

# Water use signature patterns for analyzing household consumption using medium resolution meter data

Rachel Cardell-Oliver<sup>1</sup>

Received 22 July 2013; revised 29 October 2013; accepted 1 December 2013; published 20 December 2013.

[1] Providers of potable water to households and businesses are charged with conserving water. Addressing this challenge requires accurate information about how water is actually being used. So smart meters are being deployed on a large scale by water providers to collect medium resolution water use data. This paper presents water use signature patterns, the first technique designed for medium resolution meters for discovering patterns that explain how households use water. Signature patterns are clusters (subsets) of water meter readings specified by patterns on volumes and calendar dates. Four types of signature pattern are introduced in this paper: continuous flow days; exceptional peak use days; programmed patterns with recurrent hours; and normal use partitioned by season and period of the day. Signature patterns for each household are calculated using efficient selection rules that scale for city populations and years of data collection. Data from a real-world, large-scale, smart metering trial are analyzed using water use signature patterns. The results demonstrate that water use behaviors are distinctive, for both individuals and populations. Signatures can identify behaviors that are promising targets for water conservation. Pattern discovery can be automated with an efficient and scalable computer program. By identifying relevant consumption patterns in medium resolution meter data, water use signature patterns can help to achieve the water conservation potential of large-scale smart metering.

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

## 1. Introduction

[2] Providers of potable water to households and businesses are charged with conserving water in order to achieve sustainability goals. Strategies for achieving these goals include the development of alternative water sources, water recycling, and reducing consumption [Water Corporation of Western Australia, 2009]. This paper concerns the third strategy. Reducing consumption requires an understanding of how and when water is used. Data collected automatically from smart water meters can be analyzed to explain water use and so inform decision making by water producers and consumers. This is known as smart metering.

[3] A key enabler for smart metering is the development of data mining techniques that identify patterns in metered data. The goal is to understand how households use water so that they can reduce water use where possible. To date, this has been done by detecting human activities, such as taking a shower, in high resolution meter data. However,

medium resolution meters are more appropriate for large-scale smart metering because the ongoing collection, communication, and storage of data are economical. A medium resolution, hourly meter generates 8760 readings per year, while a high resolution meter (5 s period) would generate over 6 million readings. On the other hand, activity recognition techniques cannot be used with medium resolution data because details of the underlying human activities are aggregated, and so hidden. For example, concurrent activities such as taking a shower while the washing machine is running, and sequential activities such as showering and breakfasting in the morning, are all aggregated into single, hourly volume readings. Therefore, a new method is needed to identify usage patterns in medium resolution meter data.

[4] This paper proposes *water use signature patterns*, a novel technique that is designed for medium resolution meter data. Signature pattern analysis is based on cluster analysis. This data mining technique assigns observations, such as meter readings, to subsets called clusters. The clusters are patterns with a regular and intelligible structure defined by calendar (e.g., every Monday at 5 am) and volume (e.g., more than 400 L/h) constraints. Signature clusters describe patterns of behavior. They are not directly linked to specific human activities or appliances.

[5] Data from a smart metering trial by the Water Corporation of Western Australia are used to evaluate the utility of water use signature patterns. The data set describes hourly water consumption by 187 households over 14

<sup>1</sup>CRC for Water Sensitive Cities and School of Computer Science and Software Engineering, University of Western Australia, Crawley, Western Australia, Australia.

Corresponding author: R. Cardell-Oliver, CRC for Water Sensitive Cities and School of Computer Science and Software Engineering, University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia. (rachel.cardell-oliver@uwa.edu.au)

months of use. The criteria for evaluation are that signature patterns are: (1) distinctive for individuals; (2) informative for water conservation; (3) give insight into population trends; and (4) scalable in that they can be calculated efficiently for large data sets. This paper shows how signature patterns meet these evaluation criteria. It is concluded that by identifying relevant consumption patterns in medium resolution meter data, water use signature patterns can help to achieve the water conservation potential of large-scale smart metering.

## 2. Related Work

[6] Studies of household water consumption may be categorized by the different ways they approach the problem of understanding water use. First, is the motivation for the research to predict demand or to reduce consumption? Second, which determining factors, if any, are considered for their influence on water consumption? Third, what type of data on water consumption is used, and fourth, what techniques are used to interpret that data?

[7] Water use signature patterns are designed to support water conservation. This paper addresses the question of how households consume water. Signature patterns are used to answer that question using (only) medium resolution water meter data, and using cluster analysis to discover patterns in that data. This section explains these choices in the context of related work.

### 2.1. Predicting or Reducing Demand

[8] The motivation for researching household water use can be to predict future water demand or to inform conservation strategies. *House-Peters and Chang* [2011] reviewed models for predicting urban water demand. Demand is viewed as a complex coupled human and natural system. The importance of temporal and spatial aspects of water demand is motivated, as are the difficulties of modeling these factors. *Blokker et al.* [2010] implemented a stochastic simulation for predicting demand based on end use probabilities for pulse duration, intensity, and time of day of water usage. *Aksela and Aksela* [2011] presented a probabilistic prediction model based on mixtures of Gaussians. The majority of household water use studies, however, have contributed data analytics for understanding historical water use with water conservation as the goal, rather than developing models for demand prediction.

### 2.2. Factors that Influence Consumption

[9] *Jorgensen et al.* [2009] reviewed integrated models for water consumption and the social and economic factors that influence it. Many different factors have been studied. *Fox et al.* [2009] correlated dwelling characteristics such as size, architectural type, garden presence, age, and garden aspect, with water demand. *Willis et al.* [2013] identified clusters of socio-demographic attributes that are correlated with high water consumption. *Syme et al.* [2004] studied the relationship between householder's attitudes to their garden and their outdoor water use. *Grafton et al.* [2011] analyzed the relationship between the policy levers of water pricing, subsidies, and promotion of conservationist attitudes and water consumption. *Britton et al.* [2008] studied the influence of intervention strategies for household leaks on reducing water use.

[10] Human perceptions also affect behavior. *Beal et al.* [2011] studied how well householders' perceptions of their water use are linked to their actual water use. *Russell and Fielding* [2010] reviewed theories in psychology that explain how attitudes, beliefs, habits, personal capabilities, and contextual factors all contribute to users' intentions to save water, and that intentions are believed to be the most immediate predictor of actual behaviors. *Corral-Verdugo et al.* [2002] also considered perceptions, demonstrating that the "tragedy of the commons" applies to water use: perceptions about how other people are using water have a negative effect on water conservation.

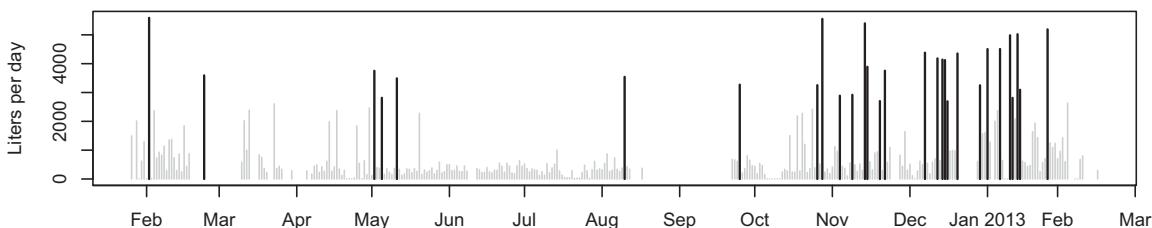
### 2.3. Water Consumption Data

[11] Obtaining quantitative data on *how* households use their water is a common requirement for empirical studies. Consumption data are collected, processed, and analyzed and then combined with qualitative and quantitative data on why the consumption occurs. Raw data on water consumption come in three forms: low, medium, and high resolution. Different analysis techniques are used to discover patterns in the raw data depending on the resolution of the data and the needs of the study.

[12] Low resolution data report annual, quarterly, or monthly water consumption. It is easily obtained for large populations from water provider billing records. But the low resolution hides significant information, and so the patterns that can be detected are limited. Analysis of low resolution data typically uses statistical techniques to summarize the data and correlate water use with other usage factors. *Corral-Verdugo et al.* [2002]; *Grafton et al.* [2011]; *Fox et al.* [2009]; *Syme et al.* [2004]; and *Jorgensen et al.* [2009] are studies that use low resolution consumption data.

[13] High resolution data report consumption with a period of seconds or minutes, using flow meters or pressure meters. Data at this level of detail can be used to identify patterns corresponding to human activities such as taking a shower. However, a system needs to be trained to recognize activities and the results necessarily have imperfect accuracy; activities that overlap in time are still hard to identify. Furthermore, with over 6 million readings generated per meter per year, it is not feasible to collect and store this type of data automatically for large populations and over long time scales. *Willis et al.* [2013] and *Beal et al.* [2011] are consumption studies that use high resolution meter data.

[14] Medium resolution data report consumption with a period of one or more hours. Smart meters to collect medium resolution data automatically are currently being deployed for large populations by water providers such as Water Corporation of Western Australia [*Cardell-Oliver and Peach*, 2013] and Wide Bay Water Corporation [*Britton et al.*, 2008]. This type of data can be obtained for large populations and over long time scales, because the costs of collecting, communicating, and storing meter readings are economical. *Britton et al.* [2008] used hourly meter data for the specific task of leak detection. However, to the best of our knowledge, to date no general techniques have been designed for pattern recognition in medium resolution data for the purpose of water consumption research.



**Figure 1.** Sporadic signature pattern of peak (black) and normal days (gray).

#### 2.4. Analysis of Consumption

[15] Statistical correlation and summary measures are used to analyze low resolution meter data. For high resolution data, activity recognition is the main approach used. Activities of interest are selected that are related to human activities and appliance use. For example, flushing a toilet, showering, running the dishwasher, or watering the garden. Each of these activities is associated with a usage signature comprising a continuous subsequence of high resolution meter readings. Next, an activity recognition program is “trained:” an expert manually labels subsequences in metered training data with activity names. A computer program is then used to “learn” rules for assigning activity labels to the remainder of the metered sequence automatically. Finally, summary results are generated such as the number of times each type of activity occurs and the total volume of water used for each activity type. In the domain of electricity consumption, *Widen et al.* [2009] employed user-annotated data to construct load profiles and *Firth et al.* [2008] observe that some appliance types such as refrigerators have recognizable cycling signatures that can be labeled automatically in metered data in order to infer trends.

[16] For analyzing household water use, *Nguyen et al.* [2013] enumerate some of the problems with the activity recognition approach and, in particular, with the commercial activity recognition system, TraceWizard<sup>©</sup>. *Nguyen et al.* [2013] proposed a novel activity recognition algorithm and compared it with TraceWizard<sup>©</sup>. They were able to improve labeling accuracy for activities such as dishwasher use, but not for activities such as garden watering and bathtub use. *Froehlich et al.* [2011] introduced HydroSense, a novel activity recognition algorithm using a pressure sensor to record water use. Estimating from Figure 7a in the paper, for isolated events the accuracy of the best algorithm was about 91% (9% error), and for overlapping events 68–82% depending on how closely the events overlap in time. *Nguyen et al.* [2013] and *Froehlich et al.* [2011] demonstrate that there is certainly scope to improve the accuracy of activity recognition techniques. But the problems remain of the large time and effort required for training; necessarily inaccurate recognition of some types of activity; and the difficulty of disaggregating events that overlap in time.

### 3. Water Use Signature Patterns

[17] This section defines water use signature patterns, and introduces four types of signature pattern for understanding household water use: continuous flows, peak days, programmed use, and normal consumption partitioned by the period of the day.

#### 3.1. Cluster Analysis

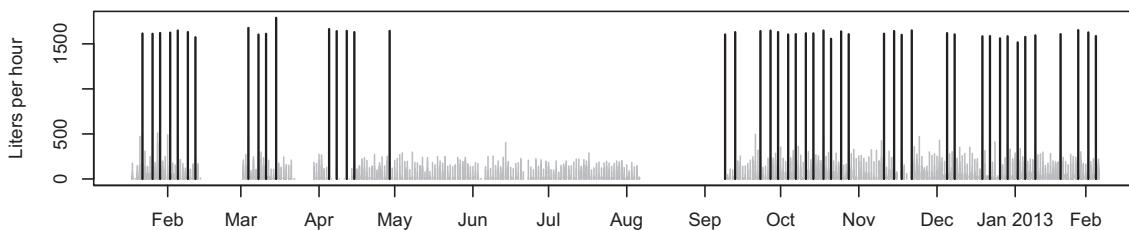
[18] A water use signature pattern is a set of selected meter readings called a *cluster*. A cluster is a subset of observations that are considered a group in that the observations within a cluster group are closer to one another, by some measure, than to observations outside the cluster. Cluster analysis is the process of identifying one or more clusters in a data set. Cluster analysis is an appropriate choice for the classification of medium resolution water meter data because it is one of the most general data mining approaches for classification, and it can discover new types of patterns without making too many assumptions about the structure of the data [Han et al., 2011]. For water use signature patterns, constraints on calendar dates and times and volumes are used to define the clusters. Of particular interest for the goal of reducing water use are clusters for which a small cause (in time) has a large effect (in volume).

[19] An example of a cluster of water meter readings is “the cluster of all days in which water use is exceptionally high,” defined as greater than 2500 L/d. Figure 1 shows an example of this type of cluster. The timing of the days in the pattern is sporadic. But a pattern is discernible in the high volumes used on the selected days in comparison with the remaining days in the data set. This exceptional use cluster is interesting, meaning it is worth identifying, for several reasons. Irregular, exceptional activities often consume more water than regular, daily activities, so they are good candidates for reducing waste. For the real-world household shown in Figure 1, the cluster of the highest 10% of water use days accounts for 43% of all the water used by this household. It is likely that the householder will have the motivation to reduce this category of use and also will be able to pinpoint the activities on the 30 days of the trial that led to that high consumption. Thus, this cluster provides the householder with the information needed to change their water use habits in future.

[20] Clusters can be also used to characterize regular, calendared water use. For example, Figure 2 shows a cluster of all observations greater than 400 L/h that recur at the same hour(s) of the week. In this case, these observations occur in the hour up to 8 pm on most Thursdays and Sundays. Regular, high-use habits often consume a significant volume of water for a small proportion of metered time: 0.6% of metered hours accounts for 22% of all water used for this household. Thus, regular patterns are promising targets for reducing water use.

#### 3.2. Notation

[21] A population of water users comprises individual households each associated with a unique water meter and



**Figure 2.** Regular signature pattern of programmed (black) and normal hours (gray).

a history of time stamped volumes corresponding to water used each hour (or other metering period). In the following,  $M$  refers to the set of meter identifiers,  $T$  the type of time stamps for meter readings and  $V$  the usage volumes corresponding to number of liters used in each metered time step. The same sequence can be viewed at different temporal granularities, from hours to days to seasons, by aggregating the volumes in each period.

[22] The raw data provided by the medium resolution meters studied in this paper are a sequence of hourly readings, denoted  $\sigma_h^m$ , for user  $m \in M$ . The elements of the sequence are pairs  $(t_i, v_i) \in T \times V$  with time stamp  $t_i$  and volume  $v_i$ . For example,  $(4/11/12 19:00, 529) \in \sigma_h^{20}$  means that 529 L of water was recorded by meter 20 in the hour up to 7 pm on 4 November 2012. Time stamps are unique, so each sequence can be treated as a set and set notation will be used throughout this paper.

[23] Water use signature patterns are clusters, that is, simply subsets of observations selected from a base set of observations. Any signature pattern  $P_m \subseteq \sigma^m$  can be defined using a set definition expression of the form:

$$P_m = \{(t_i, v_i) | (t_i, v_i) \in \sigma_g^m \wedge C(t_i, v_i)\}$$

[24] This expression is read as “signature pattern  $P_m$  for meter  $m$  is the set of all metered observations  $(t_i, v_i)$  taken from a base set of observations  $\sigma_g^m$  where the logical conditions  $C$  (given as a logical predicate) are satisfied for each selected observation.” The subscript of a sequence  $\sigma$  denotes the time granularity of the sequence: hour  $h$ , period  $p$ , or day  $d$ .

[25] Superscripts and subscripts are omitted if the context is clear. Readings are usually periodic. However, because the data are gathered in real-world contexts, there will be gaps in the time series caused by reporting errors and missing or damaged observations, as can be seen in Figures 1 and 2.

[26] Time stamps in metered sequences are assumed to conform to the ISO 8601 standard for dates and times. This assumption is made without loss of generality, since non-conforming readings can be mapped onto the standard. A single ISO 8601 time stamp has several equivalent forms from which different attributes can be accessed. Table 1 lists the attribute names used in this paper. Standard relational notation is used to identify the attributes of a time stamp: for time stamp  $t$ ,  $t.DW$  is the day of the week for that time stamp,  $t.DU$  is a unique identifier for the date,  $t.M$  is the month, and  $t.h$  and  $t.m$  denote the hour and minute of the time stamp. Time zones for readings are assumed to be local time without daylight saving. But, if required, these properties can also be specified using the ISO time zone

standards. Week of the year numbering follows the formal ISO definition with 01 as the week with the year’s first Thursday in it.

[27] Periods of the day are defined by specific sets of hours of the day:

<i>PeriodOfDay</i>	=	{dawn, morning, afternoon, evening}
<i>dawn</i>	=	{0, 1, 2, 3, 4, 5}
<i>morning</i>	=	{6, 7, 8, 9, 10, 11}
<i>afternoon</i>	=	{12, 13, 14, 15, 16, 17}
<i>evening</i>	=	{18, 19, 20, 21, 22, 23}

[28] Aggregate sequences are used for defining new patterns:  $\sigma_d^m$  sums the hourly readings of each day into a daily total;  $\sigma_{dmin}^m$  selects the minimum hourly reading from each day;  $\sigma_p^m(pod)$  sums all hourly readings from  $pod \in \text{PeriodOfDay}$ . The entry is undefined if any of the hourly observations for an aggregate are missing.

### 3.3. Peak Use

[29] Signature patterns comprise selected readings from a sequence of meter readings. A simple selection strategy is restricting the range of volumes. For example, for a set  $\sigma$  and threshold volume  $K$ , the set of all observations with volume above  $K$  is defined:

$$\text{above}(\sigma, K) = \{(t_i, v_i) | (t_i, v_i) \in \sigma \wedge v_i > K\}$$

[30] Thresholds are selected by expert opinion and analysis. For example, a constant of 1000 L could be chosen as a global threshold for peak use. Alternatively, Cardell-Oliver and Peach [2013] used the clustering algorithm  $k$ -means to partition observed volumes for each individual into three categories: exceptional, high, and low daily water use. Thresholds for each user were defined as the boundaries between these clusters. Another approach is to

**Table 1.** ISO 8601 Time Stamp Attributes

Symbol	Meaning
Y	year (0000..9999)
M	month (1 Jan. to 12 Dec.)
DM	day of month (01..31)
DY	day of year (001..366)
DU	number of days since 1 Jan. 1970 to date
DW	day of week (1 = Monday to 7 = Sunday)
W	week of year (01..52)
h	hour of day (00..23)
m	minute of hour (00..59)

set as thresholds a certain percentile of the set of all volumes either for individuals or for a whole population. There are advantages and disadvantages of each method, but the percentile method is used in this paper because: (1) it can be calculated immediately from stored meter readings in a database; (2) the calculation cost is lower than for clustering algorithms such as  $k$ -means; and (3) it is more meaningful for individual users than a global threshold.

[31] Peak days are those on which water usage is significantly higher than normal for a particular meter. These outliers are important because they have a disproportionate effect on water use. Peaks are a sporadic pattern defined for individual users by a local percentile threshold such as the 90th percentile of all daily volumes. That is, the threshold for the top 10% of daily use. Let *percentile* ( $\sigma_d^m, P$ ) denote the value of the  $P$ th percentile of volumes in the set  $\sigma_d^m$ . The *peak days* signature pattern for percentile threshold  $P$  is defined by:

$$\text{peak}(\sigma_d^m, P) = \text{above}(\sigma_d^m, \text{percentile}(\sigma_d^m, P)) \quad (1)$$

### 3.4. Continuous Flows

[32] The continuous flow signature pattern for meter  $m$  and thresholds  $F$  and  $H$  is defined as the cluster of days in which over period of  $H$  hours the minimum flow volume is at least  $F$  liters per hour. For example,  $H = 24$  specifies a continuous flow over 1 day. The continuous flow volume over each sampling period is estimated by assuming that the same minimum flow occurs during the whole period. Continuous flows are sporadic patterns: they can occur on any calendar day.

$$\begin{aligned} \text{conflow}(\sigma_{d\min}^m, F, H) \\ = \{(d, v) | \exists f. (d, f) \in \sigma_{d\min}^m \wedge f \geq F \wedge v = H \times f\} \end{aligned} \quad (2)$$

[33] Continuous flows, or leaks, are a well-known phenomena in household water use. Several similar definitions can be found in the literature. For example, Britton *et al.* [2008] define a leak as a nonzero flow over 48 h where the flow rate is estimated by the minimum flow between 1 am and 4 am.

### 3.5. Calendar Patterns

[34] Signature patterns use the calendar patterns model of Li *et al.* [2006]. Calendar, or temporal, patterns are defined over a certain time period: a set of contiguous time points defined by a start and end time and a temporal granularity. Examples of time periods include, every day in the year 2013, every month in the year 2011, or every hour in the week from 4 to 10 March 2012. A *calendar pattern* on a time period of length  $n$  is a  $n$ -bit vector that specifies a subset of the time period. For example, for a time period of a month with 31 days that starts on a Monday, the pattern 10000010000001000001000001000 defines every Monday in that month. A calendar pattern is a subset of its underlying time period. Table 2 shows examples of calendar patterns for household water use.

### 3.6. Programmed Water Use

[35] Programmed patterns are a type of calendar pattern. In dry Australian cities, many households use program-

**Table 2.** Calendar Patterns for Household Water Consumption

Pattern Description	Definition <sup>a</sup>
Every third day	$\{d   d.DU \bmod 3 = 0\}$
Every Monday and Friday	$\{d   d.DW = 1 \vee d.DW = 5\}$
Weekends	$\{d   d.DW = 6 \vee d.DW = 7\}$
Even days of the month	$\{d   d.DM \bmod 2 = 0\}$
Hour to 3 am on Tuesdays	$\{d   d.DW = 2 \wedge d.h = 3\}$
Garden watering hours	$\{d   d.h \leq 9 \vee d.h > 18\}$

<sup>a</sup>Each set is a subset of  $T$ , the set of all possible time stamps.

mable controllers to manage their garden watering systems. In Western Australia the use of these systems is controlled by law that restricts the time of year, days of the week or month, and times of day at which they can be used. Over-watering and forgetting to switch off watering systems out of season are activities that may waste a significant proportion of the potable water supplied to households. Householders may not be aware of the waste, since watering systems are often programmed to water in the early morning or late evening, and so may not be observed.

[36] A programmed pattern is one that persists above a given volume threshold for a defined calendar pattern. In this paper, the pattern is a particular time of the day and day of the week. For example, a garden watering system may be programmed to irrigate during the hour before 6 am every Tuesday in summer. Other calendar patterns, such as those in Table 2, could also be used to specify programmed patterns.

[37] The water use signature pattern rule for programmed patterns requires that water use of at least  $B$  liters occurs at the same time of day, on same day of the week, recurring at least  $R$  times. Since a programmed activity may be concurrent with other water uses, only a lower bound  $B$  is specified. The requirement for the day-hour pattern to recur at least  $R$  times defines the persistence of the pattern. Programmed patterns are detected by first selecting candidate clusters for every day of the week  $D$  and hour of day  $H$  (using the *select* function) and then filtering to retain only those clusters that meet the programmed pattern conditions (using the *programmed* function).

$$\begin{aligned} \text{select}(\sigma_h^m, D, H, B) = \\ \{(t, v) | (t, v) \in \sigma_h^m \wedge t.DW = D \wedge t.h = H \wedge v > B\} \\ \text{programmed}(\sigma_h^m, D, H, B, R) = \\ \text{if } \text{size}(\text{select}(\sigma_h^m, D, H, B)) > R \\ \text{then } \text{select}(\sigma_h^m, D, H, B) \\ \text{else } \emptyset \end{aligned} \quad (3)$$

[38] The presence of a regular hourly pattern does not necessarily indicate that garden watering is taking place, since the pattern could be the result of some other high-use, regular activity. Nonetheless, it generally is of interest when a pattern of high consumption recurs at the same hour on the same day of the week.

### 3.7. Normal Consumption Patterns

[39] Normal consumption patterns are designed to give insight into the distinctive features of day-to-day behaviors. These patterns partition all of the meter readings for a

household according to temporal patterns. A hierarchy of temporal patterns is used for partitioning. First, season of the year is recognized as the dominating temporal factor for consumption. Many long term studies have reported higher consumption in summer than in winter. Second, within a season, consumption varies according to the period of the day, day of the week, weekday, or weekend, or other patterns such as those in Table 2. Normal consumption patterns are constructed by first partitioning observations by season of the year and then by a given daily or hourly temporal pattern.

[40] The period of day pattern, for example, comprises a group of clusters. Each cluster is associated with a season, bounded by dates  $T_1$  to  $T_2$  and a period of the day  $P$ :

$$\text{podcluster}(\sigma_p^m, T_1, T_2, P) = \{(t, v) | (t, v) \in \sigma_p^m \wedge t.h \in P \wedge T_1.DU \leq t.DU < T_2.DU\} \quad (4)$$

### 3.8. Pattern Metrics

[41] Cluster metrics are used to summarize, compare, and rank signature patterns. Ranking is used to understand trends and to filter individuals according to different criteria. In this way, ranking provides support for making business decisions such as where to focus advertising campaigns or which individuals to select for personalized intervention.

[42] Metrics for water use signature patterns measure either volume or time, and may relate to an individual or the whole population. Volume metrics characterize the significance of a water use pattern in either absolute (liters) or relative (%) units. Time metrics characterize the scale of a water use pattern in terms of the number of time units (days, periods, or hours) that the pattern occurs. At the population level, the prevalence of a pattern can be measured with respect to the whole population.

[43] The *significance* of a signature pattern  $P_m$  is the sum of all volumes observed in the cluster:

$$\text{sig}(P_m) = \sum_{(t_i, v_i) \in P_m} v_i \quad (5)$$

[44] The *relative significance* of a pattern is the ratio of pattern significance to that of its parent sequence:

$$\text{relsig}(P_m, \sigma_g^m) = \text{sig}(P_m) \times 100 / \text{sig}(\sigma_g^m) \quad (6)$$

[45] The *frequency* of a pattern  $P_m$  is the size of its cluster set:

$$\text{freq}(P_m) = \text{size}(P_m) \quad (7)$$

[46] The *relative frequency* of a pattern is the ratio of the significance of the pattern to that of its parent sequence:

$$\text{relfreq}(P_m, \sigma_g^m) = \text{freq}(P_m) \times 100 / \text{freq}(\sigma_g^m) \quad (8)$$

[47] The *population measure* for a population of meters  $M$ , pattern type  $P$ , and metric function  $f$  is the sum of the  $f$  measure of all individuals' sequences for the pattern. The

metric function  $f$  is either significance (equation (5)) or frequency (equation (7)):

$$\text{pop}(M, P, f) = \sum_{m \in M} f(P_m) \quad (9)$$

[48] The *relative population measure* of a pattern is the ratio of the population measure for that pattern to that of the same measure (volume or time) for  $U$ , the complete set of observations from the whole population:

$$\text{repop}(P, U) = \text{pop}(M, P, f) \times 100 / \text{pop}(M, U, f) \quad (10)$$

[49] Prevalence captures the extent to which a signature pattern type  $P$  is common across a population of meters  $M$ . A meter is *active* for pattern type  $P$  if the pattern cluster  $P_m$  is not empty:

$$\text{active}(P) = \{m \in M | \text{size}(P_m) > 0\}$$

[50] For example, a meter with at least one continuous flow day is active. The prevalence of the pattern is the number of meters in a population that are active for that pattern. *Population prevalence* is the ratio of the number of active meters for a pattern to the population size:

$$\begin{aligned} \text{prev}(P) &= \text{size}(\text{active}(P)) \\ \text{popprev}(P) &= \text{prev}(P) \times 100 / \text{size}(M) \end{aligned} \quad (11)$$

## 4. Results

[51] In this section, the utility of water use signature patterns is evaluated using data from a real-world smart metering study. The process for computing signature patterns is explained in sections 4.2, and 4.3 evaluates the process for its scalability to large populations and over long time scales. Signature patterns are calculated for each of the case study households. The resulting patterns are evaluated using the criteria that signature patterns should be distinctive for individuals (section 4.4) and informative for water conservation (section 4.5). The discovery of previously unknown trends at the individual and population level is demonstrated for the case study population in section 4.6.

### 4.1. Kalgoorlie-Boulder Case Study

[52] Water use signature patterns are a general data analytics technique that can be applied in any setting where time stamped, medium resolution water use sequences are available. Each population will have unique water use signature patterns, dependent on properties such as climate, housing types, and population demographics. In this paper, data used for evaluation are sourced from a real-world smart metering trial run by the Western Australian Water Corporation in the city of Kalgoorlie-Boulder [Cardell-Oliver and Peach, 2013]. Kalgoorlie-Boulder is situated 600 km inland from Perth in the Eastern goldfields, with a population of 33,000. The climate is hot and very dry with mean annual rainfall of 264 mm and annual evaporation of 2943 mm. The main business is mining, from the gold-rush of 1893 to its large open cut mine today. Since 1903 the city's water has been pumped from Mundaring Dam in Perth via a pipeline to Kalgoorlie-Boulder. The long run capital cost of

supplying water for Kalgoorlie-Boulder is approximately 7 \$Aus per kiloliter. The paper uses hourly water meter readings from 187 houses selected randomly from 13,800 properties metered in the Kalgoorlie Smart Meter Trial by the Water Corporation of Western Australia. The data set spans 14 months from January 2012 to February 2013, yielding an average of 263 metered days per household after allowing for data losses during the trial period. Only properties with land use tagged as houses, and not other types such as units, parks, or clubs, are considered in this paper.

#### 4.2. Implementation Steps for Pattern Discovery

[53] This section outlines the process used to perform signature pattern discovery. It is assumed that smart meters have been used to collect water meter readings from a population of households. It is also assumed that these readings are stored in a central database as they are received. The following steps are then followed to process the metered data: (1) Determine the types of signature patterns to be analyzed, e.g., peak and programmed patterns; (2) input water meter readings for all households to a central database; (3) generate aggregate views of metered sequences, e.g., volume per day; (4) explore the input data to decide appropriate values for the parameters of signature rules, e.g., minimum volume and recurrence for programmed patterns; (5) calculate signature pattern clusters for each household and pattern type; (6) measure the significance and frequency of each discovered cluster and generate reports; (7) analyze the results by ranking individuals' clusters in terms of their significance and frequency and the prevalence of patterns in a population.

[54] Steps 1 and 4 of this process require one-off analysis by a domain expert to determine appropriate types of patterns and their parameters. Table 3 shows the patterns and parameter values that were determined for Kalgoorlie-Boulder (steps 1 and 4). This process of choosing parameters is a subjective one, although guided by criteria. For example, the parameters  $B = 400$  and  $R = 4$  for programmed clusters were decided by analysis of the data set. Visual checks (e.g., see Figure 2) suggest that the selected observations do indeed represent a repeated pattern of activity. Although, in principle, any high consumption activity at the same time of day would qualify as a regular pattern, these selected parameters capture only patterns that are (1) sufficiently significant in volume, and (2) sufficiently rare, to constitute an interesting pattern.

[55] The remaining steps 2, 3, 5, 6, and 7 were implemented as functions in the R statistical package (<http://www.r-project.org/>). R was chosen because it supports both statistical analysis and visualization of the outputs using static bar graphs, box plots, and bubble charts. Signature patterns were evaluated for each of the 187 households, and the results saved for further processing. Population summaries were also calculated and saved. Additional software was written using the d3 Data Driven Documents software library (<http://d3js.org>). This software provides interactive web-based graphics for individual signature patterns and for population patterns (step 7). The graphics are available online (<http://people.csse.uwa.edu.au/rachel/waterusesignatures-july2013/>).

[56] Signature pattern discovery could also be implemented using a standard database system that supports a

**Table 3.** Parameters for Kalgoorlie-Boulder Signature Patterns

ID <sup>a</sup>	Value <sup>b</sup>	Description (Rule) <sup>c</sup>
P	90	percentile for peak use (Pk)
F	2	minimum flow per hour for continuous flow (Cf)
H	24	minimum hours for continuous flow (Cf)
B	400	minimum hourly flow for a programmed pattern (Pr)
R	4	minimum recurrences of a day-hour event (Pr)
$T_i$	s <sup>d</sup>	season boundaries (Pd)

<sup>a</sup>Parameter identifier.

<sup>b</sup>Value determined for Kalgoorlie-Boulder case study.

<sup>c</sup>Pk = peak (equation (1)), Cf = conflow (equation (2)), Pr = programmed (equation (3)), Pd = podcluster (equation (4)).

<sup>d</sup>Summer is January–March 2012 and November 2012 to March 2013. Winter is April–October 2012.

query language such as SQL (<http://www.w3schools.com/sql/>). For these databases, SELECT-WHERE queries can be used to evaluate rules (step 5); the functions SUM, COUNT and GROUP BY used to generate views and measure properties (steps 3 and 6); and ORDER BY to rank results (step 7).

#### 4.3. Scalability of Pattern Discovery

[57] Water use signature patterns provide a new way of viewing the water use behaviors of a population. A significant advantage of the signature pattern view, compared with activity recognition, is that the process for identifying signature patterns in smart meter data is readily scalable to large populations over long time scales. Once the signature patterns and their parameters are decided, the calculation of individual signature patterns and population summaries is automatic. The amount of medium resolution metered data that needs to be measured, communicated, and stored is an order of magnitude lower than the high resolution data needed for activity recognition. Therefore, the time and memory required for collecting, storing and calculating signature patterns is manageable when scaling up to smart metering for city populations and time scales of years.

[58] One possible challenge for scalability is evaluating programmed patterns. There are 168 potential programmed patterns to be checked for each individual in the population: one for each hour of the week. For large populations, the calculation of potential patterns could be slow. Fortunately, by partitioning on the most discriminatory attribute at each stage, the search space becomes manageable in practice. For example, for the case study households, partitioning water use hours with a threshold of 400 L/h reduced the search space to an average of 3.5% of its original size. In the next stage, selecting of hours of the day with at least 4 weekly recurrences (1 month) reduced the search space to an average of a few hours per week to be considered for each individual. So the search problem for finding programmed patterns is, in practice, a manageable one.

#### 4.4. Distinctiveness of Signatures

[59] The habits, routines, and exceptional behaviors of water use are distinctive for each household. Signature patterns reflect this variety. One of the insights from this case study is the large amount of variation that exists between households. Table 4 summarizes the signature patterns of five households from the Kalgoorlie-Boulder study. Groups

**Table 4.** Selected Case Study Signature Patterns

ID <sup>a</sup>	190	53	84	10	63
<i>Overall Water Use 17–706 kL<sup>b</sup></i>					
Mean	909 L/d	912 L/d	912 L/d	1931 L/d	1931 L/d
Days	295 days	293 days	291 days	293 days	296 days
Rank <sup>c</sup>	29%	30%	30%	90%	90%
<i>Peak Usage Signature Pattern 7–270 kL<sup>b</sup></i>					
Significance <sup>d</sup>	73.9 kL	113.6 kL	66.8 kL	197.8 kL	152.1 kL
Relative significance <sup>d</sup>	28%	43%	25%	34%	27%
Frequency <sup>e</sup>	30 days	29 days	29 days	30 days	30 days
Relative frequency <sup>e</sup>	10%	10%	10%	10%	10%
<i>Continuous Flow Signature Pattern 0–346 kL<sup>b</sup></i>					
Significance <sup>d</sup>	13.6 kL	5.5 kL	5.0 kL	160.7 kL	16.6 kL
Relative significance <sup>d</sup>	5%	2%	2%	28%	3%
Frequency <sup>e</sup>	14 days	7 days	62 days	160 days	45 days
Relative frequency <sup>e</sup>	5%	2%	21%	54%	15%
<i>Programmed Usage Signature Pattern 0–365 kL<sup>b</sup></i>					
Significance <sup>d</sup>	0 kL	25 kL	0 kL	39 kL	118 kL
Relative significance <sup>d</sup>	0%	9%	0%	7%	21%
Frequency <sup>e</sup>	0 h	24 h	0 h	47 h	127 h
Relative frequency <sup>e</sup>	0.0%	0.3%	0.0%	0.7%	1.8%

<sup>a</sup>Household meter identifier.<sup>b</sup>Population minimum and maximum for this pattern.<sup>c</sup>Percentage of households with lower daily mean usage.<sup>d</sup>Absolute or relative volume significance (equations (5) and (6)).<sup>e</sup>Absolute or relative frequency (equations (7) and (8)).

of households with the same mean daily consumption were selected. Low-use households of meters 190, 53, and 84 are in the 30th population percentile, having a mean daily consumption of 909–912 L. High-use households of meters 10 and 63 are in the 90th population percentile having a mean daily consumption of 1931 L/d. The average daily use for all 187 households in the study is 1200 L/d. The full ranges of total water use across the whole population are reported in the subheadings for each pattern.

[60] Table 4 demonstrates that water consumption habits manifest distinctive individual traits. For example, meter 53 uses 43% of their water budget on extreme days, meter 94 uses only 25%, but both users have the same daily average use. Meter 53 has 4 h per week of programmed use, 190 and 84 have no programmed pattern use, but all three have similar daily average use. For the high-use households, meter 10 has 160 days of continuous flow while meter 63 has only 45 days. The programmed pattern use of meter 10 accounts for 39 kL whilst over the same period meter 63 used 118 kL.

[61] Normal day-to-day usage patterns are also distinctive. Figure 3 demonstrates this by contrasting the period of day patterns for meters 84 and 190. Neither meter has any programmed pattern use. The daily patterns of meter 84 (left) are dominated by evening use (6 pm to midnight) accounting for 46–55% per season. The daily patterns of meter 190 (right) are dominated by mornings (9 am to midday), accounting for 33–49% per season. Day-to-day patterns are biased, having higher summer than winter use. These types of seasonal and diurnal differences at the population level have been observed in previous studies, but their importance for explaining individual household habits and routines is a novel contribution of this study.

#### 4.5. Water Conservation Potential of Patterns

[62] Each signature pattern of a household gives specific details of actual water use (dates, times, and volumes) as

well as the general pattern of that use. To put it another way, a signature pattern provides an abstract view of habits or routines of water use, even though the human activities that cause these patterns are not explicit. The signature patterns of a household can be used to answer questions such as, Which of my behaviors uses the most water? Which behavior has highest ratio of cause (time) to effect (volume)? Which behavior has the highest frequency? These are the patterns that have the most impact on overall consumption and so we argue they are promising targets for water conservation.

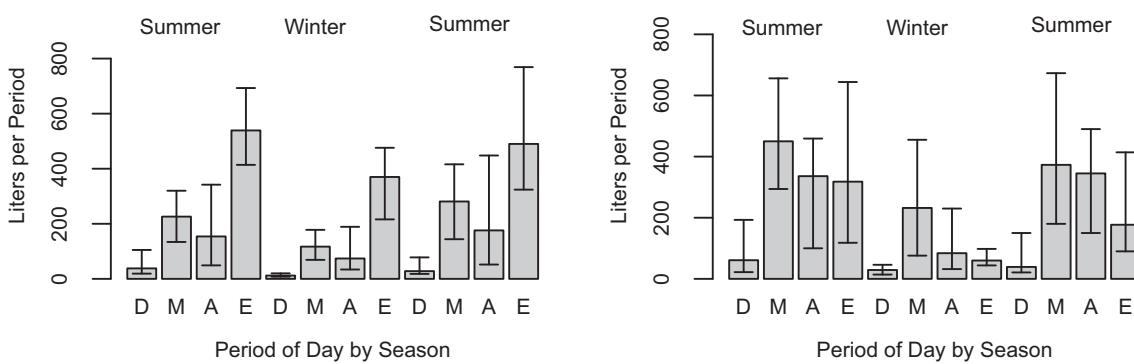
[63] For all five households in Table 4 peak use is the pattern of highest significance. Peak days accounted for 66.8–197.8 kL for these households and from 7 to 270 kL per household across the whole population. Targeting peak days for reducing water consumption has the potential for large savings in both absolute and relative terms, notably 113.6 kL (43%) for low-use meter 53, 197.8 kL (34%) for high-use meter 10, and 152.1 kL (27%) for high-use meter 63.

[64] Continuous flows are relevant for conservation because this type of use is usually preventable [Britton et al., 2009]. For meter 10, 54% of days account for 160.7 kL (28%) of its total consumption in continuous flows. In the context of a population range for continuous flows of 0–346 kL per household, this pattern is a promising target for water conservation.

[65] Many programmed patterns are promising savings targets since they have high significance and a very high ratio of cause (time) to effect (volume). For example, for meter 63, only 1.8% of all metered hours account for 118 kL (21%) of this household's total consumption. Viewing the specific times, dates, and volumes for this cluster reveals that this pattern includes 12 different day and hour combinations, covering 7 days of the week, each recurring between 5 and 21 times.

[66] Time of day patterns cover all metered usage and so can highlight regular habits and routines for water

## CARDELL-OLIVER: WATER USE SIGNATURE PATTERNS



**Figure 3.** Normal use clustered by season and period of day (D = dawn, M = morning, A = afternoon, E = evening) for meters (left) 84 and (right) 190.

conservation. In Figure 3, for example, it can be seen that the highest median water use of meter 84 (left) occurs in the evening and the highest use for meter 190 (right) in the morning and afternoon. Water use behaviors from these times of the day may be linked to routine activities that could be changed. In fact, Figure 3 shows that evening use for meter 190 falls by 44% between the first and second summer seasons from a median of 318 L to a median of 177 L for the same period. This household's evening behavior in summer has changed over the 1 year period, resulting in significant water savings.

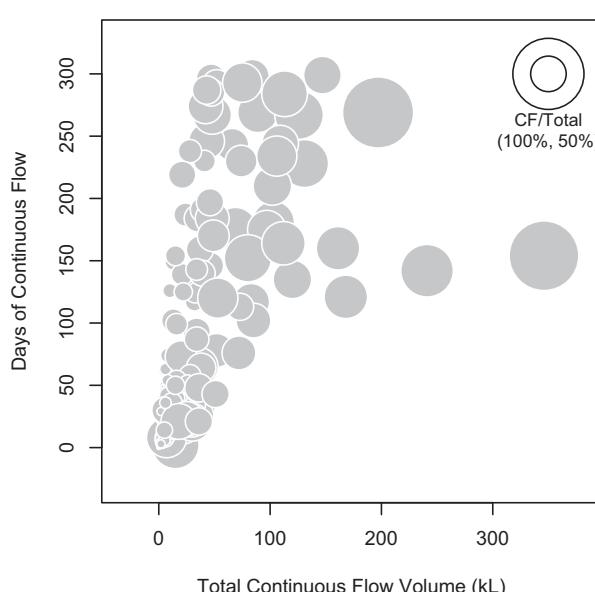
#### 4.6. Trends for Population Behavior

[67] Water utilities have an interest in the behaviors of whole populations of water users. Of particular interest are unknown unknowns: patterns of use that are surprising. Table 5 summarizes overall population patterns for the Kalgoorlie-Boulder sample. During the trial period of an average of 263 metered days per household, the 187 case study households recorded a total of 58.6 ML of water use. The prevalence of different patterns in the population is informative. Continuous flow patterns were the most prevalent at a surprising 94.1% of the population. Programmed patterns were surprising by being less prevalent than expected at 51.9% of the population. Programmed patterns have a very high cause to effect ratio: 0.7% of metered hours account for 14.8% of population volume. Continuous flows were the most frequent pattern, although they have a low cause to effect ratio: 30.2% of metered days for 10.1% of population volume. Peak use had most significant population volume and a cause effect ratio of 10.0% of metered days to 31.7% of population volume. These results demon-

strate that, at the population level, signature patterns provide clear evidence of how and where water is being used, and so where efforts for reducing water use should be focussed.

[68] A convenient way of exploring population rankings is with a bubble chart visualization. A bubble chart is a type of scatter plot in which each dot represents an individual. The x and y axes are used to represent two of the dimensions, such as significance and frequency. The size or color of dots is used to represent further dimensions, such as relative significance.

[69] Figure 4 plots the total volume against number of days continuous flow. The size of the bubbles indicates the relative significance of each continuous flow for the user. The population behaviors for continuous flows were interesting for the water utility because the prevalence of this pattern was high (94.1% of households), the behavior occurred on 30.2% of all metered days, and it accounted for 10.1% of all water used. The bubble chart representation supports targeted feedback by ranking these behaviors by volume and number of days. The bubble chart also highlighted a previously ignored class of households: those with a low overall volume of continuous flows that have high



**Figure 4.** Population summary of continuous flow signatures.

**Table 5.** Population Metrics for Signature Patterns

Population Metric <sup>a</sup>	Context	Pk <sup>b</sup>	Cf <sup>c</sup>	Pr <sup>d</sup>
Relative significance	% of 58.6 ML total population usage	31.7	10.1	14.8
Active significance	% volume for active meters only	31.7	10.5	24.5
Relative frequency	% of all metered days or hours	10.0	30.2	0.7
Prevalence	% of 187 households	100.0	94.1	51.9

<sup>a</sup>Equations (9)–(11).

<sup>b</sup>Pk = peak use pattern (equation (1)).

<sup>c</sup>Cf = continuous flow pattern (equation (2)).

<sup>d</sup>Pr = programmed pattern (equation (3)).

relative significance for that household. This group of households should be targeted because they are likely to be careful water users, who are likely to act quickly to repair a leak.

## 5. Discussion

[70] This section discusses limitations of the current study and identifies areas for future work.

### 5.1. Signature Patterns for Consumers

[71] This paper contributes a novel technique, signature patterns, for identifying patterns of water consumption. Further research is needed to understand how best to present that information to consumers in order to effect behavior change. We envisage that householders would access their household signature patterns using a web portal, a guided interview with water utility staff, information in a letter, or on the water bill. Information on how water has been used and advice on ways to reduce that water use would be provided. There are several existing intervention methods that could be used with signature patterns. In the water use domain, *Britton et al.* [2009] showed that tailored interventions were effective for leak management. For electricity usage, *Fischer* [2008] argued that in order for customers to understand and modify their behavior: “One result, at least, seems clear: feedback stimulates energy (and specifically, electricity) savings” and that feedback needs to show “a close link between specific actions and their effects.” Social-normative messages have also been used to change electricity use behavior. *Laskey and Kavazovic* [2010] found that “most people were strongly compelled to keep up with the Joneses, even if they do not realize or acknowledge it.”

### 5.2. Signature Patterns for Utilities

[72] The behavior of the water utility for this case study has already changed in response to the study’s findings. The main business benefit has been the discovery of unknown unknowns: surprising usage patterns that had not previously been identified. The prevalence of continuous flow patterns was higher than expected, and higher than observed in other studies. On the other hand, the prevalence of regular programmed patterns, linked to automatic garden watering systems, was lower than expected, and lower than observed in other studies. In response, alarms were set in the utility’s database for continuous flows. These alarms trigger interaction with the household as appropriate for the amount of the flow. For high flow rates, the householder can be contacted immediately by telephone. Medium flows may trigger a letter and the least significant flows may simply receive advice in the regular water bill. Several reasons for these unexpected patterns in the Kalgoorlie-Boulder sample have been proposed: a high incidence of evaporative air-conditioners that leak a continuous flow of water when running, a high incidence of rental properties that may have older appliance stock, and possibly a low incidence of water intensive gardens in this very low rainfall city. Developing response strategies for other types of signature patterns is an area for future work.

### 5.3. Different Populations

[73] The findings reported in this paper relate to a single case study, and for a community whose water use behaviors differ from those of the coastal and city households in

many previous studies. For the Kalgoorlie-Boulder sample, water use signature patterns identified novel and significant patterns. We are currently investigating the application of water use signature patterns to study different types of users, namely businesses such as hotels and offices, that have very high water use. A clear business case exists to understand and address the high water use for this class of users [*Atkinson and Medbury*, 2013].

### 5.4. Determinants of Signature Patterns

[74] This paper investigates how, rather than why, households use their metered water. The Kalgoorlie-Boulder case study highlights the distinctiveness of signature patterns for different households. In order to link these distinctive signatures to the human behaviors and attitudes behind them, we propose as future work a mixed-method study. Analysis of correlations between physical and socio-demographic characteristics of households and their water use would provide a richer understanding of the human behaviors underlying water use signature patterns.

### 5.5. Further Automation of Pattern Discovery

[75] Considering the technical aspects of cluster analysis, there are a few limitations of the current system. First, parameter selection for the rules is currently a subjective process of expert judgment and analysis. In future work, we plan to investigate automatic discovery of parameters using search algorithms similar to those used for rule induction in association rule mining [*Han et al.*, 2011]. Second, we plan to investigate search methods for discovering new types of signature patterns beyond the four introduced in this paper. *Li et al.* [2006] describe a method for discovering temporal rules that may be adaptable for finding new types of water use signature patterns. Third, our results demonstrate the importance of context, in this case seasonal context, for interpreting patterns. Fixed seasonal boundaries were defined for this paper. An open problem is how to learn seasonal boundaries from the data, at both the individual and population level.

## 6. Conclusions

[76] Water use signature patterns are a novel approach for characterizing household water consumption, based (only) on readings from medium resolution water meters. A case study applying water use signature patterns to data from a real-world smart metering trial demonstrates the utility of the signature pattern approach. Signature patterns are shown to highlight the distinctive usage patterns of individuals. Furthermore, signature patterns highlight behaviors that are significant, prevalent or frequent at the population level. It is also demonstrated that signatures can identify behaviors that are promising targets for water conservation. The case study shows how signature pattern discovery can be automated by efficient and scalable computer programs.

[77] Smart meters are being deployed on a large scale by water providers to collect medium resolution data on household water consumption. This paper presents the first general purpose technique designed for medium resolution meters for discovering patterns that explain how that water is used. The paper demonstrates that by identifying relevant consumption patterns in medium resolution meter data,

water use signature patterns can help to achieve the water conservation potential of large-scale smart metering.

[78] **Acknowledgments.** This work was supported by the Cooperative Research Centre for Water Sensitive Cities. The author would like to thank G. Peach, S. Atkinson, K. Medbury, S. Volotovsky, and R. Pickering of the Western Australian Water Corporation for the Kalgoorlie-Boulder smart meter data and for their advice on interpreting it. The author is also grateful for the helpful comments of the reviewers and A. Simpson, A. Ghadouani, and J. Gomes.

## References

- Aksela, K., and M. Aksela (2011), Demand estimation with automated meter reading in a distribution network, *J. Water Resour. Plann. Manage.*, *137*(5), 19–26.
- Atkinson, S., and K. Medbury (2013), Driving water efficiency improvements in large customers: Linking customer engagement and high resolution water use data, paper presented at AWA Water Efficiency Conference, Brighton Beach, Sydney, Australia, March 2013, Australian Water Association.
- Beal, C. D., R. A. Stewart, and K. Fielding (2011), A novel mixed method smart metering approach to reconciling differences between perceived and actual residential end use water consumption, *J. Cleaner Prod.*, *60*, 116–128, doi:10.1016/j.jclepro.2011.09.007.
- Blokker, E., J. Vreeburg, and J. van Dijk (2010), Simulating residential water demand with a stochastic end-use model, *J. Water Resour. Plann. Manage.*, *136*(1), 19–26.
- Britton, T., G. Cole, R. Stewart, and D. Wiskar (2008), Remote diagnosis of leakage in residential households, *J. Aust. Water Assoc.*, *35*(6), 56–60.
- Britton, T., R. Stewart, and K. O'Halloran (2009), Smart metering: Providing the foundation for post meter leakage management, paper presented at IWA Efficient 2009 Conference, Australian Water Association, Sydney, Australia.
- Cardell-Oliver, R., and G. Peach (2013), Making sense of smart metering data: A data mining approach for discovering water use patterns, *J. Aust. Water Assoc.*, *40*(2), 124–128.
- Corral-Verdugo, V., M. Fras-Armenta, F. Prez-Urias, V. Ordua-Cabrera, and N. Espinoza-Gallego (2002), Residential water consumption, motivation for conserving water and the continuing tragedy of the commons, *Environ. Manage.*, *30*(4), 527–535.
- Firth, S., K. Lomas, A. Wright, and R. Wall (2008), Identifying trends in the use of domestic appliances from household electricity consumption measurements, *Energy Build.*, *40*(5), 926–936, doi:10.1016/j.enbuild.2007.07.005.
- Fischer, C. (2008), Feedback on household electricity consumption: A tool for saving energy?, *Energy Efficiency*, *1*(1), 79–104.
- Fox, C., B. McIntosh, and P. Jeffrey (2009), Classifying households for water demand forecasting using physical property characteristics, *Land Use Policy*, *26*(3), 558–568, doi:10.1016/j.landusepol.2008.08.004.
- Froehlich, J., E. Larson, E. Saba, T. Campbell, L. Atlas, J. Fogarty, and S. Patel (2011), A longitudinal study of pressure sensing to infer real-world water usage events in the home, in *Proceedings of the 9th International Conference on Pervasive Computing*, edited by K. Lyons, J. Hightower, E. M. Huang, pp. 50–69, Springer, Berlin.
- Grafton, R. Q., M. B. Ward, H. To, and T. Kompas (2011), Determinants of residential water consumption: Evidence and analysis from a 10-country household survey, *Water Resour. Res.*, *47*, W08537, doi:10.1029/2010WR009685.
- Han, J., M. Kamber, and J. Pei (2011), *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, San Francisco, Calif.
- House-Peters, L. A., and H. Chang (2011), Urban water demand modeling: Review of concepts, methods, and organizing principles, *Water Resour. Res.*, *47*, W05401, doi:10.1029/2010WR009624.
- Jorgensen, B., M. Graymore, and K. O'Toole (2009), Household water use behavior: An integrated model, *J. Environ. Manage.*, *91*(1), 227–236, doi:10.1016/j.jenvman.2009.08.009.
- Laskey, A., and O. Kavazovic (2010), Opower energy efficiency through behavioural science and technology, *XRDS*, *17*, 47–51, doi:10.1145/1961678.1961687.
- Li, Y., S. Zhu, X. S. Wang, and S. Jajodia (2006), Looking into the seeds of time: Discovering temporal patterns in large transaction sets, *Inform. Sci.*, *176*(8), 1003–1031, doi:10.1016/j.ins.2005.01.019.
- Nguyen, K., H. Zhang, and R. A. Stewart (2013), Development of an intelligent model to categorise residential water end use events, *J. Hydro-environ. Res.*, *7*(3), 182–201, doi:10.1016/j.jher.2013.02.004.
- Russell, S., and K. Fielding (2010), Water demand management research: A psychological perspective, *Water Resour. Res.*, *46*, W05302, doi:10.1029/2009WR008408.
- Syme, G. J., Q. Shao, M. Po, and E. Campbell (2004), Predicting and understanding home garden water use, *Landscape Urban Plann.*, *68*(1), 121–128, doi:10.1016/j.landurbplan.2003.08.002.
- Water Corporation of Western Australia (2009), Water forever: Towards climate resilience, Water Corporation, Leederville, Western Australia. [Available at [www.watercorporation.com.au](http://www.watercorporation.com.au), last accessed Oct. 2012.]
- Widen, J., M. Lundh, I. Vassileva, E. Dahlquist, K. Ellegrård, and E. Wackelgard (2009), Constructing load profiles for household electricity and hot water from time-use data: Modelling approach and validation, *Energy Build.*, *41*(7), 753–768, doi:10.1016/j.enbuild.2009.02.013.
- Willis, R. M., R. A. Stewart, D. P. Giurco, M. R. Talebpour, and A. Mousavinejad (2013), End use water consumption in households: Impact of socio-demographic factors and efficient devices, *J. Cleaner Prod.*, *60*, 107–115, doi:10.1016/j.jclepro.2011.08.006.