

# Online advanced uncertain reasoning architecture with binomial event discriminator system for novelty detection in smart water networks

S. R. Mounce<sup>1</sup>, A. Fergus<sup>2</sup>, M. Weeks<sup>2</sup>, J. Young<sup>2</sup>, D. Ejimbe<sup>3</sup>, E. Goya<sup>3</sup>, M. Holburn<sup>4</sup>, T. Jackson<sup>2</sup>, and J. B. Boxall<sup>1</sup>

<sup>1</sup> Department of Civil and Structural Engineering, University of Sheffield, Sheffield, S1 3JD, UK

<sup>2</sup> Cybula Ltd, Computer Science Building, Deramore Lane, York. YO10 5GH, UK

<sup>3</sup> Thames Water, The Power House -Reading STW, Island Rd, Reading, Berkshire, RG2 0RP, UK

<sup>4</sup> Solvd, 2 Faraday Court, Rankine Road, Basingstoke, Hampshire RG24 8PF

*<sup>1</sup>s.r.mounce@sheffield.ac.uk*

## ABSTRACT

*Minimising the loss of treated water from water supply systems due to bursts is an ongoing issue for water service providers around the world. Sensor technology and the ‘big data’ they generate combined with machine learning based analytics are providing an opportunity for automated event detection. AURA-Alert has been developed as an online (Software as a Service) system which automates the training data selection (by selecting data with acceptable Match Strength and with regular retraining). The addition of a Binomial Event Discriminator service can produce alerts based on windows of thresholded match distances. A pilot deployment on over 200 live data streams in the cloud has been deployed as part of the SmartWater4Europe project. Examples of analysis for real events are presented. For a historic subset of eight data streams over a three month period up to 58% of bursts were detected (depending on window used for evaluation). It is concluded that the system is an effective and viable tool for novelty detection for water network time series data with potential for wider applicability. Key strengths include lack of per site configuration, data-driven self-learning (from periods of normality), real-time, high scalability and full automation of model retraining.*

**Keywords:** Smart water networks, neural networks, data driven, burst detection, online.

## 1 INTRODUCTION

Smart Water Network (SWN) technologies have the potential to deliver an improved service to customers and cost-effective performance improvements for the water industry. Sensor technology and the ‘big data’ they generate combined with advanced machine learning (ML) techniques are providing new opportunities for deeper understanding of water distribution systems (WDS). SmartWater4Europe (SW4EU) is a European demonstration project. Four demo sites comprise solutions incorporating sensors, data processing, modelling and analytics technologies. The UK demo site (TWIST) in Reading has been investigating how emerging technologies can be used to create a SWN with near real time notification of performance to enable proactive management and intervention (particularly for leakage). Technologies utilised during the pilot have included installation of flow instruments through full bore hydrants, instruments capable of high resolution monitoring (thus enabling the identification of pressure transients), AMR customer smart meters [1] as well as traditional sensors. Increasing amounts of SWN data is only of real business value if this

valuable resource is ultimately used to inform and support decision making i.e. data to information to insight to action.

AURA-Alert is a novelty-detection approach for condition monitoring of assets, designed to detect anomalies (also termed novelties) in time series data. This paper describes how AURA-Alert has been extended to an online (fully automated) data driven system for WDS event detection and applied to the TWIST demo site. Details are provided of the system adaptation for automatically generated detectors, data filtering and automated (re)training and example results.

## 2 BACKGROUND

Methodologies for identification of burst (or other abnormal) incidences direct from analysis of hydraulic data have been increasing in popularity and application over the last decade. A number of data driven approaches have been proposed for burst detection in WDS, classified into three categories: classification, prediction-classification and statistical [2]. However, few of these have been implemented online or demonstrated in live pilots, with some exceptions [3], as often, manual tuning or site specific configuration is required and system scale up is non-trivial. WDS are demand driven, and the system does not know its demand in advance [4]. Pressure control allows the raw 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 necessarily straightforward, and other network activity will also manifest as abnormality.

Water companies use DMAs (District metered areas) to identify particularly problematic areas and help target leakage detection efforts. Areas where expected demand differs significantly from the observed DMA flow, particularly at night, are investigated for leakage using traditional detection methods. Unexpected spikes and increases/ drops in DMA flows/ pressures can suggest a burst in a DMA. Automation of such DMA level analysis is one aspect of the self-learning algorithms developed for SW4EU as part of an overall continual Condition Monitoring aspiration.

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) [5]. AURA allows a data driven model of an asset to be developed by training via observation of 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. The offline application of AURA has been explored for both water quantity (for leakage) and quality [6]. AURA-Alert is a novelty-detection approach for real-time condition monitoring of assets, designed to detect anomalies in signal data stored in the cloud [7]. The model (detector) is built based on processing raw (historic) data from the monitored asset. AURA-Alert was originally a novelty-detection plug-in for Cybula's Signal Data Explorer (SDE) software [8], which was been developed to visualise and query large amounts of time series data, enabling the rapid searching for statically defined patterns across variables. AURA-Alert was previously applied offline to some TWIST data streams [9]. Historic periods of data were selected manually to create detectors and AURA-Alert applied to test periods to automatically calculate a continuous novelty score (AURA Match Distance) for unusual events. Periods of high deviation from normality were correlated with customer contact and leak repair information. See [6] for full details including a mathematical description of AURA and how a

TimeOfDay (also referred to as hours\_of\_day) channel is incorporated so that typical daily profile is incorporated.

## **3 METHODS**

### **3.1 EVP and SW4EU client**

The Event Visualisation Platform (EVP) is a flexible, scalable data management platform that focuses the user on detected events. The EVP system (Cybula technology) provides a mechanism for creating software services that can be exposed in the cloud, deployed via a SaaS model. Each service is a Java class that satisfied a particular contract and processes a defined quantity of data from a defined set of input channels to produce data that is written to one or more output channels. EVP offers a comprehensive capability for managing end-to-end workflow processes and providing remote data hosting. Data is managed as asset items on the system, and a front end client (EVP client or portal) on the data server infrastructure provides the means to explore, browse and manage the data. The EVP system also offers the capability to provide a bespoke instantiation of the underlying system and services, such that a private client (portal) can be set up within the environment to support individual projects. This client has been branded according to SW4E. As well as an online client for the live system, an offline client was utilised for purposes of historical replay of data and for exploring parameter settings for AURA-Alert with BED (using ‘run custom model’). The system as developed for SmartWater4Europe hosts multiple data types from the TWIST demo site (DMA meters, Syrinix loggers, Incertameters and AMR).

### **3.2 BED**

The prediction-classification approach is the most popular in the field and since AURA-Alert gives a similar output to a prediction (though not an actual prediction, instead a degree of abnormality or novelty per time instant) this approach was applied to create specific alerts over windows of time.

After the incoming data is cleaned (see section 3.3), AURA-Alert software is run as an EVP service to identify novel (anomalous) time steps given a previously-learned model of normality from previous data. The output can be sensitive to isolated anomalous timestamps as AURA-Alert processes each time step independently yet the water network operator is primarily interested in periods where many outliers occur in close temporal proximity so aggregation of outliers using a moving window is needed to arrive at the probability of an event occurring for each time step.

Outliers can be aggregated using a Binomial Event Discriminator (BED) in a similar manner to. A prerequisite is that each time step has previously been classified as novel or not, which may require thresholding of the output of the outlier detection methodology. Each time step is then considered an independent Bernoulli trial, with all trials in a moving window comprising a Binomial experiment. The cumulative distribution function (CDF) of the Binomial experiment gives the probability of an event, which can be thresholded to give a binary classification. A BED EVP service was developed in Java to achieve this which is applied to windows of thresholded match distances, with alerts produced based on the BED output. The EVP service provides the probability of an event given the number of outliers in a moving window.

### **3.3 AURA-Alert with automated model generation and online retraining**

AURA-Alert was rewritten as an EVP service for use on SW4EU thus allowing it to be utilised in a real-time fashion (deployed on the cloud) for the demo site and wider Reading distribution system. The current state of system parameters can be compared to the stored normal operating behaviour 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. Initially, manual detectors (that is, using data selected manually on a per sensor data stream basis) and with each model (for single channel) constructed by hand with a command line application. Whilst not practical for a large numbers of data streams, and resulting in 'static' models, this was nevertheless a useful stepping stone to full automation in terms of understanding system configuration and the size and quality of data sets required for training.

In the standard version of AURA-Alert, detectors are created by recording relationships between channels on known reference data (multi-channel state at each given instant is stored). Then new (unknown) data is checked against stored states to check for novelties (outliers or something unusual). Generally, a specialist with appropriate expertise in the domain creates such detectors. Training data is manually selected and cleaned. This is subjective and is not scalable to many thousands of channels (e.g. for AMI/AMR), particularly given that for flow and pressure meters the definition of normality, expressed in absolute units, may vary considerably between sensors. A second issue is that typically it is not possible to isolate enough training data to capture seasonality without also including abnormalities that it is not desirable for AURA-Alert to include in its normality envelope: the flow and pressure data of interest is noisy, non-stationary and may be affected by sensor and communications errors. Static models of normality are incompatible with signals that are non-stationary at the week or month scale.

These issues suggest that what is required is an event detection system that is:

- Adaptive, as it retrained either continually, at regular intervals or following a trigger event (thus trained to 'build in' current system characteristics);
- Automatically trained, so training is scalable and not subjective.

This approach was adopted to create a dynamic SW4E version of AURA-Alert (with automatically generated models) summarised in Figure 1. The AURA-Alert Java code was rewritten to create a dynamic version on EVP for SW4E:

- Apply a periodic (weekly) retrain per model with automated trigger.
- Use  $n$  (flexible but for example 2 to 4) weeks of the most recent data (or use existing detector model if no new data available).
- Data handling and Pre-processing:
  - NaN/null values are not present (excluded by services).
  - Date time stamp gaps are not filled and for small sections of these in the training period these should not significantly affect performance as data points will always be accompanied by ToD (hours\_of\_day) channel (hence no data filling needed).
  - Autoscaling for bin ranges (AURA-Alert default) was not used (extreme values/abnormalities will then set the limits of the bin ranges). Instead, the 5% and 95% percentiles from the training set are used for the bin range extremes.
  - The previous step helps keep data in the 'normal' range across bins (and removes any abnormal outliers such as negative pressures). However, extreme events at non peak/trough times will still build in abnormality in the detector (due to time of day

channel). Hence, a second phase runs the AURA model over the training data and excludes any data points over a certain match strength threshold of abnormality thereby removing extreme data at non-typical times of day from the model for situations where the training data has events/outliers.

- Generate the detector per model based on the final training data generated.
- Daily (modifiable and currently running at 4pm):
  - Run all detectors and multiple BED services (creates AURA output, BED probability and BED event channels per model).
  - Run the email alert service and send to recipients (including Comma Separated Values (CSV) file to integration portal).

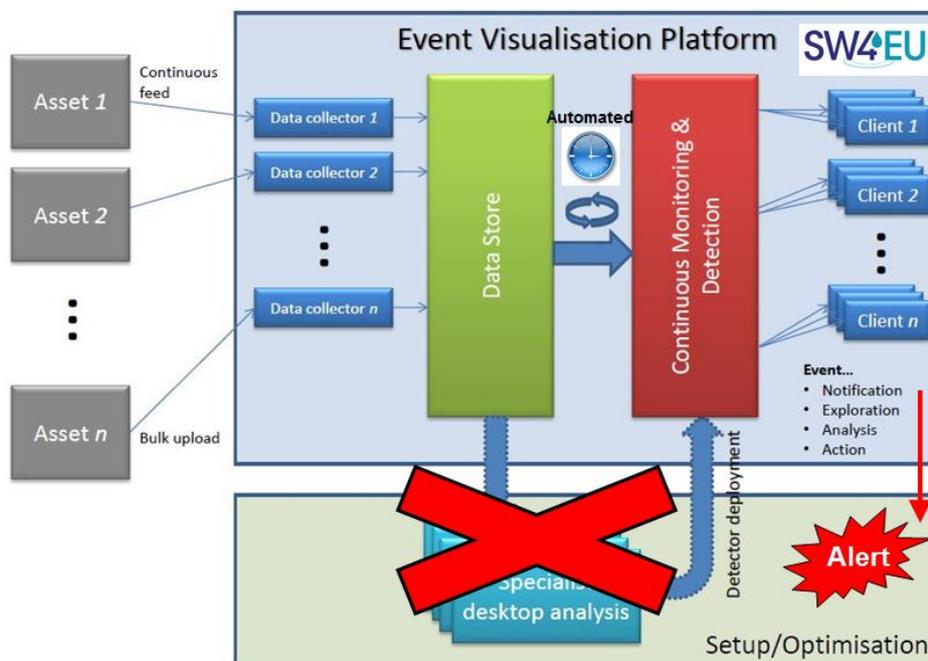


Figure 1. Automated online AURA-Alert

Parameters for the system are as follows: Number of bins in CMM, AURA Threshold, The number of training weeks, The BED window size (set as a number of sampling interval, so 8 for 15 min data is two hours), The BED probability threshold, The number of null (empty) data points allowed in a BED window (any more than this and an event detection will not be allowed).

## 4 CASE STUDY AND INTEGRATION

The demonstration site is approximately 71km<sup>2</sup> in area. It comprises 686 kilometres of distribution mains and 179 kilometres of trunk mains located in and around the city, which has many pipes over 60 years old, serving 89,000 commercial and domestic properties. Monitored data streams include 80 Network meters 14 Incertameters, and 20 Syrinix Instruments (as well as AMI infrastructure).

An integrated route allowing delivery of these data sources was implemented for a near real-time automated transfer of time series data to the cloud based EVP platform.

In order to provide details of alerts generated by the BED analysis an EVP service was developed such that the results of daily analysis (data is transferred to the cloud daily) are written to a channel specific URL, and in addition an email containing analysis of all AURA-Alert models are sent to a distribution list. Alerts take the form of an automated summary report which is read by the integration web portal being developed for alert visualisation.

## **5 RESULTS AND DISCUSSION**

### **5.1 Real events**

During 2016 - 2017, bursts and leaks occurred within the Thames Water demonstration site and many of these leaks were only found after significant time and impact. Some of these occurred in the DMAs with the extra instrumentation so these events were compared with the alerts from AURA/BED system (the system configuration was being tuned during this period). The details from these events were obtained from the corporate system and only “Main burst” events which have an active leakage rate greater than zero, were used in the test. The sample consisted of 12 events which occurred between the 08/11/2016 and the 26/01/2017, including “Active Leakage” and “Visible Leakage” events. Active leaks occur when the pipe begins to leak or has a burst but has not reached the surface remaining invisible until it is detected with specialist equipment. Visible leaks occur when the water from the leak / burst eventually makes its way to the surface. AURA/BED system generated around 200 alarms during this period; Thames Water selected 29 AURA/BED Alerts with duration of more than eight hours and compared with 12 real events. The number of opportunity windows was calculated as the difference between the earliest start date of the event and the latest start date of the event divided by the chosen length of each opportunity window. Overall the AURA/BED alerts performed better when detecting reported events with a longer opportunity window. When the window is set to one day, AURA/BED detects no events but when the window is increased to two weeks, AURA/BED detects 58% of the historic events. Figure 2 shows detections of abnormalities from the live system on a DMA flow meter.

Obtaining example real burst data is not a straightforward enterprise, unless simulated trials are performed by flushing hydrants. Records of real events in systems are not always easy to obtain, or may be inaccurate due to the systems employed to follow the lifecycle of repairs. Figure 3 shows the detection of two engineered flushes in a single DMA, conducted for the purpose of investigating detection capabilities. It shows an example of the BED in operation within the client with the raw signal, AURA output (Match distance) and BED output.

### **5.2 Discussion**

A Condition Monitoring approach for Smart Networks can allow the early detection of potential faults in assets. The integration of real time analytics can facilitate rapid determination (i.e. before customers are impacted) of abnormal flow events.

In the reality of day to day WDS monitoring, operators are unable to fine-tune parameters on per channel basis for many thousands of data streams. One of the strengths of the new system is its ability to perform data-driven self-learning with no need for classification examples (other than periods of normality used to boot-strap an automated process), manual configuration or updating.

The CMM ANNs are particularly suited to dealing with missing or spurious sensor readings and exhibit graceful degradation in such situations.

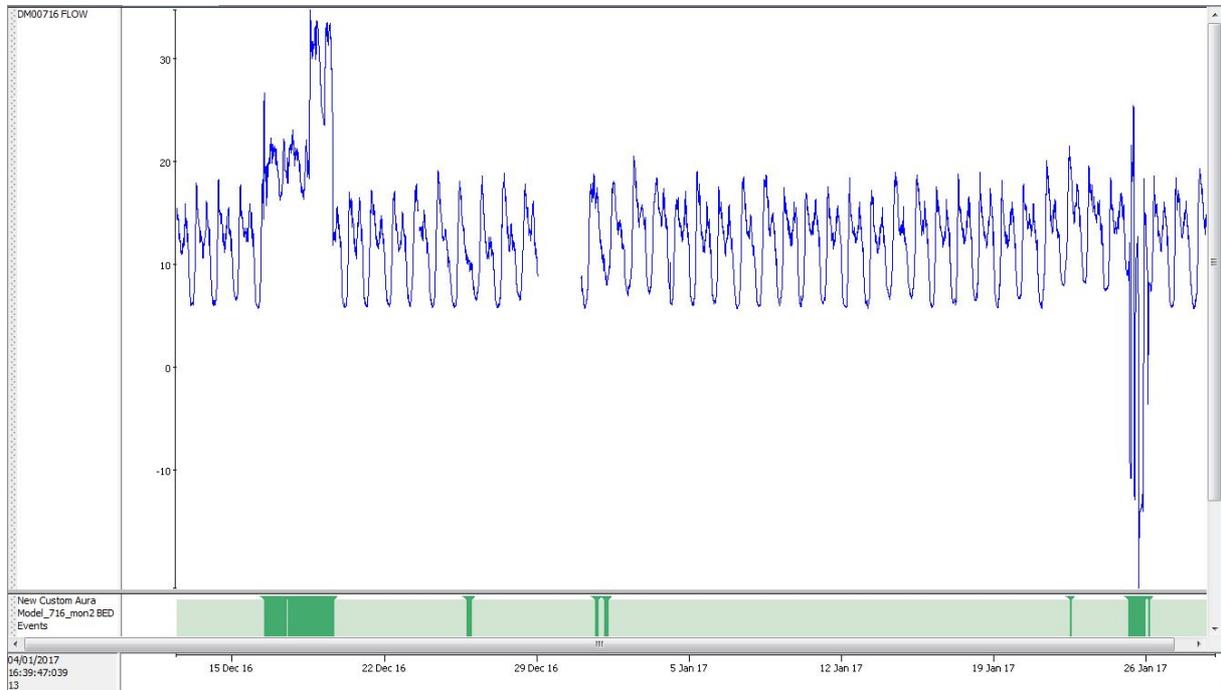


Figure 2. AURA/BED analysis of DMA Flow Meter

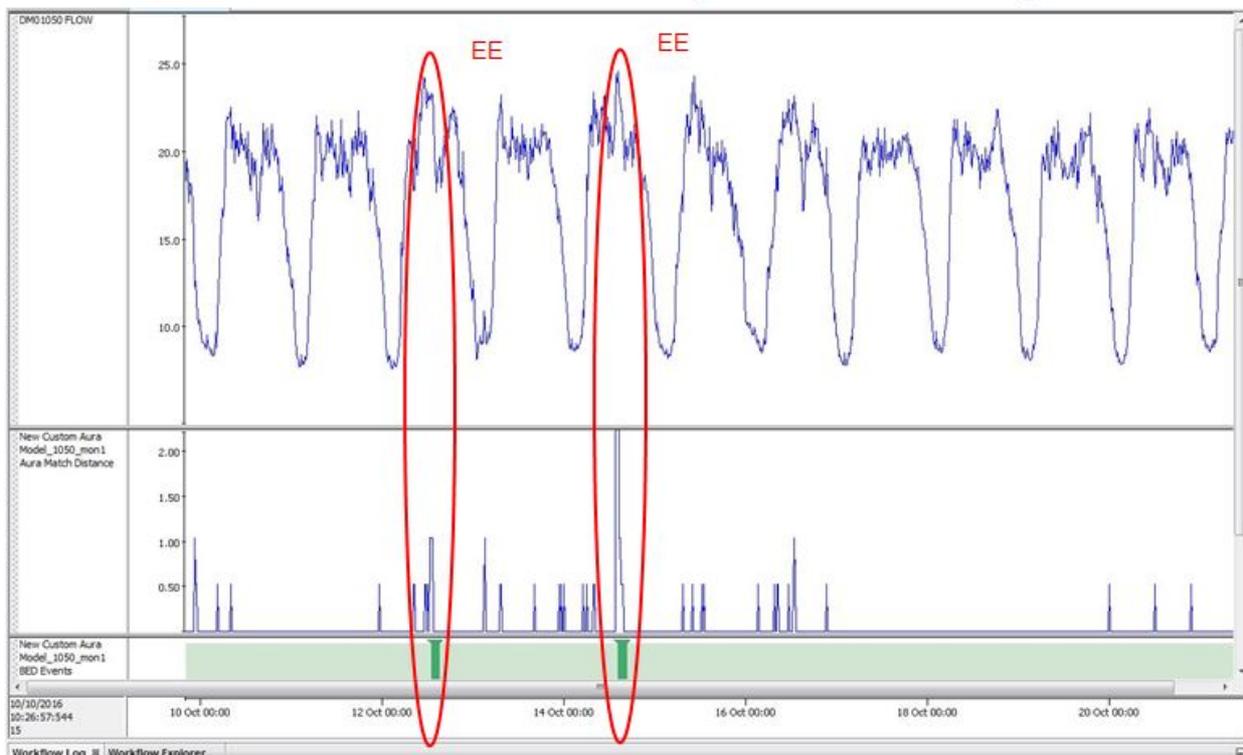


Figure 3. Engineered events detected with sensitive AURA/BED parameters

This technology has potential for use across multiple data sources and many data streams simultaneously and for other applications including water quality and wastewater. Future work will explore application of the system to AMR for customer side leak detection, by means of per customer automated models. A longer and more in depth evaluation is required for the system once parameter configurations are fully optimised and according to the trade-off required between sensitivity and specificity. Future Condition Monitoring approaches for SWN can also focus on the early detection of potential faults in the monitored system or asset, data frequency permitting, allowing preventative action to be taken before major damage occurs.

## 6 CONCLUSIONS

The condition monitoring system presented in this paper adopts a ‘bottom up’ approach to analysing WDS data from on-line monitors as this is most appropriate due to the ‘dirty’ nature of hydraulic time series data. An online self-learning algorithm (AURA-Alert) has been developed and enhanced which automates the model data selection and applies the CMM for outlier filtering. The cloud based SaaS system analyses meter data and produces an alert CSV summarising the events on a daily basis. AURA-Alert can rapidly learn and model the normal operating envelope for a system. A continuous novelty score can be utilised to enable the detection of leak, burst and other anomalous events. Key strengths include lack of per site configuration, data-driven self-learning (from periods of normality), real-time, high scalability and full automation of model retraining.

- A new client for the EVP platform was developed which allows deployment of custom code as services, chaining of services, collections of signals, display of signal outputs including AURA-Alert output. A branded SW4EU client was produced for live pilot analysis.
- AURA-Alert was integrated in the EVP platform, and the service was modified to be adaptive with automated model assembly and retraining.
- A BED service for event classification processing windows of output from AURA-Alert detectors has been developed in Java and deployed on the EVP live system. A service for delivery of alerts is being integrated into a demonstration web portal.
- Initial results for real bursts show the potential of the approach. For a historic subset of several data streams for a three month period up to 58% of bursts were detected (depending on the evaluation window size used for correlating between detections and events).

### Acknowledgments

The authors wish to thank Thames Water for data provision and assistance. In addition, thanks go to Dr. Will Furnass for his contribution to the early work. This research was funded under the EU FP7 SmartWater4Europe project, grant 619024.

### References

- [1] S. R. Mounce, W. R. Furnass, E. Goya, M. Hawkins and J. B. Boxall, “Clustering and classification of aggregated smart meter data to better understand how demand patterns relate to customer type,” Proceedings of the 16th Conference on Computing and Control for the Water Industry, Amsterdam, 9-11 November 2016.

- [2] W. Yipeng and L. Shuming, “A review of data-driven approaches for burst detection in water distribution systems”, *Urban Water Journal*, DOI: 10.1080/1573062X.2017.1279191, 2017.
- [3] S. R. Mounce, J. B. Boxall and J. Machell, J., “Development and verification of an online artificial intelligence system for burst detection in water distribution systems”, *ASCE Journal of Water Resources Planning and Management* vol. 136 (3), pp.309–318, 2010.
- [4] Olsson, G., “Instrumentation, control and automation in the water industry – state-of-the-art and new challenges”, *Water Science & Technology*, vol. 53 (4–5), pp. 1–16, Q IWA Publishing, 2006.
- [5] Austin, J., “Distributed associative memories for high speed symbolic reasoning”, *International Journal on Fuzzy Sets and Systems* vol. 82, pp. 223-233, 1995
- [6] S. R. Mounce, R. B. Mounce, T. Jackson, J. Austin and J. B. Boxall, “Associative neural networks for pattern matching and novelty detection in water distribution system time series data,” *IWA Journal of HydroInformatics*, vol. 16, no. 3, pp. 617-632, 2014.
- [7] J. Austin, G. Brewer, T. Jackson and V. Hodge, “AURA-Alert: The use of Binary Associative Memories for Condition Monitoring Applications”. In: *CM 2010 The Seventh International Conference on Condition Monitoring and Machinery Failure Prevention Technologies*, 2010.
- [8] M. Fletcher, B. Liang, B., L. Smith, A. Knowles, T. Jackson, M. Jessop and J. Austin, “Neural network based pattern matching and spike detection tools and services in the CARMEN neuroinformatics project”. *Neural Networks*, vol. 21, 1076–1084, 2008.
- [9] S. R. Mounce, C. Pedraza, T. Jackson, P. Linford and J. B. Boxall, “Cloud Based Machine Learning Approaches for Leakage Assessment and Management in Smart Water Networks”, *Computing and Control for the Water Industry (CCWI2015)*, *Procedia Engineering*, vol. 119, pp. 43-52, 2015.