



Jul 1st, 12:00 AM

# Neural Identification of Fuzzy Anomalies in Pressurized Water Systems

Joaquín Izquierdo

R. Pérez

P.A. López

P.L. Iglesias

Follow this and additional works at: <https://scholarsarchive.byu.edu/iemssconference>

---

Izquierdo, Joaquín; Pérez, R.; López, P. A.; and Iglesias, P. L., "Neural Identification of Fuzzy Anomalies in Pressurized Water Systems" (2006). *International Congress on Environmental Modelling and Software*. 383.  
<https://scholarsarchive.byu.edu/iemssconference/2006/all/383>

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

# Neural Identification of Fuzzy Anomalies in Pressurized Water Systems

**J. Izquierdo, R. Pérez, P. A. López, P. L. Iglesias**

Grupo Multidisciplinar de Modelación de Fluidos, Universidad Politécnica de Valencia, Spain  
Camino de Vera, s/n 46022 Valencia, Spain; jizquier@gmmf.upv.es

**Abstract:** The objective of a Water Supply System (WSS) is to convey treated water to consumers through a pressurized network of pipes. A number of meters and gauges are used to take continuous or periodic measurements that are sent via a telemetry system to the control and operation centre and used to monitor the network. Using this typically limited number of measures together with demand predictions the state of the system must be assessed. Suitable state estimation is of paramount importance in diagnosing leaks and other anomalies in WSS. But this task can be really cumbersome, if not unattainable, for human operators. The aim of this paper is to explore the possibility for a neural network to perform such a task. For one thing, state estimation of a network is performed by using optimization techniques that minimize the discrepancies between the measures taken by telemetry and the values produced by the mathematical model of the network, which tries to reconcile all the available information. But, for the other, although the model can be completely accurate, the estimation is based on data containing non negligible levels of uncertainty, what definitely influences the precision of the estimated states. The quantification of the uncertainty of the input data (telemetry measures and demand predictions) can be achieved by means of robust state estimation. By making use of the mathematical model of the network, estimated states together with uncertainty levels, that is to say, fuzzy estimated states, for different anomalous states of the network can be obtained. Also a description of the anomaly associated with such fuzzy state must be stored. The final aim is to train a neural network capable of assessing WSS anomalies associated with particular sets of measurements received by telemetry and demand predictions.

**Keywords:** Water supply systems; Neural Networks; Fuzzy logic; Modelling

## 1. INTRODUCTION

Most Water Companies, especially those supplying water to big cities, use telemetry systems for control and operation purposes. By considering the data provided by this system, the engineer on duty makes operation decisions trying to optimise the system utilisation. Nevertheless, the system complexity does not allow but to take a few real-time measures, which slightly represents the network state. They give indication of only certain aspects of the system, leaving out other more specific or “less relevant” ones. Thus, suitable techniques that allow more accurate network health estimation are necessary so that anomalies can be detected more rapidly and light anomalies, which develop progressively and insidiously, can be identified. This will enable to control their consequences in earlier stages, thus avoiding, among other things, losses, which can be of great importance, of this scarce resource called water. In addition, in the case of water for human consumption, it is an expensive one since additional investments in terms of quality have already been made.

When the objective is sustainable development, more efficient management and suitable water audits are completely necessary in water distribution systems in order to get better Integrated Water Management.

The state of a WSS is obtained by interrelating different measures within a mathematical model of the network with as much accuracy as possible, Martínez et al. [1995]. Different tools to analyse water networks have been developed in the last years, SARA (GMF [1998]), and EPANET (Rossman [1997]), among others.

But state estimation cannot be accurately performed if there are missing or uncertain data. Thus, system operators must not only be provided with data but also with some measure of their reliability. That is to say, they need a set of feasible states corresponding to the level of measure uncertainty. Equivalently, they need error limits of the state variables. But data are abundant since they are permanently received. Thus, the operators can evaluate errors neither easily nor in real time.

It is expected that suitable techniques borrowed from Artificial Intelligence (AI) could encapsulate the necessary knowledge to assess the network state, Abbot [1993], Dibike et al. [1999]. Different approaches would consist, for example, on artificial neural networks, or expert systems obtained from statistical or fuzzy techniques or based in Data Knowledge Discovery, Data Mining or Machine Learning Methods. Whatever the approach, the model should be able to identify the real system state, detecting anomalies, what will be used in decision making processes, necessary to optimise the water use.

There are a number of proposals in the literature to detect and identify anomalies and topological errors; yet, this still represents an important challenge facing water supply managers. In this paper we present an approach for the diagnosis and decision making process that consists on a neural network for clustering and pattern classification. Input patterns are either crisp or fuzzy. It is expected that the neural network will be able to eventually produce diagnoses of the network state when presented with new patterns.

The structure of this paper is the following. First, the mathematical model, a state estimation procedure and a mechanism for treating uncertainties, already presented in Izquierdo et al. [2004] and Izquierdo et al. [2005], are briefly presented. The mathematical model is essential for state estimation. The state estimation approach will allow handling uncertainty by using a technique based in matrix sensitivity, Bargiela et al. [2003]. The state estimator, together with the error limits provided by the matrix sensitivity approach, will be used as a surrogate of the real WSS to generate data to train and check the neural network (NN). Then, the inherent procedures to neural techniques will be described. Specifically, the NN architecture, the classification and clustering mechanisms of both crisp and fuzzy patterns and the training technique will be presented. Finally, the details of a preliminary application to a real WSS will be given.

## 2. MATHEMATICAL MODEL OF A WSS

Analysing pressurized water systems is one of the mathematically more complex tasks hydraulic engineers must face, especially for big systems, arising even for moderate size cities, since it involves solving a big number of non linear simultaneous equations. Several formulations are available. One of them considers the  $N - 1$  continuity equations, which are lineal, plus  $L$  energy equations, typically non linear.  $N$  is the number of demand junctions and  $L$  the number of lines in the system.

$$\sum_{j \in A_i} q_{ij} = Q_i, \quad i = 1, \dots, N-1 \quad (1)$$

$$H_{k_1} - H_{k_2} = R_k q_k |q_k| \quad k = 1, \dots, L$$

$A_i$  is the number of nodes directly connected to node  $i$ , and  $k_1$  and  $k_2$  represent the ends of line  $k$ , which conveys unknown flowrate  $q_k$  and is characterized by its resistance  $R_k$ , which depends on  $q_k$  through the Reynolds number.  $H_{k_1}$  and  $H_{k_2}$ , piezometric head at nodes  $k_1$  and  $k_2$ , are unknown for consumption nodes and are given for fixed head nodes. The complete set of equations may be written, by using block matrix notation,

$$\begin{pmatrix} A_{11}(q) & A_{12} \\ A_{12}^t & 0 \end{pmatrix} \begin{pmatrix} q \\ H \end{pmatrix} = \begin{pmatrix} -A_{10}H_f \\ Q \end{pmatrix} \quad (2)$$

where  $A_{12}$  is the so-called connectivity matrix describing the way demand nodes are connected through the lines. Its size is  $L \times N_p$ ,  $N_p$  being the number of demand nodes;  $q$  is the vector of the flowrates through the lines,  $H$  the vector of unknown heads at demand nodes;  $A_{10}$  is an  $L \times N_f$  matrix,  $N_f$  being the number of fixed head nodes with known head  $H_f$  and  $Q$  is the  $N_p$ -dimensional vector of demands. Finally,  $A_{11}(q)$  is an  $L \times L$  diagonal matrix, with elements

$$a_{ii} = R_i q_i + B_i + \frac{A_i}{q_i} \quad (3)$$

with  $R_i = R_i(q_i)$  being the line resistance and  $A_i, B_i$  coefficients characterizing a potential pump.

System (2) is a non-linear problem, whose solution is the state vector  $x = (q, H)^t$  of the system.

## 3. STATE ESTIMATION

For the state estimation process resource is made of the mathematical model. The non-linear relations among flowrates and heads describing the system balances are complemented by specific measurements taken by using telemetry systems. These measurements are integrated into the model by expanding system (2) to a new system, typically overdetermined, of more equations than unknowns

$$\begin{pmatrix} A_{11}(q) & A_{12} \\ A_{12}^t & 0 \\ A_{31} & A_{32} \end{pmatrix} \begin{pmatrix} q \\ H \end{pmatrix} = \begin{pmatrix} -A_{10}H_f \\ Q \\ M_r \end{pmatrix} \quad (4)$$

The new matrices in system (4) account for measurements  $M_r$ , which are incorporated this way into the model. Since measurements are often contaminated by noise and errors (coming from telemetry devices and from uncertainty in the demand predictions), system (4) has to be solved by using some optimization technique. The

selected technique defines the state estimator. Since most water systems involve a huge number of equations and unknowns, system (4) is usually solved by an over-relaxation iterative process applied to a linearized version of (4):

$$\begin{pmatrix} A_{11}^i(q^{(k)}) & A_{12} \\ A_{12}^i & 0 \\ A_{31} & A_{32} \end{pmatrix} \begin{pmatrix} \Delta q \\ \Delta H \end{pmatrix} = \begin{pmatrix} -A_{10}H_f - A_{11}(q^{(k)})q^{(k)} - A_{12}H^{(k)} \\ Q - A_{21}q^{(k)} \\ M_t - A_{31}q^{(k)} - A_{32}H^{(k)} \end{pmatrix} \quad (5)$$

where  $A_{11}^i$  is the Jacobian matrix corresponding to  $A_{11}$  (all the other matrices are constant). Let us rewrite (5) in a more compact way as

$$A_k \Delta x_k = b_k \quad (6)$$

Matrix  $A_k$  is size  $(L+N_p+N_t) \times (L+N_p)$ ,  $N_t$  being the number of telemetry measurements incorporated to the system.

The next pseudo code uses the so-called normal equations and gives the least square solution of (6)

Given  $x_0 = (q_0, H_0)^t$ , initial estimate  
 For  $k = 0$  to convergence  
   Obtain  $b_k$  and  $A_k$ .  
   Solve  $A_k^t W A_k \Delta x_k = A_k^t W b_k$   
   Obtain  $x_{k+1} = x_k + \alpha \Delta x_k$   
   Check convergence  
 Next  $k$

where  $\alpha$  is the over-relaxation factor, and  $W$  is a diagonal matrix weighting measurements.

#### 4. ERROR LIMIT ANALYSIS

Error limit analysis is a process to determine uncertainty bounds for the state estimation originated by the lack of precision of the measurements and, in general, of the used data. This representation of uncertainty produces better results than noise probabilistic representation due to the fact that most telemetry systems do not use suitable noise probabilistic characterization, Milanese et al. [1996].

To put it in a nutshell, the question is what is the reliability of the estimated state  $x^*$  if measurement vectors,  $y$ , are not crisp but may vary in some region,  $[y - \delta y, y + \delta y]$ ?

Different techniques may be used to estimate this unknown but bounded error, Milanese et al. [1996], Norton [1986], Kurzhanski et al. [1997].

We use a variant of the so-called sensitivity matrix analysis, Bargiela et al. [2003], which uses the state estimator presented above. In effect, from the linearization process it can be proven that the variation of  $x^*$  ( $\delta x^*$ ) is related to  $\delta y$  by means of

$$A_k^* W A_k^* \delta x^* = A_k^* W \delta y, \quad (7)$$

where  $*$  indicates that calculations are made for  $x^*$ .

In Izquierdo et al. [2005], (7) is used to prove that a component by component bound,  $e^*$ , for  $\delta x^*$  can be obtained by means of

$$e^* = \left| (A_k^* W A_k^*)^{-1} A_k^* W \right| |\delta y| \quad (8)$$

where the vertical bars indicate absolute values of all matrix and vector entries. Since (7) is linear the bounds calculated by (8) are symmetrical and error limit may be expressed as a multidimensional interval (see cell definition in next paragraph)  $[x^*]$  in the state space. Maximum and minimum values for the state variables may be expressed by

$$[x^*] = [x_{\text{inf}}^*, x_{\text{sup}}^*] = [x^* - e^*, x^* + e^*]. \quad (9)$$

#### 5. THE NEURAL NETWORK

A neural network for clustering and classification is a mechanism for pattern recognition. Two main approaches are generally used. We will use here the one based on multidimensional cells, Simpson [1992], Likas et al. [1994]. Another alternative makes use of the so-called Voronoi diagrams, Blekas et al. [1997].

A cell  $C$  is a region of the pattern space of  $n$ -dimensional vectors obtained as the intersection of  $n$  pairs of half-spaces of the form  $m_i \leq x_i \leq M_i$ , for  $i = 1, 2, \dots, n$ , where  $m_i$  and  $M_i$  are real numbers. Vectors  $m = (m_i, i = 1, \dots, n)$  and  $M = (M_i, i = 1, \dots, n)$  are called min and max points of  $C$  and completely determine  $C$ . Membership of patterns to a cell is defined from fuzzy grounds. For fuzzy patterns,  $P = [P^{\text{inf}}, P^{\text{sup}}]$ , like the ones obtained in (9), membership values are given by the membership function

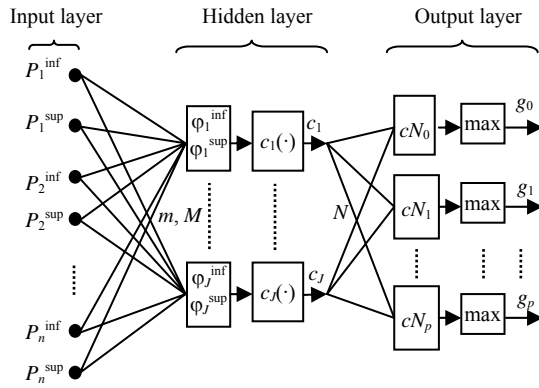
$$c(P) = 1 - \max_{i=1, \dots, n} \left\{ \max \left\{ \varphi_i(P_i^{\text{sup}} - M_i), \varphi_i(m_i - P_i^{\text{inf}}) \right\} \right\} \quad (10)$$

where  $\varphi_i(x)$ , for each coordinate  $i$ , is non-decreasing and takes values from 0 to 1: 0 for negative numbers; then it monotonically grows for 'small' positive numbers; and 1 for 'big' positive numbers. The 'size' of a positive number (and the shape of  $\varphi_i$ ) is controlled by one or more sensitivity parameters. For example,  $\varphi_i$  may be a sigmoid function or, simply, a ramp function controlled by a positive parameter  $\mu_i$ , say  $\varphi_i(x) = \mu_i x$  for  $x \in [0, 1/\mu_i]$ . Note that membership of crisp values can also be expressed by (10) considering  $P^{\text{inf}} = P^{\text{sup}}$ .

Values taken by membership function (10) are used in two ways. During the training phase they are used to decide if a cell must be or not expanded, as explained below. And during the operation phase, the NN will use them to decide the membership degree to the class associated with a cell exhibited by certain pattern presented to it and, as a consequence, to recognize the potential anomalous state of the water distribution system corresponding to the class' associated label.

Patterns presented to the network during the training phase are ordered pairs  $(P, e)$ , where  $P$  may be either crisp or fuzzy and  $e$  is a label associated to  $P$  describing the type of anomaly it represents. Fuzzy pattern  $P$  is described by its lower and upper limit vectors  $P^{\text{inf}}$  and  $P^{\text{sup}}$ , respectively, and  $e$  is an index or a label corresponding to one of the  $p$  likely anomalies in the water distribution system (for example, leak between nodes  $i$  and  $j$ , wrong valve position on line  $k$ , etc.). Note that a label must be allocated to describe normal operation of the system, as well.

The NN implementing the classification process is a three-layer network (Figure 1) that grows adapting itself to the problem characteristics, as explained below.



**Figure 1.** Neural network architecture

The input layer has  $2n$  neurons or process units, two for any of the dimensions of the patterns  $P = [P^{\text{inf}}, P^{\text{sup}}]$ . When a new pattern is presented to the network through the input layer, the components of vectors  $P^{\text{inf}}$  and  $P^{\text{sup}}$  are compared, respectively, with those of the minimum point,  $m$ , and the maximum point,  $M$ , of the  $J$  existing cells.

Specifically, numbers

$$m_i - P_i^{\text{inf}} \text{ and } P_i^{\text{sup}} - M_i$$

are calculated for the  $J$  existing cells and then mapped by functions  $\varphi_i$

$$\varphi_i(m_i - P_i^{\text{inf}}) \text{ and } \varphi_i(P_i^{\text{sup}} - M_i).$$

Thus way, each neuron on the hidden layer has two  $n$ -dimensional vectors  $\varphi^{\text{inf}}$  and  $\varphi^{\text{sup}}$  as its input, formed by numbers between 0 and 1 ready to be operated first component by component with the max operator and then with the max operator but now through all the components. Specifically,

$$c(P) = 1 - \max_{i=1, \dots, n} \left\{ \max \left\{ \varphi_i^{\text{sup}}, \varphi_i^{\text{inf}} \right\} \right\}$$

is calculated for each cell. This process of applying the membership function of the  $J$  existing cells to pattern  $P$  gives the membership degree of  $P$  to every one of the cells. Thus, membership functions may be considered as the transfer or activation functions for all the  $J$  existing hidden neurons. And the values of the minimum and maximum points of those existing cells, which will be adjusted during the training phase, must be regarded precisely as the synaptic weights between the input and the hidden layer.

The values produced by the membership functions of the existing cells constitute the outputs of the hidden layer. These values must be operated with the weights between the hidden and the output layers. And this process will produce a class, a diagnosis of the hydraulic system represented by pattern  $P$ . That is to say, the output layer must provide a class or label  $e^*$  (a diagnosis) to be associated to pattern  $P$ . During the training phase, one input pattern  $P$  will be associated with its 'correct' (target) class  $e$ , defining certain diagnosis, and comparison between target and output will allow training the network. After training, during the execution phase, the diagnosis produced by the whole combination of weights will be associated to any 'new' pattern presented to the network reflecting a real state of the hydraulic system. It would help the process of decision making by the system managers.

This functionality is achieved by assigning binary values,  $n_{jk}$ , to the synaptic weights between the hidden and the output layers: the link connecting cell  $C_j$  (hidden layer) to the neuron representing class  $e_k$  (output layer) will be assigned the value  $n_{jk} = 1$  if cell  $C_j$  belongs to class  $e_k$ , otherwise it will be assigned 0. These binary values are stored in a matrix  $N$  whose size is  $J \times p$ .

During the training phase these values are changed, according to the classification processes discussed below. This way, the output given by the  $p$  units of the output layer, which represent a class or diagnosis associated to the system, gives the degree to which the current pattern  $P$  reflects each diagnosis. This output is calculated by using the transfer or activation function

$$g_k = \max_{j=1,\dots,J} C_j n_{jk} \cdot \quad (11)$$

This provides each pattern  $P$  with either a fuzzy membership assignment to the different classes, being  $g_k$  its grade of membership to class  $e_k$ , or a crisp defuzzified membership assignment to that class  $e_k$  exhibiting the highest  $g_k$ .

During the training phase patterns are presented to the net in turn. For each pattern, the membership values corresponding to the cells already created are used for two alternative basic purposes: to create a new cell if the current pattern does not exhibit suitable membership level to any of the existing cells, or to expand a cell to include the current pattern provided it exhibits sufficient membership level. Either creation or expansion processes are carried out every time a new pattern  $(P, e)$  is presented to the NN.

After creating or expanding a cell, overlapping must be checked since it would create the ambiguous situation of a pattern fully belonging to two different cells, which could imply the possibility for the system to exhibit two different states, which would be nonsense. Overlapping must be dealt with a suitable cell contraction process. We will discuss here only creation and expansion processes. Overlapping and subsequent contraction are straightforward processes mainly of computational nature.

To create a cell,  $C_j$ , vectors  $P^{\text{inf}}$  and  $P^{\text{sup}}$  of the current pattern are used to initialize its minimum,  $m$ , and the maximum,  $M$ , points and label  $e = e_k$  is used to assign a class to  $C_j$ :

$$m = P^{\text{inf}}, \quad M = P^{\text{sup}}, \quad n_{jk} = 1.$$

The expansion of an existing cell,  $C$ , is carried out by updating its minimum and maximum points according to the values  $P^{\text{inf}}$  and  $P^{\text{sup}}$  of the current pattern. Expansion of cell  $C$  is carried out through the re-definition process:

$$m_i \rightarrow \min(m_i, P_i^{\text{inf}}), \quad M_i \rightarrow \max(M_i, P_i^{\text{sup}}), \quad i = 1, \dots, n. \quad (12)$$

A new cell is created if upon presentation of a new pattern  $(P, e)$ , with  $P = [P^{\text{inf}}, P^{\text{sup}}]$ , all the existing cells either

a) verify

$$\max_{i=1,\dots,n} \{ \max(M_i, P_i^{\text{sup}}) - \min(P_i^{\text{inf}}, m_i) \} > \sigma \quad (13)$$

where  $\sigma$  is a real number used to limit the maximum size of cells, that is to say, (all the existing cells) are too big to be expanded, or

b) have associated classes different from  $e$ , thus being absurd to expand any of them to include the new pattern.

The whole process, including overlapping test and subsequent contraction, allows existing cells to be refined progressively and, at the same time, new cells to be created as new patterns claim for them.

## 6. APPLICATION TO A WSS

Figure 2 shows a small hydraulic network to which the processes presented are been applied. It has one feed point of constant head, 50m, and is made out of 20 nodes and 32 lines.

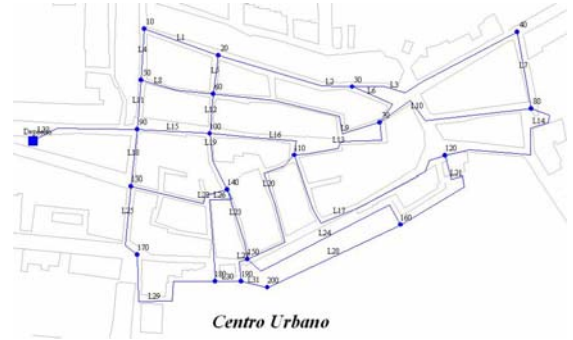
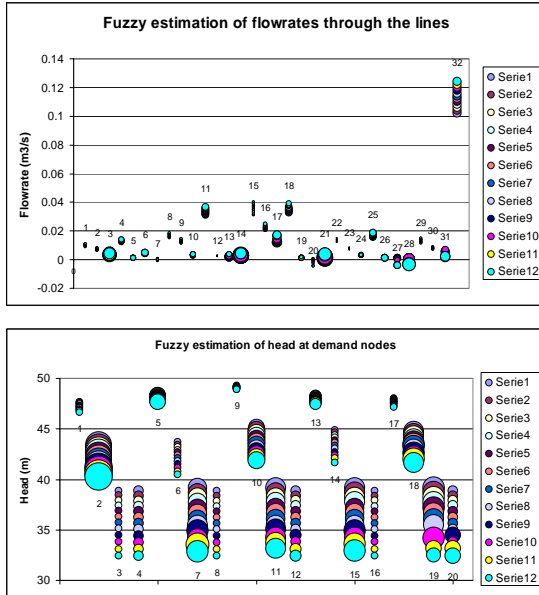


Figure 2. Hydraulic system under study

Different levels of losses have been simulated at the mid point of each line for different demands by using the mathematical model of the system. For each simulation some of the data have been used as surrogates for the measurements that certain meters and gauges installed in the system should give as a result of the simulated state of the hydraulic system. For this network, precisely 8 values corresponding to piezometric heads and 7 values corresponding to flowrates at specific nodes and lines, respectively, have been used in all the simulations. Then, these values have been incorporated into the system, but now without considering the anomaly. Thus, the state estimator must have solved systems of  $20 + 32 + 15 = 67$  equations with  $20 + 32 = 52$  unknowns. It means that a factor of redundancy of 22% has been used. The estimated states, in terms of flowrates through the lines and heads at the demand nodes, that best reconcile the simulated demand patterns within the model (now ignoring the anomaly) have been calculated by using (5) and the pseudo code given in paragraph 3. Also uncertainty values for the measurement devices have been provided and the associated fuzzy estimated states have been obtained by using (7). Also, as noted in par. 3, a weighting matrix for the different measurements has also been provided. The fuzzy states associated to 12 different losses on line 16 and average

demand load have been represented in Figure 3 by using bubble charts.

Note that bubbles are out of scale, so that some of the variables, for which the error interval is really small, may be spotted. As a reference, in these charts the maximum size of the bubbles, that is to say, the maximum error for the different dimensions of the state space, is  $0.016\text{m}^3/\text{s}$  for flowrates and  $1.35\text{m}$  for heads.



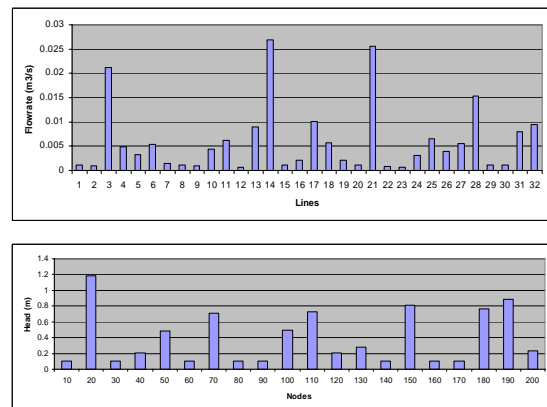
**Figure 3.** Fuzzy state estimation

When considering, for example, three demand loads and 12 levels of losses in every line, a database of patterns has been obtained that will be used to train the NN. The database thus obtained is formed by  $32 \times 12 + 1 = 385$  patterns for every demand load, what amounts to  $385 \times 3 = 1155$  patterns obtained from respective simulations representing anomalous states of the system characterized by losses in its lines, plus patterns associated to normal operation of the system.

Patterns corresponding to the three different demand loads can be easily discriminated by using the flowrate through pipe #32, which feeds the system from the point of constant head (for other more complex networks, fed from different sources, the identification of variables able to perform such a discrimination will be straightforward). As a consequence, the flowrate corresponding to line #32 has been withdrawn from all the training patterns. On the other hand, and more importantly, this enables to split the training set into different subsets that will be used to train different specialized (local) sub-neural networks each of which is responsible for a particular region or demand load of the input space, what closely relates to the idea of modular

modelling, Solomatine et al., [2006]). Accordingly, in a first stage only data for the average demand load have been used to train one of the neural networks that will eventually integrate the whole committee machine, one expert for each demand load.

For this hydraulic system it amounts to consider only the 385 patterns corresponding to this demand load. In order to normalize the data, the training set has been first reduced to the unit cell, with 0 and 1 as min and max points. For this specific system and demand load, piezometric heads have been normalized between 30 and 50 (water column) meters and flowrates between  $-0.1$  and  $0.1$  cubic meters per second. Different training sessions have been performed for different values of parameter  $\sigma$  in (13). Obviously the success of the process is strongly dependent on the value of  $\sigma$ . An early conclusion has been that of using different values of  $\sigma$  for different dimensions of the pattern space, which is in agreement with the intuitive idea that not all the pipes and nodes in a hydraulic system are equally important for the performance of the system. There are few attempts in the literature to order or rank pipes and nodes in a hydraulic system using some suitable index (for pipes, and in connection to system maintenance and design using genetic algorithms, two examples are Arulraj et al. [1995], and Vairavamoorthy et al. [2005]). In our case, we use the error limit analysis described in paragraph 4, applied to the steady state of the system (see Figure 4) to assess the ‘importance’ of pipes and nodes and, based on it, to decide suitable values of  $\sigma$  for the different dimensions of the pattern space. Obviously, the best training results are obtained in this case.



**Figure 4.** Error limit analysis for steady state

When the training data are used for a preliminary testing, the number of faults depends on the type of output one is interested in, according to (11). If a crisp defuzzified output is considered, 1.8% of the patterns are misclassified. The misclassified patterns correspond to the lower losses in the more

peripheral (and small) pipes, specifically pipes #3, 14, 21 and 28 (see upper chart in Figure 4). If a fuzzy output is considered, only 0.7% of the patterns fail to be classified into the two classes exhibiting the two top membership values (lower losses of pipes #3, 14 and 21). No patterns are misclassified when considering more than the two classes with top membership values. These results must be considered as excellent. Nevertheless, it must be noticed at once that testing has been performed for state estimates obtained from exact measurements. The reality is that in the environment of a real hydraulic system measurements are received affected by noise. As a consequence, an independent set of patterns affected with some noise has been generated to really test the performance of the recognition systems. Such a set is obtained by adding random errors (depending on the accuracy of meters and gauges) to the exact measurements calculated by the mathematical model. The estimation state process, by using these ‘corrupted’ measurements, allows producing a much bigger database of patterns, formally similar to the ones used to train the network, but now affected by noise.

## 7. CONCLUSIONS

The described neural procedure does not fit into any standard paradigm since it is made of several sub-nets that evolve by accumulating experience as new loads (peak, valley, seasonal-dependent, etc.) are presented, which mimics human knowledge acquisition.

From the reduced number of tests performed we claim that the classification ability of the NN is excellent. Also, overlapping is robustly solved, especially if using fuzzy, instead of crisp, patterns.

Since the response given by the NN is graded, as a consequence of its fuzziness (11), the information it provides is not only qualitative (pointing out an anomaly) but also quantitative (pondering the distributed importance of the problem).

The tool presented her, once completed, calibrated and implemented, will provide WSS managers with a decision support mechanism allowing early identification of anomalies and, as a consequence, better Integrated Water Management.

## 8. ACKNOWLEDGEMENTS

This work is been performed under the support of the projects Investigación Interdisciplinar nº 5706 (UPV), **DPI2004-04430** of the Dirección General de Investigación del Ministerio de Educación y Ciencia (Spain) and FEDER funds.

## 9. REFERENCES

- Abbot, M. B. The electronic encapsulation of knowledge in hydraulics, hidrology and water resources. *Advances in water resources*, Vol. 16, pp 21-39, 1993.
- Arulraj, G. P. and Rao, H. S. Concept or significance index for maintenance and design of pipe networks. *J. Hydraul. Eng.*, 121(11), 833-837, 1995.
- Bargiela, A., Pedrycz, W. *Granular Computing*. Kluwer Academia Press, Boston, Dordrecht, London, 2003.
- Blekas, K., Likas, A., Safylopatis, A. A Fuzzy Neural Network Approach to Classification Based on Proximity Characteristics Patterns. *Proc. 9th IEEE Int. Conference on tools with Artificial Intelligence*, Newport Beach, CA, 1997.
- Dibike, Y. B., D. Solomatine and M. B. Abbot. On the encapsulation of numerical-hydraulic models in artificial neural networks. *Journal of Hydraulic Research*, Vol 37, 1999, No 2.
- GMF (Grupo Mecánica de Fluidos). SARA, *Software de Análisis de Redes de Agua, Manual de Usuario*. Ed. Grupo Mecánica de Fluidos, UPV, 1998.
- Izquierdo, J., Iglesias, P. L., Díaz, J. L., Pérez, R. Identificación de estados en sistemas de distribución de agua mediante técnicas de optimización, *IV SEREA Seminario hispano-brasileño: Planificación, proyecto y operación de redes de abastecimiento de agua*, CD-Rom, Valencia, 2004.
- Izquierdo, J., López, P. A., Pérez, R., Ledesma, B. Intervalos de confianza en la estimación de estados en sistemas de distribución de agua. *V SEREA Seminario hispano-brasileño: Planificación, proyecto y operación de redes de abastecimiento de agua*, CD-Rom, Valencia, 2005.
- Kurzhanski, A. B., Valyi, I. *Ellipsoidal Calculus for Estimation and Control*. Birkhauser, 1997.
- Likas, A., Blekas, K., Safylopatis, A. Application of the Fuzzy Min-Max Neural Network Classifier to Problems with Continuous and Discrete Attributes. *Proc. of IEEE Workshop on Neural Networks for Signal Processing (NNSP'94)*, pp 163-170, 1994.
- Martínez, F., R. Pérez and J. Izquierdo. Optimum Design and Reliability in Water Distribution Systems, in *Improving efficiency and reliability in water distribution systems*. Kluwer Academic Pub. Dordrecht, Boston, London, 1995.
- Milanese, M. et al. (Eds.). *Bounding Approaches to system identification*, Milanese, M. et al. (Eds.), Plenum Press, NY, 1996.
- Norton, J. P. *An Introduction to Identification*. Academic Press, 1986.
- Rossmann, L. A. *Manual de usuario de EPANET*. Drinking Water Research Group. Risk Reduction Engineering Laboratory. US EPA. Translated by GM Fluidos, UPV, 1997.
- Simpson, P. Fuzzy Min-Max Neural Networks-Part 1: Classification. *IEEE Trans. On Neural Networks*, 3(5): 776-786, 1992.
- Solomatine, D. P. and Siek, M. B. Modular learning models in forecasting natural phenomena, *Neural Networks* 19, 215-224, 2006.
- Vairavamoorthy, K. and Ali, M. Pipe index vector: a method to improve genetic-algorithm-based pipe



optimization. *J. Hydraul. Eng.*, 131(12), 1117-1125, 2005.