Advanced Household Profiling Using Digital Water Meters

Md Shamsur Rahim[a]
[a]Centre for Artificial Intelligence, School of Computer Science
University of Technology Sydney, Sydney, Australia
e-mail: mdshamsur.rahim@student.uts.edu.au

Khoi Anh Nguyen[b], Rodney Anthony Stewart[b]
[b]School of Engineering and Built Environment
Griffith University, Gold Coast, Australia
e-mail: k.nguyen@griffith.edu.au; r.stewart@griffith.edu.au

Damien Giurco[*, c]
[c]Institute for Sustainable Futures
University of Technology Sydney, Sydney, Australia
e-mail: damien.giurco@uts.edu.au

Michael Blumenstein[d]
[d]Centre for Artificial Intelligence, School of Computer Science
University of Technology Sydney, Sydney, Australia
e-mail: michael.blumenstein@uts.edu.au

ABSTRACT

Advanced householder profiling using digital water metering data analytics has been acknowledged as a core strategy for promoting water conservation because of its ability to provide near real-time feedback to customers and instil long-term conservation behaviours. Customer profiling based on household water consumption data collected through digital water meters helps to identify the water consumption patterns and habits of customers. This study employed advanced customer profiling techniques adapted from the machine learning research domain to analyse high-resolution data collected from residential digital water meters. Data analytics techniques were applied on already disaggregated end-use water consumption data (e.g., shower and taps) for creating in-depth customer profiling at various intervals (e.g., 15, 30, and 60 minutes). The developed user profiling approach has some learning functionality as it can ascertain and accommodate changing behaviours of residential customers. The developed advanced user profiling technique was shown to be beneficial since it identified residential customer behaviours that were previously unseen. Furthermore, the technique can identify and address novel changes in behaviours, which is an important feature for promoting and sustaining long-term water conservation behaviours. The research has implications for researchers in data analytics and water demand management, and also for practitioners and government policy advisors seeking to conserve valuable potable-water resources.

KEYWORDS

User profiling, digital water meter, water conservation, water consumption data, behaviour change, Recommender System.

1. INTRODUCTION

In recent years, ensuring water supply during periods of shortage caused by drought and avoiding low pressure during the hours of peak demand have been two of the challenges

[*] Corresponding author

troubling many metropolitan water utilities (Nguyen et al., 2016). Rolling out Digital Water Meters (DWMs) can be considered a potential solution for overcoming these challenges because the data they collect can contribute to water conservation and peak demand management. In an experiment to measure the effects of feedback information on water consumption, researchers found that when households received more information with higher water consumption, they demonstrated higher water saving (Cespedes Restrepo and Morales-Pinzon, 2020). Another study reported that detailed water consumption feedback can contribute mean savings of 5.5% across 25 studies (Liu and Mukheibir, 2018). In addition, a short-term water demand forecasting model was proposed by Nguyen et al. (2016) based on the data collected from DWMs which stated significant impact on peak demand management. A healthy number of investigations have been undertaken to generate insights from DWM data. These studies can be divided into five categories: (1) water-use feedback; (2) water event categorisation; (3) water demand forecasting; (4) behaviour analysis; and (5) socioeconomic analysis (Rahim et al., 2020). However, further scope for improvement exists for the current situation through incorporating highly personalised systems, because there is a direct relationship between the level of personalisation and effects on water conservation. Personalisation can be achieved through comprehensive user profiling, which provides the opportunity to meet users needs and preferences (Eke et al., 2019).

A user or customer profile is defined as a "summary of the user's interest, characteristics, behaviours, and preferences," whereas user profiling is the "system of collecting, organizing and inferring the user profile information" (Eke et al., 2019). User personalisation through user profiling is now a widely adopted technique in various domains, such as artificial intelligence, data science and information science (Gauch et al., 2007). Recommender systems (RSs) can be considered one of the notable applications of user personalisation. Recommender systems are intelligent systems that recommend a list of items that are most likely of interest to the user (Burke, 2007). At present, RSs are deployed in various application domains as their application is no longer limited to selling more products or recommending news. The most common RS applications are employed in five domains: Entertainment, Content, E-commerce, Services, and Social (Montaner et al., 2003; Ricci et al., 2015). However, the water industry is yet to adopt RSs, although great potential exists for RSs to promote water-conscious behaviour (Rahim et al., 2019). Comprehensive user profiling based on water consumption data collected from DWMs can provide valuable information regarding users' water use patterns along with the ability to track any changes in water use behaviours. For this reason, in-depth user profiling is a must for the success of such RSs. Although a healthy number of user profiling techniques have been proposed in relevant studies (e.g., demand profiling, habit detection), they are still not suitable for RSs because of their limitations.

Many user profiling techniques have been adopted based on water consumption data collected from residential DWMs. For instance, hourly water consumption data have been used to build demand profiles, enabling water demand to be forecasted accurately. Gaussian mixture models (GMM's) based approach was proposed by McKenna et al. (2014) to represent demand patterns and then classify the demand patterns. In another approach, Padulano and Del Giudice (2018) introduced a two phases procedure: clustering and classification to detect water consumption patterns. In addition, the probability of a particular event occurring at a particular hour was used in water event categorisation (Nguyen et al., 2014). Furthermore, a habit detection algorithm (Cardell-Oliver, 2016) was proposed where the user profile was based on five constraints, which were expressed using five parameters. Moreover, behaviour analysis-related studies have mainly aimed to understand the behaviours and dynamics of consumers based on water consumption data. Although different user profiles are created to reach different goals, they have some common limitations. First, in many studies, weekends and weekdays are not considered separately for profiling, despite the consumption patterns varying. Second, all related studies have only considered hourly consumption; however, other frequency intervals such as 15 and 30 minutes may provide more insights. Third, almost all investigations have considered total consumption data rather than individual water event consumption data. Lastly, existing studies have been unable to accommodate recent changes in behaviours because total or average consumption is used for profiling. These limitations of relevant studies indicate the necessity for a new comprehensive user profiling approach to make effective recommendations for promoting water conscious behaviours.

In this study, we introduced a new comprehensive user profiling approach that overcomes the limitations of prior relevant studies. In numerous studies, weekday and weekend water consumption have not been considered separately. Hence, in our proposed approach we performed profiling separately for weekdays and weekends. In terms of the profiling interval, our hypothesis was that more frequent interval profiling provides a greater understanding of the behaviours and habits of customers. Therefore, along with hourly profiling, we created two other user profiling frames at 15- and 30-minute intervals. In addition, instead of profiling based on total water consumption data, the user profiles were created based on disaggregated end-use water consumption data (e.g. showers and taps). Furthermore, to accommodate recent changes in behaviours, profiles were created by giving higher priority to recent data. Finally, an algorithm was introduced that performs user profiling for each household. To enhance the user profiles further, other information that characterises behaviours such as shower duration, volume of water used in the shower, and washing machine usage was collected. After comparing customer profiling at various intervals (e.g., 15, 30, and 60 minutes), we concluded that more frequent intervals (i.e., 15 minutes) of profiling provided a better understanding of user behaviours that was previously unseen. Moreover, through tracking recent changes in behaviours, it became possible to identify and track habits and changes in behaviours.

The findings from our experiments implied that the proposed approach addresses two aspects of the promotion of water-conscious behaviours. First, it addresses water conservation through the collection, organisation, and inference of water usage and savings scopes (i.e., the average volume of water used in the shower, average shower duration, and number of times performing laundry). Second, the approach provides the opportunity to manage water demand by characterising the patterns in water consumption behaviours and habits by providing the likely time a particular event will occur at various intervals. Thus, this research introduces a comprehensive user profiling approach that accommodates recent changes in behaviours and has the potential for promoting water-conscious behaviours. The major contributions of this study are as follows:

- It has proposed a comprehensive user profiling approach that addresses the limitations (e.g. not considering disaggregated water consumption data, profiling at shorter intervals, tracking and reflecting recent changes in behaviour, profiling based on the type of day) of the most recent state-of-art studies.
- It has introduced an advanced profiling algorithm to create user profiles based on dis-aggregated water consumption data.

- It has identified the most suitable profiling interval among three profiling intervals.
- It has highlighted the key benefits from such profiling approach for consumers, utilities, and policy makers.

The remainder of the paper is organised as follows. Section 2 presents a critical analysis of the related works. Section 3 examines the methodology followed in this study. Section 4 discusses the findings of the study and finally, section 5 draws the conclusion of the paper.

2. RELATED WORKS

To the best of the authors' knowledge, no studies have been conducted on in-depth user profiling with the purpose of promoting water-conscious behaviours through an RS, as the concept of RS in the water industry is relatively new (Rahim et al., 2019). However, some studies have involved some sort of user profiling for solving other problems. These studies can be categorised into three: (1) behaviour analysis; (2) socioeconomic analysis; and (3) water end-use categorisation. Although the user profiles developed in these studies were used to solve specific problems, they have some limitations that must be overcome to develop comprehensive user profiling.

Numerous studies have been performed on behaviour analysis based on simple user profiling to understand the behaviours and dynamics of consumers from water metering data. These studies can be categorised into three: (1) habit detection and profiling; (2) demand profiling; and (3) customer segmentation. In case of habit detection and profiling, four types of pattern were identified by Cardell-Oliver (2013): continuous-flow days; exceptional peak-use days; programmed patterns with recurrent hours; and normal use patterns. By contrast, three profiles were identified by Cominola et al. (2016) after segmenting water consumers based on their eigen behaviours. In another study, a detailed breakdown of hourly water use by volume for different times (i.e., peak hour, day, and month) was performed by Cole and Stewart (2013) to provide an accurate estimation of indoor and outdoor consumption. In addition, a habit detection algorithm was introduced by Cardell-Oliver (2016) based on time series data from water meters; however, because of heuristics, it could not guarantee the detection of all habits. Demand profiling has been performed in many studies for predicting water demand. For instance, McKenna et al. (2014) utilised hourly consumption data for classifying demand patterns. Later, a demand profile was proposed (Gurung et al., 2015) based on diurnal patterns of efficiency-rated appliances for modelling water demand. Lastly, eight relevant usage profiles from water consumption data were identified using clustering and modelling techniques on water consumption data by Cheifetz et al. (2017). However, these studies were based on hourly aggregated water consumption data rather than disaggregated data. Only Nguyen et al. (2016) considered hourly disaggregated data. Customer segmentation or clustering techniques have been applied in a few studies as part of their behaviour analysis. For forecasting future behaviour and clustering consumption behaviour, a hybrid model was proposed by Leyli-Abadi et al. (2018). In another study, residents were divided into five clusters by family structure, job, or lifestyle using a fuzzy clustering algorithm based on water consumption data. In a recent study, customer segmentation based on eigen-behaviour analysis was adapted and used to identify three main water end-use profile clusters (i.e., showering, clothes washing, and irrigation) (Cominola et al., 2019). Furthermore, they observed the existence of time-of-use and intensity-of-use within each class. However, in this study, the authors did not consider recent changes in the behaviours.

In socioeconomic studies, consumers' socioeconomic and demographic factors have been studied to understand their effect on water conservation (Beal et al., 2011; Bich-Ngoc and Teller, 2018; Willis et al., 2013). Such studies have used consumers' socioeconomic and demographic data to identify the determinants of water consumption (Beal et al., 2011; Bich-Ngoc and Teller, 2018; Willis et al., 2013). Although these studies have helped to determine the factors behind water consumption, further information is required for users' profiles to promote water conservation.

In the case of water end-use categorisation, the use of user profiles is very rare. The probability of time-of-use for each event over a 24-h period based on total collected data was used by (Nguyen et al., 2014) to categorise water end-use. However, recent changes in behaviours were also omitted from this study.

After examining relevant studies, we identified the following research gaps that need to be overcome to create comprehensive user profiling:

- Consideration of disaggregated water consumption data: Disaggregated water consumption data provide more information about water consumption data (i.e., for which purpose the water is used, such as showering or irrigation). Because existing studies have mostly considered total consumption data instead of disaggregated data, they have missed the opportunity to utilise disaggregated water consumption data to generate useful insights. Therefore, future studies on profiling should consider disaggregated consumption data.
- Profiling at reduced intervals: Existing profiles are mostly on an hourly interval. However, profiling at reduced intervals may provide more specific information and a greater understanding of consumers' habits and behaviours. Thus, future studies should consider user profiling frames at a reduced interval (e.g., 15 or 30 minutes).
- Tracking and reflecting recent changes in behaviour: Studies based on profiling have been unable to reflect recent changes in behaviour in users' profiles. However, if changes go unnoticed and are not considered in the profiles, then the profiles will become less effective. Hence, the approach must be adapted to track and reflect recent changes in water consumption behaviour.
- Profiling based on the type of day: Water consumption patterns vary depending on the type of day (i.e., weekday or weekend). Therefore, user profiling should be performed based on the type of day to improve the effectiveness.

3. MATERIAL AND METHODS

The proposed approach in this study for creating comprehensive user profiling can be divided into three consecutive steps: (1) data set collection and preparation; (2) profile creation; and (3) feature extraction. In the first step, a raw data set was collected from 306 single stand-alone households in Melbourne, Australia, and in the second step, a profiling algorithm was incorporated to create profiles. Finally, in the last step, statistical modelling was adopted for extracting features to enhance the profiles. Figure 1 represents the workflow of the study. The following subsections discuss the steps in details.

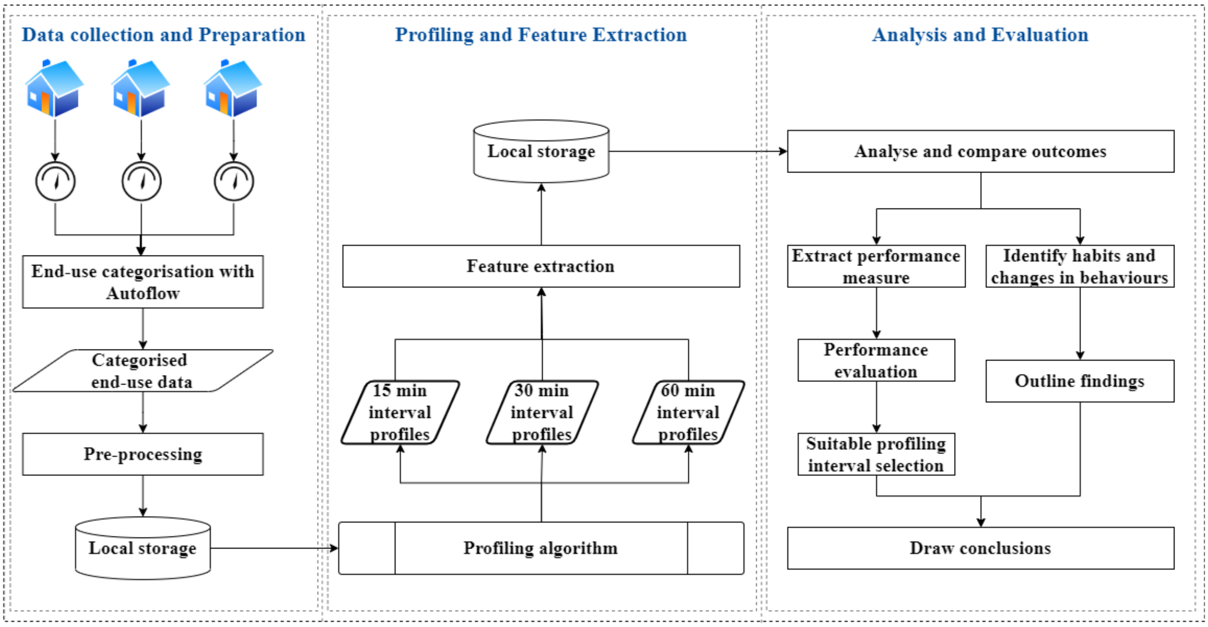


Figure 1. Workflow of the study.

3.1. Data collection and preparation

The raw data for this study originated from 306 single stand-alone households in Melbourne, Australia. The data set consists of recordings of high-resolution data at 5-second intervals that were collected for 10 months (February–December 2010). The collected raw data were then analysed with Autoflow (Nguyen et al., 2014), an intelligent metering system. Nine types of water end-use were classified by Autoflow with 90% accuracy. Autoflow provided the classifications of water end-use events: taps, dishwashers, leaks, evaporating coolers, washing machines, showers, toilets, irrigation, and bathtubs. These formed the primary data set for this research. Figure S1 depicts the steps in the primary data collection process.

After collecting the primary data set, the next step was to perform the initial pre-processing. For profiling, we considered only those events that occurred between 06h00 and 23h59 because the number of events and volume consumed from 00h00 to 05h59 are negligible in comparison. After this step, the pre-processed data were stored in a database for further processing. Table 1 presents a summary of the data set.

Table 1. Summary of the data set after initial pre-processing

Item	Description
Number of households	306
Data collection duration	11-02-2010 to 11-10-2010 (dd-mm-year)
Hours considered	06h00-23h59
Data collection interval	5 seconds
Number of end-uses	9
Number of events	56 310 576
Total shower events	181 358
Total bathtub events	18 274
Total washing machine events	30 996
Total dishwasher events	34 524
Total irrigation events	9 975
Total toilet events	1 173 902
Total tap events	4 152 386

Total leak events	50 709 161

From Table 1, it is obvious that the number of leak events dominates all other events because one long continuous leak event is a set of many tiny discrete events recorded every 5 seconds. Similarly, the number of tap events is the second highest but in each event on average 0.87 litres of water were used. Because these two events occurred frequently but an insignificant amount of water was used, these two events were not considered for profiling to reduce computational time and maintain scalability. In addition, toilet (flush) and evaporating cooler events were excluded from profiling because toilet events are spontaneous and evaporating coolers are responsible for less than 1% of total water consumption. For these reasons, we considered high water consumption events (shower, bathtub, irrigation, washing machine, and dishwasher) (Rahim et al., 2019), because when combined, these events account for nearly 70% of the total water consumption (Stewart et al., 2010).

Many studies have overlooked considering water consumption patterns on weekends and weekdays separately. However, this information is particularly crucial because water consumption patterns may differ depending on the type of day. Hence, as the last step of pre-processing, we extracted the type of day (i.e., weekend or weekday) from the given date. Figure S2 illustrates the consumption patterns of 200 sample households for 15 days, which clearly depict the difference in consumption patterns, especially for showers. During weekdays, 8:00–8:30 am is the peak time for taking showers. Conversely, on weekends, 11:00 am is the peak time for taking showers.

After the pre-processing step, the final data set consisted of 12 attributes. Table 2 presents a description of these attributes.

Table 2. Description of the attributes in the final data set

Attribute	**Description**	**Example**
Site	Unique identifier for each household in the data set.	Site001, Site002.
Start date	Start date of an event	11-Feb-2010
Start time	Start time of an event	08:38:15
End date	End date of an event	11-Feb-2010
End time	End time of an event	08:42:10
Category	Water end-use category	Shower, dishwasher
Duration	Amount of time an event took place.	0:03:55
Volume	Quantity (in litre) of water used in an event	36.81
Max flow	Maximum flow rate (Litre per minute) recorded for an event	10.83
Mode flow	Mode of flow rate (Litre per minute) recorded for an event	9.33
Cyclic event	Multiple intakes of water during one single event	C1 (Cycle 1 for washing machine), D5 (Cycle 5 for dishwasher)
Type of day	If the day is Saturday or Sunday then weekend, otherwise weekdays.	Weekend, weekday

The data set used in this study has some limitations. First of all, it does not have any socio-demographic data. Second, it does not have any weather data. Though the inclusion of these data would be interesting, however, we believe the absence of these data will not undermine the current profiling study.

3.2. Profile creation

Profile creation is the most important step in the proposed approach which is based on a profiling algorithm. The proposed algorithm is designed in such a way that it can address the limitations of existing studies. The proposed algorithm consists of three steps. First, a data structure is initialised. Next, the probability distribution (PD) is computed using probability mass function (PMF). Finally, profiles are created by concatenating multiple vectors created in the previous steps. Algorithm 1 describes the algorithm with a detailed explanation for each step.

Algorithm 1. Consumption profile creation algorithm

ConsumptionProfile(s, sd, ed, $intrvl$, td,w):creates water consumption profile for household s, based on the data where the start date is sd and end date is ed, $intrvl$ is the value of interval for profiling, td is the type of the day (i.e, weekday, weekend) and w is weight for different periods of time.

Here:

$s \in \boldsymbol{S}$, where **S**= $\{Site001, Site002 \dots Site306\}$

$intrvl \in \boldsymbol{I}$, where **I**= $\{15\ minutes, 30\ minutes, 60\ minutes\}$

$e \in \boldsymbol{E}$, where **E**= $\{Shower, Bathtub, Clotheswasher, Dishwasher, Irrigation\}$

$t \in \boldsymbol{T}$, where **T** is the set of times at *intrvl* from 06h00 to 23h59.

The major steps of the algorithm are as follows:

1. Initialise data structure
2. Probability Distribution computation
3. Profile creation

We now examine the steps in detail.

1. Initialise data structure.
   To calculate and store the distribution of an event *e* over a given time *t*, for household *s*, an n-dimensional vector $v_{et}{}^{1} = (v_{et_1},\ v_{et_2}, \dots, v_{et_{n-1}}, v_{et_n})$is taken and it is initialised to 0. Note that, the value of dimension depends on the value of *intrvl*. For instance, when the value of *intrvl*= 30 minutes, then dimension will be total hours from 06h00 to 23h59 $\times 2 = 18 \times 2 = 36$.

   (a) Set $v_{et} \leftarrow 0$.

2. Probability Distribution computation
   To understand the behaviour or habit of the users, we have employed a probability mass function (PMF). PMF can be defined as a function that provides the probability that a discrete random variable is exactly equal to some value. Formally this function p: $\boldsymbol{R} \rightarrow [0,1]$ can be defined as:

   $$p_X(x_i) = P(X = x_i)\ for - \infty < x < \infty \qquad (1)$$

   And properties of PMF are:

$\sum p_X(x_i) = 1$ (2)

$p_X(x_i) > 0$ (3)

$p_X(x_i) = 0 \, for \; all \; other \; x$ (4)

Equation 2 describes the sum of the probabilities associated with each possible value will be always up to 1. According to equation 3, the values of probabilities must be positive. And for other values, the probability will be 0.

This step will compute probability distribution (PD) for each event across time t with interval $intrvl$for the given type of the day td and assign it to v_{et}

(a) for each e in E:
   a. compute $p_X(x_i)$
   b. $v_{et} \leftarrow p_X(x_i)$

3. Profile creation
At this step, the profile is created by constructing two more PDs by repeating step 1 &2 and assigning weight w_i and merging all the PDs. The reason for creating the PDs is separating most recent, second most recent and previous historical data. In this way, it would be possible to assign a higher weight to most recent patterns as the most recent data will always capture better changes in behaviours. Finally, each weighted vector for each event is concatenated to a matrix M.
(a) ${v_{et}}^2 \leftarrow p_X(x_i)$ and ${v_{et}}^3 \leftarrow p_X(x_i)$
(b) $v_{et}^{w} \leftarrow (v_{et}^{1} \times w_1) + (v_{et}^{2} \times w_2) + (v_{et}^{3} \times w_3)$ where $w_1 > w_2 \geq w_3$
(c) $M \leftarrow (v_{e_1t}^{w} \dots v_{e_rt}^{w})$

In step one, for each event a vector is initialised with n-dimension. As we have considered five large water consumption categories (e.g., shower, bathtub, washing machine, dishwasher and irrigation), five n-dimensional vectors will be initialised. The dimension of the vectors depends on the interval value. For 15 minutes, there will be 72; for 30 minutes 36; and for 60 minutes interval there will be 18 dimensions. The reason for choosing a vector as the data structure is to measure the Cosine similarity.

Cosine similarity is a measure of similarity between two non-zero vectors to determine how similar the vectors are by calculating the cosine of the angle between them. Mathematically, the cosine of two vectors that are non-zero can be described by following equations.

$$A \cdot B = \|A\|\|B\| cos\theta \tag{5}$$

Given two vectors, A and B, the cosine similarity, $\cos(\theta)$ is measured using a dot product and magnitude as follows:

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} {A_i}^2}\sqrt{\sum_{i=1}^{n} {B_i}^2}} \tag{6}$$

Where A_i and B_i are components of vector A and B, and

$$A \cdot B = \sum_{i=1}^{n} A_i B_i = A_1 B_1 + A_2 B_2 + \cdots + A_n B_n \quad (7)$$

The derived value of similarity ranges from $[-1, 1]$ where -1 completely opposite and to 1 means completely similar. In this scenario, A and B vectors hold the PD for a particular event and the resulting similarity would provide an indication of changes in consumption behaviours based on time.

To capture and understand the timing of water consumption events, a probability mass function (PMF) is used in step two. A PMF is a function that provides the probability that a discrete random variable is equal to some value (Stewart, 2009). The computed probability based on most recent data of each event e across time t at interval $intrvl$ for type of the day td is assigned to designated vector v_{et}.

In the last step, at first, two more vectors for each event are created using PMF based on previous water consumption data. Later, these vectors are combined into one vector for each event by giving more weights to the vector holds the most recent data. Therefore, after assigning a higher weight to the first PD vector and lower weights to the other PD vectors, that is w_1 to v_{et}^1, w_2 to v_{et}^2, and w_3 to v_{et}^3, where $w_1 > w_2 \geq w_3$, a matrix M is constructed by concatenating the vectors for each household.

$$M = \begin{bmatrix} \left((w_1 \times v_{e_1t_1}^1) + (w_2 \times v_{e_1t_1}^2) + (w_3 \times v_{e_1t_1}^3)\right) & \cdots & \left((w_1 \times v_{e_1t_n}^1) + (w_2 \times v_{e_1t_n}^2) + (w_3 \times v_{e_1t_n}^3)\right) \\ \vdots & \ddots & \vdots \\ \left((w_1 \times v_{e_mt_1}^1) + (w_2 \times v_{e_mt_1}^2) + (w_3 \times v_{e_mt_1}^3)\right) & \cdots & \left((w_1 \times v_{e_mt_n}^1) + (w_2 \times v_{e_mt_n}^2) + (w_3 \times v_{e_mt_n}^3)\right) \end{bmatrix} \quad (8)$$

Each row in the matrix represents the PD of one event, whereas each column presents a time at an interval between 6 am and 12 am. Figure 2 describes the steps of the profile creation algorithm.

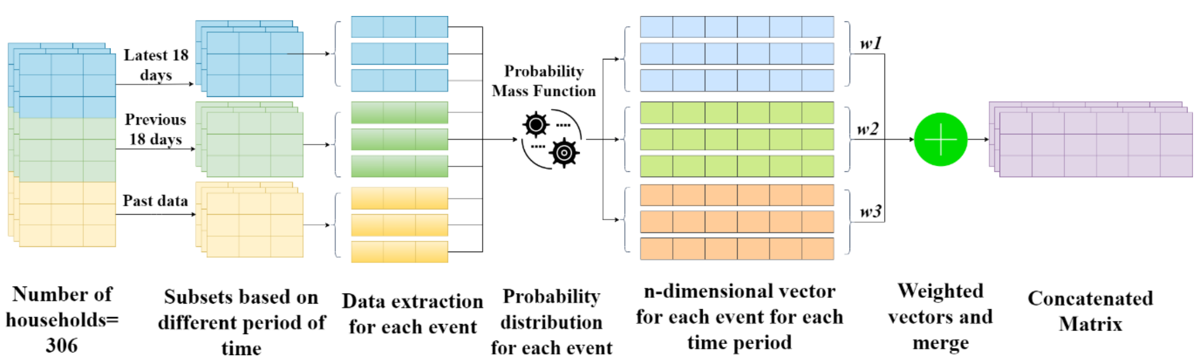


Figure 2. Illustration of the time of use and weighted probability of the use algorithm.

Such a matrix can be used to make recommendations to change a particular water consumption behaviour from a particular time. For instance, if a household seems to always perform irrigation at a particular hour of peak water demand (e.g., 8 am or 5 pm), then recommendations can be made to gradually shift this event outside of the hours of peak demand to avoid low water pressure.

3.3. Feature extraction

A customer profile with a PD that provides a possible time for an event to occur is not sufficient for promoting water-conscious behaviours. This is because more information is required. Therefore, to enhance the profiles and create in-depth user profiles, features must be extracted from the data. Depending on the goal of the water utility and items to recommend, the features to extract may vary. Table 3 lists the extracted features in this study from different events, which can be utilised to develop a personalised recommender system.

To extract these features, we mostly employed statistical methods. For instance, to derive the mean consumed volume of water for showers, bathtubs, washing machines, dishwashers, or toilets, we applied the arithmetic mean using the following equation (9):

$$\overline{x} = \frac{1}{n}\sum x_i = \frac{1}{n}(x_1 + x_2 + \cdots x_n) \tag{9}$$

where $x_1, x_2, \cdots, x_n$ are the volume of each event.

Table 3. Extracted features list from each event

Event	Feature
Shower	Mean consumed volume, mean duration, mode of flow rate, change in volume compared with the previous week, change in duration compared with the previous week, and change in flow rate compared with the previous week
Bathtub	Mean consumed volume, change in volume compared with the previous week
Washing machine	Mean consumed volume, number of loads per week
Dishwasher	Mean consumed volume, number of loads per week
Toilet	Mean consumed volume
Leak	Total volume in the last week
Tap	Mode flow rate

4. RESULTS AND DISCUSSION

4.1 Personalised recommendation system

The extracted features of each end-use category performed in the previous step are the main resource for developing a personalised RS for each household. On the one hand, the mode flow rate of shower events can be used to identify the efficiency of the shower, and recommendations to replace the showerhead can be made accordingly. Similarly, the consumed volume in toilet flushes can be used to infer whether the flush system is a dual flush system because typical single flush systems consume 11 litres per flush in Australia (Business Amenities Fact Sheet, 2009). Replacing the typical system with a 4.5/3 litre dual flush system can save approximately

11,000 litres/year/person, assuming the person uses the toilet four times a day (Business Amenities Fact Sheet, 2009). On the other hand, changes in flow rate or volume or duration can be used to infer the effectiveness of recommendations. For instance, if shower duration reduces compared with the previous week/month after a recommendation is made to reduce shower time, this would indicate that the household is responding to the recommendations. Furthermore, after making a recommendation to replace a showerhead with a water-saving one, any reduction in flow rate or volume compared with the previous week/month would imply the effectiveness of recommendations. Figure S3 is an example of the significance of the water savings that could be achieved from such a recommendation.

4.2 Effective water demand management through behaviour change and flow theory

In positive psychology, flow is defined as a subjective state where people become so completely involved in something that they forget time, fatigue, and everything else except the activity (Csikszentmihalyi et al., 2014). The flow theory of behaviour-changing tasks (Yürüten, 2017) states that if the difficulty of a task matches the capabilities of the person, then his or her engagement will be maximised. This theory can be translated into the RS field as the balance between the user's ability to perform behaviour-changing tasks and the difficulty to perform the recommended tasks. For instance, in our case, let us consider a consumer who performs irrigation or uses a bathtub at a specific time during the hours of peak demand. To shift the event time from a peak to a nonpeak hour of demand, recommendations can be made to shift the event time by 15 or 30 minutes gradually. Suggesting that behaviours be shifted by 1 hour might be difficult for many consumers, and in such cases, the users would lose interest in the system.

In this study, we created profiles at three different intervals (i.e., 15, 30, and 60 minutes) using the proposed algorithm, and performed comparisons among the profiles to identify the most suitable one. Furthermore, we extracted some features from the consumption data to enhance the user profiles, and we noted some very interesting findings.

We compared the profiling at three different intervals and observed that profiling at a lower interval (15 minutes) provided a greater understanding of the behaviour of consumers. Figure S4 a, b, and c represent the total number of each event during weekdays at 60, 30, and 15 minutes, respectively, over a period of 10 months at Site 325. If we consider that the morning peak hour of water demand (Cole and Stewart, 2013) is from 7 AM to 9 AM, then we can see that nearly 50% of shower events occurred from 7:00 AM to 7:29 AM for that particular household. It was only possible to observe this pattern as we performed profiling at 15-minute intervals. Therefore, the profiling interval of 15 minutes provided a greater understanding of water consumption patterns. The implication of this finding is that it can lead to potential effective demand management through changing the behaviour of users. This is achieved by shifting specific water consumption events from a specific time by incorporating flow theory into behaviour change.

4.3 Enhanced demand management thorough understanding customer behaviour changes

We found that profiling based on the PD of different water consumption events could be used to identify, understand, and notice changes in the habit and behaviour of users. To demonstrate this statement, we calculated the PD for three different periods of time. First, we calculated the PD for the latest 18 days, denoted as P1, and then the previous 18 days, denoted as P2, and lastly all other previous days, denoted as P3. An unchanged PD for a particular event in each period would indicate a strong habit, and any change in PD in P1 would indicate recent changes in behaviour compared with previous patterns. Figure 3 depicts the PD of shower events during weekdays across a different period of time for Site 325.

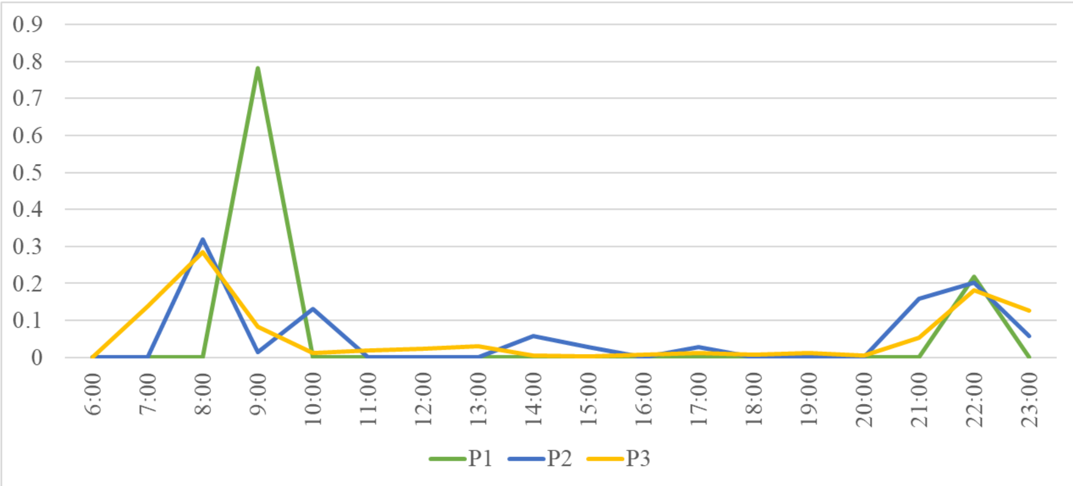


Figure 3. Comparing the probability distribution of shower events during weekdays across different periods of time at Site 325; P1 refers to the recent 18 days, P2 refers to the previous 18 days; and P3 represents all previous days' data.

From Figure 3, we can see that the probability of taking showers at 8 AM was 30% in P2 and P3. However, in the latest 18 days (P1), the shower time changed to 9 AM with the probability of 80%. This clearly indicated the change in shower time compared with previous behaviour patterns. By contrast, the probability of taking a shower at 10 PM almost remained the same, which indicated the strong habit of taking a shower at that time. Similar patterns can be found for other activities using the proposed profiling approach. The 18-day interval was chosen based on a study that reported it took 18–254 days to form a habit (Lally et al., 2010).

The main implication of this finding is that improved demand management can be ensured through identifying households with strong and flexible behaviour patterns. The weighted PD over three different periods of time will help to identify consumers with flexible demand patterns. For example, if a household exhibits fluctuations in consumption patterns for a particular event that does not have a fixed time for any water consumption activity, then the recommendation system could be developed to target that household because it is the most flexible for behaviour change.

4.4 Opportunities for system enhancement

We extracted a healthy number of interesting features from the data set to enhance profiles further, and concluded that these features provide in-depth analysis and previously unseen insights. For instance, after extracting the feature of weekly change in shower water volume, we could identify those households with a percent increase or decrease in water volume. This insight can be useful for tracking changes in behaviours and recommending relevant and effective activities. Figure 4 a, b represents the histogram of the number of households with a change in shower consumption volume compared with the previous week on weekdays and on weekends respectively.

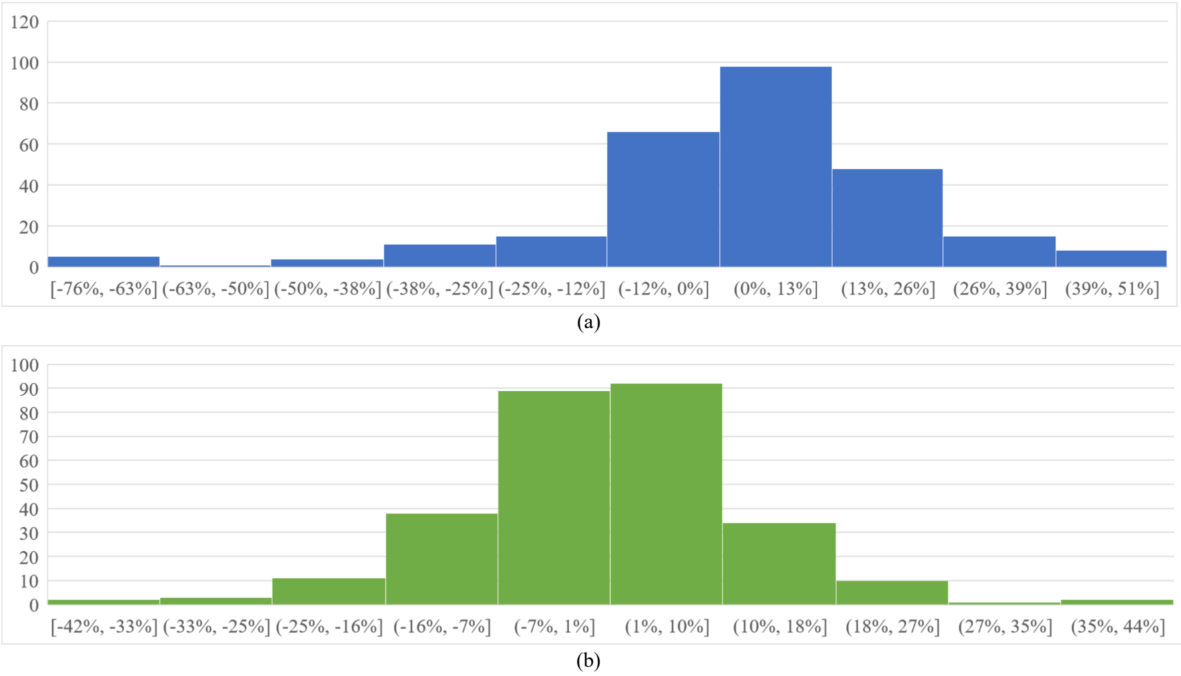


Figure 4. Histogram of the number of households with a change in shower consumption volume compared with the previous week: (a) shower volume change on weekdays; (b) shower volume change on weekends.

For instance, let us assume a household consumes approximately 70 litres of water per shower event over a week before receiving any recommendations. Based on their consumption pattern, a recommendation can be made to reduce their shower water consumption to 60 litres. Then, using the extracted features, we can determine how effective the recommendation is the next week. If the water consumed for showers drops in the following week, this would mean that the recommendation is working properly and has been accepted by the household. However, if no change occurs in the consumption or it increases, this would mean that the recommendation is not effective for that household and further calibrations can be made (i.e., reduce shower water consumption to 65 litres). In this way, it would be possible to obtain implicit preferences of consumers instead of explicit preferences (i.e., through rating or like–dislike).

4.5 Summary of key benefits

The profiling approach introduced in this study overcomes the limitations of existing studies. The anticipated benefits from such profiling are discussed here.

i. ***Identification of habits/behaviour patterns with more detail and accuracy***: The proposed profiling approach can be helpful for identifying habits/behaviour patterns of consumers in more detail and with enhanced accuracy. On the one hand, the identification of such habits or behaviours depending on the type of day (i.e., weekday or weekend) would encourage consumers to become more aware of or educated about water-conscious behaviours. On the other hand, such insights would provide an enhanced understanding of consumers to the utilities and policy makers, which can be used to determine more effective water conservation programs, campaigns, and education.

ii. ***Tracking of changes in behaviour:*** The proposed profiling approach has the ability to track changes in water consumption behaviour of customers. This would empower them to take control of their water consumption and help determine their progress towards water conservation. In addition, alerting consumers to any deviations from their usual consumption patterns would help them stay on track for sustainable water consumption. For water utilities, tracking any changes in the behaviour of customers would help them understand the various factors that influence water consumption (i.e., type of day, temperature, recommendations for water conservation, education, programs, and campaigns) at a more detailed level.

iii. ***Improvement of demand profiling:*** Demand profiling is critical for utilities and policymakers to understand water consumption patterns, enhance peak water demand management, and reduce water pumping costs as well as greenhouse gas (GHG) emissions. The profiling approach introduced in this study can play a vital role in improving demand profiling by predicting the time and probability of future events. Based on the improved demand profiling and enhanced understanding of water consumption enabled by the proposed approach, utilities and policymakers can introduce flexible tariff plans. Current state-of-the-art water end-use classification systems can classify water end-use at an accuracy rate of 95% (Nguyen et al., 2020; Yang et al., 2018). To improve the accuracy of such systems further, the profiling approach proposed in this study can be adopted. Further accuracy in water end-use would increase customers' acceptance of and trust in such systems, utilities, and policymakers.

iv. ***Grouping of households with similar consumption patterns:*** Households with similar consumption patterns can be grouped at a more detailed level based on the proposed profiling approach. This will help consumers compare their consumption patterns with other customers and increase awareness of water conservation. Grouping households will also help utilities and policymakers identify unique numbers of groups and their characteristics. Based on such insights, targeted water conservation programs and campaigns can be designed to promote water conservation further.

v. ***Foundation for a Recommender System:*** User profiling is an integral part of Recommender Systems. The profiling approach introduced in this study can be considered the foundation for a RS in the water industry for promoting water-conscious behaviours. Such a RS can be highly beneficial for consumers, utilities, and policy makers. Through the RS, consumers would be able to interact directly with water utilities through explicit and implicit preferences. This would result in improvements in customer services/satisfaction. For utilities and policymakers, the recommender system would provide the opportunity to obtain consumer feedback regarding different recommendations and policies in a cost-effective and timely, efficient manner.

Table 4 summarizes the anticipated benefits from the proposed profiling approach for different beneficiaries.

Table 4. Anticipated benefits from the proposed profiling approach for different beneficiaries

Benefit	Beneficiary		
	Customer	Utilities	Policy makers
Identification of habits/behaviour patterns	✓	✓	✓
Tracking of changes in behaviour	✓	✓	
Enhancement of peak water demand management		✓	✓
Reduction in water pumping costs (GHG emissions)		✓	✓
Understanding of water consumption patterns	✓	✓	✓
Prediction of time and probability of future events		✓	
Improvement in customer service/satisfaction	✓	✓	
Flexible tariffs	✓	✓	✓
Increased accuracy of water end-use classification	✓	✓	✓
Grouping of similar households	✓	✓	✓
Basis for a recommender system	✓	✓	✓

4.6 Challenges

In this study, we observed some challenges in user profiling. First, handling vast amounts of high-resolution water consumption data at 5-second intervals was a challenge in terms of processing and analysis even for 306 single-standalone households. This indicates further challenges when DWMs are deployed widely. Second, we observed that the profiling outcome highly depends on the accuracy of water end-use classification because the technique uses water end-use data as the primary input. Therefore, any misclassification in water end-use would result in poor profiling results. Lastly, we observed that households with no water consumption data in recent days (empty households) may not have any meaningful PD for water consumption events. In such scenarios, profiling should be performed carefully and consistency should be maintained.

5. CONCLUSION AND FUTURE WORK

The user profiling technique has been widely adopted in various domains for providing personalised services. However, the water sector is yet to adopt advanced user profiling, which could be used to promote water-conscious behaviours through a recommender system. In this study, a user profiling approach based on water consumption data from residential DWMs was proposed along with a profiling algorithm. Furthermore, profiles were created at different intervals (e.g. 15, 30, and 60 minutes). Our findings suggested that profiling at 15-minute intervals provides better insights regarding the behaviour of consumers. It can also be used to identify habits and track changes in behaviours that are important for promoting and sustaining long-term water-conscious behaviours. We believe our findings will help researchers and practitioners in data analytics and water demand management as well as government policy advisors by providing valuable insights.

Currently, we are working towards a recommender system based on the proposed profiling technique for promoting water-conscious behaviours. As for future work, clustering consumers based on the proposed profiling approach can be performed as it would facilitate grouping consumers at a more detailed level. Furthermore, because the relational database management system used in this study took a long time to execute queries, a data warehouse could be proposed for storing and analysing complex data. A carefully designed data warehouse can handle many complexities in data (Ahmed et al., 2013). In addition, profiling that includes socioeconomic data can be performed to provide further insights. Such a profiling approach can be extended to other industries or resource consumption scenarios, such as profiling of electricity and gas consumption for residential and non-residential customers.

Funding

This research was funded by the Australian Research Council (ARC), grant number LP160100215, which also includes funding from the following industry partners: Yarra Valley Water, City West Water, Southeast Water, and Aquiba.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

2009. Business Amenities Fact Sheet. Western Water.

Ahmed, T., Pedersen, T.B., Lu, H., 2013. A data warehouse solution for analyzing rfid-based baggage tracking data, 2013 IEEE 14th International Conference on Mobile Data Management. IEEE, pp. 283-292.

Beal, C., Stewart, R.A., Spinks, A., Fielding, K., 2011. Using smart meters to identify social and technological impacts on residential water consumption. Water Supply 11, 527-533.

Bich-Ngoc, N., Teller, J., 2018. A review of residential water consumption determinants, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer, Cham, Switzerland, pp. 685-696.

Burke, R., 2007. Hybrid Web Recommender Systems, The Adaptive Web. Springer, Berlin, Germany, pp. 377-408.

Cardell-Oliver, R., 2013. Water use signature patterns for analyzing household consumption using medium resolution meter data. Water Resources Research 49, 8589-8599.

Cardell-Oliver, R., 2016. A habit detection algorithm (Hda) for discovering recurrent patterns in smart meter time series, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer International Publishing, Cham, Switzerland, pp. 109-127.

Cespedes Restrepo, J.D., Morales-Pinzon, T., 2020. Effects of feedback information on the household consumption of water and electricity: A case study in Colombia. J Environ Manage 262, 110315.

Cheifetz, N., Noumir, Z., Samé, A., Sandraz, A.-C., Féliers, C., Heim, V., 2017. Modeling and clustering water demand patterns from real-world smart meter data. Drinking Water Engineering and Science 10, 75-82.

Cole, G., Stewart, R.A., 2013. Smart meter enabled disaggregation of urban peak water demand: precursor to effective urban water planning. Urban Water Journal 10, 174-194.

Cominola, A., Moro, A., Riva, L., Giuliani, M., Castelleti, A., 2016. Profiling residential water users' routines by eigenbehavior modelling.

Cominola, A., Nguyen, K., Giuliani, M., Stewart, R., Maier, H., Castelletti, A., 2019. Data mining to uncover heterogeneous water use behaviors from smart meter data. Water Resources Research.
Csikszentmihalyi, M., Abuhamdeh, S., Nakamura, J., 2014. Flow, Flow and the Foundations of Positive Psychology. Springer Netherlands, Dordrecht, pp. 227-238.
Eke, C.I., Norman, A.A., Shuib, L., Nweke, H.F., 2019. A survey of user profiling: state-of-the-art, challenges, and solutions. IEEE Access 7, 144907-144924.
Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A., 2007. User profiles for personalized information access, The adaptive web. Springer, pp. 54-89.
Gurung, T.R., Stewart, R.A., Beal, C.D., Sharma, A.K., 2015. 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, 642-654.
Lally, P., Van Jaarsveld, C.H., Potts, H.W., Wardle, J., 2010. How are habits formed: Modelling habit formation in the real world. European journal of social psychology 40, 998-1009.
Leyli-Abadi, M., Same, A., Oukhellou, L., Cheifetz, N., Mandel, P., Feliers, C., Chesneau, O., 2018. Mixture of Non-homogeneous Hidden Markov Models for Clustering and Prediction of Water Consumption Time Series, Proceedings of the International Joint Conference on Neural Networks.
Liu, A., Mukheibir, P., 2018. Digital metering feedback and changes in water consumption – A review. Resources, Conservation and Recycling 134, 136-148.
McKenna, S.A., Fusco, F., Eck, B.J., 2014. Water Demand Pattern Classification from Smart Meter Data. Procedia Engineering 70, 1121-1130.
Montaner, M., López, B., De La Rosa, J.L., 2003. A taxonomy of recommender agents on the internet. Artificial intelligence review 19, 285-330.
Nguyen, K., Stewart, R., Zhang, H., Giurco, D., Blumenstein, M., Rahim, S., 2020. Next Generation Machine Learning for Urban Water Management. Water e-Journal 5, 1-7.
Nguyen, K.A., Sahin, O., Stewart, R.A., Zhang, H., 2016. Water demand forecasting with AUTOFLOW© using State-Space approach, 8th International Congress on Environmental Modelling and Software.
Nguyen, K.A., Stewart, R.A., Zhang, H., 2014. An autonomous and intelligent expert system for residential water end-use classification. Expert Systems with Applications 41, 342-356.
Padulano, R., Del Giudice, G., 2018. A Mixed Strategy Based on Self-Organizing Map for Water Demand Pattern Profiling of Large-Size Smart Water Grid Data. Water Resources Management 32, 3671-3685.
Rahim, M.S., Nguyen, K.A., Stewart, R.A., Giurco, D., Blumenstein, M., 2019. Predicting Household Water Consumption Events: Towards a Personalised Recommender System to Encourage Water-conscious Behaviour, 2019 International Joint Conference on Neural Networks (IJCNN), pp. 1-8.
Rahim, M.S., Nguyen, K.A., Stewart, R.A., Giurco, D., Blumenstein, M., 2020. Machine Learning and Data Analytic Techniques in Digital Water Metering: A Review. Water 12.
Ricci, F., Rokach, L., Shapira, B., 2015. Recommender Systems: Introduction and Challenges, Recommender Systems Handbook. Springer US, Boston, USA, pp. 1-34.
Stewart, R.A., Willis, R., Giurco, D., Panuwatwanich, K., Capati, G., 2010. Web-based knowledge management system: linking smart metering to the future of urban water planning. Australian Planner 47, 66-74.
Stewart, W.J., 2009. Probability, Markov chains, queues, and simulation : the mathematical basis of performance modeling. Princeton University Press, Princeton, N.J.
Willis, R.M., Stewart, R.A., Giurco, D.P., Talebpour, M.R., Mousavinejad, A., 2013. End use water consumption in households: impact of socio-demographic factors and efficient devices. Journal of Cleaner Production 60, 107-115.
Yang, A., Zhang, H., Stewart, R., Nguyen, K., 2018. Enhancing Residential Water End Use Pattern Recognition Accuracy Using Self-Organizing Maps and K-Means Clustering Techniques: Autoflow v3.1. Water 10.
Yürüten, O., 2017. Recommender Systems for Healthy Behavior Change. Swiss Federal Institute of Technology Lausanne.

Appendix A. Supplementary materials

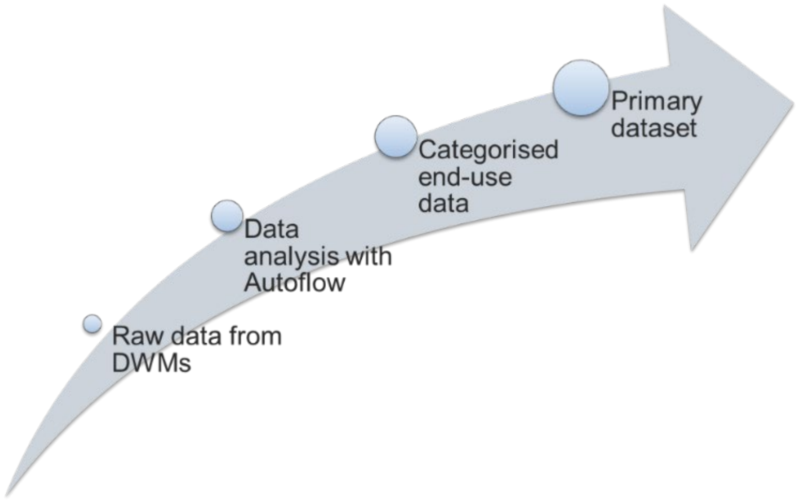
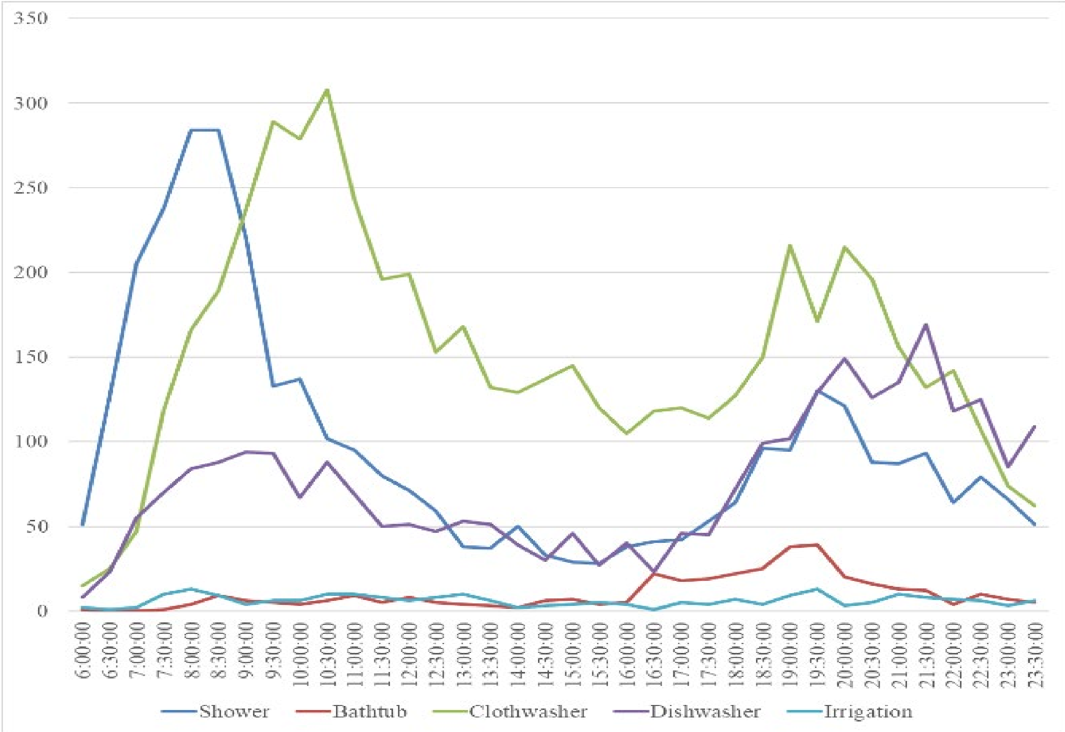
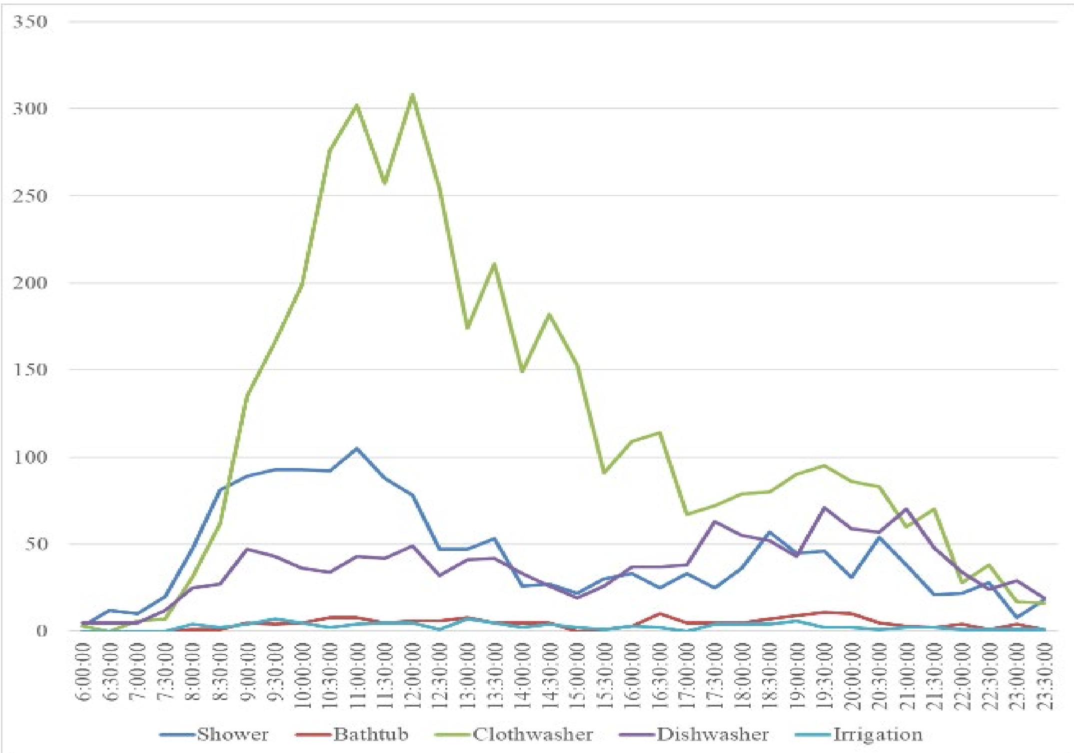


Figure S1. Primary data collection process



(a)



(b)

Figure S2. Count of water end-use events in 30 minutes interval during: a) weekdays; b) weekends.

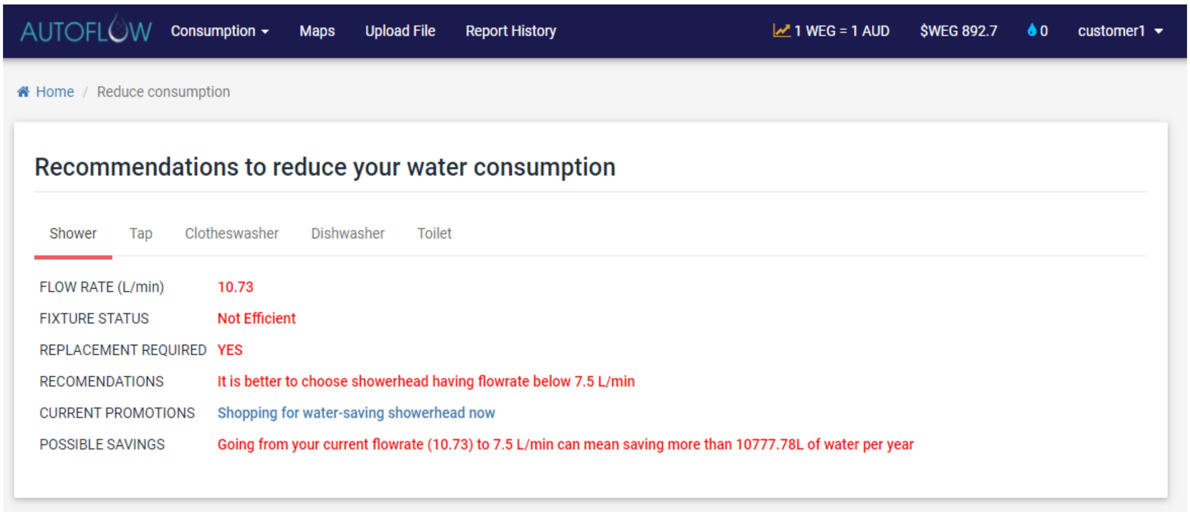
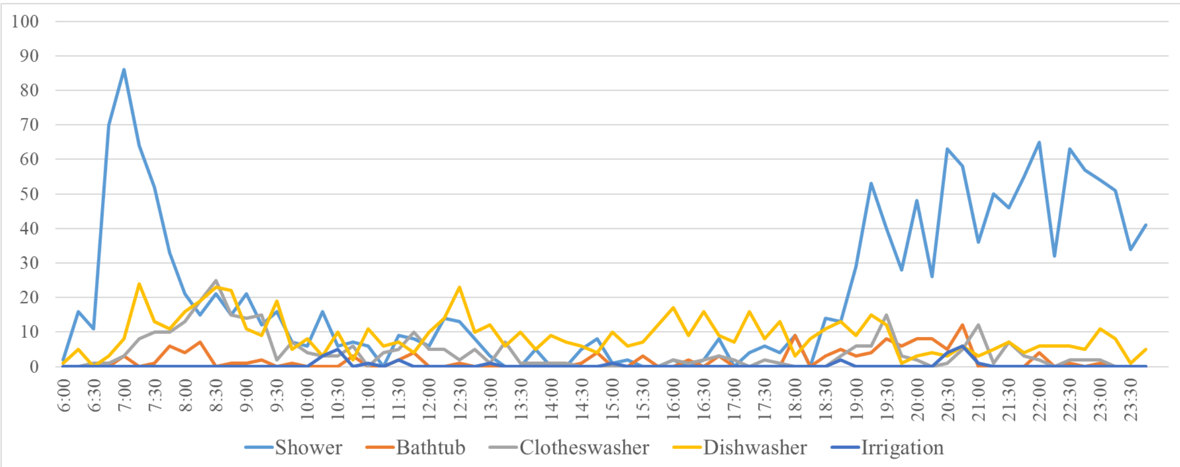
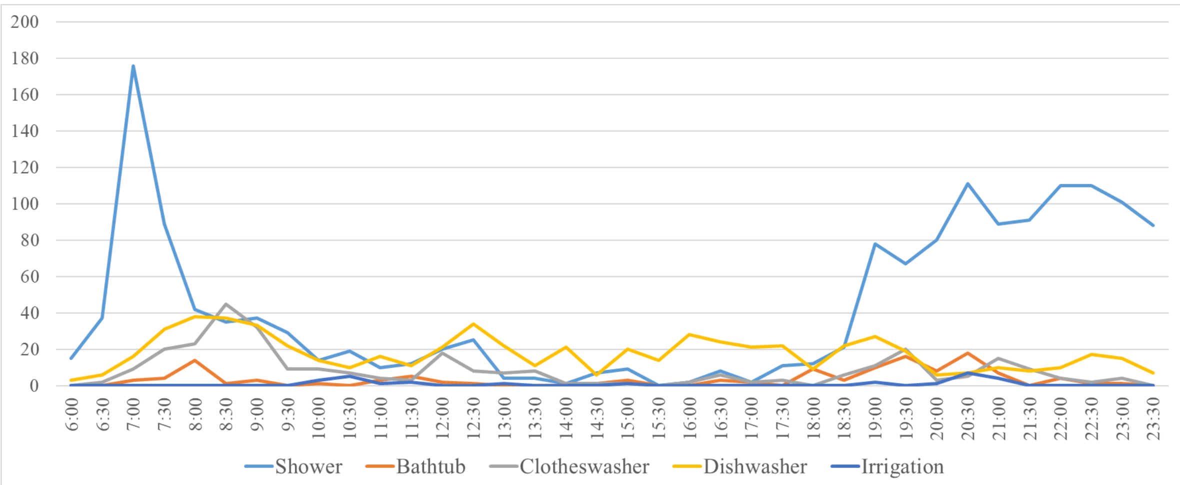


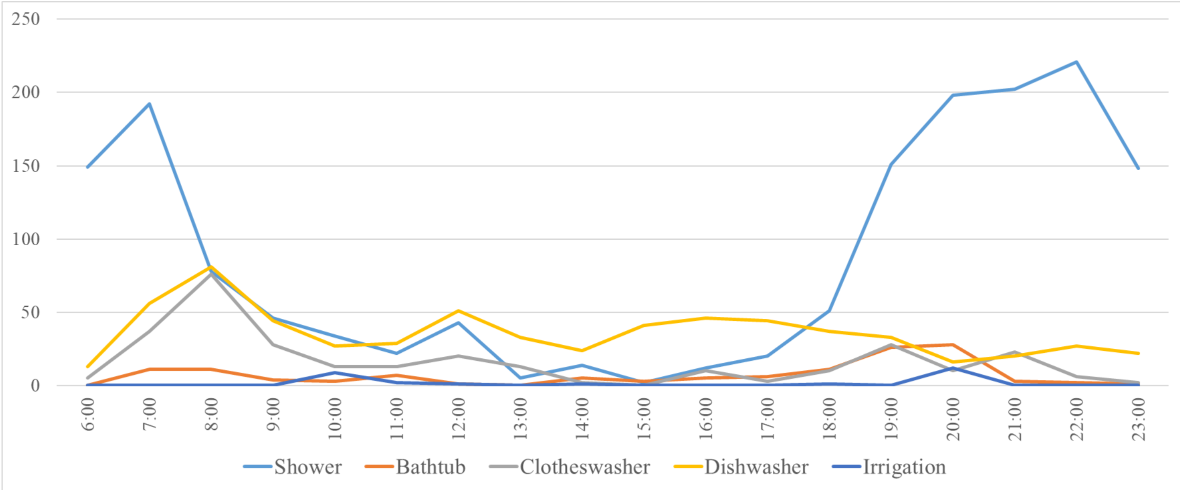
Figure S3. A personalised water recommender system.



(a)



(b)



(c)

Figure S4. Total number of each events at Site 325 at: (a) 60-minute intervals; (b) 30-minute intervals; and (c) 15-minute intervals.