AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2017WR020639

Key Points:

- Implementation of effective smart metering systems remains a significant challenge for large commercial and public buildings
- A smart water meter design and classification system was introduced based on a problem quadrant diagram and classification table
- A generalized notation of for smart meter siting analysis has been introduced to help quantify the effects of various levels of spatial and temporal aggregation

Supporting Information:

- Supporting Information S1

Correspondence to:
E. Clifford,
eoghan.clifford@nuigalway.ie

Citation:
Clifford, E., Mulligan, S., Comer, J., & Hannon, L. (2018). Flow-signature analysis of water consumption in nonresidential building water networks using high-resolution and medium-resolution smart meter data: Two case studies. *Water Resources Research*, *54*, 88–106. https://doi.org/10.1002/2017WR020639

Received 28 FEB 2017
Accepted 21 SEP 2017
Accepted article online 29 SEP 2017
Published online 10 JAN 2018

Flow-Signature Analysis of Water Consumption in Nonresidential Building Water Networks Using High-Resolution and Medium-Resolution Smart Meter Data: Two Case Studies

Eoghan Clifford[1,2], Sean Mulligan[1,2], Joanne Comer[1,2], and Louise Hannon[1,2]

[1]College of Engineering and Informatics, National University of Ireland Galway, Galway, Ireland, [2]College of Engineering and Informatics, The Ryan Institute, National University of Ireland Galway, Galway, Ireland

Abstract Real-time monitoring of water consumption activities can be an effective mechanism to achieve efficient water network management. This approach, largely enabled by the advent of smart metering technologies, is gradually being practiced in domestic and industrial contexts. In particular, identifying water consumption habits from flow-signatures, i.e., the specific end-usage patterns, is being investigated as a means for conservation in both the residential and nonresidential context. However, the quality of meter data is bivariate (dependent on number of meters and data temporal resolution) and as a result, planning a smart metering scheme is relatively difficult with no generic design approach available. In this study, a comprehensive medium-resolution to high-resolution smart metering program was implemented at two nonresidential trial sites to evaluate the effect of spatial and temporal data aggregation. It was found that medium-resolution water meter data were capable of exposing regular, continuous, peak use, and diurnal patterns which reflect group wide end-usage characteristics. The high-resolution meter data permitted flow-signature at a personal end-use level. Through this unique opportunity to observe water usage characteristics via flow-signature patterns, newly defined hydraulic-based design coefficients determined from Poisson rectangular pulse were developed to intuitively aid in the process of pattern discovery with implications for automated activity recognition applications. A smart meter classification and siting index was introduced which categorizes meter resolution in terms of their suitable application.

Plain Language Summary The phrase, "If you can't measure it, you can't manage it" is commonly voiced in the field of resource management. As the concerns for water shortages and costs at a global scale heighten, end-users are realizing that proper measurement systems, namely water meter systems, are fundamental to increasing usage transparency in order to reduce consumption. However, a problem still exists given that the range of technologies available, combined with the diverse nature of consumption, results in end-user confusion as to the suitability of one particular measurement method over another. In this study, a broad suite of water consumption data was obtained from two large building networks using 33 m sited at discrete locations in each water network. This permitted water usage evaluation at (a) personal or population wide levels for (b) various time scales simultaneously whereby usage was characterized by water consumption patterns or so-called "flow-signatures." Newly defined activity recognition coefficients analyzed in the study helped to determine the value of siting a smart meter in the water network. The overall investigation provides the basis of a generic smart metering classification and siting index which would be of use to researchers, independent water users, or practitioners in industry.

1. Introduction

1.1. Overview

Global demand for energy and water is expected to rise by 50% and 40%, respectively (Clifford et al., 2014; World Business Council for Sustainable Development, 2009; Bauer et al., 2014). By 2025, 1.8 billion people, representing almost 25% of the world's population, will live in a water scare region (Clifford et al., 2014; U.N. Water, 2013). Despite these severe global issues and challenges, water is still not being adequately considered as a resource and is thus subject to excessive and inefficient usage, particularly in developed countries. For example, 20%–40% of Europe's water is estimated to be wasted due to poor infrastructure, consumer negligence and lack of proper resource management (European Commission, 2010) and in the United States, 3 trillion gallons (11×10^9 m^3) of water is reported to be wasted every year as a result of leaks (Eichenseher, 2010). The low tariff and price elasticity of water, coupled with the lack of detailed information available on end-user consumption has translated into a general disinterest of individual and organizational stakeholders to correctly exercise water management and conservation measures.

However, in recent years, individuals and organizations have recognized that accurate, reliable and real-time monitoring of water consumption activities is essential for effective management of water networks and encouraging conservation behavior both at the treatment plant and at the tap (Buchberger & Wu, 1995; Willis et al., 2010). This reflects the well-known statement, "If you can't measure it, you can't manage it" which has been used increasingly in the water industry (Hearn, 1998; Hood, 2002; Paisley & Henshaw, 2014). In the past, water bills issued by utility managers have been an oblique form of water consumption metric employed in residential and nonresidential buildings (DeOreo et al., 1996) which tend to mask detailed water usage activities; information which is critical to inform the end-user of their consumption habits (Cardell-Oliver, 2013) and to inform bottom-up end-use models (Creaco et al., 2017). For example, knowing that a bill equates to approximately 100 m^3 per month is interesting, but it does not inform the consumer if they are being efficient or wasteful, nor if this consumption is sustainable. As a result, it has been found that individual and organizational perceptions of water usage are often not matched with actual consumption (Beal et al., 2011a, 2011b; Hamilton, 1985; Millock & Nuages, 2010;).

With the advent of smart meters, progress has been made on obtaining detailed water consumption data at a household, nonresidential or process level which can be analyzed to inform decision making at the end-use level. In the following article, we refer to the "end-user" as an individual, homeowner, management body, or other controlling entity that is responsible for the monitoring and controlling the consumption of water. A "designer" is then referred to as the individual responsible for the design planning, modifying or modeling the WDS on behalf the end-user. Smart meters permit end-users to monitor usage characteristics at all times on a digital platform allowing greater control through water usage transparency. In particular, a significant benefit of end-user access to smart meter data is the ability to intuitively recognize and self-report normal usage characteristics (baseline flows) and flow patterns which reflect specific usage activities in a process called flow-signature analysis (Cardell-Oliver, 2013; DeOreo et al., 1996).

A smart meter operates by using a sensor to detect flow in a pipeline using some physical means which is thereafter converted into a readable signal by means of pulses (for example, 72 pulses per liter was adopted by Beal et al. (2011a, 2011b)) or direct digital outputs at fixed rates. The acquisition frequency $1/\Delta t$ (i.e., number of pulses per unit time) defines the meter data acquisition temporal resolution which is an important parameter for data mining and pattern discovery. On the other hand, the spatial position of the meter in a water distribution system (WDS) is also critical in determining the various levels of activity that can be observed (Buchberger & Wu, 1995). The outcome of these two parameters: meter resolution and spatial siting, renders smart meter planning, or modeling end-use a complicated task where low-resolution aggregated data sets can mask useful information. Conversely, large numbers of high-resolution meter flow traces can be expensive, complicated and impractical. Furthermore, due to the heterogeneous nature of both the residential and nonresidential water end-user, there is no generalized understanding available to choose one method over another for a specific application.

As it will be described next, there have been a large number of past investigations in this context both in terms of end-use monitoring and end-use modeling, primarily in the residential context. Despite the significance of these studies, none have drawn a more general, holistic conclusion on the merits and weaknesses of the various orders of temporal or spatial data aggregation as applied to pattern detection for conservation monitoring or for modeling purposes. Therefore, in this study, the authors generalize an understanding and an approach to designing smart metering systems by analyzing data characteristics determined from both high-spatial and medium-spatial and temporal resolution data sets. In particular, as explicitly stated by Buchberger and Wu (1995) and Creaco et al. (2017) from the context of modeling, the correct estimation of demand in a system requires that sufficient high-resolution flow-signature characteristics be determined for the nonresidential sector. Therefore, there is a requirement to perform numerous experimental campaigns on water usage characteristics outside the household context (Creaco et al., 2017), which forms another primary focus of this article.

A comprehensive smart metering campaign of both medium-resolution and high-resolution meter systems (total of 33 m acquiring approximately 150 million data points over 6 months) was implemented on two nonresidential case studies in order to mine data and characterize pattern discovery at the various spatial and temporal resolutions. The findings of the study proposes a number of fruitful hydraulic parameters that can aid in smart meter planning to obtain minimal data sets to minimize costs and maximize the value of information obtained. The authors introduce a classification system for smart meters developed in terms of data acquisition/flow trace resolution and their respective application.

1.2. A Review of Water End-Use Studies

Water end-use studies, or “microcomponent” studies (Beal et al., 2016; Otaki et al., 2008) can provide a fundamental understanding of typical water usage characteristics in a particular setting for the purpose of establishing water demand management strategies, fault detection diagnostics (FDD) processes and forecasting usage. In the residential context, water consumption can account for up to 66% of the total water supplied by the utility and therefore has significant potential for conservation (Cahill & Lund, 2012; Willis et al., 2010). Academic discussions on residential water conservation methods have been ongoing since as early as 1910 (Cahill & Lund, 2012; Van Hise, 1910) but following the growth of smart metering technologies, recent studies have utilized various meter types, number of, resolution of acquisition (temporal aggregation), data transfer and activity recognition methods (Beal & Stewart, 2013; Cahill & Lund, 2012; Cardell-Oliver, 2013; Giurco et al., 2008; Hamilton, 1985; Larson et al., 2012; Willis et al., 2010, 2011).

Temporal aggregation (or disaggregation) of smart water meter data is classified in terms of low, medium, and high resolution. Low-resolution weekly and monthly end-use studies were performed by Grafton et al. (2011), Fox et al. (2009), and Jorgenson et al. (2009). These studies focused primarily on univariate statistics for the purpose of analyzing socioeconomic factors of household consumption (Grafton et al., 2011). Medium-resolution end-use data studies have been used to analyze diurnal flow-signatures in addition to normal and peak daily volumes. An early attempt to disaggregate usage activity in the home was carried out by Bennett (1975). Britton et al. (2008) later used hourly meter data to identify domestic leaks. Cardell-Oliver (2013) again analyzed household consumption using hourly data obtained from 187 houses from the real-world smart metering trial run by Western Australian Water Corporation (Cardell-Oliver & Peach, 2013). Beal and Stewart (2013) also adopted hourly data (aggregated from 5 s data) to disaggregate diurnal flow-signature patterns in the household. Fewer studies have been performed using high-resolution data. An early but comprehensive high-resolution flow-signature analysis study was undertaken by DeOreo et al. (1996) where more than 10,000 water end-use events in a domestic context were measured at 10 s intervals using magnetic sensors. The study revealed that high-resolution flow traces were capable of discerning very sensitive signatures associated with all major and minor water use categories (DeOreo et al., 1996). Otaki et al. (2008) adopted high-resolution metering on the basis of a low-flow water meter measuring domestic appliances water use at 1,000 pulses/L/min. Willis et al. (2011) also applied high-resolution meter recordings at 10 s intervals to disaggregate domestic end uses.

Although residential consumption accounts for a substantial share of total water supplied (Willis et al., 2010), nonresidential buildings also retain a significant portion. For example, nonresidential buildings account for 21% of the total water usage reported in Europe (European Commission, 2012) and 15–20% in the United States (Ploeser et al., 1992). As stated by Ploeser et al., (1992), “a relatively small percentage reduction in non-residential water use constitutes meaningful demand and drought management strategies” which renders nonresidential water conservation a good investment. Despite the importance of this sector, it has received less attention than the previously discussed residential sector in the development of water conservation initiatives. This is largely due to the heterogeneous nature of the nonresidential sector as well as the lack of information regarding end-uses of water (Buchberger & Wu, 1995; Creaco et al., 2017; Dziegielewski, 2000). Despite the fact that there has been a growth in the commissioning smart water monitoring systems in nonresidential buildings (Bullington et al., 2008; Dziegielewski, 2000), there is a distinct lack of academic research to establish best practice guidelines in this sector as compared to its residential sector counterpart. Moreover, current WDS demand models such as stochastic models (Buchberger & Wu,

1995) or Poisson rectangular pulse (PRP) models, still lack sufficient microcomponent behavior in the nonresidential sector to construct total municipal WDS demand models (Creaco et al., 2017).

1.3. Water End-Use Modeling Using Pulse Techniques

In WDS modeling, the rectangular pulse process has been widely used in the literature to generate synthetic time series of water demand in the residential context (Alvisi et al., 2003; Blokker et al., 2009; Buchberger & Wu, 1995). In these approaches, the issue of water demand generation is determined from small temporal and spatial resolutions through pulses featuring arrival time, duration and intensity which are then bottom-up aggregated for nodal demand reconstruction (Alvisi et al., 2003; Blokker et al., 2009; Buchberger & Wells, 1996; Buchberger & Wu, 1995; Creaco et al., 2017; Di Palma et al., 2014). In particular, these approaches have generally found to simulate actual demand levels reasonably well.

The above mentioned models fall into two different categories: (1) stochastic processes are used to reproduce overall water demand as a whole and (2) models which use the characteristics of individual microcomponents (appliances) to determine the overall consumption by summation. Because both methods operate at different spatial scales (household for the former and appliance for the latter) they require different parameters which is discussed in detail by Creaco et al. (2017). In summary, in order to parameterize these models, specific end-use statistics are required at the appliance level. For example, for the second category, the only model available SIMDEUM (SIMulation of water Demand, and End-Use Model) is parameterized from occupant data coupled to known water usage data from the household, information which is not readily available in the nonresidential context (Creaco et al., 2017).

1.4. Effect of Time Series Resolution on Flow-Signature Analysis

Flow-signature analysis can be applied for various flow trace resolutions depending on the type of end-use study being considered. For example, Cardell-Oliver (2013) applied flow-signature detection methods to medium-resolution, nonsmooth time series (water usage aggregated over 24 h periods) using cluster analysis. A seven-step algorithm was proposed to identify flow-signature patterns at an individual household level and scaled to large populations and over long time scales. For high-resolution time series, activity recognition has been used to determine water usage activity at the individual personal level. Activity recognition can, manually or automatically, be applied to a continuous high-resolution time series where typically the duration of the event and the flow of the event are the identifying metrics. DeOreo et al. (1996) applied signature detection on end-usage events occurring on the order of seconds. Activity recognition programs are often used to identify and enumerate end-usage activity. For example, Beal and Stewart (2013) and Nguyen et al., (2013) utilized TraceWizard© (Aquacraft, 2010; Mayer, 2006) which is a commercial water usage activity recognition software package. Nguyen et al. (2013) determined the problems associated with applying a flow trace software and proposed their own recognition algorithm and compared it to TraceWizard©. Beal and Stewart (2013) also applied TraceWizard© to disaggregate household activities based on a flow trace which identified peak day and hour demands.

Pressure-based activity recognition methods in the domestic setting were applied by Froehlich et al. (2011) who introduced HydroSense (Froehlich et al., 2011). Froehlich et al. (2011) described that the HydroSense, pressure-based approach, have the advantage over flow trace activity recognition in that it can identify the location and/or specific fixture or appliance that was used to create the flow event. Flow trace activity recognition software have demonstrated 80% accuracy for isolated events, 24% accuracy for overlapping events (also known as compound events; Froehlich et al., 2011) using two fixtures and 0% accuracy for three or more fixtures (Froehlich et al., 2011; Wilkes et al., 2005). On the contrary, pressure-based sensors have been shown to provide approximately 91% accuracy in the domestic setting for isolated events and 68–82% accuracy for compound events depending on how closely the events overlap in time (Cardell-Oliver, 2013; Froehlich et al., 2011).

2. Data Mining and Flow-Signature Analysis

2.1. Flow Sensor Data and Data Mining

Flow data signals from smart meters generally come in one of two formats where the time stamp t_i is used to designate the temporal position of that reading. In the first format, a volume V is determined by the meter over some small time interval Δt beginning at time stamp t_i where the flow rate in that time period is given by $Q_i = V_i/\Delta t$ (known as duration weighted average flow; Creaco et al., 2017). An example of this is a displacement meter which delivers a pulse to identify when a liter of water has passed through the meter; a common application of this is then a count of pulses within a predetermined time interval Δt which designates the resolution of the meter. The second format is that the meter generates a flow measurement reading Q_i instantaneously at time stamp t_i (for example, a magnetic or ultrasonic sensor). In either case, Δt defines the resolution of the meter which can be further aggregated into further data sets of lower resolution (Beal et al. 2011a, 2011b). It is the temporal resolution of the data (seconds, minutes, or hours) over a particular time period T (minutes, hours, or days) in data set then that defines the level of end-use activity that can be exposed. The data set may be presented as a volume aggregated discontinuous time series or a duration weighted average in a bar chart format (as used by Cardell-Oliver, 2013) or a continuous flow trace (Beal & Stewart, 2013; DeOreo et al., 1996) where the former is better suited to medium-resolution and low-resolution data visualization.

In addition to temporal aggregation (or disaggregation of data), we introduce the property of spatial aggregation (or disaggregation) as the process of siting a meter to amalgamate activities (or isolate activities) across a particular WDS. For example, locating a meter at the top of a WDS hierarchy (i.e., just downstream of the mains junction) will aggregate all the activity occurring within the site of interest into one single flow trace (100% spatial aggregation). On the other hand, placing the meter, lower down in the WDS hierarchy closer to the end-use events of interest will isolate and expose microcomponent activity (i.e., just upstream of a bathroom facility in Figure 1).

To better outline this bivariate temporal and spatial pattern discovery process, Figure 1 outlines an example of a relatively simple metered WDS comprising a mains water supply (MWS) line and a domestic hot water (DHW) line and various end-uses (toilets, sinks, showers, etc.). In this example, it is shown how group wide activity can be determined on a spatially aggregated medium-resolution flow traces. If for example, a peak day (established by exceptional shower usage) was detected in the daily water usage time series (b) and required further examination, data mining can reveal the approximate time of the day where by the usage was detected in a diurnal signature (c). Monitoring such activity can provide valuable statistical information for modeling. Thereafter, in order to better determine the source of the activity, spatially aggregated high-resolution data sets can hone in on the hourly period (d) to determine an aggregated flow-signature.

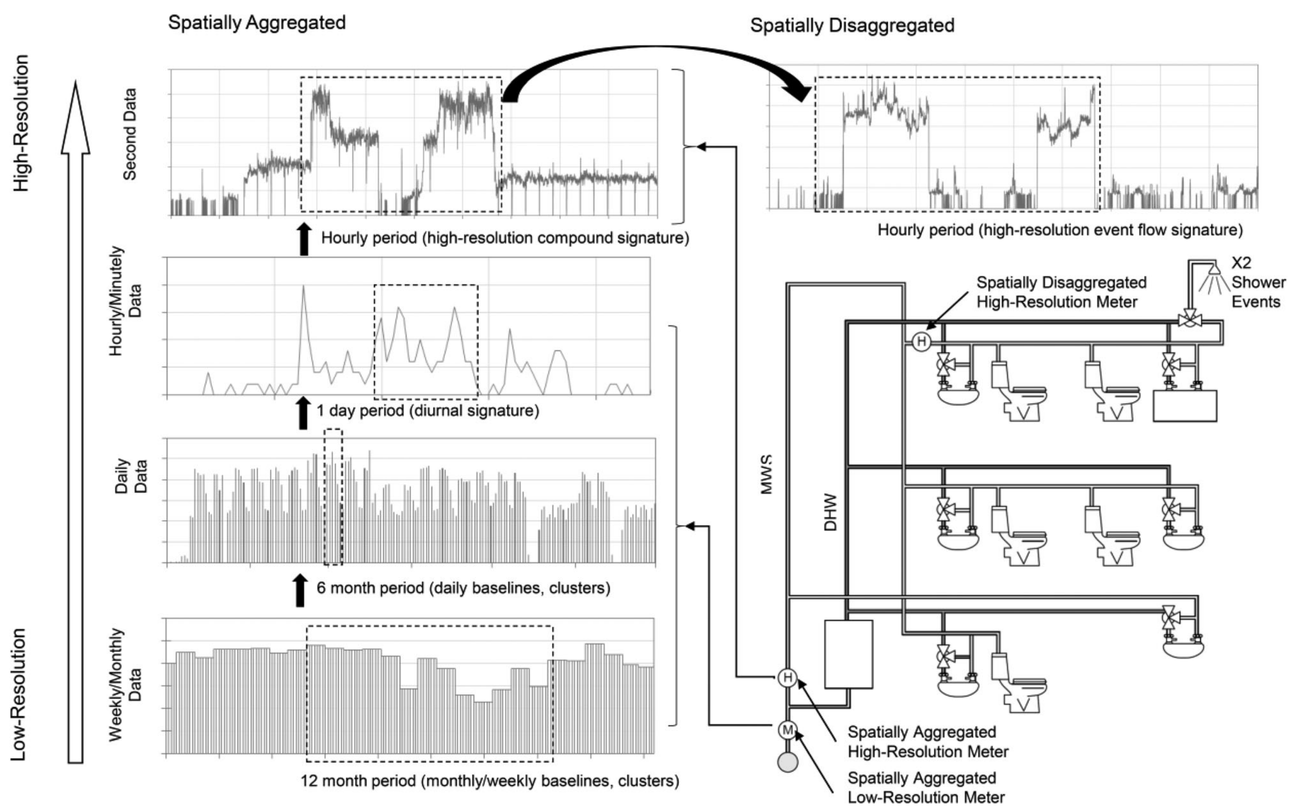


Figure 1. A summary of the process of data mining in a simple water network using water usage data of various spatial and temporal resolution.

However, because the high-resolution meter acquires 100% spatial aggregation (numerous compound events (Buchberger & Wu, 1995), a spatially disaggregated meter (2) branches deeper in the network (25% aggregation) would be capable of exposing the exact end-uses (two shower events in this case). In the following section, we generalize notation to help quantify the effects of various levels of spatial and temporal aggregation as discussed.

2.2. Medium-Resolution Flow-Signature Analysis

Due to the levels of aggregation at the meter, smart meter time series, particularly medium-resolution traces, are nonsmooth. The readings can be further aggregated from their original resolution Δt to larger observation periods Δt_{agg} to expose new patterns: (for example, hourly meter reading can be accumulated into daily readings to determine daily baselines and peak use days). The signature pattern can then be identified from a set of metered observations t_i, v_i where the data satisfies the logical requirements of a particular flow-signature or cluster. The analysis of medium-resolution in the current study is leveraged by the approach of Cardell-Oliver (2013) where cluster analysis was used to identify normal usage, peak day and continuous flow-signatures from a daily aggregated time series. The methods of Beal et al. (2011a, 2011b) were also leveraged to assess aggregated hourly time series to investigate diurnal flow-signatures and peak usage characteristics.

2.3. High-Resolution Flow-Signature Analysis

Unlike medium-resolution end-use patterns, high-resolution flow-signatures are characterized by the event duration and flow intensity as discussed by Buchberger and Wu (1995). In water end-use studies, particularly those involving nonresidential monitoring, there are large flow and temporal variations between the minimum and maximum flow events due to heterogeneous activity which renders recognition of events relatively difficult due to the frequency of compound events (Buchberger & Wu, 1995). For example, Figure 2 provides a typical example of a running tap occurring at the same time as a shower and a laboratory usage

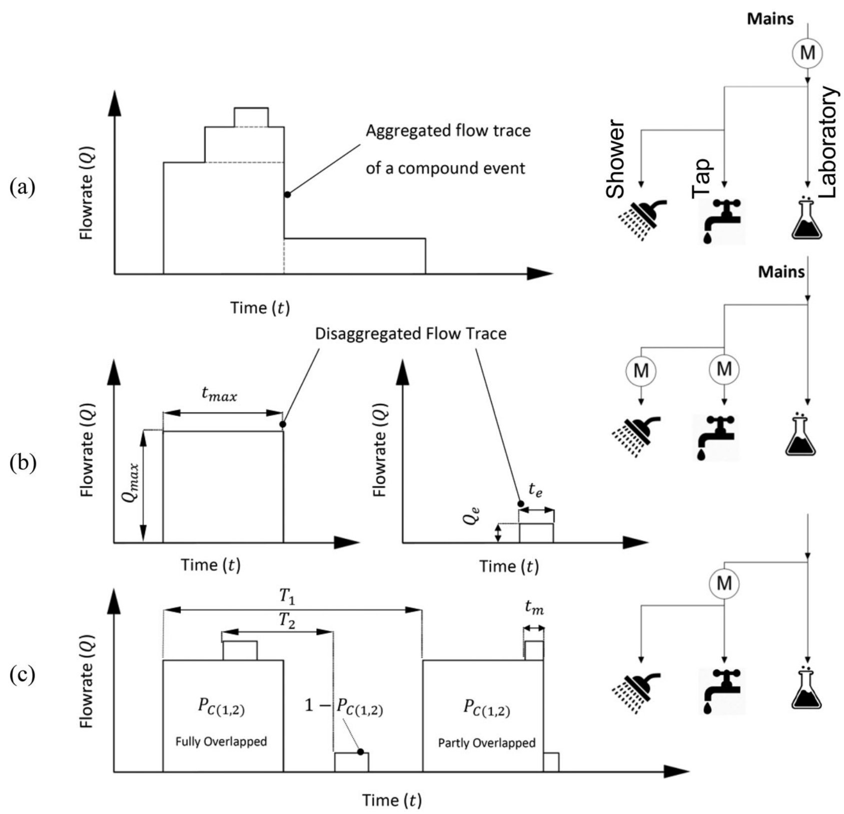


Figure 2. High-resolution smart metering approaches including (a) a spatially aggregated flow trace of a compound event comprising three different end uses, (b) spatially disaggregated using two individual submeters to identify specific end-usage information (such as Q_e and t_e), and (c) an example of the probability of two end-use events 1 and 2 coinciding.

event displayed on a spatially aggregated flow trace (Figure 2a). In such a case, the tap event may not be discernible amongst the shower and laboratory usage as shown in Figure 2a.

In order to resolve this, we can adopt methods from the Poisson rectangular pulse (PRP) process to investigate the quality of a high-resolution flow trace. The notation was developed to introduce a number of coefficients that describe the quality of a data set for activity recognition purposes. The flow C_Q and temporal C_t magnitude coefficient are used to describe the ratio between the maximum and minimum flow events on the aggregated flow trace by the following equations:

$$C_{Q(\max,e)}=\frac{Q_{\max}}{Q_e} \tag{1}$$

where $Q_{\max}$ is the average flow rate (intensity) of the largest flow event and Q_e is the average flow rate of the event of interest. Similarly

$$C_{t(\max,e)}=\frac{t_{\max}}{t_e} \tag{2}$$

where $t_{\max}$ is the average duration of the largest duration event and t_e is the average duration of the event of interest. The largest event parameters (maximum duration and intensity) were chosen as a baseline measure of variation within a particular site (i.e., a measure of the maximum level of concealment that could be experienced for the event of interest). Another parameter required for signature detection is the meter data acquisition resolution. In order to capture each event at sufficiently fine resolution that its duration and intensity characteristics are apparent in the flow trace, a meter data resolution coefficient is introduced by

$$M_t=\Delta t/t_e \tag{3}$$

which should be sufficiently small, i.e., the required event duration should be much larger than the meter resolution $(\Delta t \ll t_e)$ (hence the requirement for high-resolution flow traces as discussed before).

If say the end user deemed it necessary to disaggregate, for example, both the shower and tap flow events of Figure 2 for monitoring or modeling purposes, this may justify installing two submeters on the water supply line to spatially disaggregate the tap and shower events, respectively (see Figure 2b). However, instead of requiring two submeters (Figure 2b), the designer may instead decide to save costs and only install one submeter just upstream of the branch to both the shower and tap end-use. However, as shown by previous research, the ability of activity recognition from a single aggregated flow trace is also largely dependent on the levels of overlapping of compound events (Froehlich et al., 2011; Wilkes et al., 2005).

Thus, in order to justify whether the single submeter approach is required on the main line or whether a number of submeters is required (case b), the probability overlapping between two events of interest (1,2) is given by $P_{C(e,\max)}$ which was applied based on a model provided by Richards (1948) who determines the probability of coincidence between two periodically recurring events.

With reference to Figure 2c, if it is assumed that both events are independent of each other, they follow a Poisson distribution where $N_{(T)}$ is the number of events that have occurred up to time t (starting from zero) and recur at a constant rate of $N_{(T)}/t$. The duration of an events is denoted by Δt_e and $\Delta t_{\max}$ and the minimum satisfactory duration of overlap (coincidence) is given by t_m. Thus, the probability of at least one satisfactory coincidence between event 1 and 2 is given by (Richards, 1948)

$$P_{c(e,\max)}=P_o+wt_c\ ,\ \text{for}\ \Delta t \le Max(t_e,t_{\max}) \tag{4}$$

where

$$t_c=T-t_m \tag{5}$$

$$P_o=(t_e-t_m)(t_{\max}-t_m)/(E_eE_{\max}) \tag{6}$$

$$w=(t_e+t_{\max}-2t_m)/(E_eE_{\max}) \tag{7}$$

and where the T is the time period under investigation and E_e and $E_{\max}$ are the time intervals between repeating events of interest.

Then, the ability to distinguish an isolated event of interest from a single flow trace (using manual or activity recognition algorithms) will depend collectively on the parameters $C_{Q(e,\max)}$,$C_{t(e,\max)}$, $P_{c(e,\max)}$, and M_t. For example, if $C_{Q(e,\max)}$, $C_{t(e,\max)}$, and M_t are all large, the potential to identify the required event would be too difficult due to the levels of concealment in the spatially aggregated flow trace. If $P_{c(e,\max)}$ was high only, this would then justify the need to install a submeter to aid in the spatial disaggregation process on the line of interest. However, if $P_{c(e,\max)}$ is low but $C_{Q(e,\max)}$, $C_{t(e,\max)}$ are large, a single spatially aggregated high-resolution meter may only be required for the building network to gain a valuable data set at a minimal cost. Each of the parameters is investigated in this study in order to make design suggestions for optimal siting of smart meters.

3. Pilot Site Overview

3.1. Description of Pilot Site 1

Coláiste na Coiribe is a newly constructed secondary school (high-school equivalent with students aged between 12 and 18) in Galway, Ireland accommodating up to 720 students and 25 teaching and administrative staff. The water system (as outlined in supporting information Figure 2) is fed primarily from a single mains water supply (MWS) which conveys cold water to all floors and also tops-up a 6,000 L cold water storage tank. The cold water storage supplies the cold water supply (CWS) and provides a feed to the domestic hot water (DHW) circuit which primarily feeds showers and hot taps. The water infrastructure also houses a rainwater harvesting system which collects rain water runoff in a 37 m^3 storage tank (roof areas, pavements, etc.) and pumps it to a 9,000 L grey water storage tank in the attic which in turn feeds the grey water supply (GWS) serving the toilet and urinal flushes in the building. The MWS provides a top-up to the grey water storage tank during dry periods. The water infrastructure is monitored by 14 in-line water meters (B-Meters, Italy: meter model type varied) equipped with a magnetic pulse output. Information on the meters used and location is available in supporting information Table 1. The in-line displacement meters record data at a frequency of 1 pulse/L. The measurement error for the in-line meters is dependent on the flow with the maximum reported value of $\pm$5%. A programmable logic controller logs data at 7.5 min intervals which is stored on a cloud database.

3.2. Description of Pilot Site 2

The National University of Ireland (NUI), Galway engineering building is a state of the art educational facility accommodating 1,100 students and 100 staff with 14,000 m^2 of floor space. The existing water network (as outlined in supporting information Figure 3) comprises a system of pressurized copper pipes supplied principally by the MWS. The rainwater harvesting system comprises a 75 m^3 tank which collects rainwater and subsequently pumps it to two header tanks located on the east and west side of the building's roof. The GWS coveys grey water to the toilets and urinals in the building. The MWS feeds both the CWS and DHW and provides a top-up to the rain water harvesting system when required during dry periods. The water network is fitted with 11 in-line positive displacement meters (B Meters, Italy) fitted with a magnetic pulse output and 8 ultrasound flow (USF) meters (VTec, the Netherlands). The in-line displacement meters record data at a frequency of 1 pulse/L. The Building Management System (BMS) logs data from eight of the in-line meters at 7.5 min intervals and the remaining three in-line meters at 15 min intervals and a cloud service is used to store the data. The USF meters UFM-TM/TS (VTEC Engineering) are nonintrusive and rely on principles of ultrasound to determine the velocity and rate of flow through the pipe network. During this study, the velocity and flow were reported at a high-resolution of one second intervals and the data was stored on a BeagleBoneBlack (BBB) computer located at the meter control board. According to VTec, the meters claim an accuracy $\pm$1% at a velocity $v \geq 0.6$ m/s. Information on the meters used and location is available in supporting information Table 2.

4. Medium-Resolution Data

4.1. Establishing Normal Usage and Baselines

Figures 3a and 3b highlight examples of flow traces (flow time series) obtained for both pilots. The flow traces present daily usage which is obtainable by aggregating consumption throughout the 24 h period. When analyzing flow traces of this order for the purpose of end-use signature discovery, Cardell-Oliver (2013) found that subsets of unique observations within the data can exist which are referred to as clusters. Cluster analysis is

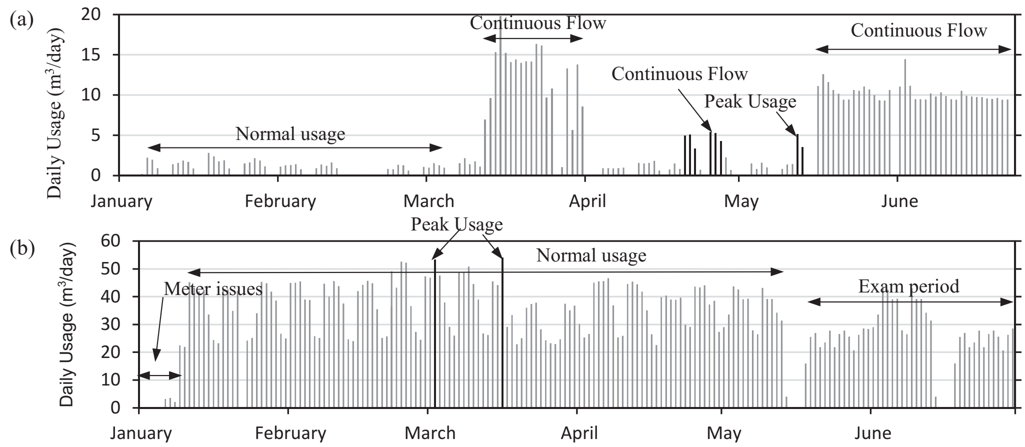
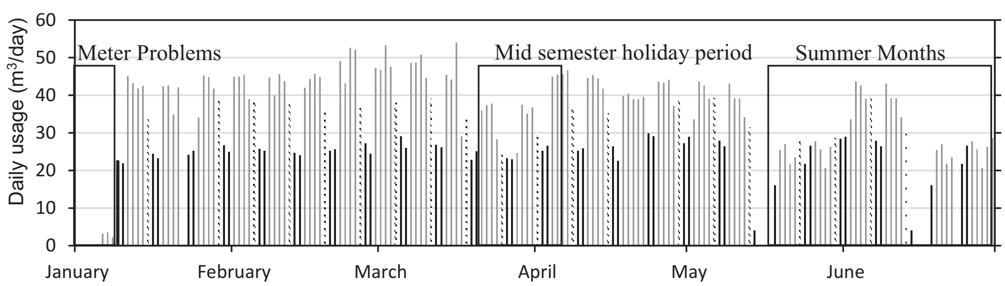


Figure 3. Medium-resolution flow traces of daily mains water usage for (a) CnaC and (b) NUIG engineering building where peak flow is outlined by the black bars.

then the process of identifying one or more clusters that satisfy the logical requirements of a particular flow-signature. Through observation of each data set, three medium-resolution clusters were identified corresponding to normal usage (baseline), peak days, and continuous flows. The normal usage or baseline flow is discernible in the flow traces of Figures 3a and 3b where a general usage characteristic appears to underpin the data set. The peak use day water usage cluster is identified as the cluster of all days where the water usage is exceptionally high (Cardell-Oliver, 2013). For the current study, we define a peak use day whereby the flow in that day exceeded the baseline plus two standard deviations. In the case of Coláiste na Coiribe, there were only 6–8 days where there were singular peak water usage incidents. In the case of the engineering building pilot, two peak water usage events were recorded on the MWS during the monitoring period. The third cluster type for the pilot designates a continuous flow which is discernible by an exceptional and relatively constant use of water been maintained for 24 h or more and are typically attributed to a leak.

Based on an expert understanding of the WDS (i.e., main water end-uses, WDS diagrams) and by removing outliers (e.g., peak and continuous flows) from the data set, it was possible to apply further data mining techniques to the flow traces to identify additional subclusters. For example, Figure 3 highlights the total building water usage at each pilot by adding the MWS and the GWS meter readings and removing the outliers (defined by fault incidents) to generate a relatively stable data set representing population behavior



Detail	Notation	Coláiste na Coiribe			Engineering Building		
		Norm	Fri	WE	Norm	Fri	WE
Average daily usage (m^3)	$V_{d(Avg)}$	4.42	2.73	0.00	43.29	36.76	25.50
Daily standard deviation usage ($\pm$ m^3)	$V_{d(SD)}$	0.4	0.27	0.00	4.74	2.40	25.50
Norm = Normal weekday cluster, Fri = Friday cluster, WE = Week end cluster							

Figure 4. NUI pilot: medium-resolution flow traces for the main water supply line highlighting each subcluster: (grey single line) normal weekday, (dashed line) Friday, and (thick single line) weekend usage. The boxed regions were excluded when calculating the baseline for each subcluster. The table outlines a summary of water usage clusters at both pilots identified in this study.

only. Within the normal usage flow traces, implementation of ANOVA revealed that there was a nonnormal distribution of data. This was supported by Quantile-Quantile (Q-Q) plots and histograms where a bimodal distribution was evident (as provided in supporting information Figures 5 and 6). By observation, it was found that the normal usage data was further underpinned by three subclusters representing calendared water usage (Cardell-Oliver, 2013). These included "normal weekdays" (Monday–Thursday) (grey), "Fridays" (pattern fill), and "Weekends" (black). The water usage for the "Weekend" subcluster equaled zero at Coláiste na Coiribe. Further application of ANOVA to "normal weekday" groups (Monday, Tuesday, Wednesday, and Thursday) indicated that there was no difference between each of the group means. A summary of the water usage clusters is provided in Figure 4. All project medium resolution data sets can be accessed via on the Aran Repository (Mulligan et al., 2017).

4.2. Diurnal Pattern Generation and Peak Water Usage Ratios: Pilot Site 1

Key water design planning parameters can be obtained through a medium-resolution diurnal flow-signature analysis. Singular diurnal time series (7.5 min resolution from between 7:00 and 20:00) of the total water usage were averaged using the available data sets within each cluster. It is highlighted in Figures 5a and 5b that the averaging process revealed routine, group wide activity at each pilot throughout the daily time-line within each subcluster. The Coláiste na Coiribe diurnal flow-signatures is presented below as an example together with the standard deviation of water usage associated with each discrete time interval. The diurnal flow-signature revealed routine activity in the pilot which is referred to in this study as activity spikes. In the specific case of Coláiste na Coiribe, 11 periodic principle activity spikes are recognizable which reflects the intermittent activity of the buildings population between teaching classes each day. This type of end-usage pattern is archetypal of water usage in schools. The Friday cluster (Figure 5b) also exhibited similar water usage behavior; however, a shorter number of seven spikes were apparent as a result of the earlier school closing times on Fridays.

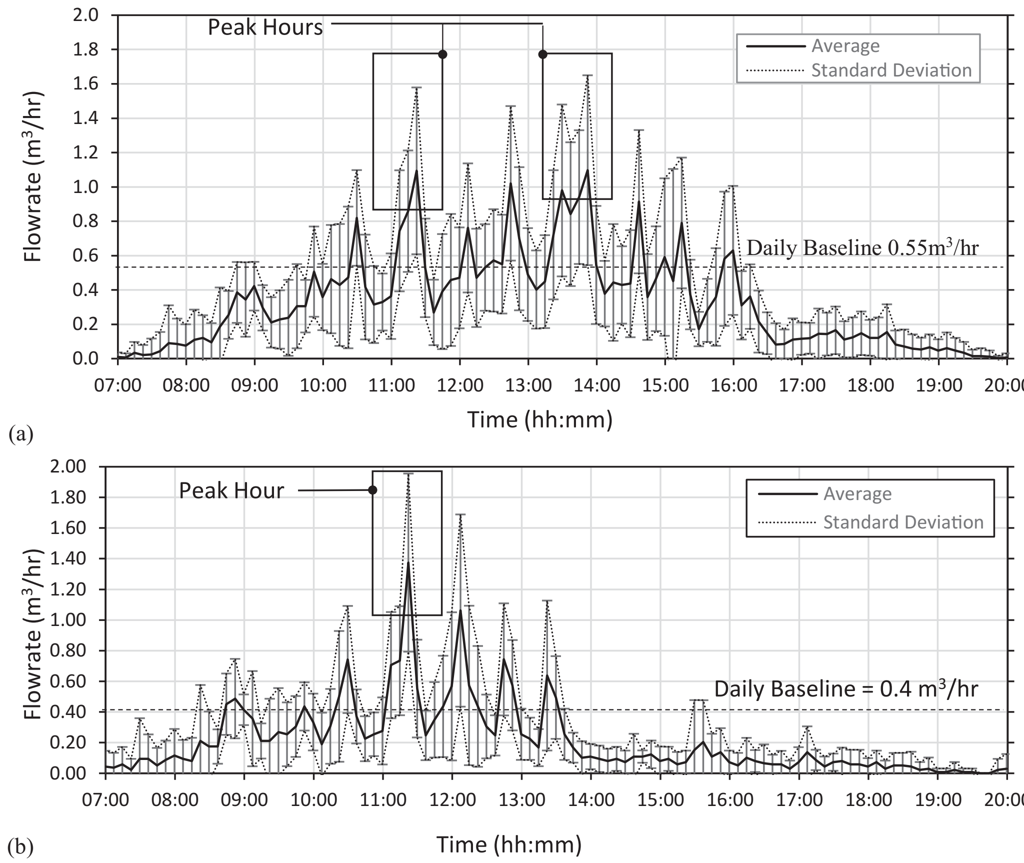


Figure 5. Averaged diurnal flow-signatures for (a) normal weekday and (b) Friday clusters for Colaiste Na Coirbe including standard deviation plotted at discrete 7.5 min intervals.

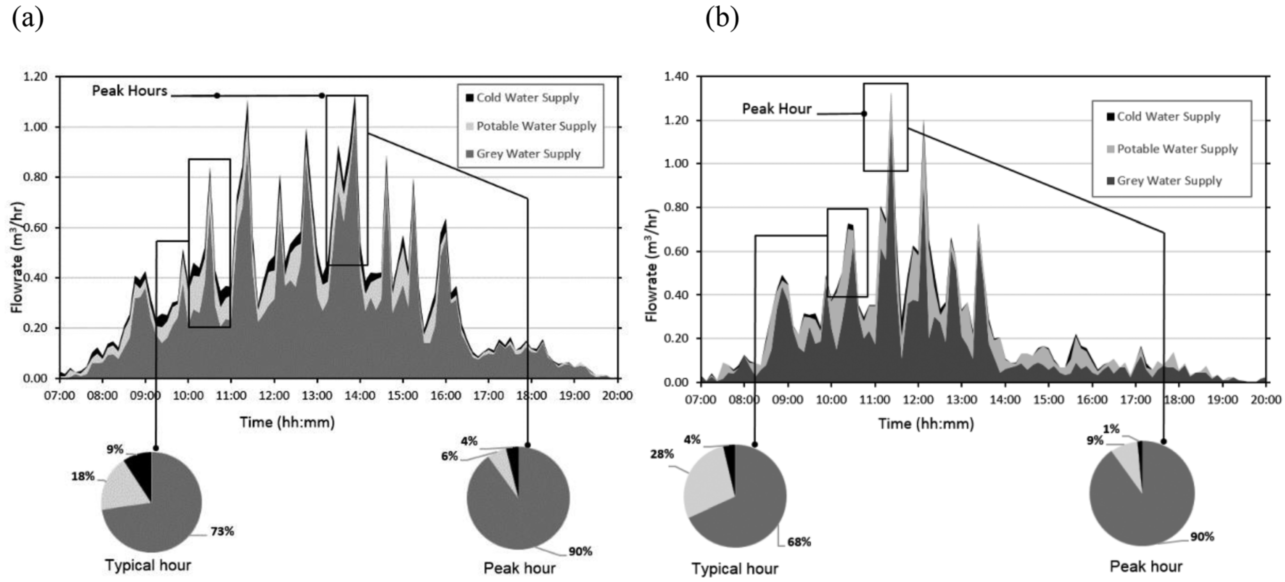


Figure 6. Spatially disaggregated average diurnal flow-signatures into major end-uses (CWS, GWS, and PWS) for (a) normal weekday and (b) Friday clusters outlining peak water usage characteristics.

The hourly peaking factor (PF) can be obtained from the peak hour demand/average daily demand PF = PD/AD. Here the hourly PF was found to be 2.5 and 3.5 for normal week days and Friday subclusters, respectively. The hourly peak flow was found to occur twice a day for the weekday subcluster (at approximately 11.20 A.M. and 2.00 P.M.). The disaggregated water consumption in each cluster using smart meter data was used to identify the primary source of the peak demand and principle activity spikes as shown in Figures 6a and 6b for both subclusters. From this, it was evident that average rain water (grey water) usage in the building can account for 70% of total usage throughout all weekdays (Mondays to Fridays Friday). The potable water was found to be the second largest usage of water ranging in magnitude from 18% to 28% of total water usage with the CWS being the lowest water user in the building. During peak hours, in both subclusters it was shown that GWS can account for up to 90% of the buildings total water usage.

5. High-Resolution Data

5.1. High-Resolution Pattern Discovery: Pilot Site 2

Six repeating flow-signature patterns could be identified from the available high-resolution flow traces: east grey water top-up, west grey water top-up (both serving urinal flushing at an exact pulse rate), male showers, female showers, and male and female bathroom tap usage activity. Based on the location of each of these end-uses, the activity should be experienced on the spatially aggregated main CWS line meter (USF_10). The CWS was therefore also monitored for comparative purposes to determine the ability of activity recognition from an aggregated meter approach (Beal & Stewart, 2013; Froehlich et al., 2011). An example of each of the discovered water usage patterns is described below.

Top-up activity in the east rain water tank was found to occur at a regular frequency of every 1.5–2.0 h ($\therefore \Delta t=$ 90–120 min) during unoccupied hours and increases to a frequency of every 0.9 h ($\therefore \Delta t=$ 52 min) during occupied hours. Figure 7 outlines a granular time series of top-up activity displaying both east and west grey water tank top-up activity simultaneously together with the spatially aggregated CWS line meter. The flow trace implies that the west and east mains top-up occur as a compound event which is traceable in the spatially aggregated flow trace by summation. The duration of the west and east top-up was determined to be 11 and 32 min with flows of 2.75 and 1.50 m^3/h, respectively. The east and west rainwater tank top-up activities were found to encompass the largest magnitude events within the building and therefore, according to the methodology outlined in section 2, the maximum flow and temporal values were found to be $Q_{max}=2.85$ m^3/h and $t_{max}=$ 32 min.

A signature for shower activity (Figures 8a and 8b) was ascertained by artificially triggering various shower events, logging the time stamp and subsequently tracking the response of the flow trace. It was found that

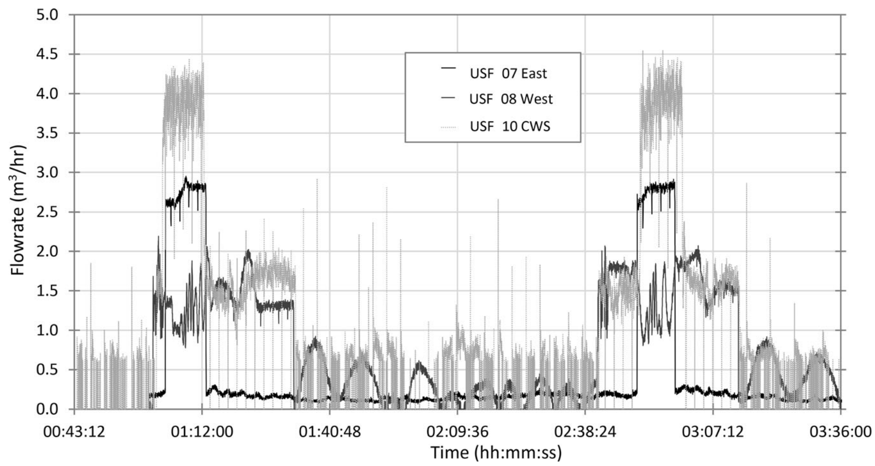


Figure 7. Example of a spatially aggregated (total CWS) and disaggregated high-resolution flow trace of east and west tank top-ups occurring simultaneously.

shower flow-signatures could be recognized if the following three conditions could be met simultaneously: (1) flow-rate in the showers meter (USF_03 and USF_04) of at least 0.8 m^3/h for (2) a duration of at least 1 min corresponding simultaneously to (3) activity on the DHW circuit with a similar duration. The patterns, in some cases, are often recognizable by an immediate increase (vertical spike) in water usage in the order of 1–3 m^3/h as is highlighted in Figure 8b.

A total of 145 shower events were isolated in the USF submeters USF_03, 04 (male and female cold shower supply) and USF_09 (domestic hot water) throughout the 8 days of monitoring. The average shower time was found to be 6 min with a minimum and maximum of 1.5 and 15 min, respectively. The flow rates varied significantly depending on the quantities of cold and hot water being used. There were approximately 12–18 showers per day (between 7.00 A.M. and 10.30 A.M.) and an average shower uses 200 L with a minimum and maximum of 50 and 660 L, respectively. Flow and temporal coefficients were thus found to be $C_Q = 1.38$ and $C_t = 5.3$ ($C_V = 7.29$). Figure 9 demonstrates male tap flow activity together with the total CWS usage. The tap flows were artificially triggered in three events for the purpose of this investigation in the following sequence (1) single tap flow for 10 s, (2) dual tap flow from 12 s, and (3) triple tap flow for 12 s. The flow of an individual tap was found to be approximately 0.25 m^3/h. The flow coefficients for the tap were

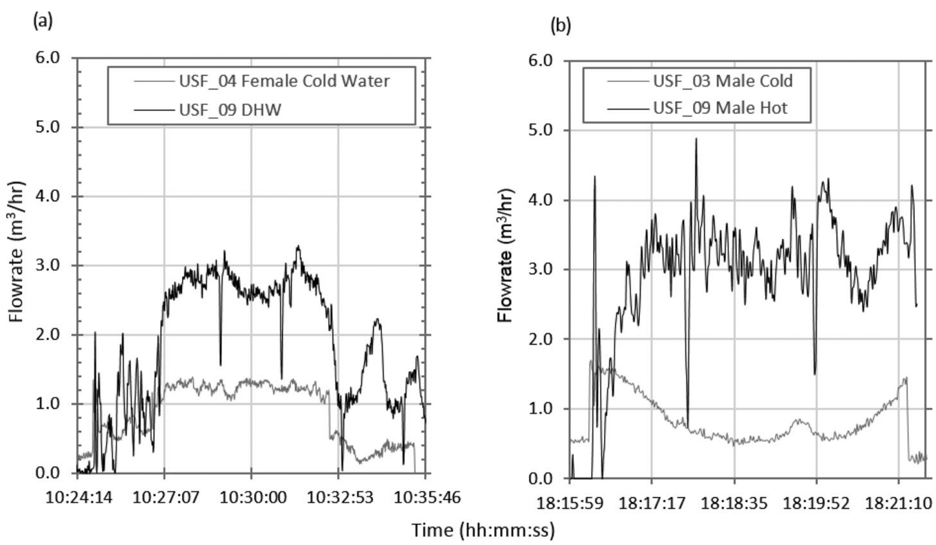


Figure 8. Flow-signature pattern for shower activity displayed on the (a) female and (b) male cold water and domestic hot water circuit simultaneously.

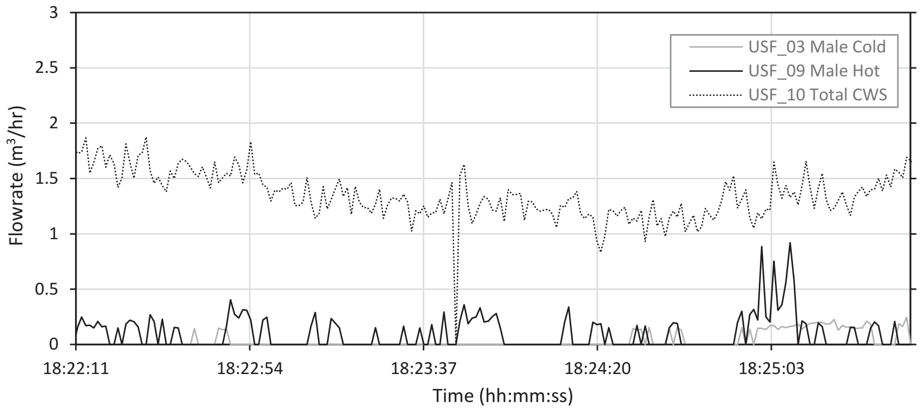


Figure 9. Example of an aggregated and disaggregated high-resolution flow trace for a number of tap events occurring together with other unknown end-usage activity ($C_Q = 11$ and $C_t = 188$). The flow-signatures for tap use are also accompanied by noise.

found to be $C_Q = 11$ and $C_t = 188$, $C_V = 2{,}068$. The three triggered tap events were identified in the DHW circuit amongst some low level noise.

6. Effects of Spatial and Temporal Aggregation on Water Metering

6.1. Spatial Aggregation

As was discussed in section 2, spatial aggregation of water usage data has the advantage of being inexpensive (in terms of number of meters required and data management) but suffers the drawback of poor quality for activity recognition purposes at high-resolution temporal scales. In this section, the ability to identify individual flow-signatures based on spatial aggregation is discussed. The flow and temporal coefficients summarized in Table 1 (as identified in the previous section) helps to identify the ability to recognize a particular flow event from a spatially aggregated data set.

As displayed in Figure 7 previously, which highlights that the east and west tank top-up activity occurring simultaneously, based on the relatively low values of C_Q and C_t, both flow events are easily discernible from the aggregated CWS meter flow trace. However it might be argued that this is an ideal case of only two events occurring simultaneously making the recognition process relatively simple. On the other hand, when monitoring events where C_Q and C_t are large and there are more than two events occurring in a compound event data set, the process of flow-signature recognition becomes cumbersome. This is displayed in the shower events of Figures 10a and 10b which outlines a shower event occurring together with the east grey water tank top-up and unmonitored hot water usage. In Figure 10a, it is relatively difficult to manually recognize the shower event in the aggregated CWS flow trace when the east top-up activates. Coincidentally, the aggregated shower usage flow rate is similar to the flow rate of the top-up and therefore it could very easily be mistaken for one water usage event if the CWS data was considered alone. On the other hand,

Table 1
Average Flow-Signature Metrics With Flow-Signature Coefficients

Event	Flow rate Q_e (m^3/h)	Duration t_e (mins)	Frequency		C_Q	C_t
East Tank Top-Up (GWS)	2.75[a]	11.00	00.00–6.00	Every 1.5 h		2.9
			6.00–24.00	Every hour		
West Tank Top-Up (GWS)	1.50	32.00[a]	00.00–6.00	Every 1.5 h	1.70	
			6.00–24.00	Every hour		
Showers	2.00	6.00	18 showers per day		1.38	5.3
Bathroom Taps	0.25	0.17	Sporadic		11.00	188

[a]Maximum value.

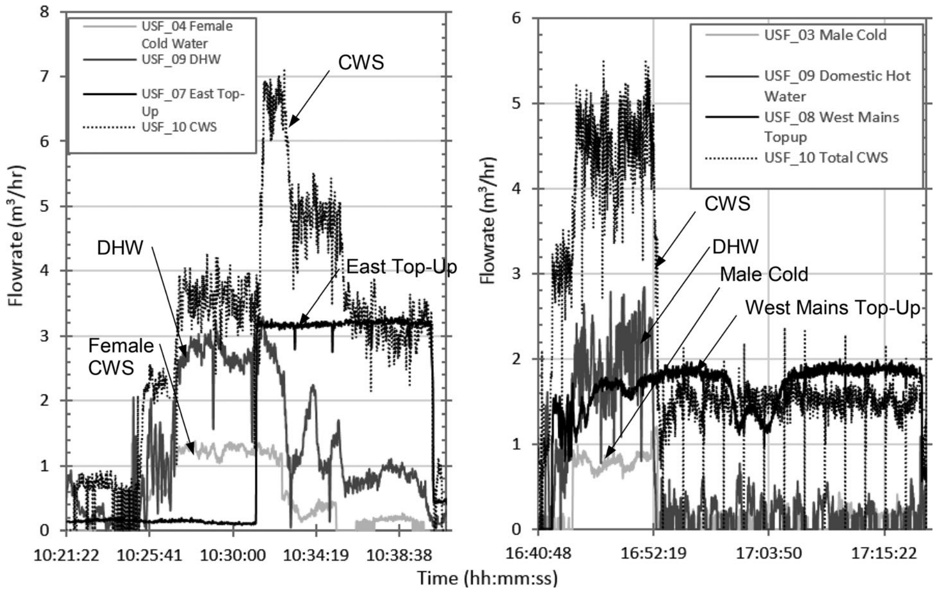


Figure 10. (a) Example of a compound event consisting of three individual events occurring on the CWS flow trace including a female shower, domestic hot water activity and east grey water tank top-up. (b) Example of a compound event consisting of a shower event together with the west tank top-up.

Figure 10b outlines a compound event consisting of an isolated shower event coinciding with a west tank top-up event alone. In this case, the shower activity was readily distinguishable in the CWS flow trace. However, automatic activity recognition methods may also mistake this for east top-up activity due to the duration of the shower and the duration of the east top-up being the same ($t_e \approx 10$ min) as highlighted in Figure 8.

In summary, it was found that a shower existed within a compound event in the CWS flow trace over 95% of the time at the pilot. Due to magnitude of the shower signature coefficients $C_Q = 1.38$ and $C_t = 5.3$, the signatures were discernible within the aggregated flow trace over 70% of those monitored. Using equation (1) to equation (7) a theoretical prediction of the ability to recognize the shower event using the worst case compound event (East and West Mains Top-Up) was made. Using the assumptions outlined in section 3.7, it can be assumed that the shower rates $r_s = 15$ showers per day/15 h = 1 showers/h (i.e., $E_{freq(e)} \approx$ every 60 min throughout the occupied period of 15 h). From the medium-resolution data flow trace, the grey water top-up was found to refill at an average frequency of 1.33. (i.e., $E_{freq(\max)} \approx 45$ min) with a duration $t_{\max} = 30$ min. Using equation (3–6), and employing an overlapping time of at least 1 min, the probability of a shower event occurring simultaneously with a grey water top-up was found to be $P_{C(1,2)} = 0.43$. Therefore, 43% of the time it can be expected that a shower event will occur simultaneously with a tank refill. From the manual shower mining assessment, the probability of compound events was found to be 95% (significantly greater than the models prediction of 43%). This was most likely to be a result of other activity occurring on the flow trace in the real system which was not included the probability of coincidence model. As can be seen from the final meter resolution coefficient, the ultra-high-resolution of this meter results in very fine details of the shower event being presented in a flow trace.

6.2. Temporal Aggregation

As discussed in section 2, temporal aggregation of meter data has the advantage of reducing the cost of meter type and expense of data management and storage. However, if temporal aggregation is excessively large, flow-signatures becomes highly smeared and unrecognizable. The aggregation is defined by the ratio of the event duration Δt_e to the data acquisition rate Δt ($M_t = \Delta t_e / \Delta t$). Figures 11a and 11b present a comparison between the high-resolution and medium-resolution flow-signatures for the same data set of the spatially aggregated CWS flow trace outlining east and west tank top-up activity. For the high-resolution flow trace $M_t = 1{,}920$ while the medium-resolution flow trace $M_t = 4.27$ (i.e., flow-signature would be

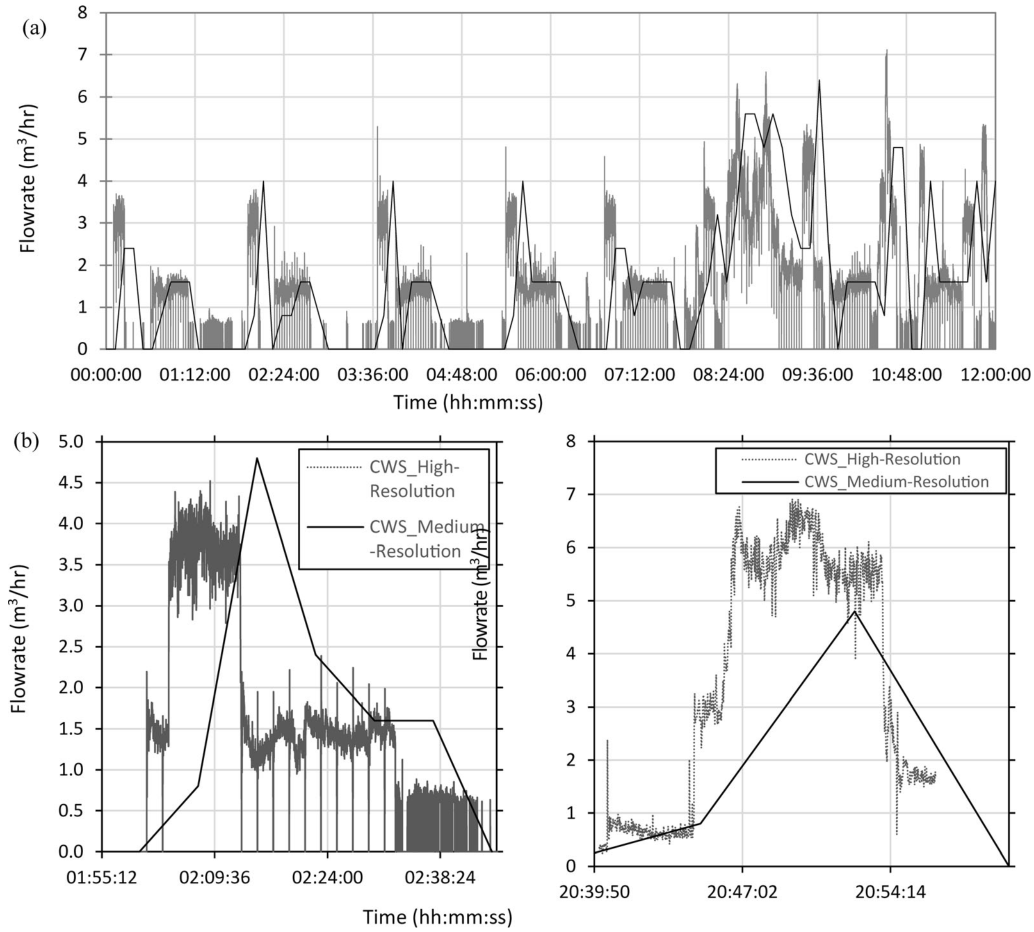


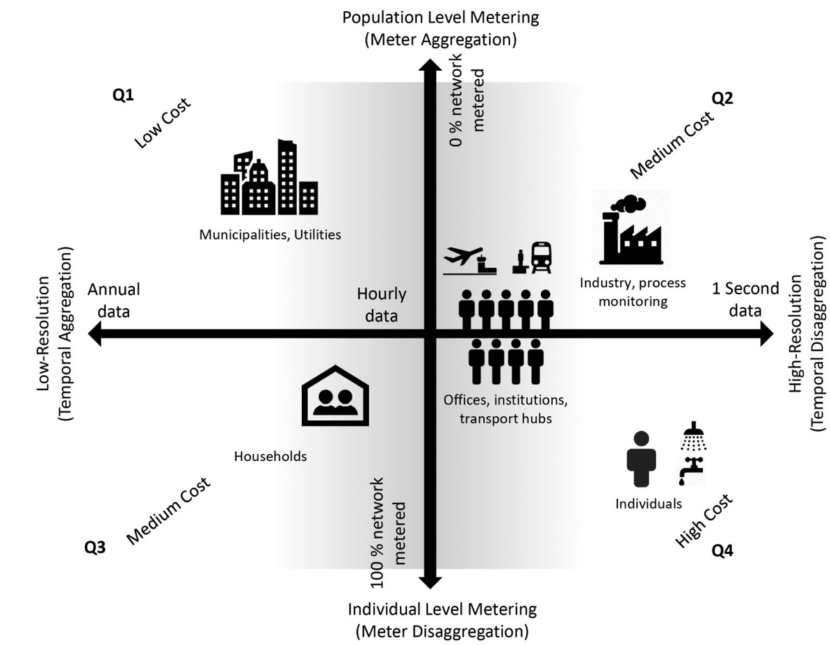
Figure 11. Comparison of a high-resolution and medium-resolution flow trace for an east and west grey water top-up compound event for (a) 00:00–12:00 and (b) magnified singular flow-signature and (c) comparison of a high-resolution and medium-resolution flow trace for a shower event occurring simultaneously with other events in the building.

represented by 1,920 data points and 4.27 data points for the high-resolution and medium-resolution flow traces, respectively).

For a large sample window ($T = 12$ h in Figure 11a), it can be concluded that high-resolution monitoring becomes ineffective as practically the same information can be obtained from the medium-resolution flow trace at this scale. When magnified, it is evident that the compound tank refill event flow-signature in the medium-resolution data becomes linearly smeared composing of a superimposed triangular and rectangular flow distribution with tapered ends as opposed to the typical rectangular signature of the actual event (depicted using the broken lines in Figure 11b). On the other hand, Figure 11c provides a comparison of medium-resolution and high-resolution flow traces of a shower event ($M_t = 540$ and 1.2, respectively). Due to a 7.5 min sample rate, the aggregated shower flow-signature is exposed as a triangular plot rendering it difficult to identify. It was expected that identification would be challenged in cases where this would occur as part of a compound event.

7. Discussion and Classification of Smart Meters

A comprehensive evaluation of both medium-resolution and high-resolution smart meter data in the trials of this study has revealed that there is a vast spectrum of data aggregates and disaggregates (spatial and temporal) which can inform building end-users of significant water usage information. The information obtained is also useful for informing the parameterization process of end-use models in the nonresidential sector. The problem in the overall smart metering sphere is that so far there has been no generic



Classification	Order	Interval Δt (s)	Application
Low-Resolution	Weekly, Monthly, Quarterly, Biannually	$1.0 \times 10^8 > \Delta t > 0.6 \times 10^6$	Billing, Water Audits, Socio-demographic and economic studies
Low to medium-Resolution		$0.6 \times 10^6 > \Delta t > 86400$	
Medium-Resolution	Minutes, Hourly, Daily	$86400 > \Delta t > 3600$	Peak days, baselines, diurnal flow signatures, peak hours, leaks and other faults, Socio-demographic studies, infrastructure modelling, sizing pumps & pipes
Medium to high-Resolution		$3600 > \Delta t > 60$	
High-Resolution to ultra-high-resolution	Seconds	$60 > \Delta t > 1.0$	Personal activity, appliances, aggregating medium-resolution data, leaks and other faults, parametrise modelling methods approaches such as PRP and SIMDEUM

Note: Classification is developed based on a synthesis of previous studies and their applications

Figure 12. Problem quadrant comparing the temporal resolution of smart meter data acquisition on the horizontal axis and levels spatial aggregation on the vertical axis together with cost characteristics. Classification is developed based on a synthesis of previous studies and their applications.

classification system to aid end-users in choosing the correct metering approach. As discussed in section two, there is no-one-size-fits-all scenarios due to the bivariate nature of water metering together with site specific requirements.

To synthesize the findings of the bivariate nature of smart water metering, a problem quadrant diagram highlighted in Figure 12 was developed where the horizontal axis depicts the temporal data resolution and the vertical axis depicts the levels of spatial aggregation. As can be seen, the choice of metering approach and the minimal valuable data set required depends on the nature of the end-use study. For example, Quadrant 1 utilizes a small number of low-resolution smart meters in a water network which would be ideal for gathering data over large time and spatial scales such as applicable to city water supply utilities. At the other end of the spectrum in Quadrant 4, large numbers of high-resolution data sets will produce extremely detailed water usage catalogues with the disadvantage of high costs of installation, high costs of smart meter (due to number and instrumentation) and data acquisition, processing and storage requirements. Such data sets are only effective at the individual level. For intermediate applications, a large number of options are available to manage a compromise between end-usage detail and cost. An example of this is provided in the engineering building trial of this study where an individual high-resolution flow trace was used to identify shower events and a medium-resolution flow trace was used to aggregate the total water usage in the building. In this way, the shower usage could be determined as a function of the buildings total water usage permitting end-user control of usage.

It would also be possible to apply the rules/lessons learned from a particular quadrant in Figure 12 and subsequently apply them to obtain the benefits of an adjacent quadrant. For example, Cardell-Oliver (2013) provides an example of where medium-resolution data determined at the household level (Quadrant 3) can be used to scale to Quadrant 1. Another application of this could be through the adoption of rules for flow-signatures (medium-resolution and high-resolution) obtained in this study and application in buildings with similar characteristics to reduce the number of meters required. It is also interesting to note from this that pulse models such as SIMDEUM perform data modeling on a bottom-up basis from the ultra-high-resolution sector of Quadrant 4 and scaling diagonally to the spatial nodal temporal aggregated data of Quadrant 1. Thus, high-resolution studies in both the residential and nonresidential sector is useful to aid parametrization of such models.

To support Figure 12, based on the findings of this study and synthesized based on the research of other studies (Beal et al. 2011a, 2011b; Cardell-Oliver, 2013), the following classification system was developed. The classification system, together with the general practice employed in the overall study, serves the purpose of aiding future designers in the choice of smart meter resolution based on their specific application.

8. Conclusions

Smart water metering is becoming the prevalent tool for water supply management across a spectrum of applications ranging from end-uses at the individual level to aggregated large population consumption. As a result, choice of meter type, number and siting is not straightforward as a result of two degrees of complexity: the temporal resolution of meter acquisition and level of spatial aggregation for the specific case in question. With that, the use of flow-signature patterns to characterize water usage behavior is becoming a popular tool across all applications in both monitoring and modeling with a distinct lack of information available in the nonresidential context.

This study presents a generalized understanding and hydraulic notation which is of use to strategizing a water metering program for monitoring or modeling purposes. The article reviewed existing methods to analyze medium-resolution flow traces in order to transfer them to the nonresidential context. The study also proposed new parameters based on the Poisson rectangular pulse (PRP) process to investigate water usage time series which can suggest required resolution and level of data aggregation required to reveal valuable flow-signatures of interest within a specific case. Based on the data obtained from two nonresidential pilot sites, both medium-resolution and high-resolution flow-signature patterns were evaluated using existing and newly proposed parameters. We show that medium-resolution data comprising minute to daily activity (86,400 s $>$ Δt $>$ 3,600 s) can expose regular (average or baseline daily flows), continuous flows (constant water usage suggesting leaks or other faults), peak day (a day where water usage was exceptional, e.g., an event, conference) and diurnal signatures (water usage patterns established throughout the day) and hourly flow patterns. These medium-resolution signatures can help explain how water utilized by populations as opposed to individuals and is therefore better suited to spatial aggregation applications.

On the other hand, the high-resolution to ultra-resolution data (60 s $>$ Δt $>$1.0 s) was capable of disaggregating specific end-usage events from analyzing their flow-signature when the event time is much greater than the meter sample time $\boldsymbol{t_e} \gg \Delta \boldsymbol{t}$. A total of six different end-use signatures were found for the building and multiple compound events were found to occur quite frequently. Because the proposed model for the probability of overlap was based on only two events, the certainty of this model when applied to flow traces where more than three events can overlap became questionable. In order to progress this further, it would be interesting to develop a method for the probability of existing compound events, and the event of interest (i.e., overlapping of three or more end-use events in time) as this would serve a greater practical application in a nonresidential context.

Finally, a smart meter classification system was proposed based on a problem quadrant diagram which categorizes smart meters into low-resolution, medium-resolution, and high-resolution categories together with their respective areas of application. In summary, the following specific findings were determined for application of flow-signature analysis in the nonresidential sector.

1. High-resolution data acquisition may not be efficient for application in commercial settings (i.e., $\Delta \boldsymbol{t} =$ 1 s). Less granular data of the order of $\Delta \boldsymbol{t} =$10–60 s will likely provide sufficient data for activity recognition without a significant cost of data storage.
2. When a compound event composes of three or more individual events, activity recognition from a spatially aggregated flow trace becomes impractical and spatial disaggregation is required.
3. Sufficient information for conservation purposes could be gained if a single medium-resolution flow trace is positioned on the mains water supply and a small number of high-resolution flow traces (designed according to the above siting rule of thumb) installed to monitor the biggest most frequent water users.

Notation

V	volume, m^3.
v_i	ith volume, m^3.
Q	flow rate, $m^3\ s^{-1}$.
$Q_{\max}$	flow rate of max. flow event, $m^3\ s^{-1}$.
Q_e	flow rate of event of interest, $m^3\ s^{-1}$.
Δt	data acquisition interval, s.
t_i	ith time stamp, s.
Δt_{agg}	aggregated time interval, s.
$t_{\max}$	duration of longest event, s.
$E_e, E_{\max}$	time intervals between repeating events of interest.
t_e	duration of event of interest, s.
t_o	duration of event overlap, s.
T	period of investigation, s.
$C_{t(e,\max)}$	temporal coefficient.
$C_{Q(e,\max)}$	flow coefficient.
$C_{V(e,\max)}$	volume coefficient.
$P_{c(e,\max)}$	probability of overlap.
$N_{(T)}$	number of events within T.
$M_{t(e,\max)}$	meter resolution quality.

Acknowledgments
The authors acknowledge funding from the European Union's Seventh Framework Programme for research, technological development, and demonstration under grant 619660 (Waternomics) and from Science Foundation Ireland (SFI) under the SFI Strategic Partnership Programme grant SFI/15/SPP/E3125 (ESIPP). The authors also thank the Waternomics project partners, building management at the NUI Galway and Coláiste na Coiribe pilots as well as project support from Informatics Research Unit for Sustainable Engineering (IRUSE). Data used in this study are available at the ARAN (Access to Research at NUI Galway) open repository at http://hdl.handle.net/10379/6737.

References

Alvisi, S., Franchini, M., & Marinelli, A. (2003). A stochastic model for representing drinking water demand at residential level. *Water Resources Management*, *17*(3), 197–222.

Aquacraft (2010). *Trace Wizard® software version 4.1. 1995-2010*. Boulder, CO: Aquacraft, Inc. Retrieved from http://www.aquacraft.com/

Bauer, D., Philbrick, M., Vallario, B., Battey, H., Clement, Z., & Fields, F. (2014). *The water-energy Nexus: Challenges and opportunities*. Washington, DC: US Department of Energy 2014.

Beal, C. D., Gurung, T. R., & Stewart, R. A. (2016). Demand-side management for supply-side efficiency: Modeling tailored strategies for reducing peak residential water demand. *Sustainable Production and Consumption*, *6*, 1–11.

Beal, C. D., & Stewart, R. A. (2013). Identifying residential water end uses underpinning peak day and peak hour demand. *Journal of Water Resources Planning and Management*, *140*(7), 04014008.

Beal, C. D., Stewart, R. A., & Fielding, K. (2011a). A novel mixed method smart metering approach to reconciling differences between perceived and actual residential end use water consumption. *Journal of Cleaner Production*, *60*, 116–128. https://doi.org/10.1016/j.jclepro.2011.09.007

Beal, C. D., Stewart, R. A., Spinks, A., & Fielding, K. (2011b). Using smart meters to identify social and technological impacts on residential water consumption. *Water Science and Technology: Water Supply*, *11*(5), 527–533.

Bennett, E. R. (1975). *Individual home wastewater characterization and treatment: Completion report submitted to Office of Water Research and Technology* (US Department of the Interior. Completion report series, No. 66). Fort Collins, CO: Environmental Resources Center, Colorado State University.

Blokker, E. J. M., Vreeburg, J. H. G., & Van Dijk, J. C. (2009). Simulating residential water demand with a stochastic end-use model. *Journal of Water Resources Planning and Management*, *136*(1), 19–26.

Britton, T., Cole, G., Stewart, R., & Wiskar, D. (2008). Remote diagnosis of leakage in residential households. *Journal of Australian Water Association*, *35*(6), 89–93.

Buchberger, S. G., & Wells, G. J. (1996). Intensity, duration, and frequency of residential water demands. *Journal of Water Resources Planning and Management*, *122*(1), 11–19.

Buchberger, S. G., & Wu, L. (1995). Model for instantaneous residential water demands. *Journal of Hydraulic Engineering*, *121*(3), 232–246.

Bullington, D., DeClair, I. M., Di Francesco, A. D., Kobrick, D., Lee, J., Newland, K., . . . Ward-Morris, C. M. (2008). *Facility manager's guide to water management*. Phoenix, AZ: Arizona Municipal Water Users Association.

Cahill, R., & Lund, J. (2012). Residential water conservation in Australia and California. *Journal of Water Resources Planning and Management*, *139*(1), 117–121.

Cardell-Oliver, R. (2013). Water use signature patterns for analyzing household consumption using medium resolution meter data. *Water Resources Research*, *49*, 8589–8599. https://doi.org/10.1002/2013WR014458

Cardell-Oliver, R., & Peach, G. (2013). Making sense of smart metering data: A data mining approach for discovering water use patterns. *Journal of the Australian Waster Association*, *40*(2), 124–128.

Clifford, E., Coakley, D., Curry, E., Degeler, V., Costa, A., Messervey, T., . . . Smit, S. (2014). Interactive water services: The WATERNOMICS approach. *Procedia Engineering*, *89*, 1058–1065.

Creaco, E., Blokker, M., & Buchberger, S. (2017). Models for generating household water demand pulses: Literature review and comparison. *Journal of Water Resources Planning and Management*, *143*(6), 04017013.

DeOreo, W., Heaney, J. P., & Mayer, P. W. (1996). Flow trace analysis to assess water use. *Journal American Water Works Association*, *88*(1), 79.

Di Palma, F., de Marinis, G., Gargano, R., Granata, F., Greco, R., & Tricarico, C. (2014). The overall pulse model to predict the end user water demand. *Procedia Engineering*, *89*, 942–949.

Dziegielewski, B. (2000). *Commercial and institutional end uses of water*. Denver, CO: American Water Works Association.

Eichenseher, T. (2010). Americans value water more than energy, and want government to fix leaking pipes. *National Geographic*, VOICES, Ideas and Insight From Explorers.

European Commission. (2010). *Water scarcity and drought in the European Union* (Ref KH-30-09-180-EN-D). Luxembourg, Luxembourg: European Commission Publications Office.

European Commission. (2012). *Water performance of buildings*. Retrieved from http://ec.europa.eu/environment/water/quantity/pdf/BIO_WaterPerformanceBuildings.pdf

Fox, C., McIntosh, B. S., & Jeffrey, P. (2009). Classifying households for water demand forecasting using physical property characteristics. *Land Use Policy*, *26*(3), 558–568.

Froehlich, J., Larson, E., Saba, E., Campbell, T., Atlas, L., Fogarty, J., & Patel, S. (2011). A longitudinal study of pressure sensing to infer real-world water usage events in the home. In K. Lyons, J. Hightower, & E. M. Huang (Eds.), *International conference on pervasive computing* (pp. 50–69). Berlin, Germany: Springer.

Giurco, D., Carrard, N. R., McFallan, S., Nalbantoglu, M., Inman, M., Thornton, N. L., & White, S. (2008). *Residential end-use measurement Guidebook: A guide to study design, sampling and technology*. Ultimo, NSW, Australia: Institute for Sustainable Futures, UTS.

Grafton, R. Q., Ward, M. B., To, H., & Kompas, T. (2011). Determinants of residential water consumption: Evidence and analysis from a 10-country household survey. *Water Resources Research*, *47*, W08537. https://doi.org/10.1029/2010WR009685

Hamilton, L. C. (1985). Self-reported and actual savings in a water conservation campaign. *Environment and Behavior*, *17*(3), 315–326.

Hearn, B. (1998). Benchmarking water use on farm—If you don't measure it you can't manage it. In *9th Australian cotton conference* (pp. 519–529). Gold Coast, QLD, Australia: Australian Cotton Growers' Research Association.

Hood, S. (2002). Rural water use efficiency-Real water use efficiency and the opportunities. In *11th Australian cotton conference* (pp. 585–595). Brisbane, QLD, Australia: Australian Cotton Growers' Research Association.

Jorgensen, B., Graymore, M., & O'Toole, K. (2009). Household water use behavior: An integrated model. *Journal of Environmental Management*, *91*(1), 227–236.

Larson, E., Froehlich, J., Campbell, T., Haggerty, C., Atlas, L., Fogarty, J., & Patel, S. N. (2012). Disaggregated water sensing from a single, pressure-based sensor: An extended analysis of HydroSense using staged experiments. *Pervasive and Mobile Computing*, *8*(1), 82–102.

Mayer, P. (2006). *Trace wizard water use analysis tool*. Boulder, CO: Aquacraft Inc.

Millock, K., & Nuages, C. (2010). Household adoption of water-efficient equipment: The role of socioeconomic factors, environmental attitudes and policy. *Environmental Resource Economics*, *46*(2), 539–565.

Mulligan, S., Hannon, L., & Clifford, E. (2017). *Water consumption data*. Galway, Ireland: ARAN Repository, National University of Ireland.

Nguyen, K. A., Zhang, H., & Stewart, R. A. (2013). Development of an intelligent model to categorise residential water end use events. *Journal of Hydro-Environment Research*, *7*, 182–201. https://doi.org/10.1016/j.jher.2013.02.004

Otaki, Y., Otaki, M., Pengchai, P., Ohta, Y., & Aramaki, T. (2008). Micro-components survey of residential indoor water consumption in Chiang Mai. *Drinking Water Engineering and Science*, *1*(1), 17–25.

Paisley, R. K., & Henshaw, T. W. (2014). If you can't measure it, you can't manage it: Transboundary waters, good governance and data & information sharing & exchange. *Indiana International & Comparative Law Review*, *24*, 203.

Ploeser, J. H., Pike, C. W., & Kobrick, J. D. (1992). Nonresidential water conservation: A good investment. *Journal American Water Works Association*, *84*(10), 65–73.

Richards, P. I. (1948). Probability of coincidence for two periodically recurring events. *The Annals of Mathematical Statistics*, *19*, 16–29.

U.S. Department of Energy Water-Energy Technology Team. (2014). *The water-energy nexus: Challenges and opportunities* (Rep. DOE/EPSA-002). Southwest Washington, DC: Office of Energy Policy and Systems Analysis.

U.N. Water. (2013). *Water scarcity factsheet*. New York City, Geneva: The UN Water Organization.

Van Hise, C. R. (1910). *The conservation of natural resources in the United States*. Basingstoke, UK: Macmillan.

Wilkes, C., Mason, A., Niang, L., Jensen, K., & Hern, S. (2005). *Evaluation of the Meter-Master data logger and the Trace Wizard analysis software* (special Appendix to Quantification of Exposure-Related Water Uses for Various US Subpopulations). Prepared for US EPA (December 2005).

Willis, R. M., Stewart, R. A., Panuwatwanich, K., Jones, S., & Kyriakides, A. (2010). Alarming visual display monitors affecting shower end use water and energy conservation in Australian residential households. *Resources, Conservation and Recycling*, *54*(12), 1117–1127.

Willis, R. M., Stewart, R. A., Williams, P. R., Hacker, C. H., Emmonds, S. C., & Capati, G. (2011). Residential potable and recycled water end uses in a dual reticulated supply system. *Desalination*, *272*(1), 201–211.

World Business Council for Sustainable Development (WBCSD). (2009). *Transforming the market: Energy efficiency in buildings* (EEB project report). Geneva, Switzerland: Author.