CRC for
Water Sensitive Cities

Using smart meters and data mining to inform demand management

Rachel Cardell-Oliver and Helen Gigney

Australian Government
Department of Industry and Science

Business
Cooperative Research
Centres Programme

Using smart meters and data mining to inform demand management
Intelligent Urban Water Systems (Project C5.1)
C5.1 – 1 - 2015

Authors
Associate Professor Rachel Cardell-Oliver
Ms Helen Gigney

© 2015 Cooperative Research Centre for Water Sensitive Cities

This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of it may be reproduced by any process without written permission from the publisher. Requests and inquiries concerning reproduction rights should be directed to the publisher.

Publisher

Cooperative Research Centre for Water Sensitive Cities
8 Scenic Blvd, Level 1,
Clayton Campus
Monash University
Clayton, VIC 3800

p. +61 3 9902 4985
e. admin@crcwsc.org.au
w. www.watersensitivecities.org.au

Date of publication: July 2015

An appropriate citation for this document is:

Cardell-Oliver R and Gigney H, 2015. *Using smart meters and data mining to inform demand management,* Melbourne, Australia: Cooperative Research Centre for Water Sensitive Cities,

Table of Contents

Executive summary

Data analytics for smart water meters, is the process of discovering valuable knowledge from usage data. Analysis of customers' smart meter consumption data can be used to underpin demand management strategies. This document has two central aims:

1. to describe a suite of data analytics methods for understanding household water meter data; and
2. to quantify potential water saving scenarios with each technique based on real-world case studies of two distinct populations.

Part 1 of this report is a tutorial introduction to a suite of techniques for smart metering that have been developed as part of project *Intelligent Urban Water Systems* (Project C5.1) by researchers at the University of Western Australia and the Water Corporation of WA. The analytics were designed to identify three customer end-use cases with applications for demand management:

- leaks and continuous flows;
- anomalous peak days of use; and
- recurrent habits of irrigation and other high volume uses.

These use cases were chosen because there is the potential to achieve significant water savings associated with each use case without negatively impacting on the user's wellbeing, by reducing the volume of water that is simply wasted. In addition, where seasonal variability can be attributed to particular end-uses water savings can be even more valuable, if they can reduce peak demand and defer capital investment.

Part 2 of this report demonstrates the business value of smart metering using results from two case studies. The Water Corporation of Western Australia has invested in a smart meter network including 13,500 customer installations in Kalgoorlie-Boulder, completed in 2012 and 14,500 in the Pilbara in 2013/14. With more than 12 months of continuous hourly water use data from both areas there is sufficient data to be able to apply a range of analytics including novel data mining techniques and automated pattern discovery algorithms to identify and quantify end use cases and profile peak demand for these populations.

Hourly consumption data has already proven valuable for early leak detection but has the potential to provide utilities with much greater understanding of water use characteristics throughout its networks, including diurnal and seasonal demand fluctuations and drivers to enable targeted demand management strategies to be implemented. However, for this potential to be realised water utilities need to have the capacity to run suitable analytics and the confidence that the analytics are addressing fundamental issues of benefit to the business.

The results demonstrate that application of knowledge on user behaviour enables more effective targeted water efficiency campaigns and engagement strategies to be used. Behaviours targeted in this study include those associated with high volume discretionary outdoor use, irrigation roster compliance, irrigation efficiency and leakage. Water use associated with these behaviours is identified by one or more of the following characteristics:

- intense demand (high rate per hour)
- recognisable temporal patterns (eg every Monday at 5am)
- population significance (high rate of penetration, perseverance or water use across a population)

Highlights of the results from the cases studies include:

1. Quantification of "unknown unknowns" such as a high proportion (90%) of customers with continuous flows in Kalgoorlie-Boulder.

2. Identification of a high proportion of inefficient garden watering in Karratha with 30% of customers applying significantly more water than the recommended rate for gardens in this region.

3. Quantification peak for garden watering in the early morning, accounting for up to 70% of the 6am demand in Karratha in summer months.

4. Potential water savings are population dependent, but significant for both the case studies. For Karratha (average use 446 kL/hh), recurrent habits offer the best target for water efficiency measures (91 kL/hh/year). For Kalgoorlie (average use 366 kL/hh), where automatic watering systems are less common, anomalous peaks are the best target (78 kL/hh/year). Continuous flows, a traditional water efficiency target are 23 kL/hh/year for each of these two populations.

Acknowledgements

The authors would like to thank Steve Atkinson, Garry Peach, Ben Jarvis, Kel Medbury, Richard Pickering and Nathan Harper at the Water Corporation of Western Australia for the smart meter datasets and for their advice on interpreting them. This research is funded by the Cooperative Research Centre for Water Sensitive Cities (CRCWSC) under *Intelligent Urban Water Systems* (Project C5.1) with additional R&D funding from the Water Corporation. The authors thank colleagues across the CRCWSC programs for their feedback on this project. This research has been approved by the Human Research Ethics Office (HREO) of the University of Western Australia (RA/4/1/6253).

Part 1: Smart metering data analytics

Figure 1. Overview of a Data Analytics Process

The aim of **smart water metering analytics** is to discover valuable knowledge about how drinking water is used by individuals and populations. The input for this activity is consumption data from water meters at each household. The output is knowledge for water utilities and for individual water users. Our approach to the analytics is based in the discipline of Computer Science in the field of **data mining**. Smart meter time series from households of a population is analysed using algorithms for identifying water use behaviours. Features of the discovered behaviours are analysed to provide actionable insights. For example, the total volume of water used in a particular behaviour, or comparisons of an activity with those of other individuals or other activities. Results for individual households can be used for personalised feedback on ways to save water. Results for a whole population can be used by water utilities to segment customers and so better understand their customer base. Figure 1 gives an overview of this process.

Household end use analysis is used to apportion water used to different activities. Analytics software such as Trace Wizard can be trained on high resolution (L/min) data to identify most common household uses. However, disaggregation of water use activities is difficult using the coarser L/hr data supplied by the smart meters, because concurrent and sequential activities can aggregated into single hourly volumes. The novel analytics presented in this report to identify water use signature patterns in hourly data. Three types of signature patterns are considered: continuous flows, anomalous peaks and recurrent habits.

The ability to determine how water is being at a range of spatial and temporal scales can be used to better understand and manage demand. Being able to then identify 'who' by clustering similar users into groups can provide the basis for tailored demand management activities pitched to different user segments. This customer segmentation approach has already proving successful in other areas of the water business.

Figure 2 summarises the four steps of data analytics: input, behaviour identification, feature selection, summarise. The following sections summarise each of these steps.

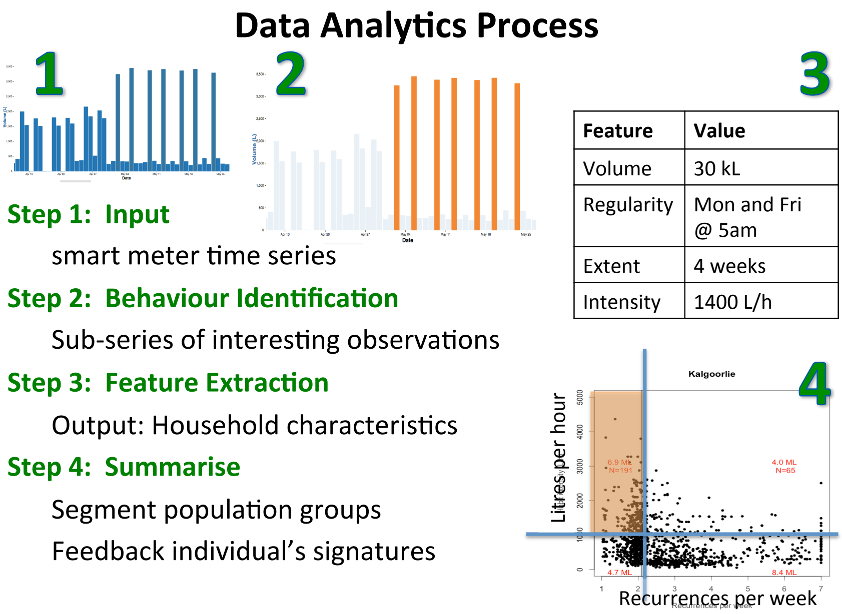


Figure 2. Four Steps of Data Analytics for Smart Water Metering

Four steps of smart metering

Step 1: Input

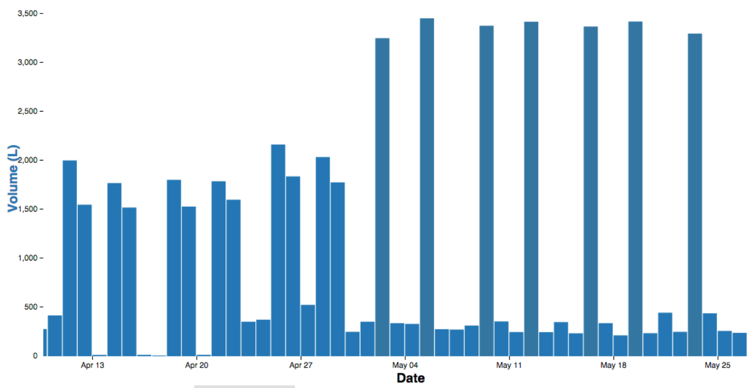


Figure 3. Raw Smart Meter Time Series

The input for the analytics process is raw data: smart meter time series of hourly water consumption from individual households. The data is first cleaned and stored as both hourly and daily consumption. Cleaning the data checks dates, removes incomplete days of data and any erroneous values such as negative meter readings. Figure 3 shows an example of a 48-day time window (x-axis in days) of daily water consumption (y-axis, in Litres) for a particular household. Both regular and irregular demand patterns can be observed.

Step 2: Data selection

Given a particular analytics goal, a typical smart meter time series contains many hours of observations that are irrelevant for that goal. The first step in smart metering is to select only observations from the raw data that are relevant. The output of this step is a smart meter **sub-series**: selected (time, volume) pairs from the full sequence.

Three types of selected sub-series are used for analysis: continuous flows, anomalous peaks, and recurrent habits.

Continuous flows occur where there is a 24 hour period with no 0-flow hours. For each day, the minimum hourly flow is used to estimate of the daily continuous flow. For example, if the minimum hourly flow over a 24 hour period is 10 L/h, then the daily continuous flow is estimated to be 240 L/d. But if the minimum hourly flow is 0 L/h for at least one hour of the day, then the daily minimum flow is 0 L/d. Days with positive continuous flow are called "continuous flows" of the customer. Continuous flows are likely, but not necessarily, caused by a leaking pipe or appliance. Because of uncertainties in estimating the daily minimum flow, we do not attempt to subtract continuous flows from other behaviours identified. Continuous flows are, therefore, double counted in other measures.

Anomalous peaks are exceptionally high days of demand that do not match a user's normal behaviour. The timing of these days is ad hoc, and infrequent, rather than regular. Exceptional peaks can signal the onset of a major event such as a burst pipe, or an unusual activity such as filling a swimming pool. It is important to identify anomalous peaks because they can account for a high proportion of use.

Recurrent habits are another type of high-magnitude use. Habit behaviours comprise *frequent* hours that have similar high magnitudes that occur at *regular* and *predictable* times. For example, 1500L/h used every Monday and Thursday at 5am. Habits differ from peaks because peaks are irregular and rare, but habits occur relatively frequent and with a regular pattern. Occasionally, there can be an overlap between detected peaks and habits.

Figure 4 illustrates the three types of behaviour for a 2 month segment of a time series of daily demand. Continuous flows, shown in green, occur on 41 of the 60 days. The maximum flow is 408 L/day. Two anomalous peak days during this period are shown in orange, with consumption of 4316 L/day and 6996 L/day. A regular, high-magnitude habit of 1200 L/h at 5am on alternate days can be seen on the right hand side of Figure 4. This habit is one part of the daily total of 2100 L/day. A longer history for this household is shown in Figure 6.

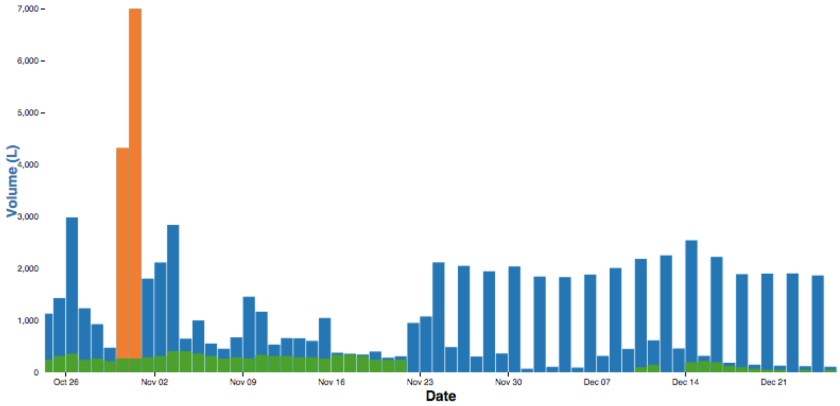


Figure 4. Behaviours discovered in a smart meter time series

Step 3: Feature extraction

Step 2 produces behaviours, which are collections of time-stamped observations. Step 3 performs **feature extraction**: the process of identifying summary characteristics of a behavior that are useful for analysis. For example, adding all volumes from the pairs is the **significance** of the selection and the number of pairs is the **frequency** of the selection. Other features such as the length of the longest contiguous subsequence (the longest period of consecutive days of continuous flow) or the temporal recurrence pattern of a habit are also useful. The feature extraction step produces a table of summary information about the behaviours (see Table 1).

Type of Behaviour	Feature	Value
Continuous flow	Aggregate volume	9 kL
	Frequency	41 days
	Longest continuous run	27 days
Anomalous Peak	Maximum peak	6996 L/day
	Number of days	2
Recurrent Habit	Aggregate volume	25 kL
	Recurrence pattern	Even days at 5am
	Extent	1 month

Table 1. Examples of extracted features for the behaviours in Figure 4

Step 4: Summaries for populations and individuals

Step 4 converts the results of steps 1 to 3 into actionable knowledge. The knowledge can be presented as summary information for a whole population of customers, or as feedback to individual customers. For example, Table 2 shows selected features of the continuous flow patterns of 440 households from Kalgoorlie-Boulder during 363 days in 2014.

Continuous flow features	Value
Significance	6.2% of all demand
Prevalence	90% of customers
Frequency	65 days/hh/year
Water use	25 kL/hh/year

Table 2. Summary of Continuous Flows for Kalgoorlie-Boulder data (N=440, days=363)

Population data can also be used for customer segmentation. For example, the intensity (L/h) and frequency (recurrences per week) of a recurrent habit can be represented as a point in 2D space as shown in Figure 5. This scatter plot summarises population results for the habits of 446 households in the Pilbara region of WA. The recurrent habits are partitioned by their intensity: more or less than 3.5 times per week. This threshold represents every alternate day as mandated by watering restrictions. Habits are also partitioned by the intensity of hourly use: either less or more than 1000 L/h. This creates four partitions. The top right hand segment of Figure 5 shows 92 high-frequency, high-intensity habits. These suggest inefficient garden watering at a higher rate than recommended for lawns in the Pilbara region. The bottom right and top left regions constitute the main volume of demand. These habits are undesirable for either high frequency or intensity. The lower left region represents habits of efficient garden watering, at least for the hours reported.

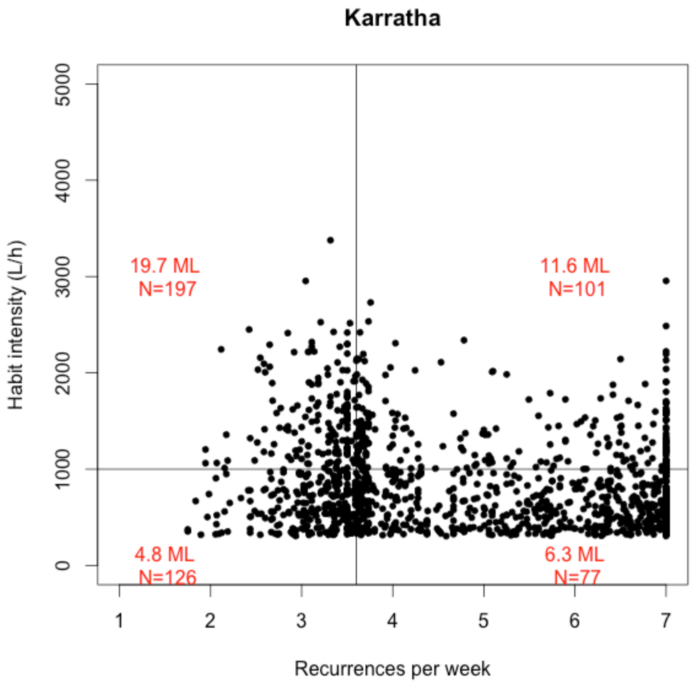


Figure 5. Population summary of habit frequency and intensity

Population analytics can be used to identify the average demand or spread of demand for different behaviours across a population. But in reality there is no "average household": each household has its own unique water use signatures. These individual characteristics that need to be tackled to change behaviours in order to reduce water use.

The data analytics generate unique water use signatures for each household. This information can be presented to individual customers to enable them to identify activities and prioritise actions. A sample of the water use history for this household is shown in Figure 6. Orange indicates anomalous peak days, red is recurrent habits, and green is continuous flows. Blue shows remaining use. For this household:

Annual demand	= 786 kL (top 10% of households)
Continuous flows	= 17 kL
Anomalous Peaks	= 241 kL (28 days > 5.4 kL/day, maximum 14 kL/day)
Recurrent Habits	= 392 kL ((usually 4 times per week during 11pm to 3am and 7am to 10am)

Cross checking one year of smart-meter data with customer survey data, this household is known to be a family of 2 adults and 2 school aged children, 60 months in their current residence. The garden comprises lawn and native or low water garden beds irrigated for "lush green lawn". Exceptional peak days occur in an ad hoc way throughout the year.

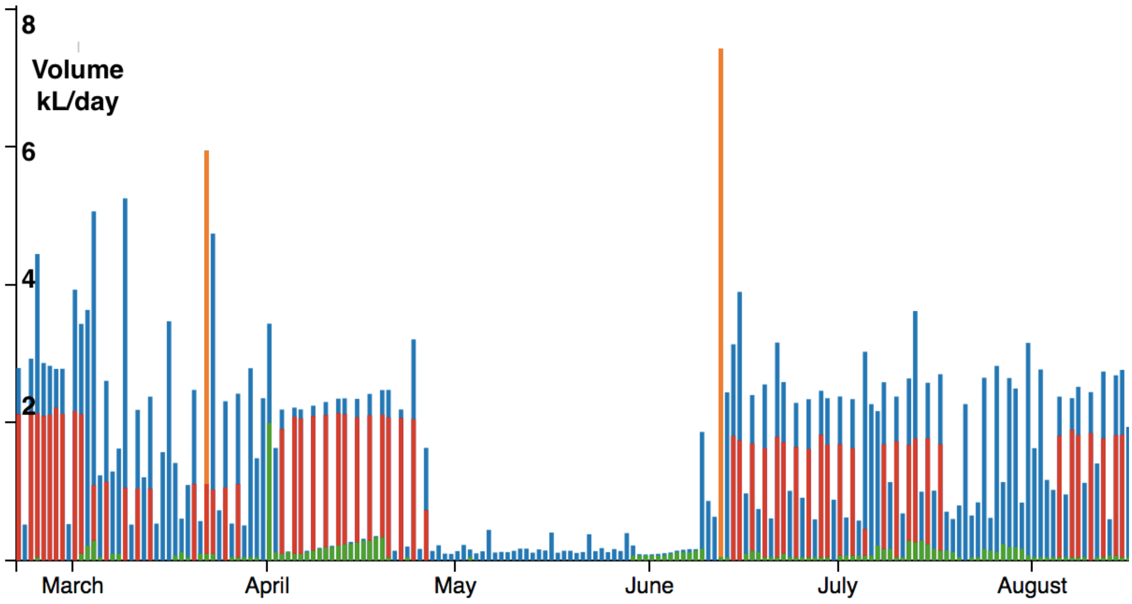
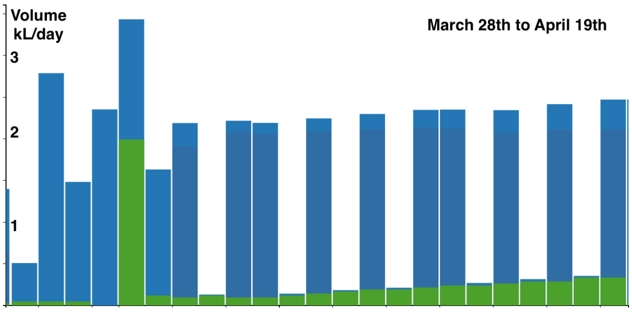


Figure 6. Daily water use history showing component behaviours for a single household in the Pilbara

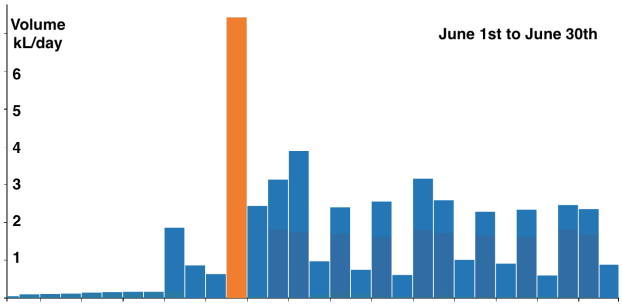
Annual use is informative and provides information for high level messaging about the scope for change, but it assumes stable occupancy and relies on a good memory. For engaging with residential customers for personalised water coaching about water use it is important to focus on recent water use as something they are more likely to connect with: what am I doing now? what can I do now to change? how successful were my actions to reduce consumption? A Water Corporation Pilbara coaching pilot in 2014 has demonstrated the potential of such immediate feedback. Figure 7, using 30 day segments for the household of Figure 6, shows how results from data mining could be used to enhance customer feedback.



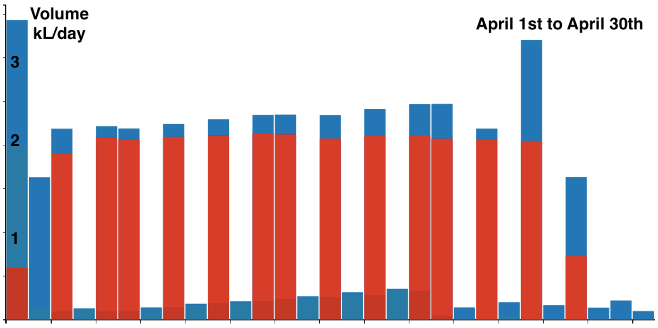
Feedback on 19 April 2014
Your smart meter has detected a constant flow of water over the past 19 days. In the last month your home may have lost:

5,640 LITRES

If you suspect you have a leak, there are some simple things you can do to find it or contact your licensed plumber to find and fix it.



Feedback on 30 June 2014
During the past month, your household's highest water use was **7,424 LITRES on Thursday 12th June**
Can you think what was causing this high use?



Feedback on 30 April 2014
During the past month, your smart meter recorded recurrent high water use on **Tuesday, Thursday, Saturday and Sunday at 8pm and 9pm** which is likely to be irrigation.
Many people we have spoken to have reduced their irrigation station run times. Is this something you might try?

Figure 7. Examples of water use signatures for 30 day smart meter samples with possible feedback messages for customers.

Frequently Asked Questions

How do the proposed analytics differ from a simple threshold-based spread-sheet analysis?

Simple spread-sheet analyses can be performed using threshold values to select and group data. For example, selecting all hourly values greater 300 L/h (Cole and Stewart, 2012) could be used to approximate garden watering activities. Some disadvantages of this approach that are addressed by data analytics are:

1. The data analytics algorithms are fully automated for running on large population data sets. Spread-sheet analysis is cumbersome and time consuming for more than 500 customers or more than one year of data.

2. Data analytics can be used to *learn* parameters that are appropriate for each data set. Spread-sheet analysis typically requires a domain expert to select suitable thresholds (e.g. 300 L/h for outdoor use).

3. Data analytics provides special purpose algorithms for clustering data into meaningful groups and for identifying patterns.

4. Data analytics provides algorithms for automatic pattern discovery that allow results to be presented to users in a meaningful way (e.g. every Monday and Friday at 5am). In the spread-sheet analysis such patterns must be detected by human inspection of the data tables or a visualisation.

5. As well as algorithms for data mining, we have developed software for web-based visualisation of the results.

How would these data analytics be used in a business setting?

The analytics described in this reports provide intelligence for utility planning purposes. Potential water savings can be identified and monitored by breaking down residential water use into segments that can be targeted for specific water efficiency opportunities. The algorithms can be run on demand over large data sets. The results provide evidence-based support for decision-making on demand management strategies designed to address capacity.

For customer feedback, a web application, such as MyWater, is an ideal medium for customer feedback. Special-purpose web or standalone applications can be developed by consultants and WaterCorp staff for data exploration and querying. We have developed a prototype web-based graphics interface for visualising water use signatures: http://people.csse.uwa.edu.au/rachel/waterusesignatures-july2013/

What is the confidence for the accuracy of the detected behaviours?

It may seem counter intuitive, but accuracy in terms of precisely identifying what one household is doing, is *not* the most important criteria here. The *goals* are to be able to identify which households and which trends are significant for at a population level, and to provide meaningful feedback to customers.

Recurrent, high-use habits can be responsible for a significant proportion of the water used. They are therefore good targets for demand management activities and customer feedback. However, whether the identified recurrent activity turns out to be garden watering or some other automated activity (such as topping-up a swimming pool) is not critical, so long as the user can identify what activity lies behind the recurrence.

At a population level, we are confident that most of the recurrent habits identified are garden watering since:

- they always occur at the same hour of the day;
- the recurrence of the behaviour is on regular days, which is consistent with an automatic controller;
- we detect many habits in the early hours of the morning, when other household activity is unlikely;
- we have tested the sensitivity of the parameters to minor changes, so the results are known to be robust.

We are also aware that some high-magnitude garden-watering activities are *not detected* using standard settings of the algorithm's parameters. We study this through detailed analysis of individual cases using data inspection and customer interviews. Improving the accuracy of detection for individuals is ongoing work. For customer feedback, the algorithm parameters can be set to more generous bounds than for population analysis. For customer feedback, a high-magnitude, recurrent behaviour is significant even if it is not related to garden watering. Giving details about a behaviour, including its volume and the times it occurs, enables the customer to decide what the water usage actually is, and how or whether they will change that use.

A continuous flow may correspond to any one of: a leaking appliance (tap, toilet or shower); a continuously running evaporative air-conditioner; or a household where people are active during both day and night by preference, sickness or shift work. The behaviour detected does not claim to be a specific human activity, but rather a family of activities with the same water use signature. Again, the customer is empowered to identify what human activity lies behind the behaviour and what action to take.

How much data is needed?

A certain number of supporting observations are needed for confidence that there is a consistent behaviour, rather than a one-off event. For example, in order to be confident that a habit is recurrent, at least one month of data is desirable, preferably more. Feedback to customers can be based on the previous month's activity. Similarly, for confidence that results are representative of a whole population, a sample of at least 100 customers is desirable.

What software and expertise is needed to use these methods?

The algorithms presented in this report could be implemented in any general-purpose programming environment. That could be a spread-sheet (if new macros are written); or a database management system; or in a general purpose programming language such as Python, Java or R. For the results presented in this report, the algorithms have been implemented in the open-source R system (www.r-project.org) using the R-studio development environment (www.rstudio.com).

The data exploration phase is a type of feasibility study. This phase requires two types of experts: data mining and water efficiency management. In this phase, standard and novel types of usage behaviour are explored. For example, we developed a novel algorithm for detecting recurrent habits. Feature engineering is the process of defining features that give good insight into the data. This exploratory process requires data mining expertise. Domain experts are essential during the exploratory phase, to ensure that the behaviours and features make sense and are relevant for business use cases. For example, is it relevant to detect continuous flows as low as 48 L/day (2 L/hour), or should a higher bound be used?

Once the behaviour selection and feature extraction methods have been decided, then data querying can be performed using a standard database, and built-in to business systems such as customer web portals. Querying by business and customers can be performed as required using these systems.

Algorithms for identifying behaviours

This section summarises three algorithms for identifying different types of behaviours in smart meter time series data. Details can be found in the reference papers.

Continuous flows

A normal day of household water use includes some hours of zero water use, typically during the night-time when residents are asleep. A continuous flow day is a day in which there are *no* zero-flow hours. The minimum hourly flow observed during the day is taken as the baseline hourly continuous flow. This value is used to calculate a 24 hour continuous flow amount for the day. Figure 8 outlines an algorithm for identifying all continuous flow dayss. It's parameter MINFLOW is the minimum hourly flow volume to quality as a continuous flow day.

```
MINFLOW = 2 L/h
conflowlist={}
FOR (every day d) {
        minhour = minimum flow volumes from hours 0:23 of day d
        IF minhour > MINFLOW THEN conflowlist[d] = minhour*24 ELSE conflowlist[d] = 0
}
RETURN conflowlist
```

Figure 8. Algorithm for detecting continuous flows

Anomalous peaks

There are several possible methods for identifying exceptional values in time series. The Tukey box-and-whiskers boxplot (see appendix) is chosen because it can identify outliers without making unjustified assumptions on the underlying statistical distribution of the data. The upper and lower quartiles of a data set partition the ranked values: q3 is the lowest of the top 25% of values and q1 is the highest of the bottom 25% of ranked values. The *upper fence* is a limit of extreme values q2 + k(q3-q1) where k=1.5 is usual choice. An *upper outlier,* value is defined to be any value above the upper fence limit. A data set has no outliers if its maximum value is no more than the upper fence. Figure 9 outlines an algorithm for calculating a list of peak days, using a parameter PEAKLIMIT to define the daily upper fence for each individual's daily water use.

```
peaklist={}
FOR (every day d) {
        dayvolumes[d] = sum of volumes for hours 0:23 for each day d
}
PEAKLIMIT = median(dayvolumes) + 1.5*(q3(dayvolumes) – q1(dayvolumes))
FOR (every day d) {
        IF dayvolumes[d] > PEAKLIMIT THEN peaklist[d] = dayvol ELSE peaklist[d] = 0
}
RETURN peaklist
```

Figure 9. Algorithm for detecting anomalous peaks

Recurrent habits

Recurrent habits are an important part of human behaviour. High-use habits, such as garden watering using an automatic sprinkler system, can account for a high proportion of household demand. A habit is defined as a set of observed water-use *hours* that 1) occur at the same hour of day 2) have similar, high volume, 3) are connected in time and 4) that recur on a particular pattern of days.

Two examples of habits are:

~ 1500 L/h every day at 5am from Nov to Apr

~ 500 L/h every We,Fr at 10pm from Jan to Dec

Figure 10 outlines an algorithm to generate habits from a household time series. More details about this algorithm can be found in the reference papers.

```
LOWERLIMIT = 300 L
VOLRADIUS = 150 L
MAXTIMEGAP = 10 days
MINSUP = 7 observations
MINMATCH = 0.75 minimum F-measure for pattern match

habitlist = {}
FOR (hour h in 0:23) {
       Partition volumes above LOWERLIMIT for hour h into clusters Ch1 … Chk
  with time distances no more than MAXTIMEGAP and
  volumes within VOLRADIUS of cluster centre

  FOR (each Chi ) {
      IF more than MINSUP observations in Chi THEN
             Generate candidate temporal patterns Pj for the days of Chi
             matchj = maximum match ratio for Chi and Pj
             IF (matchj > MINMATCH) THEN {
                   Add Chi , Pj to habitlist
             }
      }
  }
}
RETURN habitlist
```
Figure 10. Algorithm for detecting recurrent habits

Part 2: Case studies

This section demonstrates the data analytics of Part 1 in action. The results are taken from two case studies of samples of 500 customers from detached homes in the Pilbara and the Goldfields of Western Australia analysed over a one year period.

Data

Data specification

A smart meter time series comprises hourly observations each with a date, time of day, and volume of water consumed. The resolution of the meters is one Litre per hour, so the recorded value is the actual water use for one hour rounded down to the nearest Litre.

The examples in this document are taken from two contrasting populations in Western Australia, Karratha and Kalgoorlie. The 500 Karratha properties were randomly selected from the one new suburb (Baynton) and two established suburbs (Bulgarra and Millars Well), the only three Karratha suburbs with smart meter available from February 2014. The 500 Kalgoorlie properties were randomly selected from a greater range of suburbs as smart meter installations were completed in the area in 2012. These suburbs included Broadwood, Sommerville, Lamington, West Lamington, Kalgoorlie, South Kalgoorlie, Hannans and Boulder.

The sample of households provides a reasonable representation of the whole population of 'house' properties based on water use. For example, for the 8000 (apx) metered properties recorded as houses in Kalgoorlie, average water supplied per year from 2010 to 2014 has been 412.5, 362.6, 371.1, 369.4 and 342.8. kL/house/year (respectively). For the 440 houses in our sample, the average water use is 366.5 kL/house/year for the 2014 calendar year. Differences may be due to a number of factors including the timing of the water use reading year which ends in July August. Developing techniques for selecting samples of meters that closely mirror properties the full population is a current research task in CRC Water Sensitive Cities project C5.1.

Water consumption was measured over a period of one year from January 2014 to 2015. The smart water meter data set comprises hourly smart meter data readings from the 500 properties for each population.

As well as meter readings, contextual information for each property is available, including its land use, lot size, garden size, watering roster, whether there is a swimming pool or evaporative air-conditioner, and whether the property is owner occupied or leased.

During March 2015, the Water Corporation commissioned an study to interview a customers in Karratha and Kalgoorlie about their household characteristics and outdoor use of irrigation controllers, swimming pool, car washing. The samples of 167 and 157 were selected to be representative of the whole population, and of the smart metering datasets of 500 customers. Table 3 summarises characteristics of the interviewed sample.

Household Characteristic	Karratha (N=167)	Kalgoorlie (N=161)
People per household	3.4	3.4
Adults per household	2.3	2.2
Shift workers per household	0.3	0.4
Automatic irrigation controllers	75%	46%
Swimming pool	25%	32%
Evaporative air-conditioner	1%	90%
Interview sample size	167	161

Table 3. Household characteristics for Karratha and Kalgoorlie from interview data

Pre-processing

Real world data is noisy and so pre-processing is required. Four steps are required for the smart water meter data: assessment of missing readings, identification and removal of erroneous readings, removal of data from unoccupied properties, and filtering by land use type. Erroneous values, such as negative consumption, are occasionally reported but are rare.

Of 8.76 million readings in the dataset only 108 were out of range and so were removed during pre-processing. Point errors of missing hours within a day's readings occurred with a frequency of 2.6% to 4.8% of all readings. Data is normalised by the number of meters active in any particular hour.

A few properties were empty or had a high rate of missing values, giving them an artificially low annual consumption. Properties with annual consumption below 25 L per day were removed from the sample. Contextual information on the land use of each property was used to identify houses, duplex, triplex and quadruplex units, home units and flats.

Houses comprise 92% of the sample. For consistency, the time series for the other property types were removed from the data set. After this pre-processing, the data sets had 466 water meter time series for Karratha and 440 for Kalgoorlie. For comparison with national trends, throughout this report demand is normalised to kilolitres per household per year: kL/hh/year. Table 4 summarises the results of pre-processing.

Data	Karratha	Kalgoorlie
Total water use (full days)	208 ML	161 ML
Sample population (houses only)	466	440
Days in sample	365	362
Error rate (missing readings)	2.6%	4.8%

Table 4. Summary of pre-processing for the Karratha and Kalgoorlie data sets

Choosing parameters

Each of the signature patterns covered in this report depend on parameters that can be set by a domain expert, or discovered for a particular data set. There is usually a trade-off to be made. On one hand, if the parameter settings are too general then more activities are identified but the significance and reliability is lower (for example, detect many small, short period continuous flows). On the other hand, if the parameter settings are too constrained then only highly significant activities are detected, whilst others will be missed. The choice of parameter is, thus, a business decision depending on how the results are to be used (e.g. one-to-one interaction with customers with large leaks, or for population-wide estimates of all continuous-flow activity). Figure 11 illustrates the effect of different parameter settings.

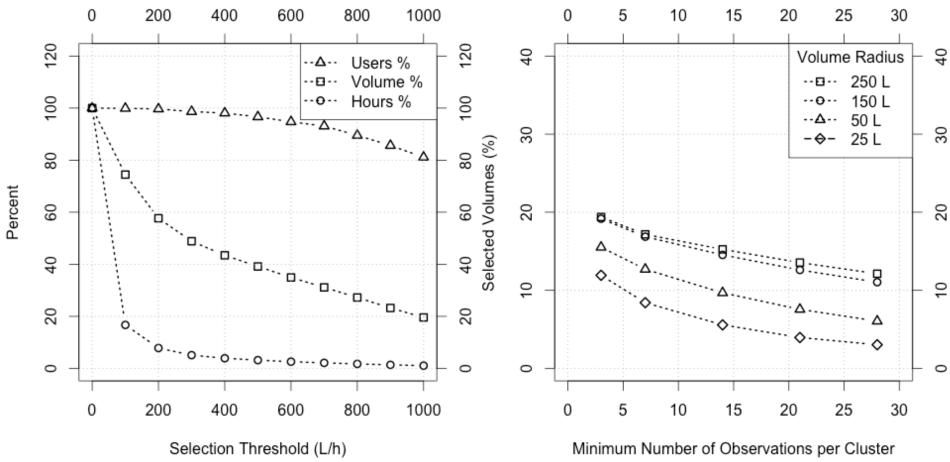


Figure 11. Sensitivity of habit detection algorithm to changes in its parameter settings

Context of water use: climate and situation

It is well known that there are significant differences between the water use behaviours of urban populations around Australia, and that water use can change in response to situations such as droughts. Figure 12 shows results from the National Water Commission's 2012-13 National Performance report for Urban Water Utilities. There is less known about water use by non-urban populations and in more challenging Australian climate zones.

This section analyses the effect of climate and situation on water use for the case study populations in Western Australia's Pilbara and Goldfields regions. Karratha in the Pilbara and Kalgoorlie in the Goldfields both have challenging climates for water use, with high summer temperatures and very low rainfall. Figure 13 summarises the weather conditions and water demand during 2014 when the case study data was collected. Climate data throughout this section is sourced from the Bureau of Meteorology at www.bom.gov.au.

Major urban area	2008–09	2009–10	2010–11	2011–12	2012–13	% change from 2011–12
Sydney	198	205	197	193	198	2.6%
Melbourne	147	142	138	142	152	7.3%
South-east Queensland					156	
Perth	277	276	264	250	249	–0.6%
Adelaide	190	191	180	179	193	7.6%
Canberra	201	199	177	180	199	10.2%
Darwin	491	458	405	471	454	–3.6%

Melbourne figures are the weighted average of three retailers (that is, total residential water supplied ÷ total residential customers).

South-east Queensland figures are the weighted average of Queensland Urban Utilities, Unitywater, Gold Coast Water and Logan Water.

Figure 12. Urban water use statistics of kL supplied/household/year [National Water Commission Report 2015]

Is response to daily temperature an important factor for water efficiency?

Yes. Figure 13 and Figure 14 demonstrate that, as expected, the trend of daily water demand follows seasonal temperature trends. For almost the same number of households, Kalgoorlie shows more variance than Karratha both between and within seasons. Kalgoorlie also shows a greater variation in day to day temperature maxima and minima, again both between and within seasons.

Pearson's product moment correlation co-efficient for daily demand vs mean temperature is 0.62 for Karratha and 0.68 for Kalgoorlie with p-value < 2.2e-16. The correlation coefficients for daily demand vs minimum or maximum temperature alone are slightly lower.

Thus, forecast mean temperature could be used to predict daily demand for a population, but the prediction uncertainty would be high. For example, for the hottest days in Kalgoorlie (i.e. mean temperature greater than 30 degrees) the average demand is 1592 Litres per household, but the minimum demand for these days is 958 L/hh and the maximum 2081 L/hh. Identifying relevant features for predicting demand, together with sampling strategies to drive it, is a topic of our ongoing research.

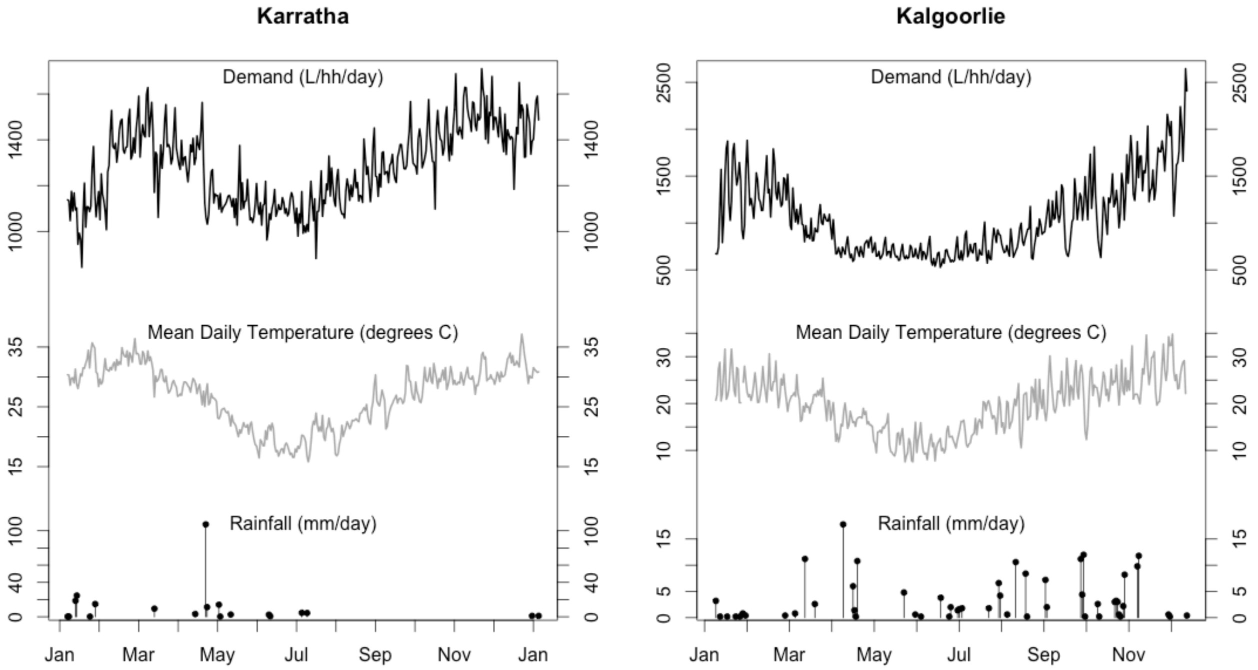


Figure 13. Summary of daily water use/household, mean daily temperature and rainfall during 2014 case study period.

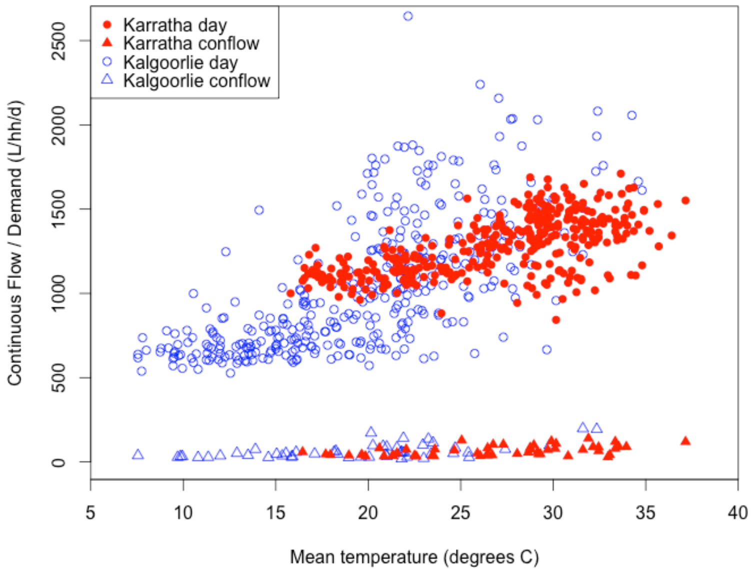


Figure 14. Daily water use (circles) and continuous flow (triangles) *vs* mean temperature for the 440 Kalgoorlie properties (black) and 466 Karratha (red) properties.

Is response to rainfall an important factor for water efficiency?

No. Having efficiency measures during periods of rain is only effective where there are sufficient days of rain. For example, switching off a garden irrigation system should be done where there are days of rain over 12mm. But Kalgoorlie and Karratha have no more than 5 such days per year. Table 5 shows rainfall statistics for the trial period. Rainfall in the 2014 trial period was lower than the BOM 30 year annual averages (Figure 15). Kalgoorlie had one exceptional day of 94.0 mm of rain in Jan 2015, but this was outside the trial period.

	Karratha	Kalgoorlie
Days of rain / days of trial	19 / 365	52 / 362
Days > 12mm	5	1
Maximum day in trial	107.4 mm	17.8 mm
Exceptional days	107.4 mm in May	94.0 mm in Jan after trial
Annual average rainfall (30 year average)	285 mm	264 mm
Average daily evapotranspiration	9.3 mm/day	7.2 mm/day
Average annual days of rain	28	68

Table 5. Bureau of Meteorology rainfall statistics for Karratha and Kalgoorlie

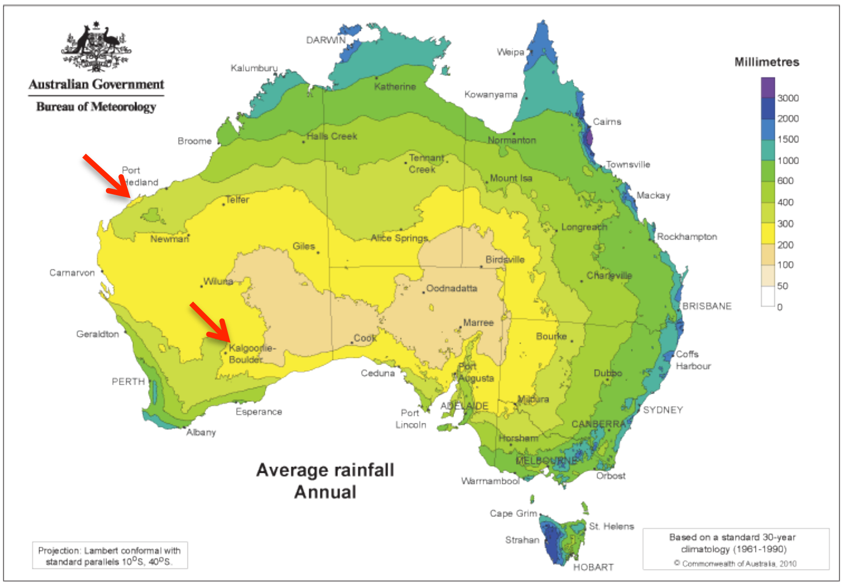


Figure 15. Annual average rainfall for Australia with Karratha and Kalgoorlie indicated

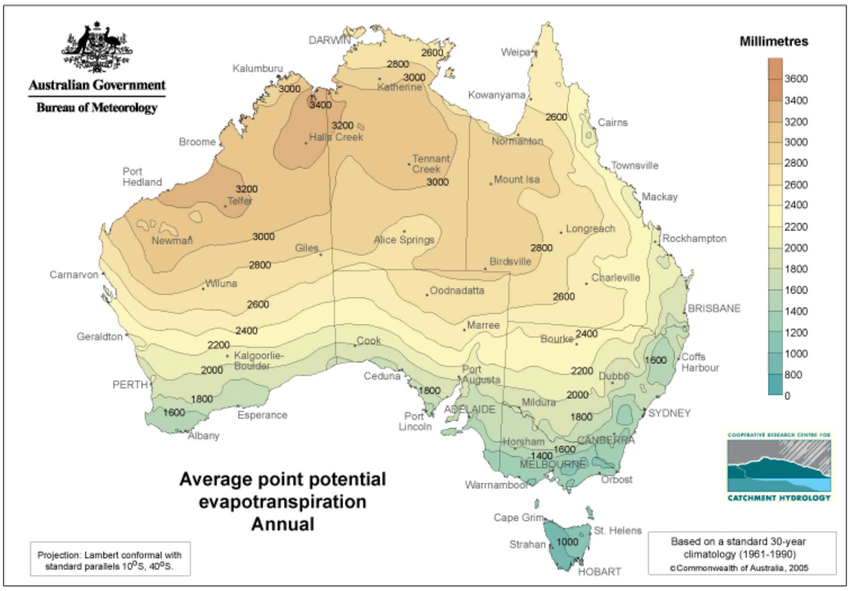


Figure 16. Average point potential evapotranspiration in Australia

Are there differences between water use on weekdays and weekends?

Not much. In urban populations, household water demand on weekends is typically higher than on weekdays. However, for the Pilbara and Goldfields case studies we found little difference between demand on different days of the week. Figure shows the range of daily demand for each population sample separated by the day of the week. Karratha has marginally higher weekend than weekday use. The range of daily use across the population is higher in Kalgoorlie than Karratha.

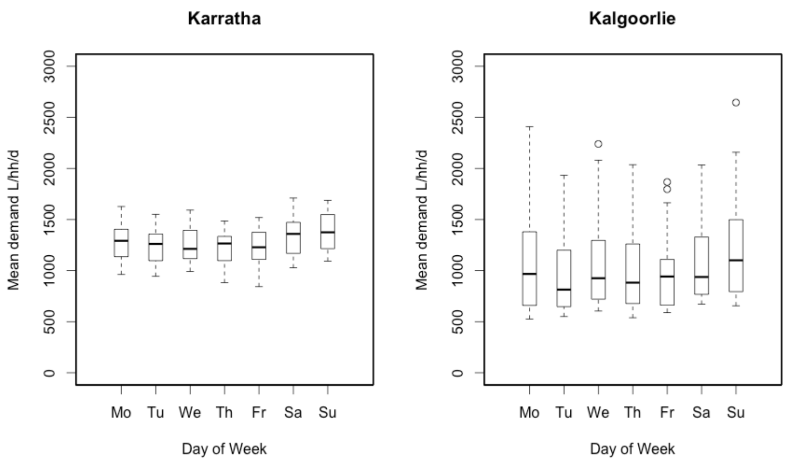


Figure 17. Daily water use by day of the week for Karratha and Kalgoorlie

Overview of three behaviours

Table 6 summarises key metrics for each of the three behaviours we analysed on the two case study populations. Cells shaded in red indicate the most promising targets for water efficiency campaigns. The following sections of the report unpackage these headline measurements, to show findings on specific business questions.

Behaviour	Feature	Units	Karratha	Kalgoorlie
Summary	Total demand	ML	208	161
	Population Size	households (hh)	466	440
	Demand per hh	kL/hh/year	445.9	366.5
Continuous flows	Significance	kL/hh/year	23.4	22.8
	Prevalence	hh (%)	363 (78%)	**395 (90%)**
	More than 30 days of CF	hh (%)	**193 (41%)**	**198 (45%)**
	Average days of CF[1]	days/hh/year	54	58
Anomalous peaks	Significance	kL/hh/year	57.9	**78.5**
	Prevalence	hh (%)	405 (86%)	**433 (98%)**
	Mean number of peak days	days/hh/year	13	**22**
Recurrent habits	Significance	kL/hh/year	**90.8**	46.3
	Prevalence[2]	hh (%)	**294 (63%)**	196 (44%)
	Frequency > Roster[3]	kL/hh/year	**64.8**	9.6
	High Intensity[4]	kL/hh/year	**38.3**	24.6

Table 6. Data Analytics highlights for the case study households (hh) of Karratha and Kalgoorlie

[1] Average for all households, including those with no continuous flows.
[2] Close to the survey reported penetration rate of automatic irrigation controllers: 75% & 46%.
[3] Watering roster for Karratha alternate days (3.5 days/week) and Kalgoorlie 2 days/week
[4] High intensity habits defined as regular activity with > 1000 L/hour

Water savings potential of continuous flows

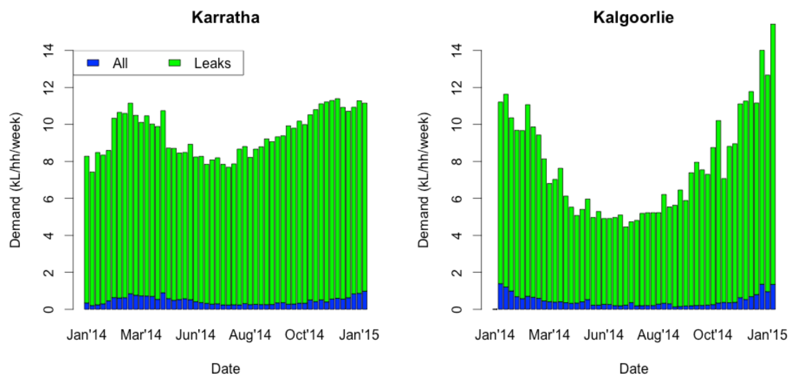


Figure 18. Continuous flow per week (blue) vs all use (green)

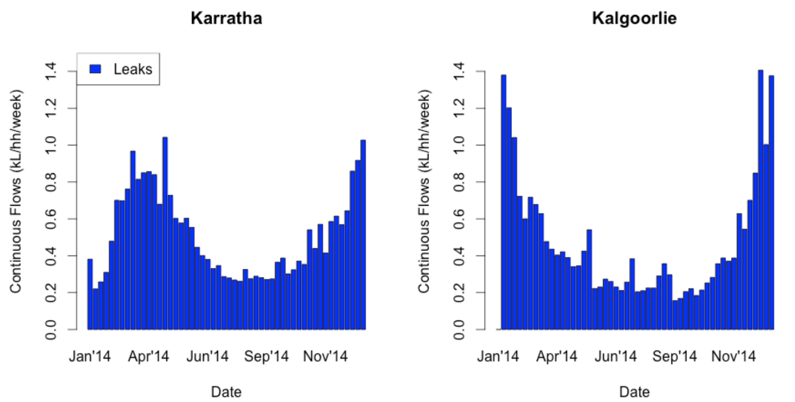


Figure 19. Continuous flow per week

Figure 18 and Figure 19 show aggregated weekly totals of continuous flows per household. Around 90% of houses in the Kalgoorlie sample have evaporative air-conditioners, which can use up to 10 L/min through evaporation, bleed and dumping. We would therefore expect to see this reflected in high incidence of continuous flows in Kalgoorlie during the warmer months, which can be seen in the figures from November to February. However, the summer bias is also in Karratha where evaporative air-conditioners are not used[5].

Karratha's highest continuous flows are in December, March and April and also appear to be associated with high temperature months, although this is because of evaporative air conditioners. For both Karratha and Kalgoorlie continuous flows and demand are correlated with temperature as was seen in Figure 14 in the previous section.

Pearson's product moment correlation co-efficient

[5] The student t-test shows that there is no statistically significant difference between the means of properties in Kalgoorlie without (N=16) air-conditioning, those with (N=388) and unknown (N=36) ($p=0.2811 > 0.05$ for volume and $p=0.2877 > 0.05$ for days).

Continuous flow vs temperature: cor = 0.45, p<0.001
Daily demand vs temperature: cor = 0.69, p<0.001

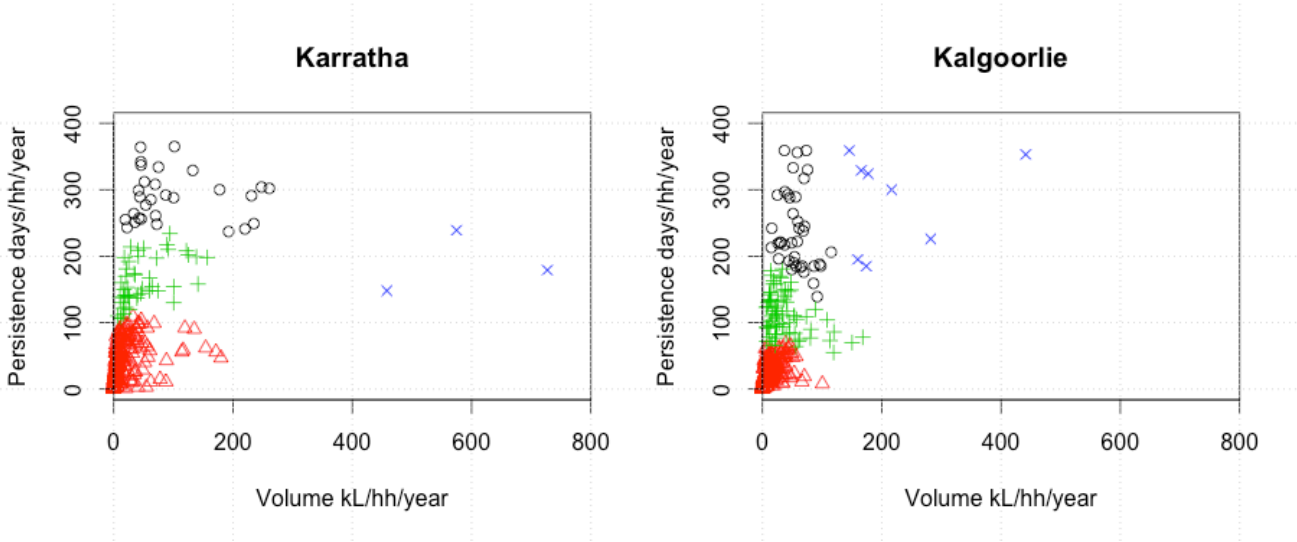


Figure 20.Continuous flow clustered by significance and persistence

Figure 20 segments households by the volume and number of days of their continuous flows. Each point in the scatter plots represents a single household. The colours and point markers show a clustering of these values into four groups. This example used a hierarchical clustering algorithm to group users. However, a similar segmentation can be achieved by using expert judgement to choose simple threshold values for partitioning. For example, 200 kL/hh/year for volumes and 100 and 200 days per year of continuous flow days. The partitions can be used to answer questions such as:

What is the most common type of continuous flows?

Low-volume continuous flows for only a few days are the most common. The red triangles in Figure 20 indicate the most common type of continuous flows. For Karratha, this group has 388 households with 10 kL/hh over 23 days/hh. For Kalgoorlie, this group comprises 317 households with 10 kL/hh over 18 days/hh.

What are the worst-case continuous flows?

A few households have very high, persistent continuous flows. The blue crosses in Figure 20 indicate these households. In Karratha there are 3 households with continuous flows of more than 300 kL/hh/year and over 100 days/hh/year and 8 in Kalgoorlie of more than150 kL/hh/year and more than150 days/hh/year.

Water savings potential of anomalous peaks

What is the overall significance of anomalous peaks?

Exceptional peak days are a significant factor for household use, especially in Kalgoorlie. Peak use is not often considered for water efficiency programmes. Our results suggest that implementing peak warnings, similar to leak warnings, could be very effective for saving water. As shown in Figure 21, peaks account for 57.9 kL/hh/year (13% of all demand) in Karratha and more at 78.5 kL/hh/year (23%) in Kalgoorlie. Over 85% of properties have some exceptional peak days of use.

What are typical peak ratios between anomalous peaks and average daily water use?

Peak days are often 10x higher than the average daily use of a household. Figure 21 plots maximum peak day against average daily demand for each user. Karratha has 94 users with >10:1 peak day ratio and Kalgoorlie has 80. These single peak days for each of the users account for 1455 kL overall in Karratha and 1103 kL in Kalgoorlie.

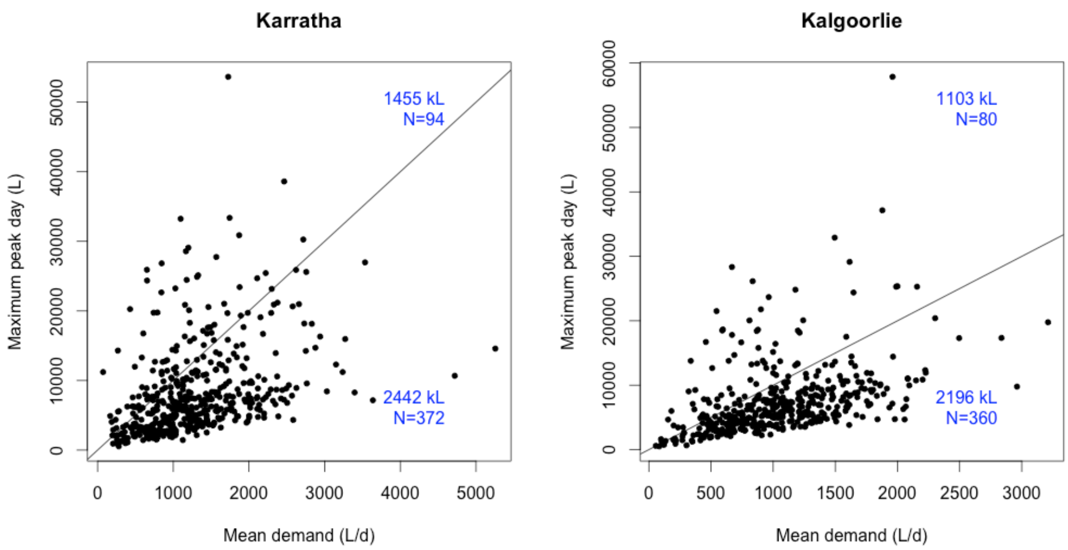


Figure 21. Household average daily demand *vs* maximum peak day, partitioned by the 10:1 peak day ratio line.

Are peaks concentrated at particular times of the year?

Peaks are more prevalent in summer, particularly in Kalgoorlie. Figure 22 shows some examples. Irregular peaks are likely to be associated with outdoor activities such as hand watering or washing down a car.

What are some examples of individual household peak patterns?

Every household has a different pattern of peaks. Figure 22 shows peak days (in orange) compared to other days (in blue) for two households. Peaks occur at irregular times, and some households have more exceptional peak days than others.

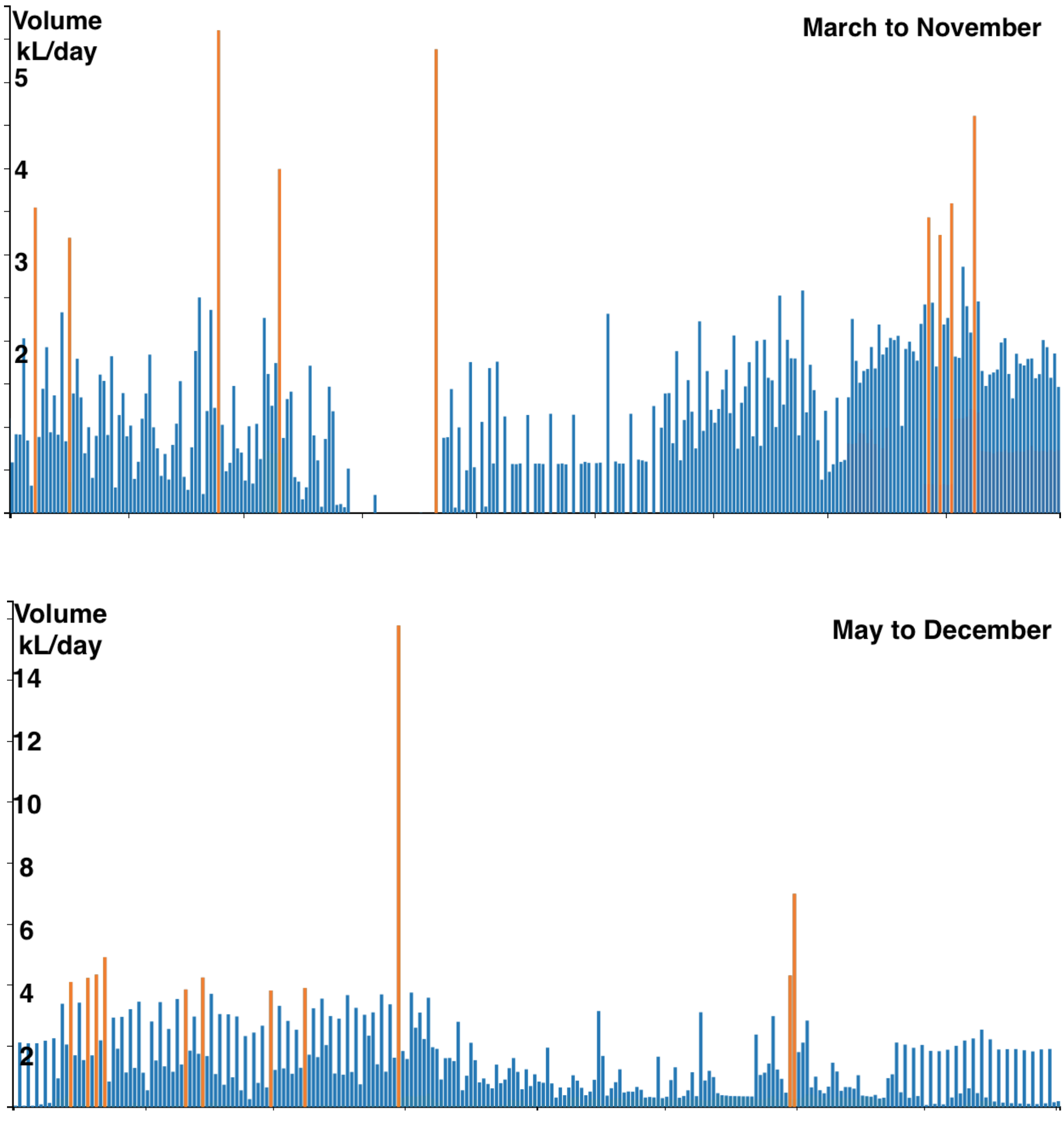


Figure 22. Examples of peak distributions (orange days) for two households

Water savings potential of recurrent habits

Habits are water consumption hours that have similar volume, occur at the same hour of day, and have a persistent, regular recurrence pattern (eg. every day or every Mon, Wed and Fri for at least 4 weeks). Habits represent human water-use behaviours such as automatic watering systems programmed to run at a fixed hour on certain days. A single household typically has several different habits over the course of a year. An example rule summarising a habit is:

HABIT 850 L/h OCCURS ON Mo, Th days AT 5am
FROM 2 Jan 2014 to 9 Apr 2014

What are the most common habit patterns?

Table 7 characterises the most common habit patterns for Karratha and Kalgoorlie. The median habit in Karratha is 4 times per week at 711 L/h over a period of 35 days. Habits do not typically persist for a long time, such as the whole summer season. Kalgoorlie's median habit is 2 times per week at 796 L/h over a period of 53 days.

Median Habits	Karratha	Kalgoorlie
Median volume (L/h)	711	796
Frequency (times per week)	4	2
Persistence (elapsed days)	35	53
Total volume per habit (kL)	17	15

Table 7. Median characteristics of all habits

Are households watering inefficiently for their garden size?

Irrigation can be a large proportion of residential water use, making irrigation efficiency a significant target for demand management activities such as watering rosters. For Perth, the average for properties without access to a bore was reported as 45% of total water use per annum in the 'Perth Residential Water Use Study 2008/09'. State-wide permanent water efficiency measures require customers in each area to adhere to watering rosters, where they can water once before 9am or after 6pm on their rostered days. Karratha's roster is every alternate day and Kalgoorlie's roster is twice a week, with watering days determined by the street number of the property. Roster compliance can be difficult to assess using traditional observations, but identifying irrigation events using the data can provide a good indication of the scope for compliance and community engagement activities to reduce water use. Karratha and Kalgoorlie both have high evapotranspiration and low rainfall, which leads to a high water requirement for irrigation of traditional gardens. Encouraging waterwise gardening practices appropriate to the Pilbara and Goldfields can reduce irrigation demand.

The Water Corporation used GIS information to estimate the garden size (within the lot size) for each property in the case study samples. Using this information along with recommended irrigation budgets[6] based on evapotranspiration rates (Figure) irrigation efficiency can be assessed. For a conservative estimate the budget recommended for domestic lawns in both areas was applied to the Kalgoorlie and Karratha properties to provide an indication of over-watering. More detailed spatial data on garden types is likely to identify significant areas with lower water requirements and would result in higher volumes being attributed to over-watering.

	Karratha	Kalgoorlie
Recommended irrigation rate for turf lawn (kL/m2/year)	1.358	1.051
Average garden size (m^2)	193	316
Average lot size (m^2)	637	883
Recommended irrigation budget for average garden with turf lawn (kL/hh/year)	263	333
Measured annual water use (kL/hh/year)	446	366
Average habit use (kL/hh/year)	91	46
Number of households with habits > recommended rate	72	13

Table 8. Irrigation efficiency summary statistics based on recommended evapotranspiration replacement (ET) irrigation budgets for domestic lawns and gardens

[6] Karratha ETturf = 0.4 x 9.3 mm/day = 3.72 x 365 = 1357.8 mm/yr (/100) = 13.58 ML/ha/yr = 1.358 kL/m^2/year and Kalgoorlie ETturf = 0.4 x 7.2 mm/day = 2.88 x 365 = 1051.2 mm/yr (/100) = 10.51 ML/ha/yr = 1.051 kL/m^2/year.

The scatter plots of Figure 23 show aggregate habit demand for each household in kL per year against the garden size of each property. The diagonal line on each figure is the recommended budget for watering lawn for each climate. For a given garden size, properties to the left of the line are efficient (lower annual habit use than the recommended rate), and those to the right are inefficient (higher annual habit use than recommended).

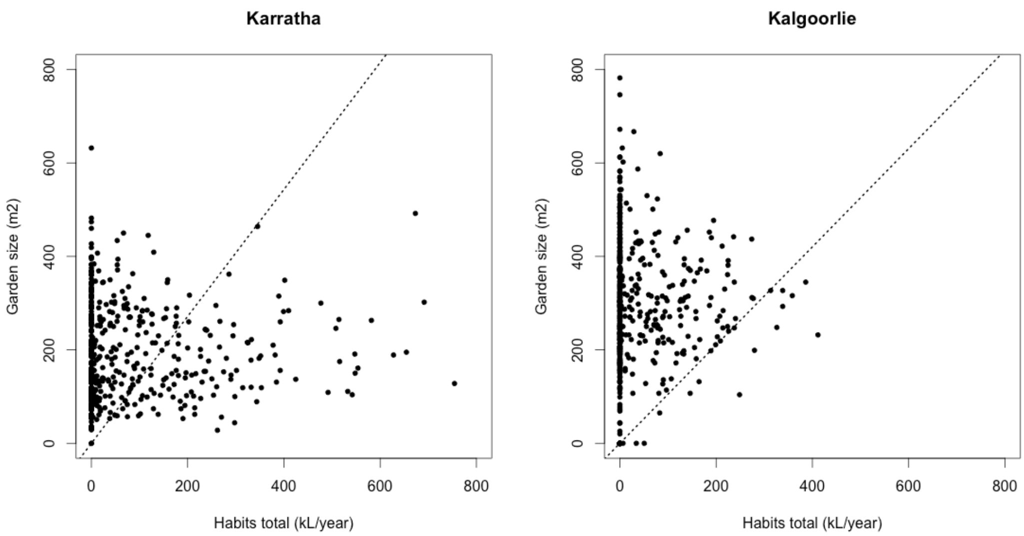


Figure 23. Habit demand *vs* garden size with recommended irrigation levels shown by the diagonal line. For a given garden size, consumption of properties to the left are efficient and those to the right are inefficient.

What are the worst cases for habits, and how prevalent are they?

Intense (>1000 L/h) and over-frequent (>3.5d/w or 2d/w) habits account for around 16% of all habit demand. A habit is described by a rule such as:

HABIT 1511 L/h OCCURS Mo, We, Fr, Su- day AT 6:00 FROM 16 Mar'14 to 6 Apr'14

The *intensity* of a habit refers to its hourly rate, in this case (approximately) 1511 L/h. The *frequency* of a habit refers to the number of times it occurs per week, in this case 4 times a week. The significance of a habit is the aggregate of all water used during the life of the habit. The example habit has a significance of 19.9 kL (13 x 1511 L).

Water users are obliged to comply with watering rosters and watering days are allocated based on street number of the property. The roster for Karratha provides for alternate day watering using odds and evens, which broadly equates to 3.5 days per week.

Thus Karratha intensive habits that occur 4 times a week or every day may be garden watering activity that is not compliant with the roster. For Kalgoorlie the watering roster is two times a week, so intensive habits 3 or 4 times a week or every day are the target. Customers in these segments are in the top right region of Figure . These customer segments could be targeted for water efficiency programs.

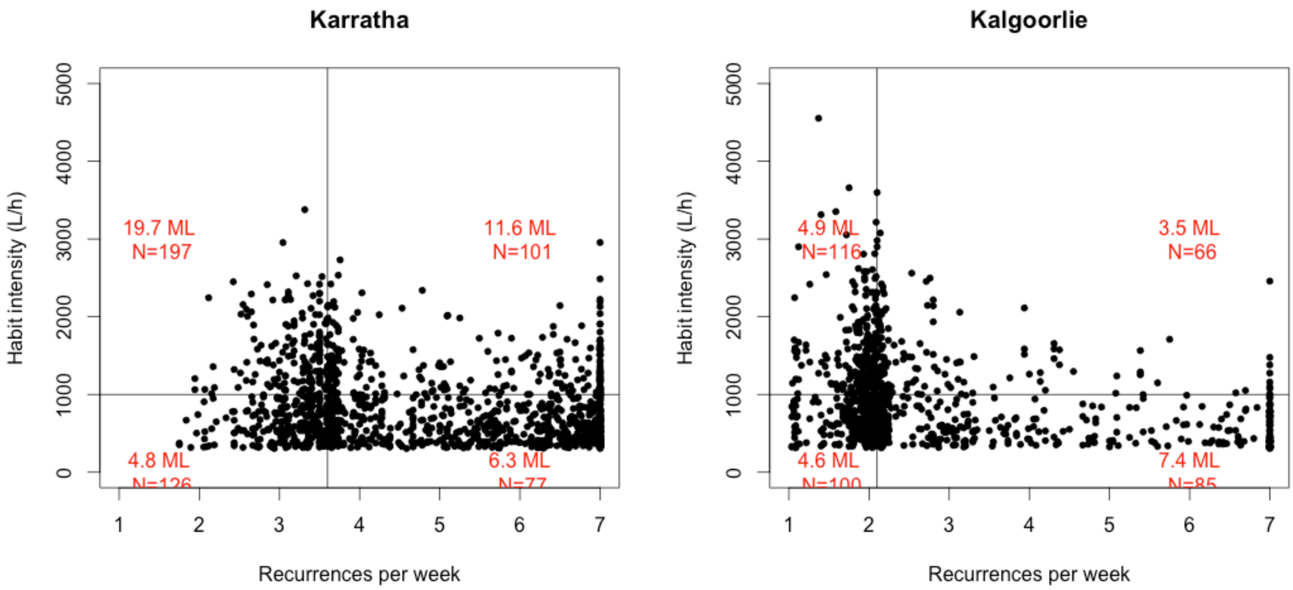


Figure 24. Distribution of habit frequency *vs* intensity

To what extent are identified habits in compliance with the watering rosters for each household?

Table 9 compares the days of the week of identified habits with the rostered watering days for each household. The level of compliance with the spirit of the watering rosters, that is habits with no more than the rostered number of days per week, is just under 70% of households for both populations. On the other hand, only 30% of households with habits do so in exact compliance with Karratha's odd or even day of the month roster. It should be remembered that identified habits may not correspond to garden irrigation, and so the results summarised in Table 9 are indicative, rather than a basis for policing compliance.

	Karratha		Kalgoorlie	
	Number of houses (N=466)	Total volume (ML)	Number of houses (N=440)	Total volume (ML)
Households with at least one regular habit	294	42.3 ML	196	20.3 ML
Households with at least one habit in the spirit of the watering rosters (i.e. does not exceed the rostered number of days per week)	200	17.5 ML	137	13.6 ML
Households with at least one habit in exact compliance with the watering rosters	89	3.8 ML	115	10.9 ML

Table 9. Compliance with watering rosters

Which hours of day have peak habit demand? How do these fit with the overall diurnal water use cycle?

6am is the peak habit time for both populations, which coincides with the diurnal peak in Karratha, but not in Kalgoorlie.

Time of day segmentation partitions users according to the timing of habits. Figure shows the diurnal cycle of demand, and the contribution of habits to that cycle. The grey section of each bar represents non-habit demand and the coloured segments are habit activity. Habit demand is partitioned into the contribution of the 25 customers with the largest contribution to the habit demand (red) and all remaining habit customers (yellow) for that hour. Customers in the red segments are candidates for intervention since this segment represents an ideal situation for customer engagement: a relatively small number of customers (25 is approximately 5% of the population) contribute a large proportion of demand, in what is likely to be a discretionary activity. December is shown because it is a peak month of demand in both populations.

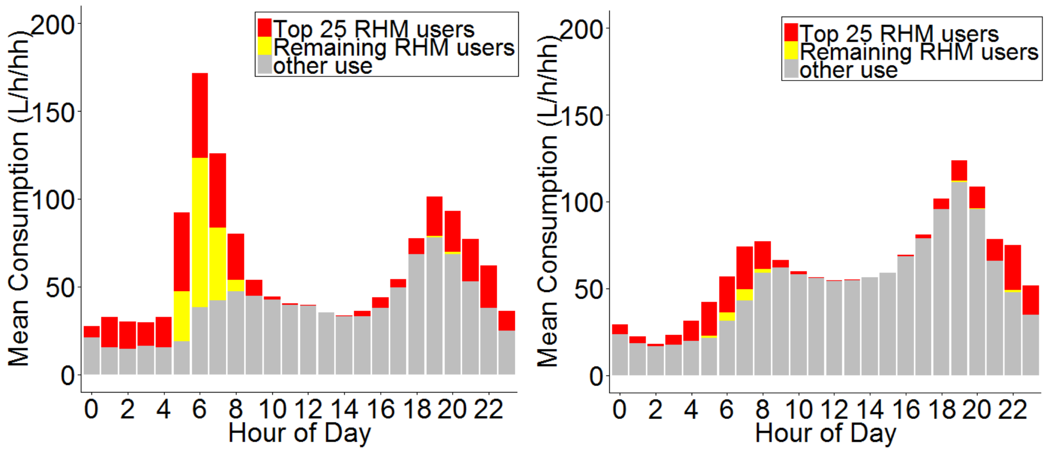


Figure 25. Habit demand as a proportion of all demand for Karratha (left) and Kalgoorlie (right) in the peak month of December 2014.

Conclusion

This report is intended to illustrate the potential of data analytics for smart metering. Case studies were carried out for two populations from WA's Pilbara and Goldfields regions. Three different types of water use signatures were analysed: continuous flows, anomalous peaks, and regular habits. Each of these uses is likely to have a high discretionary component, which makes them good targets for water efficiency actions (e.g. reducing garden watering times or fixing a leak). We also identified important contextual features that lead to different water use behaviours of populations.

Figure 26 shows that potential water savings are population dependent, but significant for both the case studies. For Karratha, recurrent habits offer the best target for water efficiency measures. For Kalgoorlie, where automatic watering systems are less common, anomalous peaks are the best target. Continuous flows, a traditional water efficiency target, are around 5% of all demand for each of the two populations.

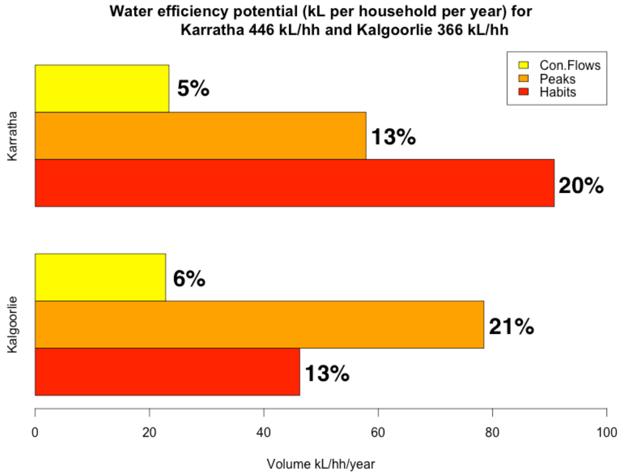


Figure 26. Water efficiency potential per household for two populations

Interactive visualisations

Cardell-Oliver, R, 2015. Karratha and Kalgoorlie 2014 smart metering case study examples *Water Use Signature Patterns*, http://people.csse.uwa.edu.au/rachel/waterusesignatures-june2015/

Cardell-Oliver, R, 2013. Kalgoorlie 2012 smart metering case study examples *Water Use Signature Patterns*, http://people.csse.uwa.edu.au/rachel/waterusesignatures-july2013/

References

Cardell-Oliver, R, 2013. *Water use signature patterns for analyzing household consumption using medium resolution meter data.* Water Resources Research, 49(12), 8589–8599, doi:10.1002/2013WR014458.

Cardell-Oliver, R, 2014. *A habit discovery algorithm for mining temporal recurrence patterns in metered consumption data.* 1st International Workshop on Machine Learning for Urban Sensor Data (SenseML) 15 September 2014, <https://www.tk.informatik.tu-darmstadt.de/en/senseml-2014/>.

Cole, G. and Stewart, R. A., 2013. *Smart meter enabled disaggregation of urban peak water demand: precursor to effective urban water planning.* Urban Water Journal, 10(3), 174-194.

CRC for
Water Sensitive Cities

Cooperative Research Centre for Water Sensitive Cities

8 Scenic Blvd, Level 1
Monash University, Clayton,
Victoria 3800, Australia

info@crcwsc.org.au

www.watersensitivecities.org.au