Water Science & Technology: Water Supply—WSTWS | 10.3 | 2010

Implementation of an on-line artificial intelligence district meter area flow meter data analysis system for abnormality detection: a case study

S. R. Mounce and J. B. Boxall

ABSTRACT

Faster detection of bursts saves water, minimises the inconvenience of interruption to customers and decreases the damaging consequences to infrastructure. Flow monitoring techniques are used by water service providers to monitor leakage, generally through offline application of mass balance type calculations and manual observations of change in night line values. This paper presents the combination of real-time data collection (using cello loggers with General Packet Radio Service communications) and a self-learning, online Artificial Intelligence system for detection of bursts at the District Meter Area level. The system components consist of communications software, a data warehouse and a MATLAB application. The online system continuously analysed a set of 146 DMAs in a case study area every hour generating automated alerts in response to abnormal flow. Specific examples are given, including a validation field test, and overall results are presented for a one year period. 36% of alerts were found to correspond to bursts confirmed by repair data or customer burst reports with only 18% ghosts. The results indicate that the software tool has the potential to reduce lost water and improve customer service hence enhancing water service provider's reputations.

Key words | artificial neural networks, data analysis, leakage, on-line monitoring, water distribution systems

S. R. Mounce (corresponding author)
J. B. Boxall
Pennine Water Group,
Department of Civil and Structural Engineering,
University of Sheffield,
S1 3JD,
UK
E-mail: *s.r.mounce@sheffield.ac.uk*;
j.b.boxall@sheffield.ac.uk

INTRODUCTION

If present trends continue, 1.8 billion people will be living in countries or regions with absolute water scarcity by 2025, and two thirds of the world population could be subject to water stress (UNEP 2007). A rapidly growing world population combined with a scarcity of water arising from localised drought is highlighting the need for optimal use of existing water supplies. The UKCIP02 Scientific Report on Climate change Scenarios for the United Kingdom predicted shortages in water supply, especially during summer (Hulme *et al.* 2002). Simultaneously, demand for water is increasing because of population growth, a decreasing average household size and growing use of water-intensive appliances. Hence, minimising the loss of treated water from water supply systems due to leakage is an ongoing issue for water service providers around the world. Yet in England and Wales, leakage as a percentage of total input has been stable at around 23% for over a decade.

Until recently, the state of the art for the management and derivation of meaningful information from water distribution system data has been limited, time consuming and of inadequate accuracy. This is primarily due to a reliance on human data analysis and interpretation, which is unfeasible and inefficient for the growing volume and complexity of data involved. Real time leakage estimation is now increasingly receiving more attention as water utility companies, especially in the UK, are set ever increasing targets for regulatory compliance and standards of service delivery. Real-time burst detection will be imperative if such

doi: 10.2166/ws.2010.697

targets are to be attained. The development of new leak detection equipment and techniques using microprocessors and wireless communications is now making it possible to find leaks more efficiently and reduce losses. It is expected that the next few years will see a rapid increase in the development and use of new technologies for detecting leaks in transmission mains including low-frequency hydrophones, signal analysis, in-pipe acoustic technology and ground penetrating radar (Farley & Hamilton 2008).

Water utilities generally first obtain awareness of bursts through customer reports, via call centres where the details of various service problems are collected such as complaints of low pressure, discolouration and possibly signs of visible surface water. Although large bursts in a water supply system are usually found and fixed quite rapidly due to multiple customer reports, not all bursts result in visible surface water and therefore can go undetected for several months, or longer.

A widely adopted standard to facilitate leakage management is the sub-division of water distribution systems into District Meter Areas (DMAs). A DMA will typically have flow and pressure measured at the inlet (and possibly outlet) of the hydraulically partitioned area typically containing approximately 200–2000 properties. Jankovic-Nisic *et al.* (2004) proposed a methodology for optimal positioning of flow meters and selection of a monitoring time step, and recommended that the size of the monitored DMA area should be smaller than it is typically in operational practice of water utilities, and further that it should be determined for every distribution network independently. The measurement interval for these meters is generally every fifteen minutes in the UK. In the past, and often still currently, the mechanism for detecting events has involved offline nightline analysis by water company personnel inspecting for unusual changes in water volumes can be detected. The night flow minimum (NFM) is the lowest flow supplied to a DMA, usually measured between midnight and 5:00 am. Night flows are used because water use is at a minimum and it is easier to identify and subtract the legitimate flows. If the night flow minus the legitimate flow is close to zero, the leakage must also be close to zero. On the other hand unusual jumps in volume will signify leakage in the absence of any other factors (such as industrial usage). Estimating distribution losses from NFM relies on the accurate estimation of the additional components that contribute to night flows (McKenzie & Seago 2005). The system operator will generally decide, based on experience with the particular water system, if the flow increase is due to a leak. Even then, although this technique can be effective, nightline analysis is not necessarily conducted routinely. It is generally part of a leak detection survey (e.g. Covas *et al.* 2008).

This paper describes the practical application of an online self-learning Artificial Intelligence (AI) system to a full-scale distribution network by conducting automated hourly analysis for a case study area in the UK consisting of 146 DMAs over a one year period. This AI system is not reliant on any special hardware or network configuration and produces intelligent 'smart alarms'. The system is designed to provide detection of bursts and leaks as they occur but not existing leaks or background leakage. Results are evaluated by comparing and correlating alerts with customer burst reports and subsequent main repairs. The research presented here was a part of the NEPTUNE Project, a £2.7 m UK research council and industrially sponsored project with seven academic partners and three industrial partners; two water utility companies and an automation company, ABB (2007–2010). The core deliverable will be an integrated risk-based Decision Support System (DSS) for evaluating intervention strategies to inform decision-making for sustainable water system operation. The flow analysis system described here is one component of this deliverable, providing alerts for further analysis and aggregation. Other software components include incident-likelihood, hydraulic, economic and customer impact models and a GIS visualisation environment (http://www.neptune.ac.uk).

METHOD

Algorithms

There has been recent interest in on-line monitoring of water distribution sensors and corresponding event detection. Event detection algorithms work in near real-time by obtaining SCADA data, performing some analysis and then returning a binary classification (i.e. sounding the alarm or not). The existing state of the art operational systems in the UK are implemented using flat line alarm levels on key monitoring sites in the control room and allow near real-time identification of large bursts.

Water distribution systems are demand driven, and the system does not know its demand (Olsson 2006). Pressure control allows the treated water pumping to match the output flow. A significant part of this output flow consists of leakage, and thus the challenge is to distinguish between leakage and true user demand–this is not at all straightforward. An approach that applies Artificial Neural Network (ANN) and Fuzzy Logic (FL) technology has been developed for automated online analysis of DMA flow data. An offline application has previously demonstrated how analysis of flow data can identify bursts (Mounce *et al.* 2002). This system was developed and applied to case studies of simulated bursts in Mounce *et al.* (2003). An ANN model, a mixture density network, was trained using a continually updated historic database that constructed a probability density model of the future flow profile. A Fuzzy Inference System was used for classification, comparing the latest observed flow values with predicted flows over time windows such that abnormal flow conditions generate alerts (Mounce *et al.* 2006). From the probability density functions of predicted flows the fuzzy inference system provides confidence intervals associated with each detection. Additionally an accurate estimate of abnormal flow magnitude is produced to further aid in ranking of alerts (Mounce *et al.* 2007).

Online implementation

The offline system was modified to work in an online environment (a MATLAB® application). Key software challenges included implementing a data warehouse solution, dealing with multiple data sources, alarm handling, automation of training and testing and, most significantly, data preprocessing and quality issues. It is widely recognized that approximately 80% of the resources in data mining applications are spent on cleaning and preprocessing the data (Fayyad *et al.* 1996) and this is of particular challenge when designing an automated system.

Figure 1 gives a summary of the algorithm structure for assembling training data. The system assembles the potential training file and then conducts several tests for amount and quality of data. Firstly, the data obtained via SQL from the database is transformed into internal MATLAB time series objects consisting of both time series data and events (the events are loaded from an event database). Events that can be applied to the time series include 'reset' (do not use any data for this logger before this date e.g. in the case of a network change such as rezoning), 'alarm on' and 'alarm off'. The data is then subjected to a number of logic checks based on overall percentage of valid data (that is, data not missing or periods corresponding to an alarm state) and also the total number of days of valid data. Based on user definable parameters, the site is either then passed, failed or tagged as 'low data period–treat with caution'. For further details see Mounce *et al.* (2010).

System integration

Figure 2 provides a scheme of the system integration for the online AI system. The automated analysis system is data driven starting from the logger units which initiate calls to the telemetry software every thirty minutes, GPRS signal permitting. This data is then mirrored to another server and

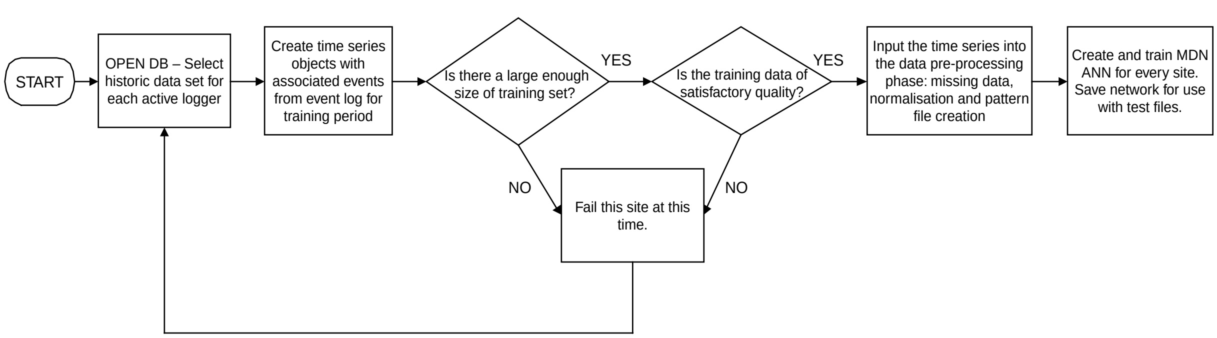

Figure 1 | Online AI system algorithm structure for assembling training data.

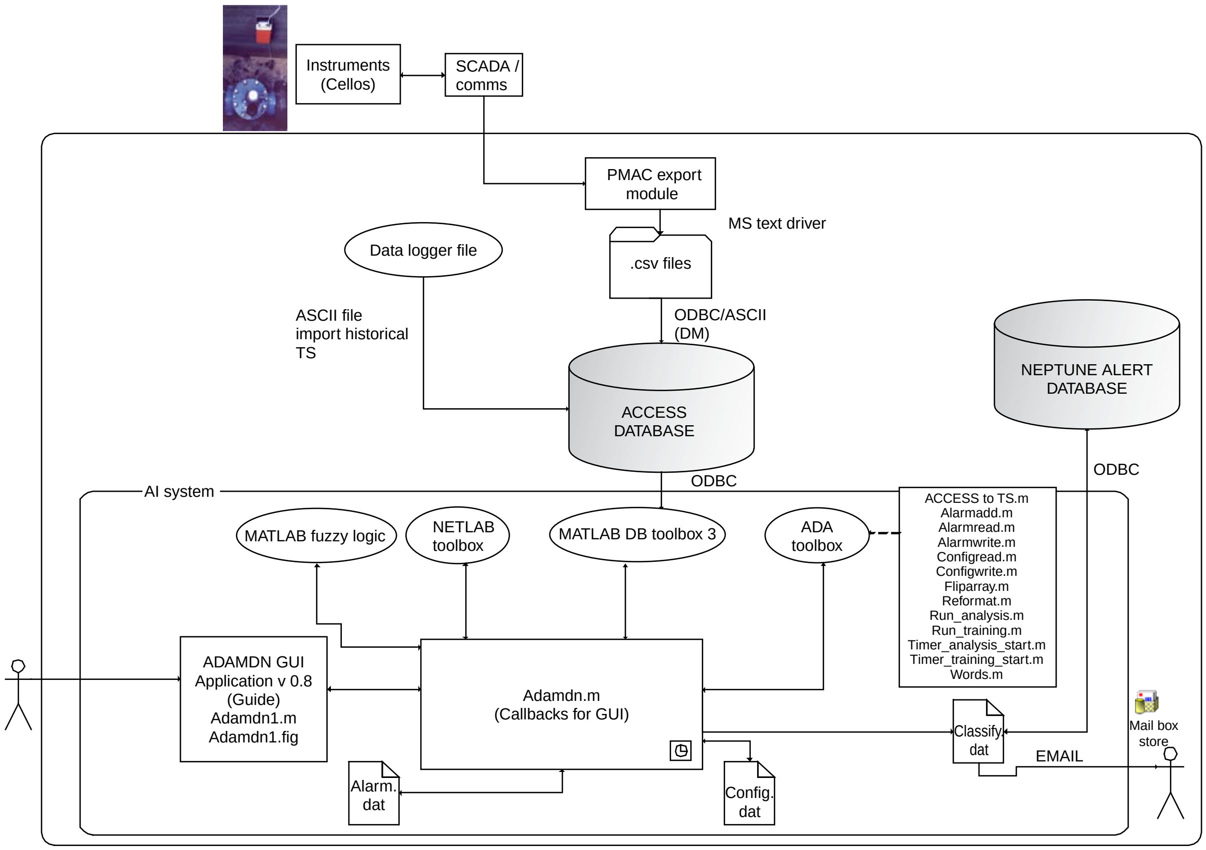

Figure 2 | Online system integration.

export functionality in the communications software is used to automatically update a set of CSV files every hour (which is a limitation of the communications and monitoring software update). An Open Database Connectivity (ODBC) text driver is used to interface to an MS Access database storing the time series history, and current logger values are appended by a data warehouse application every hour. The AI application is a MATLAB application which uses functionality from the Database toolbox, the Fuzzy Logic toolbox and NETLAB (Nabney 2001). The online system accesses the database via an ODBC driver. The ANN modelling each logger is retrained at some regular time interval (e.g. weekly) with several weeks' recent data and then runs the analysis phase every hour. Abnormal classifications by the FIS result in automated emailed alerts and an entry being added to the NEPTUNE alerts database using the ODBC standard. These source alarms, if deemed to be a significant event when amalgamated with additional information, are further evaluated by the DSS. An impact model component (incorporating a hydraulic model) can be used to assess the affected properties as well as the pipes which are likely to experience an increase in discolouration potential due to the failure, allowing operators better insight into the adverse effects of a burst and their development in time (Bicik *et al.* 2009).

RESULTS AND DISCUSSION

The case study

The Harrogate and Dales (H&D) area in the North Yorkshire region in the UK consists of nearly 200 DMAs (excluding trunk main and industrial user DMAs) and includes approximately 122,000 properties. Yorkshire Water Service, the water utility company responsible, conducted a network service pilot called RTNet in the H&D area by installing 450 Cello Loggers equipped with a General Packet Radio Service (GPRS) communications infrastructure, to dramatically improve data transfer for both flow and pressure data. Data is communicated every thirty minutes and two readings are obtained (fifteen minute sampled data). The pilot provided the case study data (flow only) sources to trial the AI data analysis system for detecting more subtle change from DMA flow data, such as leaks and bursts as defined in the introduction. Access to all DMAs within the H&D region was provided, resulting in 146 suitable sites. The online system was operational and generating alerts for the whole of 2008.

It should be emphasised that data was from a 'live' network and not a desktop study using historical data sets. Alerts were emailed automatically to a dedicated email account accessible by the control room and, in the latter part of the year, the alerts were also written automatically using ODBC to a NEPTUNE alert database. Figure 3 shows the NEPTUNE live alerts browser for the H&D case study area, showing the latest alarms in a one week period.

Although network control room operators had access to the AI alerts, the information was not directly integrated into operational practice. However, in many cases leakage teams were notified about a problem in a DMA following AI alerts, who conducted a leak survey and located a leak. Sources of information for correlating alerts included the following:

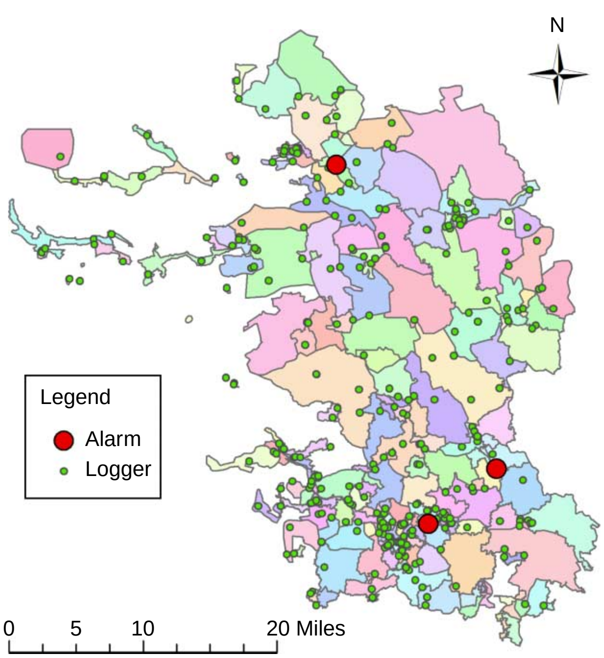

Figure 3 | H&D case study area.

- Daily emailed briefings on major events
- Customer contact database (customer reports of visible leakage)
- Work Management System (WMS) record of repairs database
- Ad hoc liaison with water utility personnel

The analysis of these data sources was usually performed a posteriori, so as to ensure all available databases were fully updated.

Example events

Two example events for the automated AI system are now provided and summary results are then presented for a one year period of analysis from 1st January 2008 to 31st December 2008. The alerts generated were produced by the online system previously described in the Method section. An ANN model, a mixture density network, was trained using continually updated flow data thereby constructing a probability density model of the future flow profile. A Fuzzy Inference System produces a classification, by comparing the latest observed flow values with predicted flows over time windows such that in the event of abnormal flow conditions alerts are generated. From the probability density functions of predicted flows the fuzzy inference system provides confidence intervals associated with each detection. An estimate of abnormal flow magnitude is generated for a particular burst classification by comparing, over the window, the difference between the actual observed value and the expected (i.e. predicted from the mixture model) value.

Figure 4 illustrates a situation in which the AI system correctly produced an alert for a burst which subsequently resulted in customer reports of leakage, and then was found and fixed. The alert was received at 15:15 on the 13th December 2008, with a confidence of 95% and a size estimate of 0.4 l/s. As can be seen from Figure 4, flow was abnormal in the 12 hour window up to this time (the period of FIS analysis). The DMA has 124 domestic properties.

The control room received contacts from customers, the first being more than two and a half days after the AI alert, at 09:45 on the 16th December. WMS information confirmed that a service pipe repair was conducted on the 22nd December. Note that the burst start time, indicated on

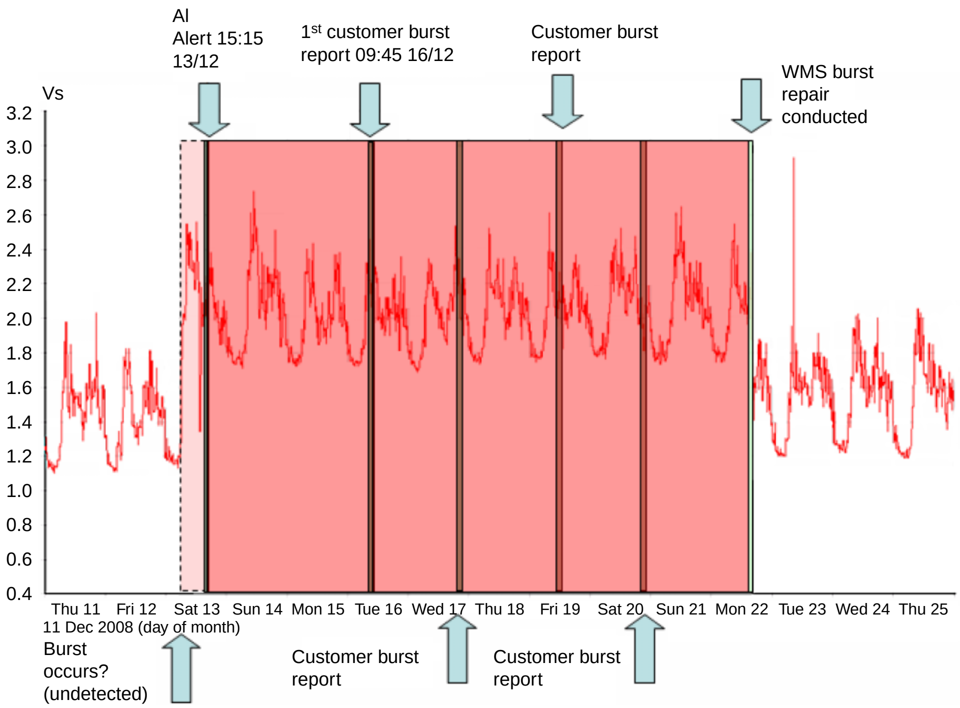

Figure 4 | Event 1—AI alert for burst (13th to 22nd December 2008).

Figure 4, is an estimate based on retrospective visual inspection.

In a second example, the sequence of events for another example of the live system is given. This event occurred in a DMA located in a cascaded section of the distribution system. A Water Pumping Station (WPS) feeds DMA A (270 Properties) which in turn feeds DMA B (213 properties). Figure 5 shows the flow and pressure for DMA B. During a period of operational fieldwork in the H&D area, a fire hydrant was opened producing a flow to waste of

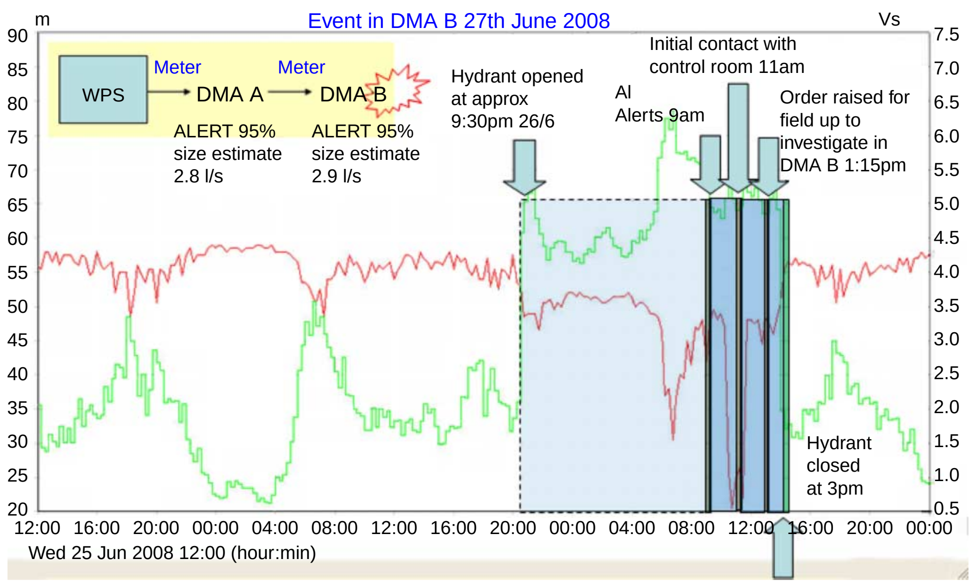

Figure 5 | Event 2—Hydrant flushing event detection in DMA B (25th to 27th June 2008).

approximately 3 l/s. The nightline reveals the hydrant was opened at 21:30 26/6/08. The hydrant opening was detected at 9am by the online AI system the following morning with two alerts being generated. The alert was followed up in the control room via a water company contact at 11:00. This was the first notification in the control room of a problem. The only customer contact was a 'No water' contact for DMA A at 12:32. An order was raised in the control room at 13:15 and a field operative located the hydrant and closed it at 15:00.

This validation event demonstrates how the system can give important online notification of significant abnormal flow which would have otherwise gone undetected.

Summary of all alerts for one year

A total of 195 alerts were produced by the AI system for the period 1/1/08 to 31/12/08 (see Figure 6). 146 sites were in the system but due to data quality and subsequent failure of data checks typically 90–100 were under analysis at any one time. The system was running more or less continuously 24 hours per day over the year, with only infrequent short periods of downtime for minor version upgrades. Of the 195 alerts:

- 70 are leakage or suspected leakage:
    - 35 have been correlated with WMS repair record.
    - 25 have been correlated to customer reports of bursts
    - 10 have been correlated to apparent repairs on nightlines or rezoning
- 7 have been correlated with known industrial events or network activity (such as flow tests)
- 67 are 'abnormal'–large unusual demands or short term increases in nightline (these events have a similar initial profile to bursts)

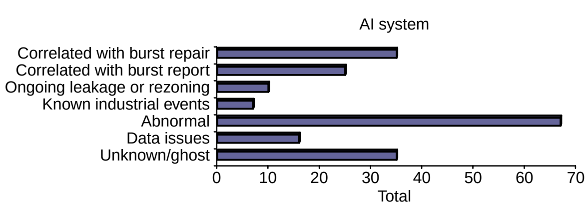

Figure 6 | Summary of all alerts for 2008.

- 16 are known to be caused by data corruption issues/logger failures
- 35 are ghosts/unknown

A comprehensive review was conducted for every alert using the available WMS and associated information. However, it should be noted that records for the year period were not exhaustive in terms of exact details on the totality of operations that had occurred in all DMAs such as industrial events, network alterations (like rezoning) and even scheduled maintenance. Consequently it was not always possible to correlate many alerts from Figure 6 (especially in the category 'Abnormal') to a definitive cause due to missing information.

CONCLUSIONS

A full-scale Water Distribution System in the UK was used for a case study with near real-time flow data provided by General Packet Radio Service. An online system has been operating for a one year period continually analysing this data (every hour) and providing timely alerts via automated emails (and using ODBC to a database) for 146 DMAs. The system provides a confidence estimate, in the form of a percentage, of the abnormality of the flow event and an estimate of the burst size. Results for the system for a one year period are reported. By bringing together varied information sources such as customer reports of visible leakage and retrospective records of repairs it was possible to show that the AI system successfully identified many abnormal events and alerts were often raised prior to their detection in the control room through customer leakage reports. Major findings include:

- technique proven in a live environment of a real system over a one year period
- 36% of alerts were correlated to definite bursts, using WMS and customer contact information, with the majority of other alerts likely to be high abnormal demands or network activity (38%)
- relatively few ghosts alarms produced (18%)

One comparison not possible was with results arising from automated nightline analysis, which is in use by some UK water service providers though not the partner in this work. Future comparison with the AI system would be useful.

The results obtained from the case study have shown that the AI system has the potential to be used as a useful tool for real-time identification of leakage of small to medium sized bursts. Its use promotes a more proactive approach to leakage management, with a possibility of fixing of leakage incidences soon after they occur and before the customer is seriously impacted. The system makes it feasible to identify and hence find and fix leaks that would previously have become background leakage, providing the potential to reduce the so called 'economic level of leakage' which has otherwise remained roughly stable in the UK as a whole for over a decade. Phase two of the RTNet pilot will see 4500 loggers deployed across a wider region during 2009. This higher footprint of data and alarms will allow a more detailed and rapid interpretation of network activity than ever before, and a faster and more appropriate response to customers. The water utility company are exploring a wider roll out of the AI analysis system in the next phase.

ACKNOWLEDGEMENTS

This work is part of the NEPTUNE project supported by the U.K. Science and Engineering Research Council, grant EP/E003192/1, and Industrial Collaborators. The authors would like to particularly acknowledge the assistance of Mr. Ridwan Patel (Yorkshire Water Services). Integration with the NEPTUNE DSS was facilitated by Mr. Josef Bicik (University of Exeter).

REFERENCES

Bicik, J., Kapelan, Z. & Savic, D. A. 2009 Operational Perspective of the Impact of Failures in Water Distribution Systems, *Proceedings of the World Environmental & Water Resources Congress 2009* (WDSA 2009).

Covas, D., Jacob, A. & Ramos, H. 2008 Water losses assessment in an urban water network. *Water Pract. Technol.* **3**(3), http://www.iwaponline.com/wpt/003/0061/0030061.pdf (accessed 14 September 2009).

Farley, M. & Hamilton, S. 2008 Non-intrusive leak detection in large diameter, low-pressure non-metallic pipes: are we close to finding the perfect solution? *Proceedings of the IWA World Water Congress*, September 7–12, Vienna, Austria.

Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P. 1996 From data mining to knowledge discovery: an overview. In Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. & Uthurusamy, R. (eds) *Advances in Knowledge Discovery and Data Mining*. AAAI Press/The MIT Press, pp. 495–515.

Hulme, M., Jenkins, G. J., Lu, X., Turnpenny, J. R., Mitchell, T. D., Jones, R. G., Lowe, J., Murphy, J. M., Hassell, D., Boorman, P., McDonald, R., & Hill, S. 2002 Climate Change Scenarios for the United Kingdom: The UKCIP02 Scientific Report, Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich, UK. p. 120.

Jankovic-Nisic, B., Maksimovic, C., Butler, D. & Graham, N. 2004 Use of flow meters for managing water supply networks. *J. Water Resour. Plann. Manage.* **130**, 171–179.

McKenzie, R. & Seago, C. 2005 Assessment of real losses in potable water distribution systems: some recent developments. *Wat. Sci. Technol.: Water Supply* **5**(1), 33–40.

Mounce, S. R., Day, A. J., Wood, A. S., Khan, A., Widdop, P. D. & Machell, J. 2002 A neural network approach to burst detection. *Water Sci. Technol.* **45**(4–5), 237–246.

Mounce, S. R., Khan, A., Wood, A. S., Day, A. J., Widdop, P. D. & Machell, J. 2003 Sensor-fusion of hydraulic data for burst detection and location in a treated water distribution system. *Int. J. Inf. Fusion* **4**(3), 217–229.

Mounce, S. R., Machell, J. & Boxall, J. 2006 Development Of Artificial Intelligence Systems For Analysis Of Water Supply System Data. *Proceedings of the 8th Water Distribution System Analysis Symposium*, August 27–30, Cincinnati, USA.

Mounce, S. R., Boxall, J. B. & Machell, J. 2007 An artificial neural network/fuzzy logic system for DMA flow meter data analysis providing burst identification and size estimation. In Ulanicki, B., *et al.* (ed.) *Water Management Challenges in Global Change*. Taylor and Francis, pp. 313–320, ISBN 978-0-415-45415-5.

Mounce, S. R., Boxall, J. B. & Machell, J. 2010 Development and Verification of an Online Artificial Intelligence System for Burst Detection in Water Distribution Systems. *ASCE Water Resour. Plann. Manage.* **309**, 309–318.

Nabney, I. 2001 *NETLAB: Algorithms for Pattern Recognition*. Springer-Verlag, UK.

Olsson, G. 2006 Instrumentation, control and automation in the water industry–state-of-the-art and new challenges. *Water Sci. Technol.* **53**(4–5), 1–16.

UNEP 2007 Fourth Global Environment Outlook: environment for development assessment report, United Nations Environment Programme, Nairobi (Kenya), ISBN 978-92-807-2836-1.