sensors are added the integration does not need any modification of existing models.

III. APPLICATION

An application of the methodology on water networks is explained in this section. First, some scenarios with different faults will be simulated in order to obtain data. These simulations are always necessary because data of different faults cannot be obtained from the real network.

Next, a fuzzy model for all the combinations of simulated leakages and available sensors will be generated, as well as a

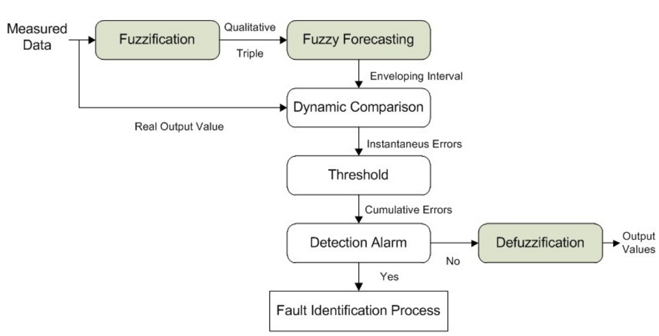

Fig. 2. Fault detection process

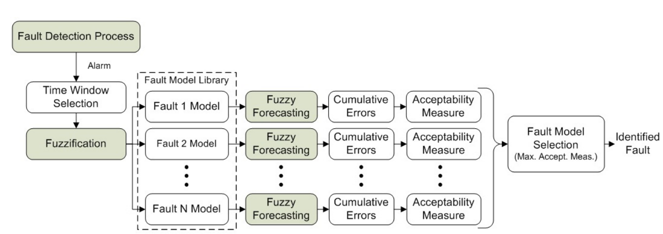

Fig. 3. Fault identification process

978-1-4673-2529-5/12/$31.00 ©2012 IEEE

fuzzy non faulty model for each sensor. This modeling stage will be done with the help of the software Visual-FIR.

Finally, a stage of leakage classification will be performed in order to identify which model predicted sensor pressure fits best with the real one (test data). Software VisualBlock-FIR will be used in this stage.

In a real case, the test data would be provided by sensors installed in the water network.

A. Fault Definition and Data Generation

First of all, the network is divided in six different zones. These zones represent groups of nodes that produce similar effects to the sensors, and have been defined geographically. A representative leakage in each of these groups of nodes is defined. Fig. 4 shows the division of the DMA, as well as the simulated leakages. Location of the sensors is depicted too, being sensor 1 located in zone B, and sensor 2 located in zone D. In a real application more leakages would be defined in each zone in order to make the identified fuzzy models more representative. In the same way, the use of more sensors will increase the isolation precision. This leads to a problem of sensor placement for fuzzy applications which is not treated in this paper.

In order to generate the data for identifying the fuzzy models and the test data for the fault classification stage, some simulations of the network are performed. Each of these simulations has one of the defined leakages with a loss of approximately 6.3 l/s. An extra simulation with no leakage is performed too. Input data formed by pressures and inflows at the network inlets are saved. In the same way, output pressures from the two sensors are collected. Fig. 5 shows the pressure obtained in one sensor for each defined scenario. Each simulation lasts 30 days. First 3 weeks are used for model identification and validation, whereas remaining days will form the test data used in the classification stage. Samples are taken every hour.

B. Model Generation

Once data are generated, a total of 14 fuzzy models are identified using Visual-FIR software. Each model is composed by the membership functions of the inputs and outputs, the optimal mask and the pattern rule base. In Fig. 6 the optimal mask for each predicted pressure of one fuzzy model is depicted. It can be seen that predicted pressure of sensor 1 depends of pressure of one input at the same sample and flow of the other input at two previous samples. On the other hand, predicted pressure of sensor 2 depends of pressures of both inputs at the same sample. Optimal masks obtained for each sensor are the same in all models. In the same figure, landmarks of the membership functions of each variable are signaled as dashed lines. These landmarks are used to fuzzificate each variable in three qualitative classes.

There is not a methodology for choosing the design parameters like number and shape of membership functions, complexity of the optimal mask, etc. The results of the application depend on these parameters, so a previous study of the best combination has been done using the model identification period (first three weeks).

C. Classification Stage

VisualBlock-FIR allows determining which model predicted pressure represents better the real pressure. The faulty models generated in the previous section are very similar, so doing a simultaneous classification does not return interpretable results. Consequently, this classification has been divided in 12 classification stages. Each of these stages consists in classificating test data among a faulty model for one sensor and the non-faulty model for the same sensor, as seen in Fig. 7. This results in a table indicating the number of times that a set of data belongs to each faulty model.

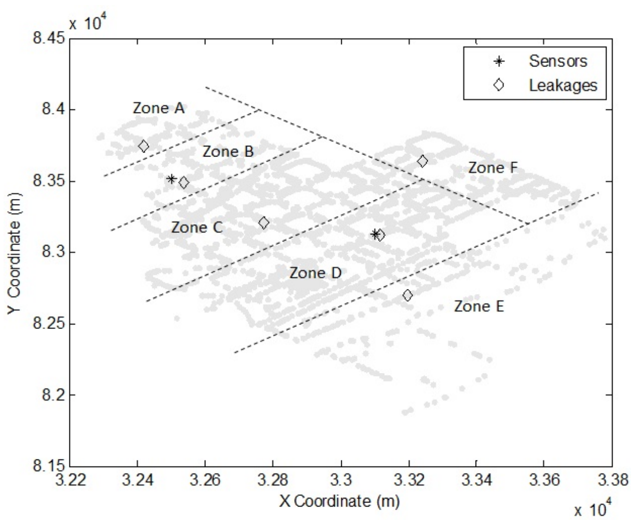

Fig. 4. DMA zone division

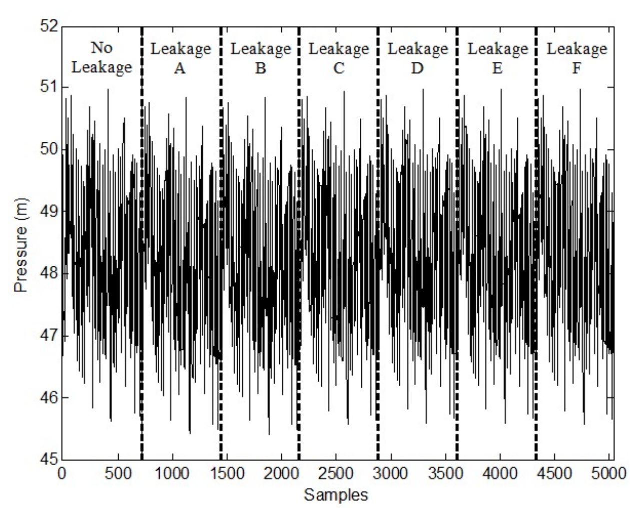

Fig. 5. Pressure in sensor 1 on different scenarios

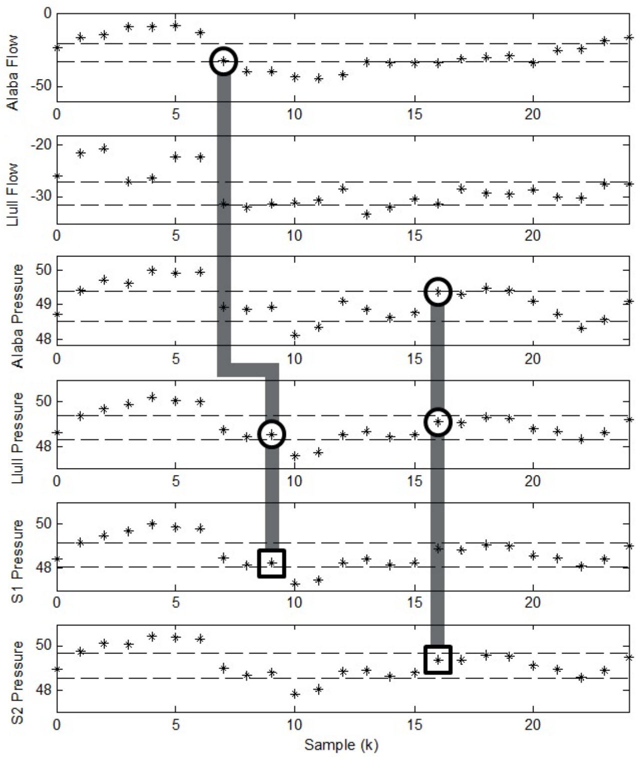

Fig. 6. Optimal masks for predicted pressures of sensor 1 and 2

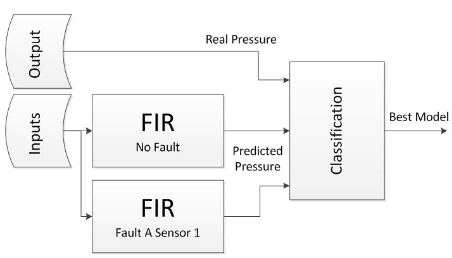

Fig. 7. Example of one classification stage

IV. RESULTS

A. FIR Approach

After computing the test data generated in section III-A in all available fuzzy models, the number of times that each faulty model has been identified in each dataset is shown in Table I. Columns represent the test scenarios, whereas rows represent each faulty fuzzy model.

The value on a cell corresponds to the number of samples (hours) that a model has been recognized in a dataset. Each value is over 216 hours. For instance, the faulty model of the zone D predicting pressure of the sensor 2 identifies 140 times the dataset with the leakage D. It can be seen that models of leakages near a sensor have null count when the simulated leakage is far (zones A and B for sensor 1, and zones D, E and F for sensor 2). This is due to the low pressure change on the sensors when the leakage is far.

In order to classify the leakages independently of the sensor, the product of the values from the same models is done for each dataset. This eliminates all the zones where one of the counts is zero, even if the other count is non-zero. Table II shows the new values. Grey cells point out where the simulated leakage is, whereas bold values indicate the model with higher probability of containing the leakage.

TABLE I

MODEL COUNTS PER SENSOR

Model Zone	Model Sensor	No Leakage	Leakage Location A	B	C	D	E	F
A	1	0	80	140	0	0	0	0
A	2	20	20	40	80	141	99	153
B	1	0	60	140	0	0	0	0
B	2	20	20	40	80	141	99	153
C	1	0	193	167	20	80	37	43
C	2	0	40	20	20	101	59	140
D	1	0	160	160	120	101	76	96
D	2	0	0	0	0	140	0	0
E	1	40	160	160	80	101	116	96
E	2	0	0	0	0	181	39	80
F	1	60	160	160	120	80	57	76
F	2	0	0	0	0	181	19	40

TABLE II

MODEL COUNTS

Model Zone	No Leakage	Leakage Location A	B	C	D	E	F
A	0	1600	**5600**	0	0	0	0
B	0	1200	**5600**	0	0	0	0
C	0	**7720**	3340	**400**	8080	2183	6020
D	0	0	0	0	14140	0	0
E	0	0	0	0	**18281**	**4524**	**7680**
F	0	0	0	0	14480	1083	3040

The analysis of the results has to be done column by column, as they represent each of the studied datasets. The count of each model will be plotted in a grey map. The darkness of each zone depends on the count obtained in the classification stage for that zone.

First of all, the non-leakage column of Table II shows that when there is no fault on the network all models have null count, thus the non-faulty scenario is perfectly identified.

Leakages C and E (Fig. 8 and Fig. 9) would be rightly classified, as the higher count obtained for them corresponds to the zone where they belong. Focusing on leakage C it can be seen that it is only identified by his own model, because the leakage location is in the middle point between the two sensors.

Taking a look at column B, the count of the faulty models of zones A and B are the same. That means that both zones have the same probability of containing the leakage. This is a good result, although the identified zone where the leakage would be searched is bigger.

Similar to the previous case is seen when the leakage is located in zone A. The model with higher count is C, so the leakage would be searched in this zone. As the leakage will not be found in this area, the search will continue with the next zone with higher count (zone A), which in this case

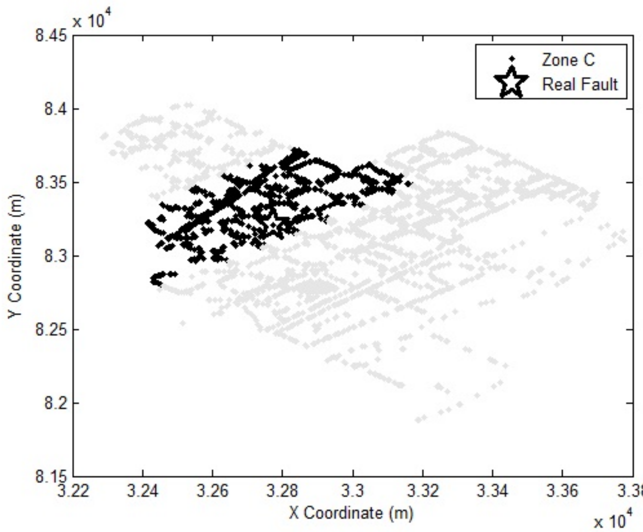

Fig. 8. Zone count representation when leakage C is simulated

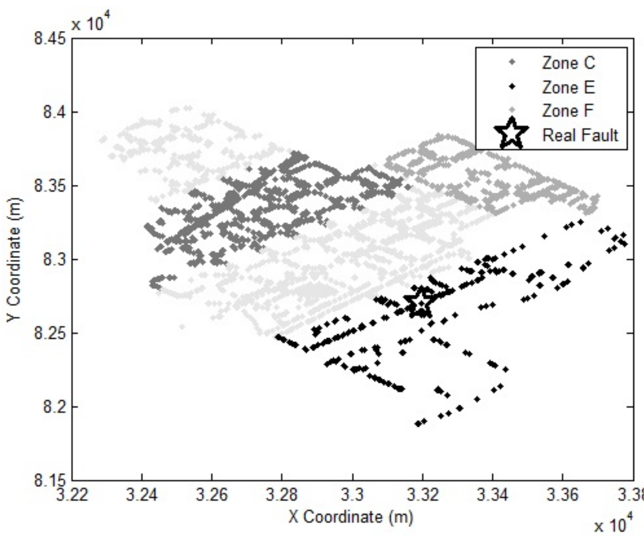

Fig. 9. Zone count representation when leakage E is simulated

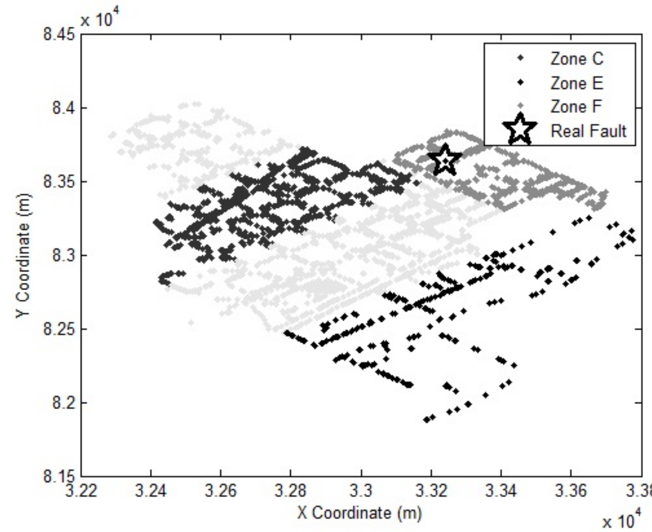

Fig. 10. Zone count representation when leakage F is simulated

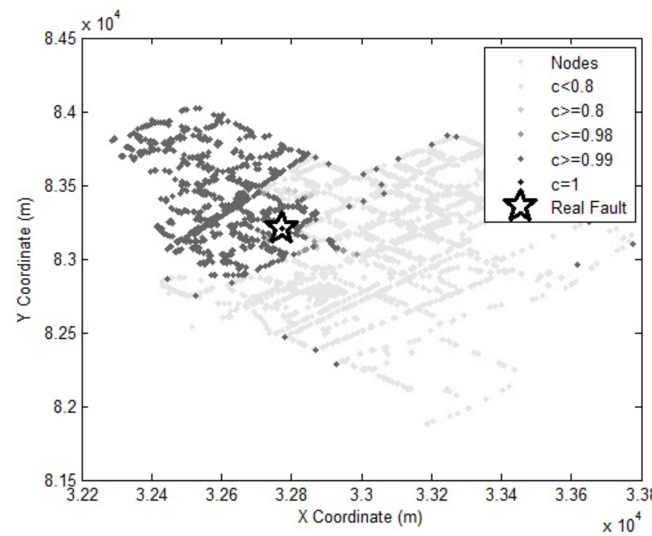

Fig. 11. Potential faulty nodes when leakage C is simulated using correlation method

contains the leakage.

Finally, leakages D and F are located, but with worse count than other models. However, the faulty zone is delimited and always contains the leakage which, sooner or later, would be found. Fig. 10 shows the grey map with the count of each zone and the location of the simulated leakage for the case F.

B. Results Comparative

In order to evaluate the FIR approach, a comparative with the correlation method for isolating leakages [10] is done. This methodology is based on correlation measurements of pressure sensors. Same test data are evaluated under the same conditions.

Results obtained with correlation method isolate correctly the leakage (as FIR approach), but with less precision. Zones where leakage would be searched are bigger, so a major effort would be necessary in order to find them. Fig. 11, Fig. 12 and Fig. 13 show the potential faulty nodes using the correlation method when leakages are in zones C, E and F, respectively. In the first and third case, the number of potential nodes are approximately the same in both methodologies. Whereas in the second case the FIR approach isolates better the leakage, as the number of potential nodes is less. It can be seen that the low number of sensors limits the precision of the correlation method.

Increasing the number of fuzzy faulty models in FIR approach (reducing the number of nodes in each zone) may generate better results and would increase the difference between methods.

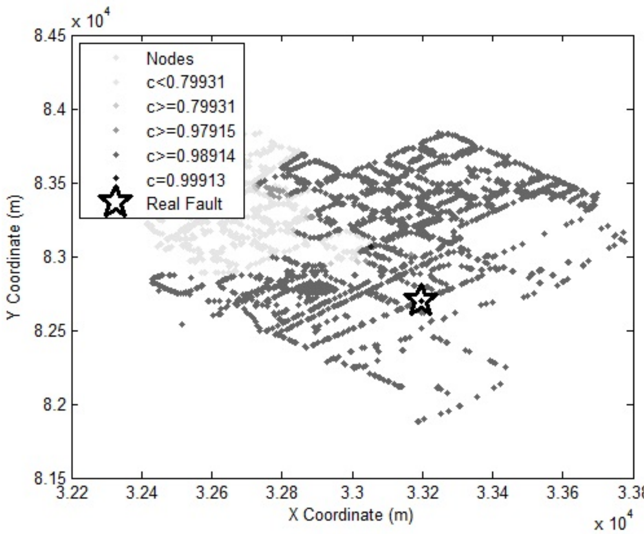

Fig. 12. Potential faulty nodes when leakage E is simulated using correlation method

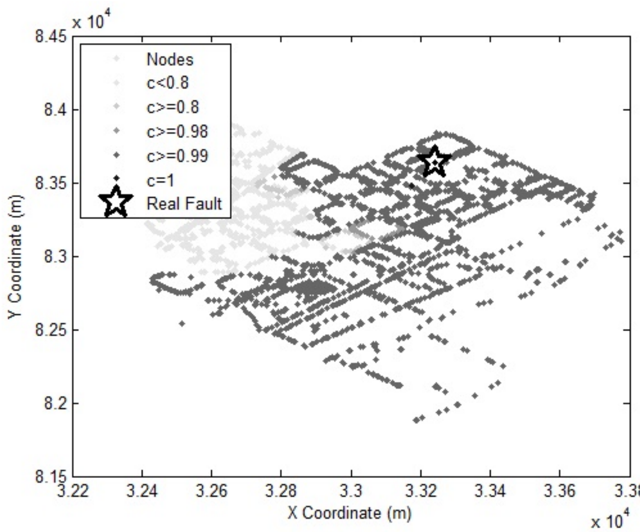

Fig. 13. Potential faulty nodes when leakage F is simulated using correlation method

V. CONCLUSIONS AND FUTURE WORKS

This paper has presented a methodology for leakage detection and isolation using fuzzy logic. A faulty fuzzy model for each leakage and sensor has been generated using Visual-FIR. Individual classification among faulty and non-faulty models has been performed by means of VisualBlock-FIR. Once the model recognition has been carried out for each model, the number of incidences has been used to give a probability of the leakage to be in each of the six areas.

Results presented in section IV are on simulation. They show the computation of each fuzzy model that delimits a zone where the leakage may be found. The precision achieved is low, although the use of more sensors may improve it, reducing the size of faulty potential areas. Apart from this improvement, the location of the sensors is significant too in fault isolation. In the same way, location of the leakages used for fuzzy model identification is also relevant. A study of the effect of each leakage over each sensor may improve results. This study would suggest a new zone division depending on network topology. Furthermore, using more leakages in the same zone to identify each model would make them more reliable. Application with real data will be done in the future.

As seen in the results section, faultless case is perfectly identified as all faulty models get null or insignificant count in the first dataset. This has to do with the interest of avoiding false alarms.

The comparative shows that the FIR approach suggest a search zone with less nodes than the correlation approach. However, there are some cases where the search zones are not adjacent using the FIR approach.

Finally, results may improve by tuning the modeling parameters on the Visual-FIR software. The number of membership functions can be modified, as well as the complexity of the optimal mask of each fuzzy model. Data classification may improve too by dealing with recognition parameters on the VisualBlock-FIR application.

REFERENCES

[1] A. Lambert, ”Accounting for Losses: The Bursts and Background Concept”, *Water and Environment Journal*, vol. 8, no. 2, pp. 205-214, Apr. 1994.

[2] M. Farley, S. Trow, *Losses in Water Distribution Networks*, IWA Publishing UK, 2003.

[3] R. Pérez; V. Puig; J. Pascual; J. Quevedo; E. Landeros; A. Peralta, ”Leakage isolation using pressure sensitivity analysis in water distribution networks: Application to the Barcelona case study”, *12th IFAC Symposium on Large-Scale Systems: Theory and Applications*, Villeneuve D'Asq, France 2010.

[4] M.A. Brdys; B. Ulanicki, *Operational control of water systems: Structures, algorithms and applications*, Prentice Hall International, UK, 1994.

[5] R. Pérez; V. Puig; J. Pascual; A. Peralta; E. Landeros; Ll. Jordanas, ”Pressure sensor distribution for leak detection in Barcelona water distribution network”, *Water Science & Technology*, vol. 9, no. 6, pp. 715-721, 2009.

[6] F. Cellier, ”General system problem solving paradigm for qualitative modeling”, *Qualitative simulation modeling and analysis*, (P.A. Fishwick and P.A. Luker, eds.), Springer-Verlag, New York, pp. 51-71, 1991.

[7] A. Escobet; A. Nebot; F. Cellier, ”Visual-FIR: A tool for model identification and prediction of dynamical complex systems”, *Simulation modelling practice and theory*, vol. 16, no. 1, pp. 76-92, January 2008.

[8] J.J. Gertler, *Fault Detection and Diagnosis in Engineering Systems*, Marcel Dekker, 1998.

[9] A. Escobet; A. Nebot; F. Cellier, ”Fault diagnosis system based on fuzzy logic: Application to a valve actuator benchmark”, *Journal of intelligent and fuzzy systems*, vol. 22, no. 4, pp. 155-171, May 2011.

[10] J. Quevedo; M.A. Cuqueró; R. Pérez; F. Nejjari; V. Puig; J. Mirats, ”Leakage location in water distribution networks based on correlation measurement of pressure sensors”, *IWA Symposium on Systems Analysis and Integrated Assessment*, San Sebastian, pp. 290-297, 2011.