Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 119 (2015) 43 – 52

Procedia Engineering

www.elsevier.com/locate/procedia

13th International Conference on Computing and Control for the Water Industry, CCWI2015

Cloud based machine learning approaches for leakage assessment and management in smart water networks

S. R. Mounce[a]*, C. Pedraza[b], T. Jackson[c], P. Linford[d], J. B. Boxall[a]

[a]Department of Civil and Structural Engineering, The University of Sheffield, Sheffield, S1 3JD, UK
[b]The Power House -Reading STW, Island Rd, Reading, Berkshire, RG2 0RP, UK
[c]CYBULA, Computer Science Building, Deramore Lane, York, YO10 5GH, UK
[d]Syrinix Limited, Units 19 and 20, Hethel Engineering Centre, Chapman Way, Hethel, Norwich NR14 8FB, UK.

Abstract

One-third of utilities around the globe report a loss of more than 40 percent of clean water due to leaks. By reducing the amount of water leaked, smart water networks can help reduce the money wasted on producing or purchasing water, and the related energy required to pump water and treat water for distribution. A UK demo site is presented focusing on leak management, integrating fixed flow and pressure instrumentation, advanced (smart) metering infrastructure and novel instruments (capable of high resolution monitoring). Example data analysis results for this site using the AURA-Alert anomaly detection system for Condition Monitoring are presented.

Peer-review under responsibility of the Scientific Committee of CCWI 2015

Keywords: Leakage; Smart Networks; AMR; Neural Networks; Cloud computing.

1. Introduction

European water utilities face many problems related to their 3.5 million km's of distribution networks. Large parts of water distribution networks have to be rehabilitated requiring investments of € 20 billion/year. Prioritization and optimisation of investments is needed urgently. The European Innovation Partnership on Water

* Corresponding author.
E-mail address: s.r.mounce@sheffield.ac.uk

Peer-review under responsibility of the Scientific Committee of CCWI 2015
doi:10.1016/j.proeng.2015.08.851

has established priority areas related to the challenges in water supply distribution networks, focusing on resource efficiency, Smart Water Management and decision support systems. It is estimated that implementing Smart Water solutions could save water utilities and their customers up to $12.5 billion a year (Sensus, 2013). Smart Cities need to bring together hard infrastructure, social capital including local and community institutions, and technologies to fuel sustainable economic development and provide an attractive environment for all. The concept of a Smart City within the context of water means using technologies for optimising water resources and waste treatment, monitoring and controlling water, and providing real-time information to help water companies and households manage their water better. Although the technology components for Smart Water Cities are available, the route to application is uncertain. The main hurdles are: lack of integrated and open solutions; difficulty to comply with user and integration requirements; lack of clear and validated business cases for solutions; lack of business intelligence awareness and lack of political and regulatory support.

The quantity and complexity of sensor and environmental data is growing at an increasing rate whilst the demands for new solutions and tools to utilize and interpret this data are likewise growing due to financial and regulatory pressures. The phrase ‘Big Data’ may then become a reality for the water sector particularly on the customer side, since when smart metering becomes more prevalent a huge amount of data will be collected. If the UK goes to a point where the entire water industry is universally metered with Smart Metering there will be approximately 25 million water meters for customers. Organising, managing and supporting such massive ICT network infrastructure, however, is a substantial technical challenge. This data could be used, in conjunction with mapping software and hydraulic models to map consumption in DMAs where there are spikes in usage.

As demand for clean water increases with population growth in the coming decades and supply remains stagnant or shrinks due to climate change, solutions to manage and minimise leaks will become increasingly critical. Many water utilities are struggling to measure and locate leaks in their distribution networks beyond the Economic Level of Leakage, and there is a drive to efficiency by implementing leak reducing solutions. Leakage results in wasted energy costs (such as spent pumping water), water treatment costs (energy and chemicals), misdirected repair activities, regulatory penalization and environmental damage to city infrastructure. Smart water networks offer the potential to identify leaks early thus reducing the amount of water that is wasted and saving utilities money. These solutions include the use of flow and pressure sensors to gather data, analyse the data using algorithms to detect patterns that could reveal a leak in the network, and provide real-time data on the location of a leak. Mounce et al. (2014) review approaches for event detection in WDS measured time series data, with a focus on data driven methodologies for leak detection.

SmartWater4Europe (SW4E) is a four-year FP7 demonstration project (2014-17) funded by the European Commission (Demonstration of integrated smart water supply solutions at four sites across Europe, grant 619024). SmartWater4Europe consists of 21 participants including three water utilities, SMEs, research organisations and platform organization. The four demo sites (of varying in scale and located in the United Kingdom, Spain, The Netherlands and France) will allow demonstration of solutions incorporating sensors, data processing, modelling and ICT technologies.

The UK demo site (a large town on the outskirts of London) is focusing on leak management (as well as exploring energy optimisation and customer interaction): in particular the use of advanced metering infrastructure (AMI) smart metering. The demonstration sites objectives are to identify the pathways to transform network and customer property instrumentation to maximise the sustainable use of existing infrastructure, maximise network performance and improve water resource and energy efficiency. The demonstration site will investigate how new and emerging technologies can be used to create a ‘smart network’ with real time notification of performance and even asset condition to enable proactive management and intervention and a step change in social awareness of water used and infrastructure. This is no easy task and necessitates interdisciplinary working between researchers using techniques such as wireless distributed sensors, Grid and Cloud computing, hydro-informatics, data mining and machine learning and for understanding the implications of these in a social context in terms of privacy, trust, acceptability and utilisation. The IT aspects of integrating Grid and Cloud computing, hydro-informatics and machine learning for this case study are particularly explored in this paper.

The pervasive instrumentation and data mining that will be demonstrated in this project may well allow a move away from current DMA structure, with closed boundary valves, leading to more adaptive and resilient networks, and a step change in customer awareness of this essential infrastructure. This paper presents background to the demo site area and some initial pilot deployment of instrumentation, along with initial integration with the Cloud based portal and data analysis services and some preliminary results.

Nomenclature

AMR	Automatic Meter Reading
ANN	Artificial Neural Network
AURA	Advanced Uncertain Reasoning Architecture
CMM	Correlation Matrix Memories
DMA	District Meter Area
SDE	Signal Data Explorer

2. Demo site

2.1. Overview

The Thames Water demonstration site is situated in Reading, a large town in the royal county of Berkshire, England (Reading is 58 km west of central London). Key SMEs and research organisations from the sector are working together to demonstrate innovative new solutions. These solutions will be tested, validated and evaluated and business cases will be generated. Technologies considered include revolutions in instrumentation associated with networks, such as installation of flow instruments through full bore hydrants, novel instruments (capable of high resolution monitoring and logging of flow and pressure and thus enabling the identification of pressure transients) as well as at the customer's meter. Key technological developments to be exploited at the household scale are AMR and AMI. AMR is the technology of automatically collecting consumption, diagnostic, and status data from a water meter and AMI describes networking technology that goes beyond AMR into remote utility management. With such technology it could be possible to supply meter readings daily leading to the potential for household leak reports for every customer. The 'smart' integration of such technology with network instrumentation will be explored as part of the demonstration site.

The demonstration site comprises 870 km of distribution mains and 172 km of trunk mains located in and around the city, which has many pipes over 60 years old, serving 89,000 commercial and domestic properties. The majority of these mains are of ferrous material at varying levels of degradation, with plastic (PE) mains now used as a standard for full replacement, and ductile iron used generally for larger diameters. The trunk mains network consists of 172 km of mains which convey about 45 Ml/d of chlorinated potable water from the treatment works into the distribution network. They include installations varying from 4" (100 mm) up to 32" (800 mm) in size, with majority of larger diameters constructed of iron. Currently available demonstration infrastructure includes around 1,000 smart AMI water meters. The demonstration site is adopting a holistic approach where optimisation using smart technologies balances conflicting interests such as energy consumption, leakage and pressure management.

TW aims to understand the correct level of instrumentation and analysis required to deliver optimised smart water networks, including factors such as cost, risk, energy consumption, level of leakage and customer interaction. One of the main innovation themes is in the area of leak detection: detecting and localising leaks soon after they occur, or ultimately detecting failure mechanisms before they occur. For leakage, the main focus is on two DMAs which have been heavily instrumented with AMR. These two DMAs have an advanced fixed network (bidirectional) that doesn't need any repeaters as the smart meters have a long range radio attached which transmit to antennas in 5 physical sites. Meters communicate via radio to either repeaters /concentrators or directly to an antenna, depending on the technology being used on the DMA. Data files (csv) are transmitted via secure FTP to Thames Water servers every four hours. They contain the meter readings and alarms from the meters. In addition to standard flow and pressure zonal measurements and the AMR, novel Syrinix instrumentation has been deployed (Trunkminder and BurstMinder) in the trial DMAs (E8 and E14). Figure 1 shows a simplified schematic of the surrounding zones and measurement points. There are 5 DMAs with flow (and pressure) measured at the inlet which are used for the DMA level historic analysis (one of these has two inlets). Figure 2 shows the deployment of BurstMinder instruments in E8. The primary function of BurstMinder is to continuously monitor the pressure in a pipeline for significant transient events. It is a battery powered pressure transient detection and reporting device. When powered from a small Lithium battery packs it is able to monitor a water main for pressure transient events for 3 years. Events are detected locally (at the pressure fitting) and transmitted via a GPRS link to Syrinix's servers for further analysis. The system also stores the full pressure data history internally. Upon detection of an event the BurstMinder can send an SMS alert, send a window of data using the GPRS modem or store the event data locally. In addition to pressure transient detection and reporting, BurstMinder may also be configured to continuously log data to an internal memory at a maximum sample rate of 128 S/s; send a daily summary including the minimum, maximum and mean pressure for each 15 minute window.

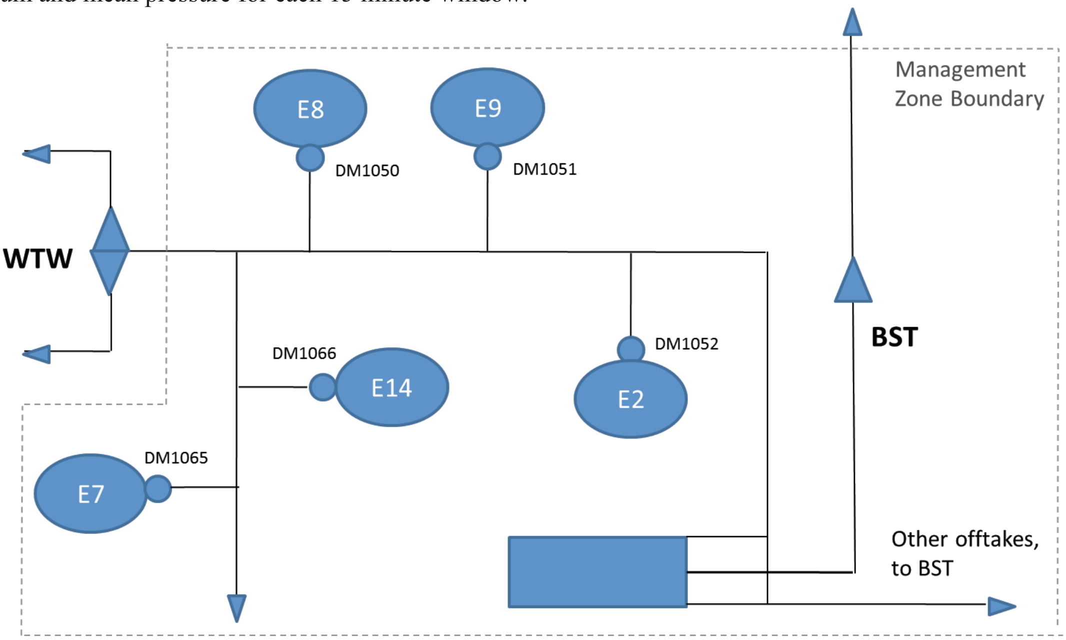

Fig. 1. Schematic of case study used in AURA

Fig. 2. Deployment of BurstMinder in E14 and data graph of high sample rate pressure from live system

3. Methods

3.1. youShare

A Cloud based portal called youShare has been developed and is being customised for SW4E, with a move to the HADOOP environment, integrated with one of the leading data technologies for time-series and event based data sets: Cassandra. This platform allows rapid access to large data sets and a framework for running analytical methods on the data on a parallel, high performance computing infrastructure (Hodge et al. 2014). Storage and searching of the data is very rapid, and as necessary, Grid technology can be leveraged. Data at different sampling rates and incomplete and noisy data can be managed. Because analysis applications need to be executed on a wide range of data formats a uniform file and format structure has been specified for data sharing and communication between applications. YouShare provides a mechanism by which researchers can port their application code onto the computing platform as an interactive service. This allows user code to be deployed as part of a catalogue of applications available to any registered user within the system. There is a simple wrapping interface which allows end users to encode their software or applications in the format required for the system. Once uploaded, and approved for deployment by the system administrators, the code becomes a 'service', which can be executed using any of the data held on the system. If required, the service can also become part of a workflow of services, orchestrated by the workflow execution engine. An added benefit of this software delivery model is that it abstracts away all of the underlying computer resourcing issues so that the user is not concerned with the low level detail of how a job gets processed or the underlying HPC. The youShare architecture allows for the hosting and application of various algorithms. The objective will be to populate this portal with water engineering specific metadata and services for the demo site, including data handling, cleaning pre-processing and analysis.

A number of approaches from the fields of artificial intelligence and statistics have been applied for detecting abnormality in WDS from sensor time series data, both for quantity and quality, for example for leak detection based on pressure and flow (providing usable information in the form of timely alerts). Mounce et al. (2010) describe an online system pilot implemented with a UK water company using an artificial neural network (ANN) and fuzzy logic system for detection of leaks/bursts at DMA level.

3.2. AURA-Alert

The Advanced Uncertain Reasoning Architecture (AURA) provides analysis capability for anomaly detection in time series data. AURA is a class of binary neural network, developed and implemented in Correlation Matrix Memories (CMMs). The approach and technology provides efficient, scalable pattern recognition in complex and large scale condition monitoring (CM) applications (Austin 1995, Hodge and Austin 1995). AURA allows a data driven model of a system or an asset to be developed by training via observation on normal operating condition data. The CMMs can be trained on multi-dimensional data time series data to build up a complex internal representation of the system characteristics from operational data. This model can be built with reference or historical data sets. The system can quickly spot 'out of normal' operating values in complex high volume data, and is scalable to industrial size data volumes with negligible performance impact. Mounce et al. (2014) successfully explored the offline application of AURA (Advanced Uncertain Reasoning Architecture) for both water quantity (for leakage) and quality. Examples demonstrated successful early detection of abnormality in systems using multi-parameter data as well as significant potential for precursor event detection beyond typical outlier detection approaches. These precursors could be linked to appropriate maintenance requirements for water infrastructure.

Cloud computing offers an opportunity to make the results of CM readily available to a range of stakeholders responsible for the maintenance of an asset. Data from sensors distributed across one or more assets at one or more sites are uploaded to the cloud compute resource for continual analysis. Hickinbotham et al. (2012) propose a new system for CM using the cloud. The system combines state of the art pattern search capability based around AURA-Alert, designed to detect anomalies in signal data from complex assets. AURA-Alert is a novelty-detection plug-in for Cybula's Signal Data Explorer (SDE) software, which has been developed to visualise and query large amounts of time series data, enabling the rapid searching for patterns and events across variables and times. Current applications include EEG analysis and incident event analysis in engineering. AURA-Alert may be used on historical and real-time data feeds, with the potential to be utilised in a real-time fashion (deployed on the cloud in youShare) for the demo site. Figure 3 provides an overview of the AURA architecture. During CM for real time alerting, the current state of the system parameters can be compared to the stored normal operating behaviour (in the CMM) to see if that combination of variable values is a known state. If not, this could be indicative of a problem with the system and lead to an alerting event.

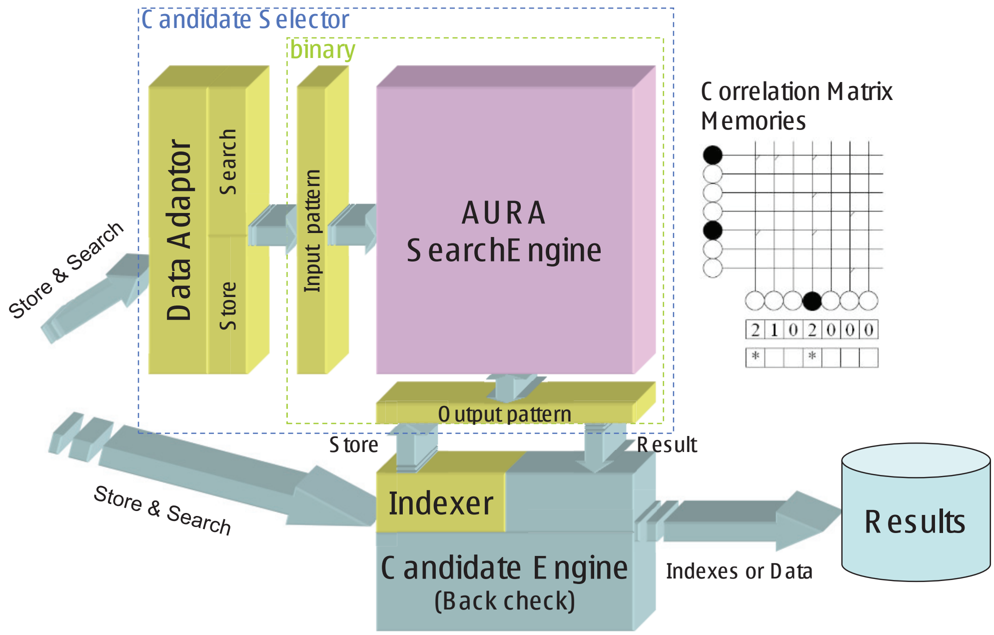

Fig. 3. AURA Alert architecture

3.3. AURA Application

AURA was applied to historical data for the DMA inlet flows shown in Figure 1. In order to use AURA-Alert on data with periodic (e.g. daily) cycles such as DMA inlet flow and pressure, it is necessary to introduce an extra 'time of day' variable (e.g. the number of elapsed hours of the day). This enables AURA-Alert to detect patterns in the data that are unusual at that time of day. With sufficient quality and quantity of data, this should ideally be a weekly index. However, data quality is a continuing issue for fixed WDS sensor networks. One study found up to a fifth of the data collected by the evaluated sensor system to be either missing or erroneous over a three year period (Ediriweera and Marshall 2010). The data collected from sensors are first formatted into input files for a MATLAB pre-processing program which identifies and fills in missing timestamps or values (for the latter only zero order hold was used to flag missing data periods as anomalous), deals with non-unique timestamps and resamples to a 15 minute sampled frequency (occasional oversampled periods, such as at one minute sample period were present) so as to provide a continuous stream of data. The data are finally reformatted into an appropriate comma delimited format required by the SDE. Data was generally available for each site between January / March 2013 to June 2014, though in practice gaps and sensor drop outs were present. Nevertheless, the whole set of available data was used for testing. AURA-Alert was trained by provided it with data from a period of time (generally 2 weeks) during which the sensor was visually confirmed to be performing correctly (no missing data) and with 'normal' conditions in the distribution system (in terms of stability nightline). This pattern is then stored in an AURA associative memory. Note though, that over such a long period 'normality' may well drift. AURA detectors were created per site with fixed width binning and default width parabolic kernels. Flow data was used for anomaly detection for this case study, being more sensitive to leak/burst events than pressure data (Mounce et al. 2011).

4. Results and discussion

4.1. DMA inlet meter flow analysis

The AURA-Alert system utilising CMMs has been used for the novelty detection in highly complex assets in a variety of industries. AURA-Alert was applied to combined data streams from the demo site to automatically calculate a novelty score for any type of unusual event, and results obtained are now described. A match score global threshold of 70% was used to identify reasonably large deviations from normality (in practice this figure can be adjusted, and also provided as an output is the novelty score per channel). Figure 4 shows output for the E8 flow meter. It can be seen highly novel flow corresponds to significant periods of zero (sensor failure) and a large abnormal increase 23/8-25/8/2013 on the flow (top) channel. Using a 70 Global Threshold, 1787 novelties were present out of 42912 states (~4%). The AURA Alert output can be seen in the 'Match Distance' channel (100 – Match Strength), which has a value of 0 when in a previously seen state and increases when a novelty is detected.

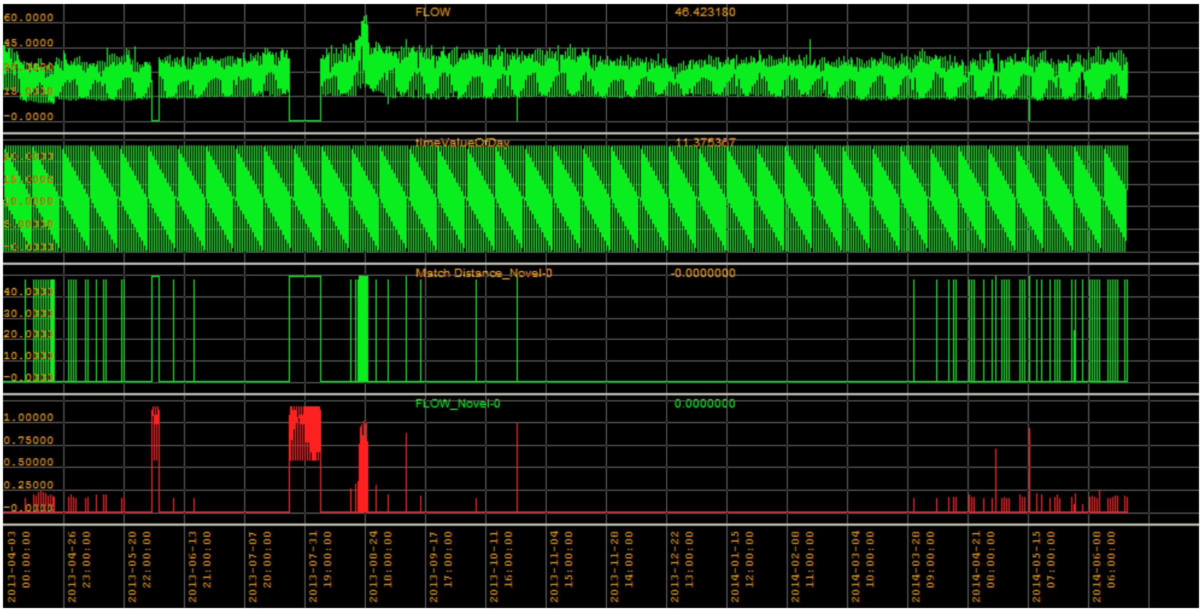

Fig. 4. AURA Abnormality detections (E8 inlet flow) March 2013 – June 2014

One of the inlet meters was not operational during the data period (E9) hence was not analysed. The results for DMA E7 are provided in Figure 5. Again, the sensor drop outs are highlighted as novel as well as potential minor bursts in early 2013, and again in mid-November 2013. Using a 70 Global Threshold, 2165 novelties were present out of 51744 states (~4%). DMA E14 had similar levels of anomaly detection with 1994 novelties out of 51757 states (~3.8%). The final DMA E2 provides similar results with 1981 novelties of 51744 states; Figure 6 shows a zoom of a detected event that was correlated with customer contact and leak repair information.

4.2. High frequency BurstMinder event

Figure 7 provides an example of the type of event which can be detected in live sensor network system (from a large city in the UK). An alarm was generated automatically from an algorithm analysing the hydraulic data which was discovered to be an unauthorised pump shutdown caused by a power failure. Both 15- minute average data and high sample pressure are plotted, and note the over-pressure spike event when power is restored which is not apparent on averaged 15-minute 'normal' data making the event appear far less serious than it actually was.

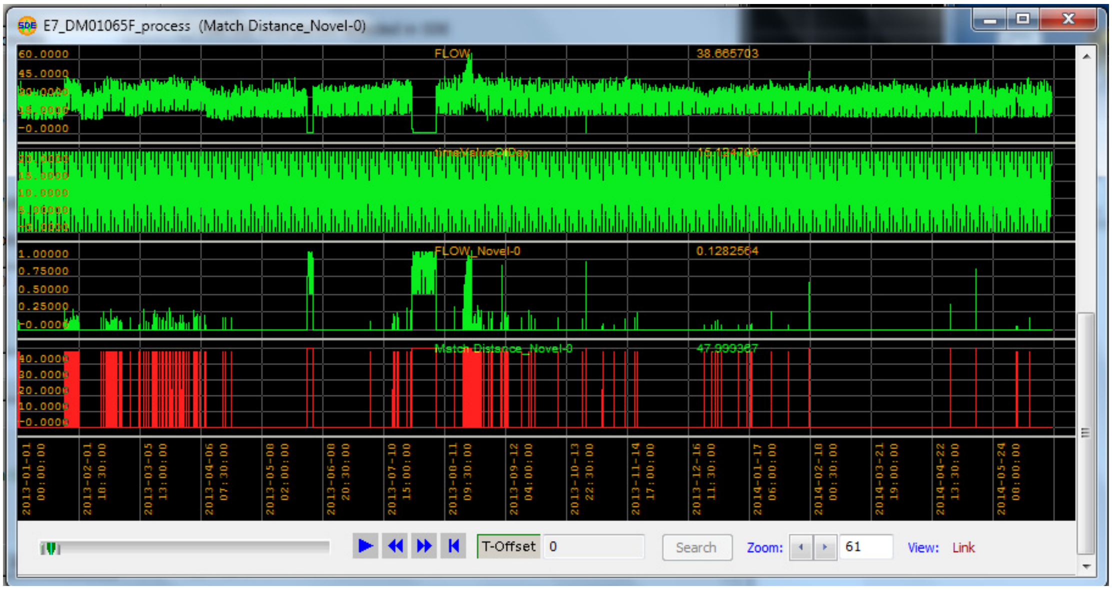

Fig. 5.AURA Abnormality detections (E7 inlet flow) January 2013 – June 2014

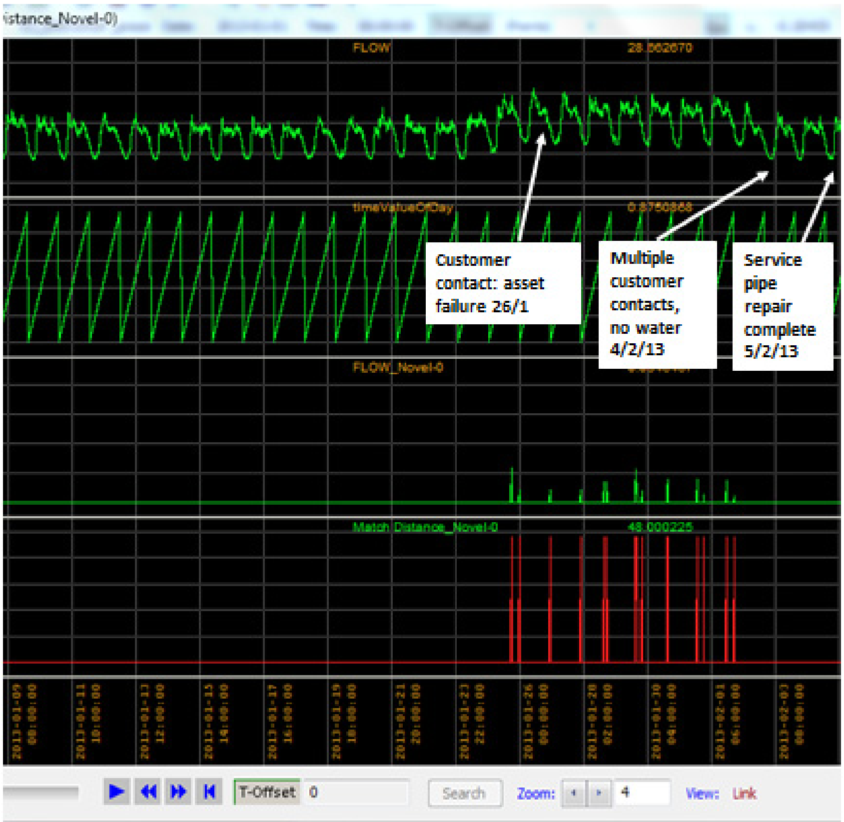

Fig. 6.AURA Abnormality detections (E2 inlet flow) – zoom of burst event

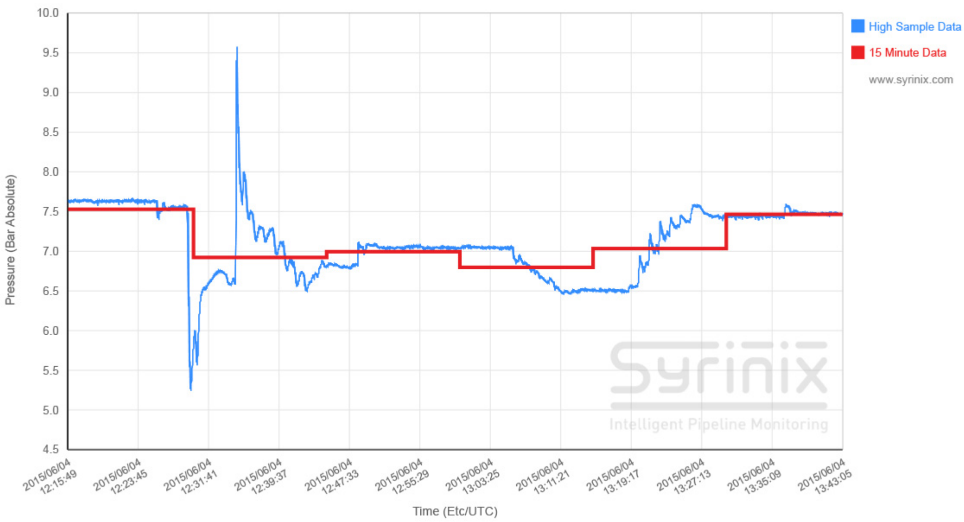

Fig. 7. BurstMinder event detection with averaged and high frequency data

4.3. Discussion

The full possibilities of Smart Network technologies to deliver improved service to customers and cost-effective performance improvements in the water industry are yet to be realised. Smart metering technologies need to be able support decisions at both the household and utility levels (Savic et al. 2014). Sensor technology and the 'big data' they generate combined with advanced machine learning techniques are providing exciting new opportunities for new scales of understanding of WDS. With increased DMA flow and pressure measurements comes the possibility of leak localisation at the sub-DMA level (Farley et al. 2013, Romano et al. 2013, Candelieri et al. 2014). AMR data expands information availability even further and has the potential to be used for customer profiling at the WSP side, allowing urban water planning and management based on consumer types and for informing customers on their water end-use patterns. Nguyen et al. (2013) present a methodology using Hidden Markov Model and Dynamic Time Warping Algorithm techniques disaggregating customer data into its end-use categories, including for rapidly alerting consumers of occurring leak events. By utilising AMR for demand forecasting, the possibility for reducing costs for treatment, storage and distribution arises such as through optimisation of pump scheduling. Candelieri et al. (2015) present a data-driven, fully adaptive self-learning algorithm for short-term water demand forecasting utilizing AMR data. In the future, water and energy use could be more efficiently managed through smart meter adoption and changing the 24 hour diurnal demand profile.

5. Conclusions and further work

Projects such as SmartWater4Europe (SW4E) allow the exploration, at demo and full WDS pilot scale, of deploying multiple Smart Network technologies, both hardware and software, and the multiplicative synergy between them. Successful demonstration of the return on investment business case will ultimately allow full-scale roll out of Smart DMAs. This paper presents the UK SW4E demo site and some preliminary analysis results to illustrate obtaining improved understanding from various deployed sensors as follows:

- Condition Monitoring (CM) applications in the environmental sensing systems world suffer from the challenge of substantial increases in the volumes of data being produced ('Big Data') in the remote monitored systems and the subsequent increasing complexity. A CM approach for Smart Networks can focus on the early detection of potential faults in the monitored system or asset (based on sensor readings), allowing preventative action to be taken before major damage occurs. For example, in a

WDS the utility will ideally wish to renew or rehabilitate a failing pipe just in time before a burst and avoid any unplanned service interruption / disruption.
- AURA-Alert (Advanced Uncertain Reasoning Architecture) can rapidly learn and model the normal operating envelope for a system. Using AURA-Alert, time series data from sensors are converted into vectors using a quantisation. Vectors are then stored in a historical database in the correlation matrix memory. A measure of novelty (termed Match Strength) can be generated for new data presented as vectors. This continuous novelty score can be utilised to enable the detection of leak, burst and other anomalous events as presented in examples here from demo site flow meters.
- AURA-Alert is being developed as an online (Software as a Service) system for SW4E which automates the training data selection (by selecting data with high Match Strength and with regular retraining) and use of validation data for selecting Match Strength thresholds for alert generation. It has potential for application to multiple data sources including across many data streams simultaneously.
- High sample rate data allows much greater understanding to be made available by enabling models and event detection algorithms to function at a level of data quantity and resolution previously impossible.

Further work will include incorporating pressure data, customer complaints and repair information, AMR dataset analysis and the outputs from BurstMinder and Trunkminder instruments for truly integrated Smart DMAs.

Acknowledgements

The authors wish to thank Thames Water for data provision, particularly Matt Hawkins. This research was funded under the EU FP7 SmartWater4Europe demonstration project, grant 619024, and part funded by the Pennine Water Group Platform Grant, UK Engineering and Physical Sciences Research Council (EP/I029346/1).

References

Austin, J.,1995. Distributed associative memories for high speed symbolic reasoning. International Journal on Fuzzy Sets and Systems 82, 223-233.

Candelieri, A., Soldi, D., Conti, D. and Archetti, F., 2014. Analytical leakages localization in water distribution networks through spectral clustering and support vector MACHINES. The icewater approach. Procedia Engineering 89 (C), 1080-1088.

Candelieri, A., Soldi, D. and Archetti, F., 2015. Short-term forecasting of hourly water consumption by using automatic metering readers data. Proceedings of CCWI 2015, Leicester, UK, 2nd - 4th September.

Ediriweera, D. D. and Marshall, I. W., 2010. Monitoring water distribution systems: understanding and managing sensor networks. Drink. Water Eng. Sci., 3, 107–113.

Farley, B., Mounce, S. R. and Boxall, J. B., 2013. Development and field validation of a burst localization methodology. ASCE Journal of Water Resources Planning and Management 139 (6), 604-613.

Hickinbotham, Simon, James Austin, and John McAvoy. Interactive graphics on large datasets drives remote condition monitoring on a cloud. Journal of Physics: Conference Series. Vol. 364. No. 1. IOP Publishing, 2012.

Hodge, V. and Austin, J., 2005. A Binary Neural k-Nearest Neighbour Technique. Knowl. Inf. Syst. (KAIS) 8(3), 276-292.

Hodge, V., Turner, A., Fletcher, M., Jessop, M., Weeks, M., Jackson, T. and Austin, A., 2012. Enhancing YouShare: the online collaboration research environment for sharing data and services, Digital Research 2012, Oxford, UK, 10 - 9 - 2012 .

Mounce, S. R., Boxall, J. B. and Machell, J., 2010. Development and verification of an online artificial intelligence system for burst detection in water distribution systems. ASCE Journal of Water Resources Planning and Management 136 (3), 309–318.

Mounce, S. R., Mounce, R. B., and Boxall, J. B., 2011. Novelty detection for time series data analysis in water distribution systems using support vector machines. Journal of HydroInformatics 13 (4), 672–686.

Mounce, S. R., Mounce, R. B., Jackson, T., Austin, J. and Boxall, J. B.,2014. Associative neural networks for pattern matching and novelty detection in water distribution system time series data. IWA Journal of HydroInformatics 16 (3), 617-632.

Nguyen, K. A., Stewart, R. A. and Zhang, H., 2013. An intelligent pattern recognition model to automate the categorisation of residential water end-use events. Journal of Environmental Modelling and Software 47, 108-127.

Romano, M., Kapelan, Z., Savic, D. A., 2013. Geostatistical techniques for approximate location of pipe burst events in water distribution systems. Journal of Hydroinformatics 15 (3), 634-651.

Savić, D., Vamvakeridou-Lyroudia, L., Kapelan, Z., 2014. Smart meters, smart water, smart societies: The iWIDGET project. Procedia Engineering 89 (C), 1105-1112.

Sensus. 2013. Water 20/20: Bringing Smart Water Networks into Focus. Sensus, Raleigh, N.C. http://sensus.com/documents/10157/1577608/Sensus_Water2020-USweb.pdf/d67d0a75-255a-4a20-86f1-d4548bfcdf78 (accessed Apr. 24, 2015).