Artificial Neural Network Model for a Low Cost Failure Sensor: Performance Assessment in Pipeline Distribution

Asar Khan, Peter D. Widdop, Andrew J. Day, Aliaster S. Wood, Steve, R. Mounce, and John Machell

Abstract—This paper describes an automated event detection and location system for water distribution pipelines which is based upon low-cost sensor technology and signature analysis by an Artificial Neural Network (ANN). The development of a low cost failure sensor which measures the opacity or cloudiness of the local water flow has been designed, developed and validated, and an ANN based system is then described which uses time series data produced by sensors to construct an empirical model for time series prediction and classification of events. These two components have been installed, tested and verified in an experimental site in a UK water distribution system. Verification of the system has been achieved from a series of simulated burst trials which have provided real data sets. It is concluded that the system has potential in water distribution network management.

Keywords—Detection, leakage, neural networks, sensors, water distribution networks.

I. INTRODUCTION

AN "event" in a water pipeline distribution system is defined when something unusual or non-standard in the measured or operational characteristics of the flow is detected. Detection can be visual evidence of a leak, or, more often, changes in flow, pressure or some other parameter which is being measured. A leak can be major and catastrophic ("burst" – in which case the evidence and location is clear to see) or minor. Locating the position of this latter type of leak can be very difficult. A leak survey is often conducted, and can be divided into two phases [1]. In the first phase, the whole distribution network is examined for leaks during minimum flow conditions. The usual approach is district flow metering which is the most common method in England and Wales. Martin and Farley [2] commented that computer technology has provided the opportunities for continual monitoring of leakage levels through district metering. The principle of district metering is based on the subdivision of the distribution system into discrete district meter areas (DMAs), by the permanent closure of valves, and the measurement of the flows into each zone. A DMA generally comprises an area containing approximately1000 properties. By examination of the measured minimum night flows (NFM) for each zone the existence of a leak can be confirmed. The NFM is the lowest flow supplied to a hydraulically isolated supply zone, and is usually measured between midnight and 5:00 a.m. because water use is at a minimum and it is thus easier to identify the legitimate flows. If the night flow minus the legitimate flow is close to zero, the leakage must also be close to zero. In contrast, discrepancies will signify leakage in the absence of any other factors.

In phase 2 of a leak survey the aim is to pinpoint the position of the leak. Detection crews (or contractors) will enter an area and perform a systematic listening for the sound of a leak at valves and fittings where the main can be reached without any excavation. Leaks make noise because as pressurised water is forced out through a cavity somewhere in the pipe wall, flow energy is lost to the pipe wall and to the surrounding soil area. This energy creates sound waves in the audible range which can be sensed and amplified by electronic transducers, or in some cases, by simple mechanical means. Thames Water alone has more than 100 two-man teams working full time looking for leaks [3]. Sometimes these operations will be combined with other activities, for example routine maintenance or rehabilitation. This two-stage leak survey methodology has significant room for improvement. Many of the processes are inconvenient, expensive, time-consuming and unreliable [4].

This paper presents a solution to this challenge by providing technology in which low-cost sensors and appropriate signal processing by Artificial Intelligence techniques are the major distinguishing features. The concept of a "failure" sensor is introduced, in which the output is not necessarily proportional to some hydraulic parameter, but is unmistakably affected when an unusual event occurs. Thus the basis for a pipeline monitoring system has been established: a spatial array of time series data provided by a low-cost sensor installation. Interpretation of the low-grade signal data in the spatial array can then be achieved by the use of an ANN system in a pattern recognition mode. The paper describes the sensor design, development & deployment, system verification, associated simulation analysis and an ANN approach for burst detection, ending up with conclusions.

World Academy of Science, Engineering and Technology
International Journal of Physical and Mathematical Sciences
Vol:2, No:9, 2008

II. SENSOR DESIGN, DEVELOPMENT AND DEPLOYMENT

The basic working principle of the low cost sensor which has been researched and developed is that the local opacity or cloudiness of a water flow can be affected by a change in the flow regime such as turbulence caused by flow velocity change, including reversal, aeration of the water flow, or even colour changes as two different water streams combine. An opacity failure sensor was therefore designed to project a narrow beam of infra red (IR) light into the water flow and then measure the intensity of the light which is scattered sideways by the water flow. The design of this sensor has been previously reported [5].

A dual channel data logger (Lascar Electronics Ltd., EL-3-12bit), having a storage capacity of 8000 readings per channel in its on-board memory, was selected. A laptop computer was required for the configuration of the channels and the downloading of the recorded data.

III. NEURAL NETWORK ANALYSIS

The totality of a water distribution network as a distributed, non-linear dynamical system may not be effectively or satisfactorily described using purely deterministic and linear methods. Techniques such as neural networks, which are capable of performing non-linear discriminant analysis, appear more suitable for categorising time series data from such sources. The opacity sensor output was one useful type of time series data which could be analysed in this way (other relevant variables include flow and pressure), and an ANN was implemented using these data for abnormality detection within a sub-zone. This implementation demonstrated that an automated online system utilising ANNs could, in principle, be capable of monitoring a large number of sensors for abnormal signal detection.

A. Analysis and Design

Initial analysis of opacity sensor time series (for the multi-prototype installation) and actual abnormal flows in the area under examination revealed abnormal peaks in the opacity output which appeared to correlate with known events such as bursts and flushing. Detailed analysis of this correlation indicated several characteristics of the response:

- An initial abnormally high peak in opacity shortly after the commencement of the burst was followed by a gradual drop over a short duration to normal levels. Approximate durations for initial peaks were in the range of 20 minutes to several hours – hence a sensor sampling rate of 5 minutes was considered to be sufficient to capture the events.
- The size of the peak and the lag between the event start and the opacity response was correlated to the proximity of the sensor to the event. Simulated burst trials subsequently showed that the response to in-zone events was usually limited to sensors in the locality of the event.

- Major events were often found to lead to a persistent change in opacity level over a long time period (even after the event was terminated). Smaller events resulted in a 'settling' back to normal levels of opacity even while the burst or flushing persists.
- The installation site of the sensor (location within the pipe network, size/material of pipe and duration since the last cleaning) was a critical factor for performance.

Knowledge of the causal relationship between an event and the opacity response was not available (no water distribution network uses opacity simulation tool), nor was sufficient training data for known events collected in order to adopt a comprehensive pattern recognition approach. However, it was empirically apparent, from both the monitoring of everyday operation and the simulated burst trials, that the opacity level could produce supplementary information to the normal hydraulic (flow and pressure) data for leak location. Hence, these factors led to the adoption of a time series prediction approach, in conjunction with a classification module for monitoring the opacity data. The set of outputs could then be analysed to supplement sub-zone location, and this was the approach adopted in the work presented here.
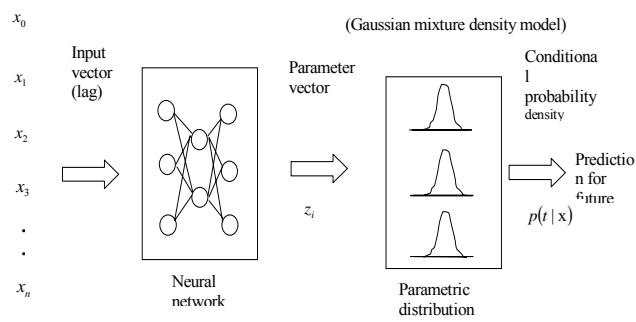

Fig. 1 MDN architecture for time series prediction

B. Algorithms

A Mixture Density Neural Network (MDNN) [6] was applied in a manner similar to that used previously for flow data [7] with little modification. In summary, precise point predictions were not required for the application, instead, a distribution of likely values was more appropriate. Hence, the core of the system was a Mixture Density Neural Network used as a time series predictor. The inputs were lagged normalized raw values (processed for rogue and missing values) and the output was a mixture model [8] of Gaussian distributions for the prediction some time step in the future (one day in this case). A 'most likely' value along with variance for the value was calculated from the mixture model. These data, along with the actual observed value, were processed by the abnormality detection module.

A traditional feed-forward ANN such as a Multi-Layer Perceptron (MLP) [9] could be used for time series forecasting by employing a sliding window over the input

World Academy of Science, Engineering and Technology
International Journal of Physical and Mathematical Sciences
Vol:2, No:9, 2008

sequence. The aim was to forecast future developments of the time series by finding a function $f : \Re \rightarrow \Re$ such that:

$$t(\tau + d) = f(x(\tau), x(\tau - 1), ..., x(\tau - N - 1)) \qquad (1)$$

In equation (1) τ represents current time, and d is the number of time steps ahead to predict (so that $d = 1$ is next step ahead prediction). The standard static ANN approach to perform this prediction is to induce the function f as an MLP or Radial Basis Function architecture, using a set of N-tuples as inputs, and a single output as the target value of the network (sometimes referred to as the sliding window technique in which the input layer represents a moving window through the time series). Fig. 1 shows the basic architecture of the MDN applied to time series prediction in this fashion.

A time series prediction on its own will not provide a classification. A classification module was therefore developed to detect discrepancies between actual and predicted values, over some time window. A value for the prediction (either centre of the highest component or an average across distributions) along with variance for the value was calculated from the mixture model. These data, along with the actual observed value, was then processed by the burst detection module. The module operated by analyzing the actual observed value in the context of the predicted value and a user definable error sensitivity level. If the observed value was outside the defined threshold for a windowed period, then the state passed from normal to abnormal (0 to ±1). In the case of opacity data, the signal type required a short period, high threshold filter – based on analysis of historic data the threshold was set at six standard deviations and a time window of 20 minutes.

C. Training

Seven sensors from the trial installation of 10 were selected for monitoring. Unlike previous work applied to flow data, the characteristic of the opacity data over a historic time period of weeks/ months was found to be far from stationary. The evolving dynamics of the system produced frequent step changes in the opacity level as well as in some cases an adapting diurnal cycle (some signals exhibited no diurnal cycle).

The MDN can be extended to non-stationary problems, provided that the model is treated as continuously adaptive. In other words, for non-stationary data, the model must be re-estimated within a relatively short time interval. Therefore, in operational conditions, the ANNs would require very regular retraining to capture the current state of the sensor output. To simulate this, the ANN module for each sensor was trained with just over one week's data immediately preceding the simulated burst trials. No major abnormalities were present during this interval in the majority of cases thus providing training data representing normal operating conditions. Data

collected for seven days including the simulated burst trials formed the unseen test data set.

The MDN was trained on the pre-processed data set, in order to learn a one-day ahead time series prediction (mixture model distribution of the prediction). A network was trained for 100 cycles on the training set, a relatively low figure to attempt to prevent over-fitting on the small training set. Three hidden units and two Gaussians proved sufficient for the problem. The test sets were then presented and the classification module applied.
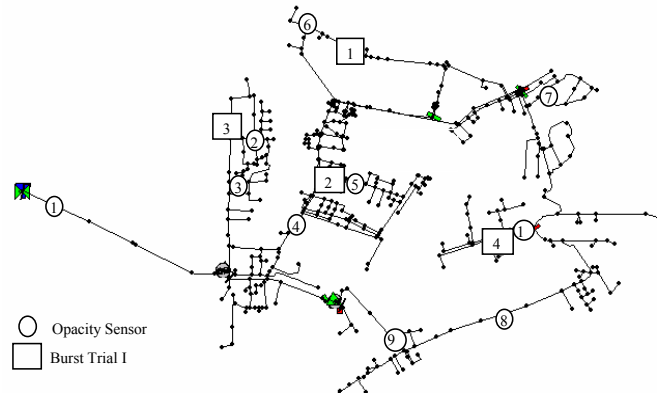

Fig. 2 Sensor locations and simulated burst sites in the DMA

IV. SYSTEM VERIFICATION BY EXPERIMENTAL BURST TRIALS

In order to test out the ANN system for event detection and location, experimental burst trials were carried out in the DMA under investigation. From the pressure variation and simulation viewpoints, the DMA could be divided into three sub-zones, viz;
(i) close to the reservoir,
(ii) central portion,
(iii) tail end of the distribution network.

Two opacity failure sensors (1 & 4) were located close to the reservoir, five (2, 3, 5, 6 & 7) were located in the central portion and three (8, 9 & 10) were located in the tail end (see Fig. 2).

A series of two burst trials was planned with the co-operation of the Water Company. The results of the simulations were used to help choose the best locations for the simulated bursts, though only sites that would not cause obstructions to the public could be used. The sites chosen for both simulated burst trials are shown in Fig. 2

The Water Company specified the maximum flow rates that were permissible at each site for the simulated bursts, which were created by fitting a standpipe to a fire hydrant and then slowly opening the valve, with an in-line flow meter connected, until the desired flow rate was reached. It was essential that there was minimum risk of damage to any of the pipes, or discoloration of water in the pipe due, for example, to sediment agitation. Once the required flow rate was achieved the flow meter was removed. These flow rates were relatively low, so the burst trials were carried out during the

World Academy of Science, Engineering and Technology
International Journal of Physical and Mathematical Sciences
Vol:2, No:9, 2008

night when water usage and therefore flow rates in the pipes was low.

Each burst trial was carried out on consecutive nights so that any measured response could be attributed to a particular event. The logging intervals of each failure sensor was set to 1 minute and the loggers carefully synchronized so that the phase difference between the opacity changes at the ten sensors could be identified.

A. Trial I Procedure and Results

The first three bursts (1st at 5.5 l/s, 2nd at 5 l/s & 3rd at 6 l/s) were simulated in the central portion of the distribution network, while the last burst (4th at 3 l/s) was conducted at the tail end of the network (see Fig. 2). Fig. 3—5 show the data for the first 3 individual burst simulations. Fig. 3 shows the responses at all ten sensors to the simulated Burst 1. Sensors 6, 9 and 10 appeared to respond to the simulated burst along with Sensor 8. However the trace for Sensor 8 showed an increase in opacity prior to the valve being opened, while the rise in opacity at Sensor 6 was only around 5%, and was slow, taking over one hour to reach peak opacity.
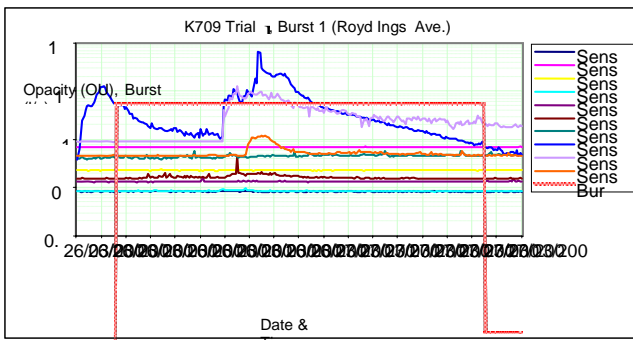

Fig. 3 Recorded Opacity Data for Burst Trial I - Burst 1

Fig. 4 shows the opacity response from all ten sensors to simulated Burst 2. Sensor 5 and Sensors 8—10 showed a response, while sensors 8 and 9 appeared to respond to an event that occurred before the "burst" was initiated.
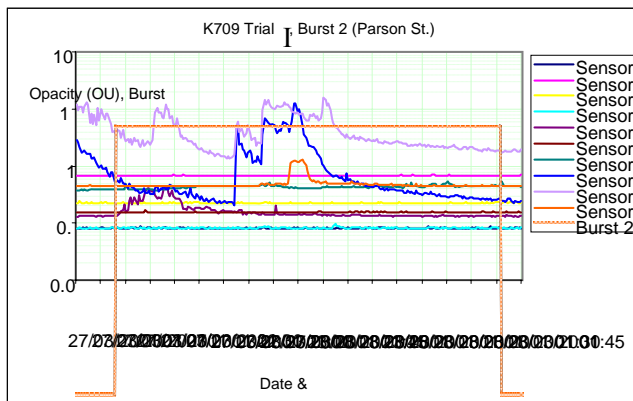

Fig. 4 Recorded Opacity Data for Burst Trial I –Burst 2
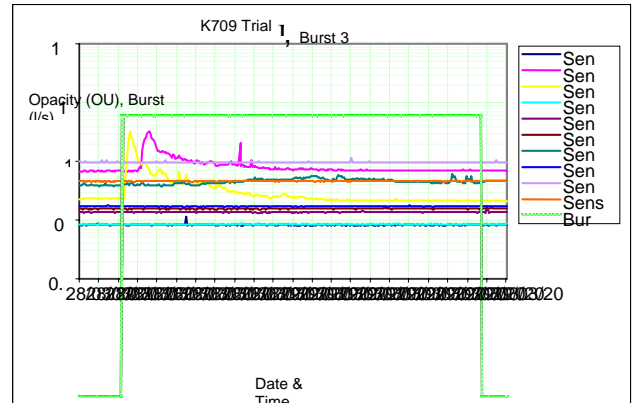

Fig. 5 Recorded Opacity Data for Burst Trial I –Burst 3

Fig. 5 shows the response at the ten sensors to a simulated Burst 3. Responses were only seen at Sensors 2 and 3 and both showed a rapid rise in opacity of the water in the pipelines followed by a more gradual fall once a peak in opacity was reached. Sensor 3 responded almost immediately to the burst whereas fifteen minutes elapsed before Sensor 2 responded. None of the ten sensors appeared to respond to burst 4.

B. Trial II Procedure and Results

It was found that whilst the Trial I experiments were being carried out in one section of the network, cleaning and relining of pipes was unknowingly being carried out in another part (the 2 sections were said to be isolated, but on investigation after the Trial I results had been analysed, it was found that this was not so. So, a further series of burst trials was carried out at almost the same locations where the first trail was conducted. Sensor 1 response was invalid as it showed both positive and negative opacity values of very high magnitude. This type of response was not evident in any other data sets, and investigation showed a fault to have developed in the data logger.
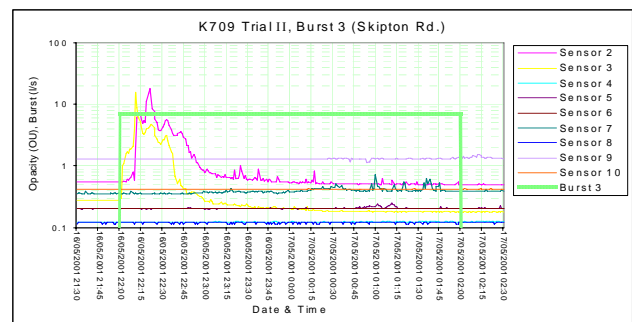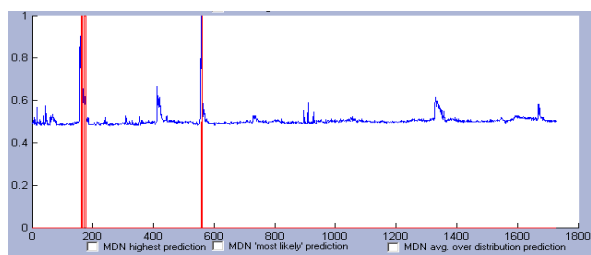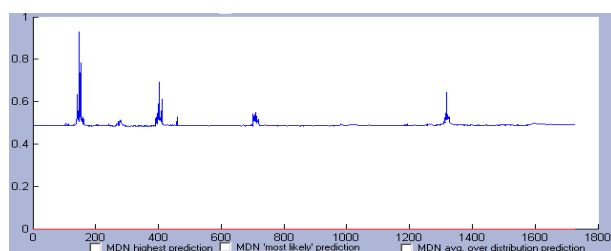

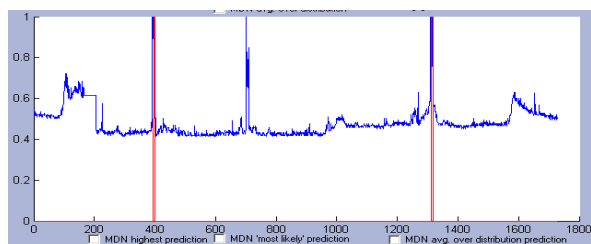Fig. 6 Recorded opacity data for Burst Trial II – Burst 3

Fig. 6 shows the plot for the simulated burst 3 which is very similar to Fig. 5, the plot for the simulated burst 3 recorded during Trial I. Once again, Sensors 2 and 3 showed a significant response to the simulated burst, with Sensor 3
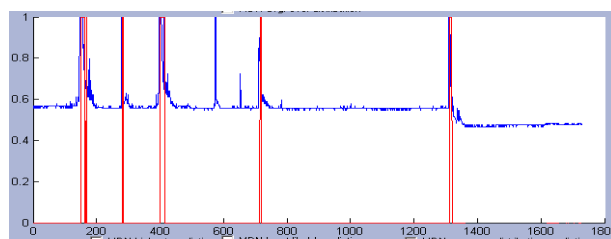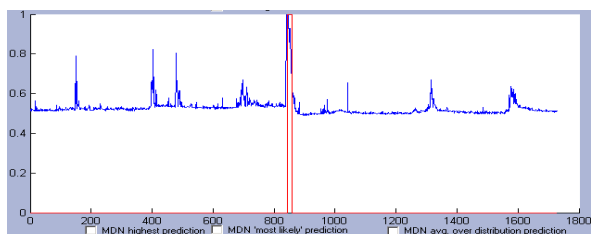
World Academy of Science, Engineering and Technology
International Journal of Physical and Mathematical Sciences
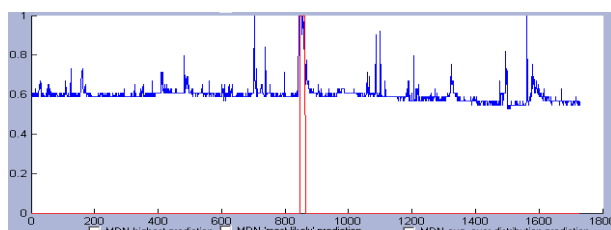Vol:2, No:9, 2008

Sensor 5

Sensor 6

Sensor 4
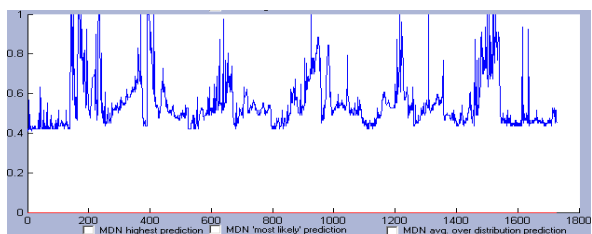

Sensor 10

Sensor 3

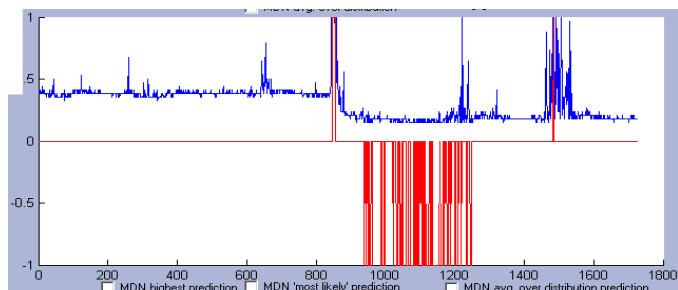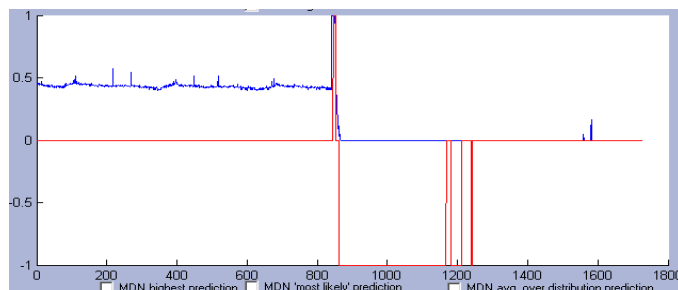Sensor 2

Sensor 7

Fig. 7 Classification plots for Trial I

Sensor 3

Sensor 2

Fig. 8 Classification plots for Trial II

World Academy of Science, Engineering and Technology
International Journal of Physical and Mathematical Sciences
Vol:2, No:9, 2008

responding almost immediately and Sensor 2 responding ten minutes later. There was possibly a small response at Sensor 10. None of the sensors responded significantly to any other burst in the trial.

V. ANN ANALYSIS AND RESULTS

Results from the interactive MATLAB MDNGUI system for Trial I are shown in Fig. 7. The classification is superimposed on the original graph, covering one week's data.

The system also produced an output file containing a summary of the classification, as described in Table I. The same classification process was also applied to the Trial II data. In this case, only two of the burst sites corresponded to those in the first trials (Burst 1 and 3 from Trial I). The only abnormality detected in response to any of the four bursts were for sensors 2 and 3, for the simulated burst corresponding to Burst 3 in Trial II. Fig. 8 illustrates the output.

TABLE I
CLASSIFICATION MODULE OUTPUT

Sensor	Abnormality detected (1st window)	Corresponding event
No.5	26/3/01 13:25-13:45	Unknown
	27/3/01 22:30-22:50	Burst 2 – flush commenced 22:01 27/3/01
No.6	None	
No.4	27/3/01 08:55-09:15	Unknown
	30/3/01 13:20-13:40	Unknown
No.10	5 events 26th-30th March 2001	Expected to be related to relining work performed during this period, when valve operations were carried out with the Gresley PRV section.
No.3	28/3/01 22:20-22:40	Burst 3 – flush commenced 21:50 28/3/01
No.3	28/3/01 22:35-22:55	Burst 3 – flush commenced 21:50 28/3/01
No.7	None	

Like the previous burst in this locale, the signal was easily classified by the system (Table II). However, in this case, the opacity level had subsequently undergone a step drop due to the flushing which effectively cleaned the pipes in the sub-zone. Consequently, the system detected an abnormal drop in opacity for the period after the burst and would require retraining with data after the flushing to update the MDN.

TABLE II
CLASSIFICATION MODULE OUTPUT FOR SENSORS 2 AND 3

Sensor	Abnormality detected (1st window)	Corresponding event
No.2	16/5/01 22:40-23:00	Burst 3 – flush commenced 22:00 16/5/01
	17/5/01 00:45-01:05 Negative	Cleaning effect of flushing
No.3	16/5/01 22:25-22:45	Burst 3 – flush commenced 22:00 16/5/01
	16/5/01 23:45-00:05 Negative	Cleaning effect of flushing

VI. CONCLUSION

The low cost failure sensor has been developed to provide input to an ANN. The opacity failure sensor was designed to measure the opacity of water flow in a pipeline which has given repeatable results in the laboratory and in situ in a water distribution pipeline network [5].

A Neural Network monitoring system based on a time series prediction and classification for abnormality detection has been successfully developed. The system has been applied to the two sets of burst trials and has demonstrated that the opacity response in some sensors could provide additional information on burst location to that supplied by hydraulic sensors.

The field results obtained from the multi-prototype sensors are encouraging for the monitoring of abnormalities (burst, leakage and flushing) in a DMA. It is concluded that a low-cost sensor technology of this type can contribute to a system for monitoring and leak detection in water distribution pipelines with appropriate computer-based interpretation, e.g. ANN, which is potentially a most valuable asset management tool.

Data collection and transfer remain major issues. Technologies for data collection from low cost sensors were not within the remit of this research programme. However, the importance of an efficient method for data collection is understood, for which the rapidly developing technologies associated with mobile communications are of great interest. This must represent a subsequent stage of the research. There will be the possibility of using a PSTN (Public Switched Telephone Network) link as a data download solution.

REFERENCES

[1] O. Hunaidi, W. Chu, A. Wang, and W. Guan, "Leak detection methods for plastic water distribution pipes, Advancing the Science of Water", *Fort Lauderdale Technology Transfer Conference, AWWA Research Foundation,* 1999, 249-270, Florida.

[2] M. Martin and L. Farley, "Improving water management in leakage control", *Proceedings of Hydrotop,* 1994, Marseille, France.

[3] J. Grolby and T. Woodward, "Find that leak", *IEE Review,* September 1999.

[4] G. Bridges and M. MacDonald, "Leakage control—the neglected solution?", *20th WEDC Conference,* Colombo, 1994, Sri Lanka.

World Academy of Science, Engineering and Technology
International Journal of Physical and Mathematical Sciences
Vol:2, No:9, 2008

[5]   A. Khan, P. D. Widdop, A. J. Day, A. S. Wood, S. R. Mounce, and J. Machell, "Low-cost failure sensor systems for water pipeline distribution systems", *Water Science and Technology,* 2002, Vol. 45 (4/5) 207-216.

[6]   C. M. Bishop, "Mixture density networks", *Technical report NCRG/94/004*, Department of Computer Science and Applied Mathematics, Aston University, Birmingham, 1994, UK.

[7]   S. R. Mounce, A. J. Day, A. S. Wood, A. Khan, P. D. Widdop, J. Machell, "A neural network approach to burst detection", *Water Science and Technology,* 2002, Vol. 45 (4/5) 237-246.

[8]   G. J. McLachlan and K. E. Basford, "Mixture models: Inference and Applications to Clustering", *New York: Marcel Dekker,* 1988.

[9]   D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning representations of back-propagation errors", *Nature (London),* 1986, Vol. 323, pp. 533-536.