High-resolution domestic water consumption data – Scope for leakage management and demand prediction

Paul Wills, Fayyaz Ali Memon and Dragan Savic

Centre for Water Systems, Exeter University, UK

***Abstract*: Challenges such as water scarcity and ever-increasing demand put an additional strain onto water distribution networks. Better asset management through leakage mitigation and demand forecasting can offset the current and future implications of these challenges. This paper shows how new high-resolution logging (e.g. 1 litre ticks) is able to enhance traditional methods of investigating leakages (e.g. minimum night flows) and instantiate novel methods for demand prediction (through micro-component analysis). Machine learning or other statistical analytical techniques coupled with the high-resolution data can be used in an adaptive way for leakage detection and demand forecasting. As a proof of concept, this paper investigates example datasets obtained from a UK based water company. The analyses suggest that it is possible to: extrapolate leakage from night flow time series data; predict water consumption patterns for different types of households and create consumption profiles based upon water user characteristics/behaviour.**

***Keywords*: leakage, water distribution, micro-component analysis, demand forecasting**

I. INTRODUCTION

Water stress is an issue that could affect two-thirds of the world population by 2025, with 1.8 billion of those people living in countries or regions with absolute water scarcity [1]. To tackle such an immense problem, it is necessary to have a range of approaches to deal with the decrease in accessible clean water and the increase of water demand. There are two areas of development that this paper will focus upon: The first is the prediction of leakage in domestic properties, using enhanced minimum night flows (MNFs), helping to reduce the water losses and therefore increase the water available for genuine consumption. The second is a better understanding of demand, how the water is being used, giving the ability to inform customers if a particular water use is considered to be excessive and suggest water saving measures. There is also the potential to have a better understanding of leakage through the flow analysis, by identifying different types of leak and potential causes for them.

In the past, the focus for water companies with respect to water losses has been on the main network, since revenue drivers provide an incentive for research into anomalies [2] in the trunk mains, such as burst pipes, with the quantification of lost water, based on using MNFs. This is normally done by balancing the water entering and leaving a district-metered area (DMA) and subtracting the estimated consumption [3]. The use of high-resolution loggers in residential properties offers an opportunity to identify leaks. The use of high resolution loggers provide two potential benefits: Firstly, a more accurate estimate of consumption, used in the leakage calculations at a DMA level, and secondly the detection of water losses within the property, which is more of an issue when water is scarce, since water companies still generate revenue from water lost at a property level.

At a domestic level, generally water meters are used that just keep a running total of water consumption, requiring a person to collect the information periodically, (e.g. monthly to 6 monthly). For end-use analysis of residential properties, surveys [4] or flow loggers [5] are the main methods of data collection. The use of higher-resolution domestic loggers has gained prominence in recent years, with technological developments in batteries

and communications making the cost of implementing real-time logging a realistic prospect, certainly for test areas. The studies conducted using high-resolution loggers to date, have shown that this is a promising field for exploration, particularly for end-use micro-component analysis (e.g. segregation of water consumption in toilets, washing machines, showers, baths using patterns within the flow trace time-series) [6]. The enhancement of the night flows analysis and demand prediction is enabled due to the installation of higher-resolution water loggers (1L pulse up to every second). Using this data, it is proposed to use machine learning and statistical methods to draw meaningful conclusions from the time–series data on water consumption.

II. METHODOLOGY

The aim of this paper is to demonstrate the usefulness of high-resolution water consumption data for a range of investigations. This has been achieved using data collected from a water consumption monitoring programme being undertaken in a region of the UK.

A. Loggers

Figure 1 – Flow data logger [7]

In order to monitor water consumption at different scales, a water company in England has initiated a data collection programme through the deployment of high resolution water meters/data loggers, as in Figure 1. The monitoring programme was commenced in June 2016 and so far around 1200 of high resolution loggers have been installed in variously distributed properties. This has resulted in the collection of 700 million water consumption values in total, logged approximately for 12-18 months. The project aims to install 1700 loggers overall, around 700 properties already have a years' worth of data. The meter data can be transmitted as frequently as every 15 minutes, giving the potential for leakage models to be tested in real-time for the future. Once the data has been collected in log-files, it is passed into a database for making analysis.

B. Survey

There is also survey information available for the demographics of the customers that are logged with these flow meters. The survey asks for information on:

- Number of occupants within each age range (16 years or less, 17-29 years, 30-44 years, 45-64 years, 65 years and over)
- their employment status (full time and part time)
- Number of residents that stay at home during the day and how often (Less than once a week, 1-3 days, more than 3 days)
- Type of property (detached house, semi-detached house, bungalow, terraced house, flat, other).
- Garden size (larger than a tennis court, ½-1 tennis court size, less than ½ of a tennis court size, none)
- Dishwasher ownership.

The survey covers both metered and unmetered customers.

C. Data extraction

To extract the data from the database, SQL queries were used to determine the information to extract and was exported into comma separated value (CSV) files. The CSV files were read into Python, using the Pandas library and then visualised using the plot.ly library, which creates html files of the graphs which can be screenshotted for use in documentation. The extracted has been anonymised.

III. RESULTS AND DISCUSSION

This section presents and discusses examples demonstrating the usefulness of high resolution data for leakage detection; micro-components analysis and data logging issues identification. Only a selected amount of data from six different properties has been analysed as a demonstration.

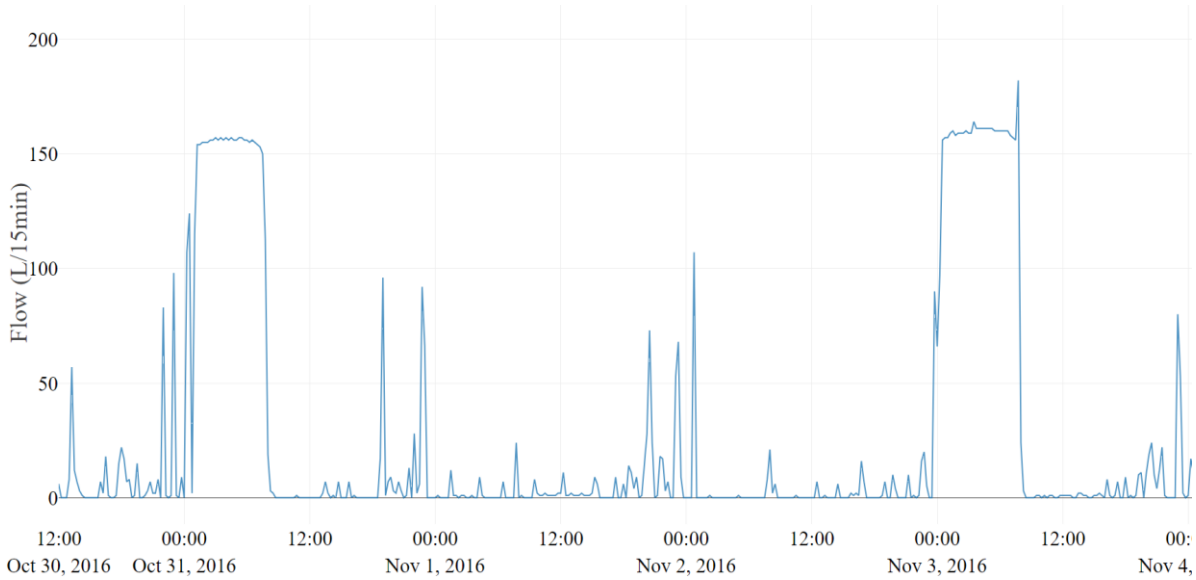

Figure 2 – Flow for property 31383, (L/15 min), with repeated night consumption.

A. Leakage

The standard practice for determining water losses is using MNFs. When looking at residential properties on an individual basis, care must be taken since properties can have unusual night flow usage.

Figure 2 shows a property having occasional high night time usage, but since it is not continuous, it is valid consumption and not due to losses. Given that it is occurring in October and November, it is difficult to speculate on the reason for such consumption, as for irrigation / sprinklers it would be expected to have periods of high usage in summer, such as the property in Figure 3, which shows continuous spikes throughout the summer as well as two periods of extended overnight use, which is consistent with sprinkler usage.

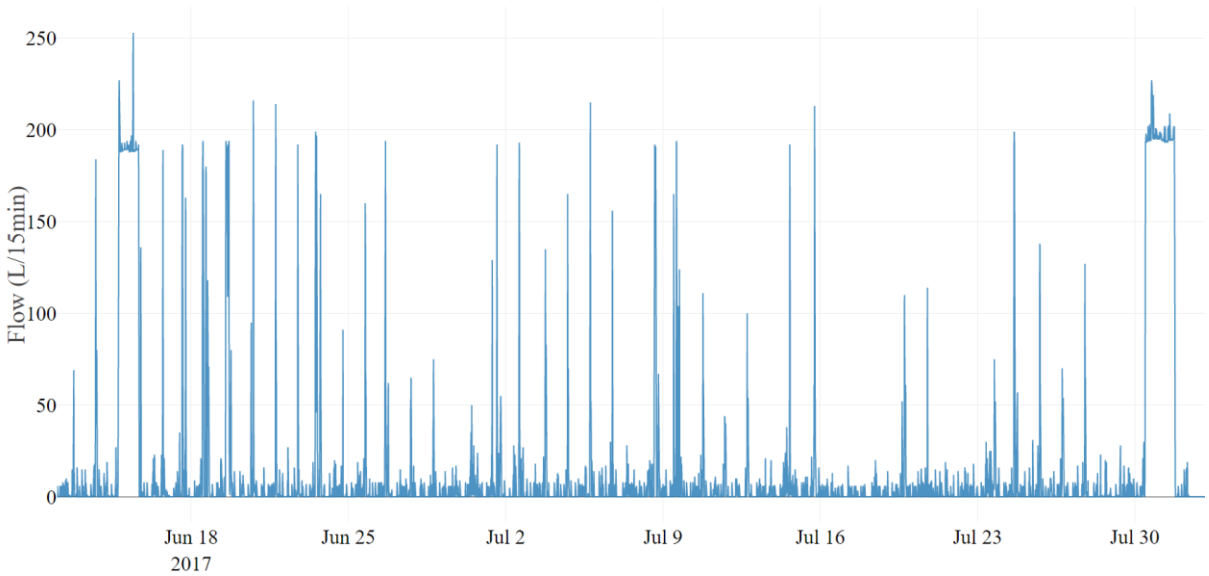

Figure 3 – Flow for property 21777, (L/15 min)

Properties that have genuine water losses would expect to have a constant flow of water with a normal consumption pattern on top of the constant flow [8], Figure 4 is an example of such a property, showing a constant flow and only returning to zero on occasion, which would be expected in the case of a leak.

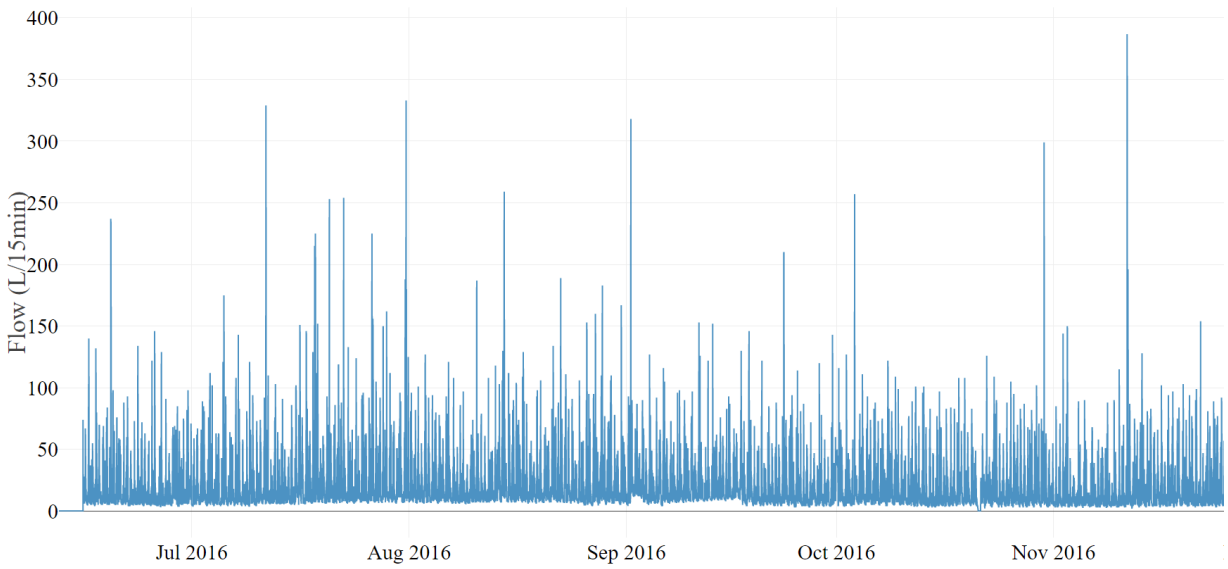

Figure 4 – Flow for property 31560, (L/15 min)

The losses here appear to be only a trickle, so would not be obvious to the owner, hence why it has continued for many months. There is a point in October where it appears to go to zero, but it is likely to be a gap in communication, since a drawback of the loggers employed in this project is that when no water is being used, no signal is sent; so, it is impossible to determine if there is genuine no use of water or if it is a gap in communication. Further analysis on the volume changes over time may provide a method of determining if constant flow is due to losses or consumption. Higher volume leaks are more likely to be noticeable, such as a toilet overflow. Such an instance would be expected to be noticed and fixed in a relatively short amount of time, such as for the property in Figure 5, for which the flow never drops below 3L/minute in a period of 40 hours.

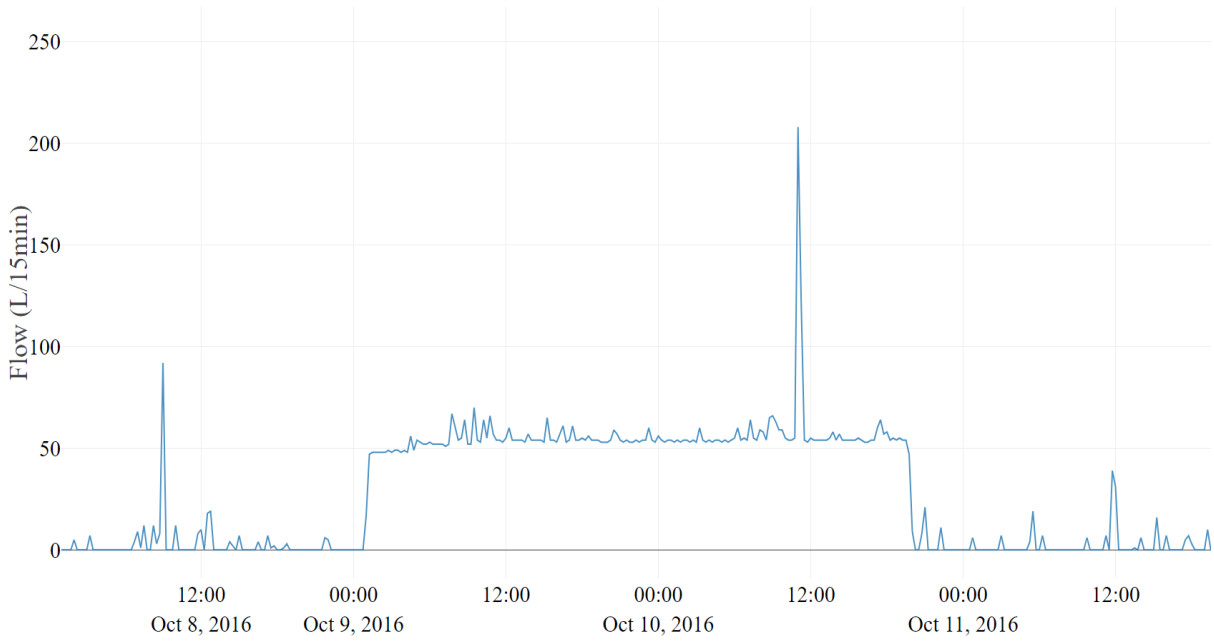

Figure 5 – Flow for property 41608, (L/15 min)

B. Micro-component Analysis

Micro-component analysis breaks down the flow trace into different components, i.e. the different appliances in the home, e.g. washing machine, dishwasher, shower and toilets. The way this is achieved for a flow time-series, is to compare the shape made on the flow-trace graph against that of a paradigm, such as in Autoflow [9]. End-use segregation through micro-component analysis is an established practice and there are already software packages that can perform this task [10].

As part of this research, the authors aim to develop a novel method using machine learning and statistical techniques that is able to automatically detect water consumption by different appliances / micro-components with an enhanced level of accuracy. In order to identify water use by different micro-components, the high resolution data logging (1L pulses every second) can be sufficient, as shown in Figure 6, which shows a typical morning pattern of toilet (same height of peak) and shower usage (constant use for 10 minutes).

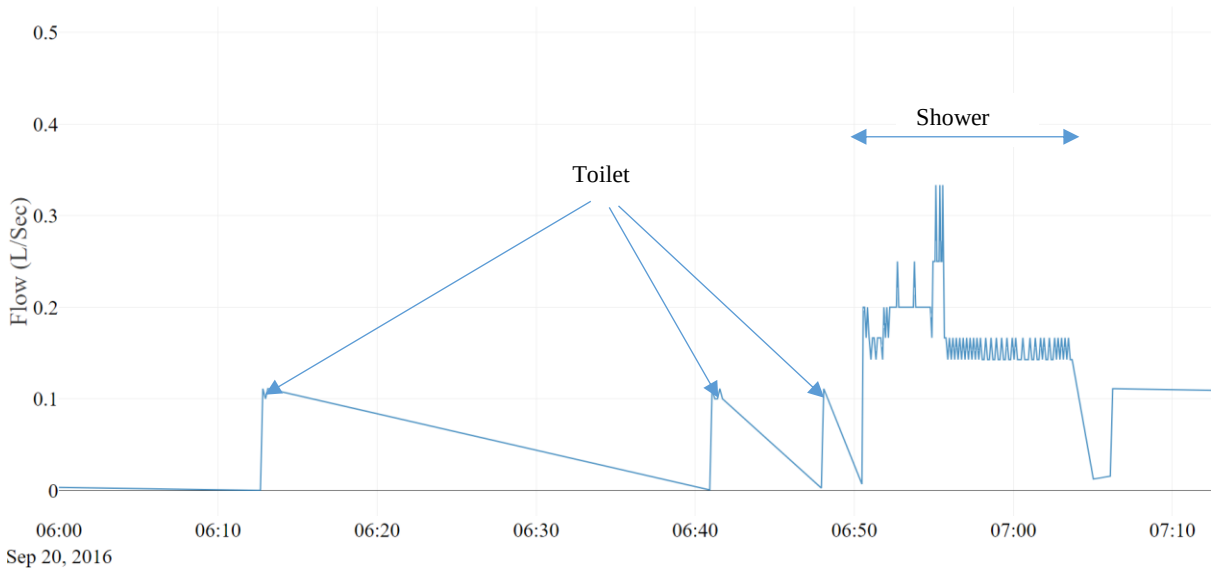

Figure 6 – Flow for property 21657, (L/s) – micro-component analysis

Demographic profiling is a method that has shown to be effective in demand forecasting when combined with micro-component analysis [11] . Combining demographic profiling with other factors, such as temperature, may help to improve both the identification of usage and the demand forecasting.

C. Logger issues

One of the main issues with the loggers is gaps in the data, through either a fault in communication or in the logger itself. These gaps in data will need to be accounted for, discussed later in future work. Unregistered water in flow meters, a combination of non-registered and under-registered volumes [12] is difficult to account for as it will vary from meter to meter and changes over time. The other potential issue is with meters having a total malfunction, such as in Figure 7, where the flow amount is constantly high (8L/min) and should be noticeable if it were a leak. There is an offline period with no readings where the problem was fixed and thereafter a normal pattern of usage is registered.

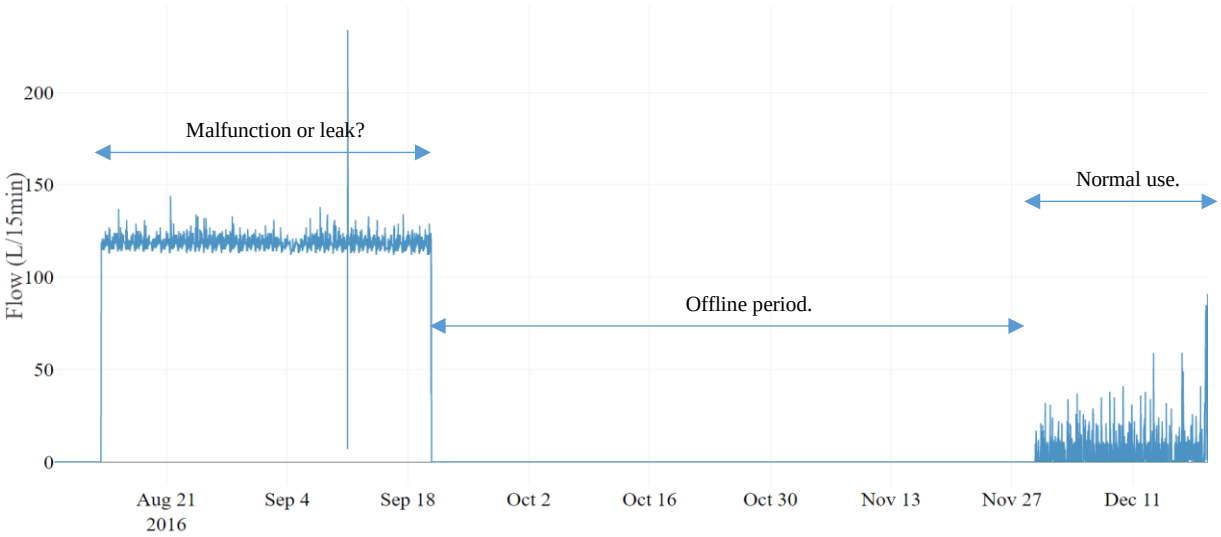

Figure 7 – Flow for property 21444, (L/15 min)

IV. FUTURE WORK

In this paper, only a segment of data from six properties has been analysed. The core of the future work will involve detailed analysis of the larger set of data from the monitoring programme with a view to develop computationally efficient data driven tools for leakage detection, micro=component analysis and demand forecasting. The future work includes the following:

- *Account for errors* - the data must be processed to account for any errors, since erroneous data will produce a substandard model. There are numerous methods for handling missing data such as large-sample maximum likelihood inference, Bayesian inference or Parametric multiple imputation [13].
- *Leakage detection* - using anomaly detection decision functions, for example, support vector machines [14], with many factors (e.g. weather data) used to decide if it is genuine consumption or water losses.
- *Micro-component analysis* – This will require multiple techniques for both pattern recognition and decision (including separation of overlapping events), for instance, the hidden Markov models and artificial neural networks used in Autoflow [9]. This will require sets of data that have already been identified, for calibration and testing purposes.
- *Leakage type determination* – Using the micro-component flow trace to extract the leak volumetric profile, analysis over time presents the opportunity to determine potential causes for a leak
- *Demand forecasting* – The information extracted, consumption and leakage, can be scaled up to create a forecast on a DMA level.

V. CONCLUSION

The main points from the analysis presented above are:

- The high-resolution data with 1 litre pulses up to every second is valid for both leakage detection and micro-component analysis.
- The data will need to be screened and adjusted to account for any errors in the flow meter or from gaps due to communication issues.
- Water loss estimates should be more accurate, since unlike the DMA level balance equations, minimum night flows will not be used blindly, therefore genuine consumption can be identified.
- Leakage detection will require a function that can decide between genuine consumption and losses, by taking multiple factors into account.
- Micro-component analysis needs to utilise pattern recognition in order to segregate the end uses, with a decision function to find overlapping events.
- All of the models will require pre-identified sets of data for calibration and testing purposes.

VI. ACKNOWLEDGEMENTS

The authors would like to thank the EPSRC funding from WISE Centre for Doctoral Training. The authors also acknowledge the provision of anonymised data and financial support from South West Water, UK.

VII. REFERENCES

[1] U. N. Water "Coping with water scarcity: challenge of the twenty-first century." Prepared for World Water Day 2007.

[2] D. Laucelli, M. Romano, D. Savić, & O. Giustolisi. Detecting anomalies in water distribution networks using EPR modelling paradigm. *Journal of Hydroinformatics* 18, no.3, pp. 409-427. May 2016.

[3] J. Morrison. Managing leakage by District Metered Areas: a practical approach. *Water 21*, 44-46. Feb 2004.

[4] W. A. Hussien, F. A. Memon, and D. A. Savic. "Assessing and modelling the influence of household characteristics on per capita water consumption." *Water Resources Management* 30, no. 9, pp. 2931-2955, Jul 2016.

[5] P. W. Mayer, W. B. DeOreo, E. M. Opitz, J. C. Kiefer, W. Y. Davis, B. Dziegielewski, J. Nelson. Residential end uses of water. 1999.

[6] T. R. Gurung, R. A. Stewart, C. D. Beal, and A. K. Sharma. "Smart meter enabled water end-use demand data: platform for the enhanced infrastructure planning of contemporary urban water supply networks." *Journal of Cleaner Production* 87, pp. 642-654, Jan 2015.

[7] *Textlog-Flow-Datasheet* [Online], Ashridge Engineering, 2017 Available: http://www.ash-eng.com/download/textlog-flow-datasheet/

[8] T. Britton, G. Cole, R. Stewart, and D. Wiskar. "Remote diagnosis of leakage in residential households." *Journal of Australian Water Association* 35, no. 6, pp. 89-93, Sep 2008.

[9] K. A. Nguyen, R. A. Stewart, H. Zhang, and C. Jones. "Intelligent autonomous system for residential water end use classification: Autoflow." *Applied Soft Computing* 31, pp. 118-131, Jun 2015.

[10] J. Morrison, & E. Friedler. (2015). A critical review of methods used to obtain flow patterns and volumes of individual domestic water using appliances. *Urban Water Journal* 12, no.4, pp. 328-343. May 2015

[11] K. Rathnayaka, H. Malano, M. Arora, B. George, S. Maheepala, and B. Nawarathna. "Prediction of urban residential end-use water demands by integrating known and unknown water demand drivers at multiple scales II: Model application and validation." *Resources, Conservation and Recycling* 118, pp. 1-12, Mar 2017[12] P. Mukheibir, R. A. Stewart, D. Giurco, and K. O'Halloran. "Understanding non-registration in domestic water meters: Implications for meter replacement strategies.", *Water*, 2012.

[13R. J. A. Little, and D. B. Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2014.

[14] S. R. Mounce, R. B. Mounce, and J. B. Boxall. "Novelty detection for time series data analysis in water distribution systems using support vector machines." *Journal of hydroinformatics* 13, no. 4, pp. 672-686, Oct 2001.