A neural network approach to burst detection

S.R. Mounce*, A.J. Day, A.S. Wood**, A. Khan**, P.D. Widdop** and J. Machell*****

* Department of Computing, Phoenix Building, University of Bradford, Bradford BD7 1DP, UK

** Department of Mechanical & Medical Engineering, University of Bradford, BD7 1DP, UK

*** Yorkshire Water, ROCC1, Western Way, Halifax Road, Bradford BD6 2LZ, UK. (E-mail: *www.laps.ac.uk*)

Abstract This paper describes how hydraulic and water quality data from a distribution network may be used to provide a more efficient leakage management capability for the water industry. The research presented concerns the application of artificial neural networks to the issue of detection and location of leakage in treated water distribution systems. An architecture for an Artificial Neural Network (ANN) based system is outlined. The neural network uses time series data produced by sensors to directly construct an empirical model for predication and classification of leaks. Results are presented using data from an experimental site in Yorkshire Water's Keighley distribution system.

Keywords Bursts; detection; leakage; neural networks; online monitoring; water distribution networks

Introduction

Problem definition

The issue of leakage management is of serious consideration to water companies because major water loss can have negative economic, environmental and political consequences. In the past, a relatively high level of leakage has been typical as the loss of treated water was considered by water utilities on a solely economic basis and the significant reduction of leakage level was not recognised as being economically viable. It has been estimated that the ten major UK water companies lose on average 30 per cent of their treated water through leaks (Butler and West, 1987). However, in the UK the Office of Water Services (OFWAT) has in recent years imposed company specific mandatory leakage targets which are currently reviewed annually. The water companies additionally now have a statutory duty to conserve water and publish details of leakage reduction performance. Consequently, the requirement for reduction in leakage levels has led to an interest in technologies for pipeline-integrity monitoring and failure detection. Current leakage management methodologies within the water industry tends to be predominantly a manual process which is resource intensive. For example data logging, data collection and transfer to point of use, data analysis, reporting and then mobilisation of leakage teams (one or more leakage teams may require several days to locate the leak). Subsequently a contractor must be employed to effect repair. This overall process may take weeks or in extreme cases months. Assuming an automated system performed daily data analysis, a realistic target for significant burst lifetime would be a reduction to 2 days or less. Ideally, a state of the art detection system should be able to:

- detect and localise bursts within an entire distribution system
- differentiate between bursts and unusual demands or system events
- differentiate between bursts which occur instantly and those which develop gradually (leaks).

District flow metering is the most common method of leakage control in England and Wales. Martin and Farley (1994) commented that in recent times, due to the opportunities afforded by computer technology, continual monitoring of leakage levels through district

Water Science and Technology **Vol 45 No 4–5 pp 237–246** © IWA Publishing 2002

metering had become commonplace. The principle of district metering is based on the subdivision of the distribution system into discrete zones or district meter areas (DMA), by the permanent closure of valves, and the measurement of the flows into each zone. A DMA will generally comprise an area containing 1,000 properties. The flow will usually be measured by a fixed flow meter at the inflow point although temporary insertion meters and data loggers can be used instead. In general, flow sensors log the flow every fifteen minutes and assumptions are made about the average usage. The most widespread operational use of this flow data is in analysis of measured minimum night flows (NFM). The minimum night flow is the lowest flow supplied to a hydraulically isolated supply zone. It is usually measured during the night hours between midnight and 5:00 am and includes leakage as well as a certain minimum night consumption. Night flows are used because water use is at a minimum and it is easier to identify and subtract the legitimate flows (e.g. 24 hr process loads). If the night flow minus the legitimate flow is close to zero, the leakage must also be close to zero. In contrast, unusual jumps in volumes will signify leakage in the absence of any other factors. It is common to identify a nominal target nightline for each DMA. This target implies a certain level of leakage activity with regard to the find and fix of hidden leaks. If it is established that leakage has increased sufficiently to warrant further investigation, the next stage involves a manual leakage detection which is carried out using methods such as sounding, leak noise correlation and step testing.

This paper describes research into an automated computing system, based on artificial neural network and sensor technology, for the detection and location of leakage in a water pipeline distribution network. The methodology involves application of neural network(s) to time series data generated by flow, pressure and other sensors (such as failure sensors developed for the purpose). The target of the research is to automate district meter analysis as well as allowing further localisation of the burst within the DMA prior to manual detection by leakage teams.

The complexity of a model for a water pipeline distribution network for a large urban area is such that no single neural network can effectively monitor the behaviour of the entire pipeline distribution network. Typically the distribution networks comprise several DMAs with each zone containing one or more main pipelines, each of which feeds into one or more sub-zones and finally into individual industrial and domestic consumers. Other features such as a reservoir or pumping station may also be present. To manage such complexity, it is necessary to build a number of neural networks arranged in a parallel and hierarchical fashion to allow different types of information processing on the same data or different subsets of data and at different levels of abstractions. This approach enables the decomposition of a complex problem into a number of sub-problems. Each sub-problem is then assigned to a functionally suitable network. At the top global level, level 1 (DMA level), a neural network monitors zones and can detect leaks to within a zone (e.g. DMA) but is unable to specify the exact position. Previous work has demonstrated the feasibility of detection at this level using hydraulic data (Torsun *et al.* 1999). In the next level of abstraction, level 2 (sub-zone level), a neural network will monitor sub-zones and can locate leaks to within a sub-zone. Finally, at level 3 (pipe level), a neural network will monitor one or more pipes and is able to accurately locate the position of the leak. Recent work has seen attempts to link hydraulic models with on-line data from telemetry systems (Skipworth *et al.*, 1999; Orr *et al.*, 1999). This type of online model could provide the pipe level input. A very precise identification of leak location would rely on input to a neural network of signals from mobile acoustic sensor instrumentation.

Initial work has focused on hydraulic data (from existing flow and pressure sensors) as the principal data from which to provide features for leakage detection. Consultation with domain experts, as well as analysis of historical data, had revealed hydraulic variables to be good indicators of abnormal flow in the pipeline distribution system (including bursts and leaks) at the DMA level. The accuracy of leak detection and location can be improved by analysing signals from a greater number of sensors in a DMA, but the installation, operation and analysis resource costs are generally seen as prohibitive. Pressure sensors may be deployed temporarily within a zone during field tests in order to provide data for hydraulic model calibration. It has been proposed that if a burst occurs the variation in pressure signal in association with the flow increase will enable more accurate location of leak (Machell, 1997). Research conducted at Bradford University for an EPSRC research project (as part of the WITE programme) in association with Yorkshire Water is investigating the feasibility of using a number of failure sensors within a zone to improve leak localisation. Consequently, we are now in a position to consider opacity (referring to the degree to which a substance is opaque) of the water flow as an additional data source because low cost failure sensors have been designed and installed as described in an associated paper ("*Low-cost failure sensor design and development for water pipeline distribution systems*"). These sensors are providing data which can be correlated with actual events, with the ultimate goal of leak localisation.

Experimental site

Yorkshire Water's Keighley distribution system provides water services for around 26,000 customers. It contains approximately 110 km of pipeline and seven service reservoirs. The distribution network has more recently been partitioned into 15 DMAs. Six of these cover the city centre and immediate environs and one (K709) has been selected as the test bed for a leakage detection case study. As well as state of the art hydraulic sensors (flow and pressure are both logged at all DMA level sites in Keighley) the facilities were available to install failure sensors designed and manufactured at Bradford University. The zone is effectively disconnected from neighbouring zones resulting in an isolated environment for tests – the area is supplied solely by Highfield SR. Figure 1 illustrates Zone K709 (from the hydraulic model) showing the pipe geometry and sensor placement. Table 1 summarises the sensors employed for the zone.

Methods

Data acquisition

Hydraulic data for the case study network was made available by Yorkshire Water. During the initial research period this had to be via manual download from company databases but later the opportunity to download directly using remote telemetry access was provided. The data consists of 15 minute values for flow and pressure. Data collection for the opacity failure sensors was via manual download on site (with data loggers providing capacity for 25 days data at reading intervals of five minutes). In either case, time stamped files of readings form the basis for further analysis.

Data pre-processing

Once data is acquired it is then passed through a number of filters (written in Matlab or JAVA) as follows.

Table 1 Sensors in experimental site

Sensor type	Number	Logging interval	Location
Flow	3	15 minutes	Highfield SR, Albert St. PRV, Gresley Rd. PRV.
Pressure	3	15 minutes	As above.
Opacity failure sensors	10	5 minutes	See Figure 1.

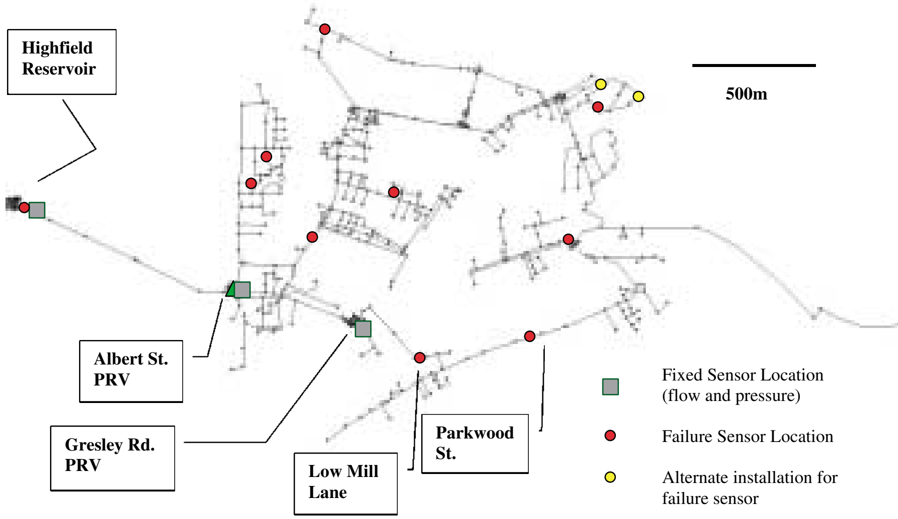

Figure 1 Zone K709 model with sensor locations

Data correction. First a filter removes any rogue values. Then, to ensure a continuous data stream an ARIMA based filter fills in any periods of missing data. These missing entries can be due to sensor failure or temporary power outages for example.

Normalisation. The input is normalised by means of linear re-scaling with mean and standard deviation (Z-score) to a range of 0 to 1.

Reformatting. Finally, we reformat the input stream into a tapped delay line format in order to prepare for neural network presentation.

ANN architecture and algorithms

For DMA level detection, a particular sensor (typically at the zone inlet) is modelled by one neural network which learns to forecast the sensor output (time series prediction). A number of architectures were investigated for forecasting, including static and recurrent networks (Elman and Jordan networks) in a conventional manner. However, review of the literature concerning Takens' embedding theorem (Takens 1981) applied to forecasting resulted in an alternative approach. Of course, the main result of Takens is that we can predict future values of a time series $x_{t+\tau}$ solely on the basis of the observed past data $\{x_t, x_{t-\tau}, \ldots\}$ without explicitly knowing the actual dynamics in the true system. However, when we have a stochastic time series the best we can do is predict the conditional probability $P(x_{t+1} | x_t)$ which is the probability that we observe x_{t+1} at the next time step t+1 given that we have observed the recent time series $x_t = (x_t, x_{t-1}, \ldots, x_{t-m+1})$. Consequently neural networks have been applied to the task of the prediction of the entire conditional probability distribution of an unknown data generating process. These techniques have been applied to time series prediction problems (Husmeier, 1999). A neural network model called the mixture density network (MDN) as presented by Bishop (1994) has been selected for development work. Essentially, the MDN is a mixture density model (semi-parametric model) combined with a neural network. It consists of a two-hidden layer network – the first layer with sigmoidal units and the second with Gaussian units. The target vector is clamped directly to the Gaussian nodes. The network is trained using maximum likelihood (minimising the negative log of the likelihood). The mechanism for adapting the network parameters is still backpropagation of error: the only difference being the error function. The various parameters of the mixture models (mixing coefficients $P(j)$, means μ_j and variances σ_j) are governed by the outputs of the neural network which takes $\mathbf{x}$ as its input. Once the network has been trained it can predict the conditional density function of the target data for any given value of the input vector i.e. we have a complete description of the generator of the data, as far as predicting the value of the target vector. A GUI has been developed for the MDN code developed using the NETLAB library (MATLAB toolbox). The prediction generated by the MDN network is then used with the actual observed data to form a comparative which is analysed by a classification module which indicates abnormalities in its output (based on a modifiable error sensitivity level). The values from the classification stage from the various zones are analysed by a rules based system which indicates overall system state. Figure 2 illustrates the sensor modelling sub-system.

Training

Algorithm. In the NETLAB library the mixture model used is Gaussian with a single covariance parameter for each component. The mixture coefficients are computed from a group of softmax outputs, the centres are equal to a group of linear outputs, and the variances are obtained by applying the exponential function to a third group of outputs. The network is trained with a scaled conjugate gradient optimiser with the error function being the negative log likelihood of the training data:

$$E = -\ln L = -\sum_{n=1}^{N} \ln p\left(x^n\right) = -\sum_{n=1}^{N} \ln\left\{\sum_{j=1}^{M} p\left(x^n \mid j\right) P\left(j\right)\right\}$$

$$\text{where } p\left(x \mid j\right) = \frac{1}{\left(2\pi\sigma^2\right)^{d/2}} \exp\left\{-\frac{\| x - \mu_j \|^2}{2\sigma_j^2}\right\}$$

The input is a lag vector and the target is a one day ahead prediction. The prediction evaluated can either be the most "likely value" (centre of the highest component) or we can take an average across distributions.

Training data. The MDN network is trained with a period of historical data for the sensor in order to learn the distribution. This data set should be significant (e.g. at least several weeks, preferably months) and should show reasonable well behaved behaviour. Every so often (either automatically or at the behest of a user) the network is re-trained with a slightly updated data set. In this way, current conditions are continually built into the model. If

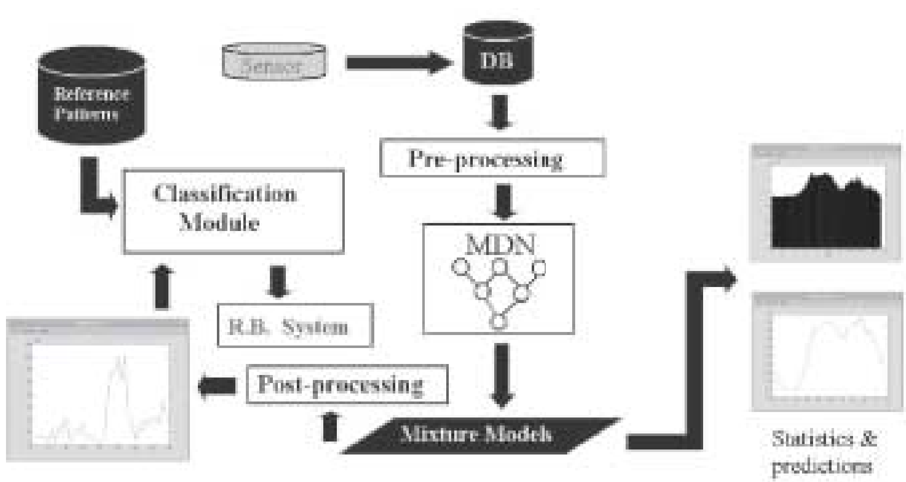

Figure 2 Sensor modelling sub-system for DMA monitoring

conditions for the area being monitored change drastically (e.g. system alterations such as valve modifications) then the training set will need to be started anew. In the example we will consider the Albert St. PRV flow meter. This is the critical flow meter for monitoring Zone K709. Figure 3 shows the training set used in the example covering a period of 10 months. Missing data has been filled (missing data segments appear dark). As can be seen, the minimum night flow is fairly constantly just below 10 l/s for this period. First, the file was normalised and reformatted as required for a tapped delay input. This data set was then used to train a dedicated MDN. Experimentation revealed that a suitable number of hidden units was 8 and a set of three Gaussians was sufficient. The main effect of variation of hidden units was convergence time and increasing the number of Gaussians above 3 had negligible effect on performance. There was little appreciable reduction of the error term after around 100 complete cycles of the training set (containing 26,208 presentations). The one day ahead prediction for the seen data was satisfactory at this stage. To test the performance a test set of data was used for a period of approximately one month (26th December 2000 to 30th January 2001). Results for this test file and supplemental sensor data will be described in the next section.

Results and discussion

The test file was presented to the MDN network. Essentially, the output consists of a file of one day ahead prediction for a moving one day window. The comparative formed by use of the actual observed reading one day later should indicate the significant in-front of a burst. Figures 4–9 illustrate the test data. Figure 4 illustrates the MDN one day ahead unseen prediction output for a number of days in the test set for a well behaved period with no abnormal events (hence the prediction is very close to the actual observed data). In Figures 5 and 6 the output is illustrated for burst events which occurred in the zone. As can be seen,

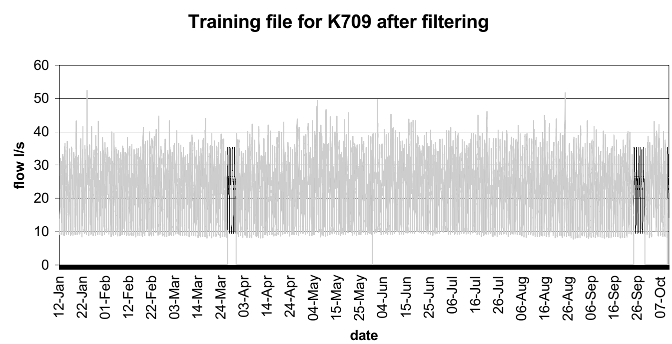

Figure 3 Training data for K709 Albert St. flow meter

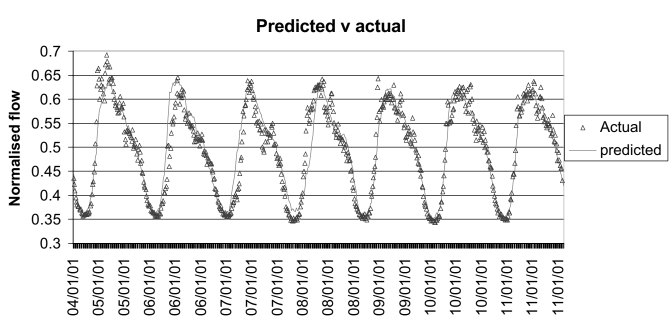

Figure 4 Predicted v actual unseen data for standard conditions

a discrepancy immediately results (due to the one day ahead nature of the prediction) since the abnormally high flow does not tally with the predicted level. Figure 7 shows the calculated comparative ($Q_{ONLINE}/Q_{OFFLINE}$) for the whole test period. These signals form the input to a general classification module which analyses the signal and outputs a class, for example burst/ no burst (Figure 8). This module has been trained on a selection of comparatives from various meters. Both the two major events were classified as significant bursts commencing 31st December and 22nd January respectively. Table 2 lists the recorded repairs of mains conducted in the zone in January.

During the test period analysed, data from opacity failure sensors was also logged at a number of sites around the zone (see Figure 1). Due to technical problems in several of the installations, a full data set was not available for the whole of January. However, significant

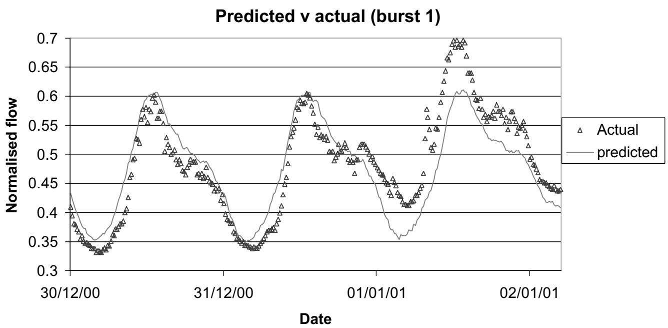

Figure 5 Burst occurs 31st December

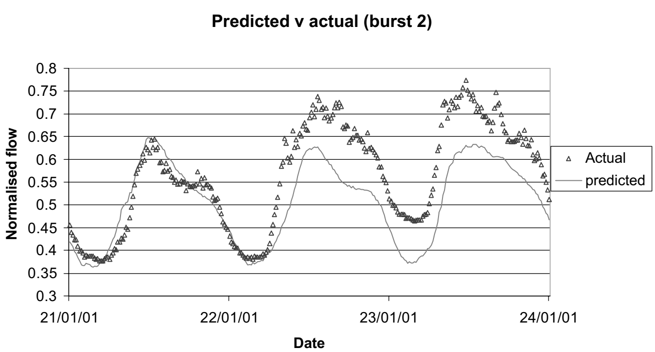

Figure 6 Burst occurs 22nd January

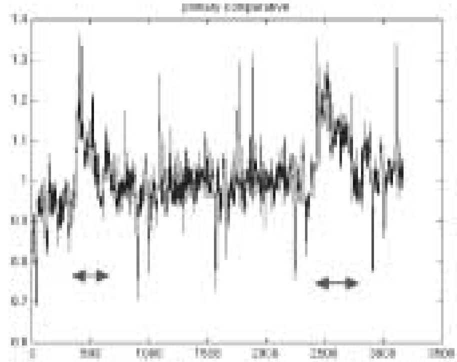

Figure 7 Comparative used for test period

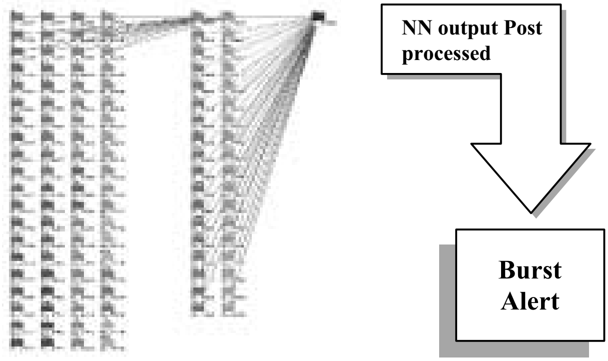

Figure 8 Classification module decision

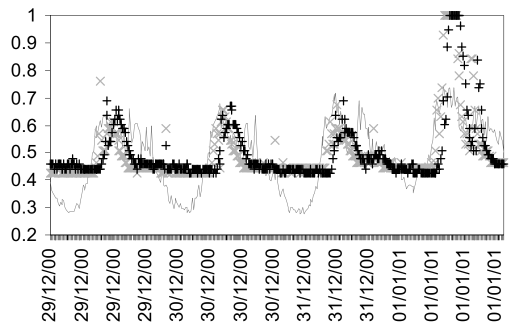

Figure 9 Correlation for diurnal cycle between opacity and flow in Gresley Rd. sub-zone (normalised data from Figure 10 29/12–01/01)

Table 2 Repaired mains January 2001 in K709

	Burst 1	**Burst 2**	**Burst 3**
Type	Repaired main	Repaired main (dowel piece)	Repaired main
Emergency allocated	4th January	16th January	25th January
Task finished	7th January	16th January	29th January
Priority	3	1	1
Description	Leakage (visible)	Leakage (visible)	Leakage (not visible)

correlation between abnormal flows and the opacity was witnessed on several sensors during the period. Figures 9 and 10 illustrate the output for two of the failure sensors, along with the Gresley Road flow meter which monitors a sub-area in K709. The data has been sampled at a 15 min interval rate. The graph shows two interesting events. Firstly, the 31st December to 3rd January abnormal flow previously discovered is noticeable on the Gresley Road flow meter. Note that the opacity failure meters have a visible diurnal cycle which initially increases significantly in response to the velocity increase through the pipe. Even with 15 minute reading interval there is a noticeable phase difference for these two geographically close sensors. Secondly, observe the sudden flow surge on the 3rd of January and the corresponding spikes in the opacity data. The source of this event is unknown, though it is possibly industrial usage as this section of the zone has several industrial customers. Figure 9 shows the first 4 days data of Figure 10 which has been normalised to illustrate the high correlation for the diurnal cycle. Note that the opacity cycle is lagged behind

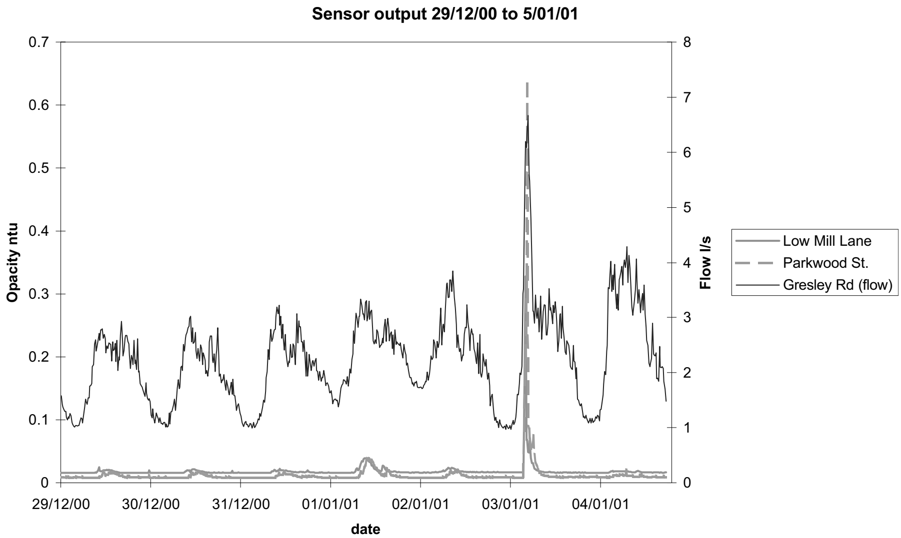

Figure 10 Correlation between opacity and flow in Gresley Rd. sub-zone

the flow. Also, observe the high opacity response the morning following the commencement of abnormal flow.

Conclusions

Many water utility companies are beginning to amass a great deal of data by means of remote sensing of flow, pressure and other attributes. In the past, this information has mainly been used for district level management including water audit and calibration of hydraulic models. However, the opportunity exists to build on existing analysis of minimum night flow for leakage detection using an automated neural network based pattern recognition system. In this paper, by means of an ongoing experimental field study, we have indicated how this may be accomplished at the district meter level. Further, we have demonstrated that additional in-zone failure sensors can provide a set of associated signals possessing magnitude and phase information which, with suitable quantities of example training data, will enable further understanding and localisation of events within a DMA. Future work will extend the system by constructing a neural network for analysis of failure sensor output. This will require an expanded data set: to be accomplished both by continuation of monitoring within the zone (along with discovery of system events) and by specific controlled trials in the area in which leak events are simulated. A multi-input TDNN is the expected architecture for this component of the system. Finally, analysis of an online hydraulic model's pipe level data will be attempted for a more refined location.

Acknowledgements

The authors wish to acknowledge the support given for this research by the Engineering and Physical Sciences Research Council (UK) under its Water Infrastructure and Treatment Engineering (WITE) Initiative. The authors also acknowledge the support of Yorkshire Water and are grateful to the directors of the company for permission to use the data and publish this paper. Finally all the authors wish to acknowledge the intellectual lead provided to this project by the late Professor I.S. Torsun.

References

Bishop, C.M. (1994). Mixture density networks. Technical report NCRG/94/004, Department of Computer Science and Applied Mathematics, Aston University, Birmingham, UK.

Butler, M. and West, J. (1987). Leakage prevention and system renewal, *Pipeline Management Seminar, Pipeline Industries Guild*, 1987.

Husmeier, D. (1999). *Neural Networks for Probability Estimation*, Springer-Verlag.

Khan, A., Widdop, P. D., Day, A. J., Wood, A. S., Mounce, S. R., and Machell, J. (2002). Low-cost failure sensor design and development for water pipeline distribution systems. *Wat. Sci. Tech.* **45**(4–5) 207–215 (this issue).

Machell, J. (1997). "Research and development proposal for an on-line leakage detection system", YWS internal report.

Martin, M. and Farley, L. (1994). Improving water management in leakage control. *Proc. of Hydrotop 94*, Marseille, France.

Orr, C., Bouulos, P., Stern, C. and Liu, P. (1999). Developing real-time models of water distribution systems. *Modeling and optimisation applications (Water engineering and management series, 3–4)* by D. Savic (Editor) and G. Walters (Editor). Vol. 1, Research Studies Pr.

Skipworth, P.J., Saul, A. and Machell, J. (1999). Practical applications of real time hydraulic modelling of water distribution networks. Comadem, 1999.

Takens, F. (1981). Detecting strange attractors in fluid turbulence. *Dynamical Systems and Turbulence*, Berlin, D. Rand and L. Young Eds. Springer.

Torsun, I.S., Day, A.J., Wood, A.S., Khan, A. and Mounce, S.R. (1999). Neural network for leak detection in pipeline distribution networks. Presented at the *1st Annual WITE Workshop, part of International Conference on Computing and Control for the water industry CCWI'99* held at University of Exeter, UK.