DEPARTMENT OF CIVIL AND ENVIRONMENTAL ENGINEERING

Residential Water Use in New Zealand – End Use Disaggregation 2.0

Dr Colin Whittaker, Teresa Scott, Professor Kobus van Zyl

July 2022

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

ENGINEERING

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

ENGINEERING

Residential Water Use in New Zealand – End Use Disaggregation 2.0

Authors

Colin Whittaker, Teresa Scott, Kobus van Zyl

Reference

Whittaker, C.N., Scott, T., van Zyl, J. 2022. Residential Water Use in New Zealand – End Use Disaggregation 2.0. Department of Civil and Environmental Engineering, Faculty of Engineering, The University of Auckland, Auckland, New Zealand.

THE UNIVERSITY OF AUCKLAND
Te Whare Wānanga o Tāmaki Makaurau
NEW ZEALAND

ENGINEERING

Contents

Figures

Tables

Executive summary

This project undertook a classification of residential water end use as part of a national-scale study by BRANZ. The study captured water use at a sampling frequency of 0.1 Hz, providing high-resolution data for the end use separation. Three methods were used to identify the end use types:

- Time series disaggregation
- Event-based disaggregation
- Support Vector Machines (SVM) supervised learning model

The mean daily use was 543 litres per dwelling (237 litres per person), while the median daily use was 397 litres per dwelling (165 litres per person). The mean hourly flow was 26 litres per hour, while the median hourly flow was 20 litres per hour. The average minimum night flow was 8 litres per hour, while the median minimum night flow was 2 litres per hour. These mean flows and minimum night flows exhibited seasonal variation. Dwellings that were metered and volumetrically charged used less water, on average, than those dwellings that were not metered and volumetrically charged.

The time series analysis provided an approximate calibration for the event-based disaggregation of residential water use, although calibration with field monitoring data would improve this process. The event-based disaggregation, which was also informed by household survey results where possible, was subsequently applied to all dwellings. The estimated contributions of different appliances to the total water use were generally in agreement with previously reported studies, although washing machine use was approximately 10% lower in this study. This may be caused by an increase in front loader use reported in the household survey or by limitations in the relatively simple disaggregation approach. The average use breakdown was:

- 24% Toilet
- 31% Shower
- 19% Tap
- 13% Washing Machine
- 3% Dishwasher
- 7% High Flow or Outdoor Use
- 2% Leaks or Drips
- 1% Undefined

The different end use types varied considerably between dwellings and fluctuated during the year, with significantly larger high flow and/or outdoor use during the summer months. Total water consumption also fluctuated between seasons.

The SVM model performed very well, with over 95% of events correctly classified using standard regularisation values. When using optimised parameters obtained using a grid search, this improved to over 99% of events correctly classified. However, it should be noted that, lacking appliance signatures or pre-labelled events for training purposes, the SVM model was trained using data from the event-based disaggregation approach, thus was not tested for independently labelled data.

Data completeness was a significant issue for this study, with large numbers of missing data points from many of the dwellings. The lack of appliance signature data or pre-labelled events was an additional limitation in this study; engagement with end users and the use of smart instrumentation to obtain these data represents an opportunity for future research.

Change Summary – Version 2.0

Page(s)	Section(s)	Change	Explanation/justification
6	Exec. Sum.	Inserted text	Added text on variability between dwellings.
8	1.2	Inserted text	Comparisons to WEEP and AWUS results, inclusion of standard deviations in the results.
10	2	Inserted text	Discussed standard deviations.
10	2	Updated Table 2	Added standard deviations, corrected values.
12	2	Inserted text	Slightly revised existing text to mention the standard deviations in Figures 5-7.
13	2	Updated Figures 5-7	Updated diurnal curve figures to include confidence intervals (±1 standard deviation).
14	2	Inserted text	Included reference to (new) Table 3 and a short discussion on the timing of the morning and evening water use peaks and the variability in peaking factors across the dataset.
14	2	New table: Table 3	Includes numerical peaking factor values (mean, median and standard deviation) for overall, weekday/weekend and seasonal data.
22	3.5	Inserted text	Added text to discuss the standard deviations.
23	4.1	Updated Table 4	Updated Table 4, including standard deviations and slightly correcting some of the average and median values where required.
24	4.1	Inserted text	Added text to discuss the standard deviations in the mean flow and MNF results.
27	4.3	Inserted text	Added text clarifying the relative contributions to use summarised in Figures 22 and 23 and referencing (new) Table 5.
29	4.3	New Table: Table 5	Includes summary numerical comparison with WEEP and AWUS datasets.
29	4.3	Inserted text	Added text referencing (new) Table 6 and discussing the variability in the results.
30	4.3	New Table: Table 6	Includes overall and seasonal breakdown of relative use, with standard deviations.
31	4.3	Inserted text	Added text referencing (new) Table 7 and comparing the daily resampled data in Figure 25 and the data in Figures 22 and 23.
31-32	4.3	New Table: Table 7	Includes mean, median and standard deviations of daily volumes associated with the different use types, overall and seasonal.
34	4.4	New Section, inserted text	Added new section and text to discuss the frequency of use results reported in the new Table 8.
34	4.4	New table: Table 8	Includes frequency of use for each disaggregated use type, reported in terms of mean, median and standard deviation.
39-44	Appendix A	Updated Figures 30-35	Updated diurnal curve figures to include confidence intervals (±1 standard deviation), separating seasonal figures into separate sub-plots for ease of interpretation.

1 Introduction

1.1 Purpose of this report

This report summarises research undertaken to support the broader project by the BRANZ titled Residential Water Use in New Zealand (Pollard, 2022). Quantifying residential use is important for councils and other entities responsible for water supply, assisting them in making informed decisions about the allocation of an increasingly scarce resource. Pollard (2022) summarises previous research projects undertaken to disaggregate residential end use into components for Winter and Summer periods; the reader is referred to this reference for the full details of these projects.

The current project (Pollard, 2022) utilised flow meters that collected water at 10-second intervals over relatively long timeframes, providing further opportunities for the interrogation of the collection data than had been possible in previous studies. The instrument data were supplemented by the results of a survey to gather information on household water use and appliances. Due to issues with the instrumentation and collected data, the project scope was narrowed to a reduced sample size and disaggregation into different end use types was not undertaken. The University of Auckland was contracted to undertake this end use disaggregation and to report on the results.

It should be noted that the instrumentation and data issues outlined in Pollard (2022) also affected this component of the overall research project. Many data records were too short or contained too many gaps to be used for the end use disaggregation or seasonal analysis (although they were included in the overall use statistics). The survey questions were not always answered fully, leading to uncertainties regarding household size or the types and numbers of appliances. Very limited appliance signature data were available, such that the identification of end uses required a somewhat iterative approach (described in Section 3). These issues and some of the inherent limitations in the current data analysis create some opportunities for future research; these opportunities are discussed briefly in Section 0.

1.2 Project objectives

The objective of this project is to quantify the components of the residential water use dataset presented in Pollard (2022). This included the disaggregation of water leakage from the overall dataset, or at least the identification of common signatures of different sources of leakage. It also included the disaggregation of water use into its components, similar to the WEEP (Heinrich, 2007) and AWUS (Roberti, 2010) projects but with a different analysis approach (lacking signature data and without using commercial software packages). Where appropriate, the disaggregation results are compared to the findings of these previous studies. Although an additional objective was the identification of influences of residential use where possible, this has not been undertaken in detail due to limitations in the dataset.

Version 2.0 of this report also includes additional information on the variability in the results, with standard deviations included in the diurnal curve figures and in the tabulated results.

2 Overall data summary

Pollard (2022) summarises the data collected within this project, particularly the information about each property and their self-reported water use as collected within the survey. This section contains some additional information about the overall dataset, which informed the methodology described in Section 3.

The mean record length is 291 days, with a minimum record length of 49 days and a maximum record length of 564 days. Although many of the records were nearly complete, a number were missing large amounts of data. The maximum percentage of missing values was 46%, while the mean number of missing data values was 15% across all dwellings. However, the records tended to cover different time periods, making direct comparisons difficult. The dwellings cover a range of water use behaviours, both in terms of the water use per person and the water use per dwelling. Both will be discussed in this section, although the sample size for the two definitions of water use is slightly different. Occupancy information was only available for 55 of the 66 dwellings considered in this study, or 83% of the dwellings; these are summarised in Table 1.

The effects of occupancy on the distribution of water use are shown in Figure 1. The average daily water use per dwelling was 543 litres, while the median daily water use was 397 litres. The equivalent average daily water use per person was 237 litres, while the median daily water use per person was 165 litres.

Table 1 Reported occupancy levels for all dwellings in this study.

Number of occupants	Count
1	12
2	22
3	6
4	8
5	6
7	1

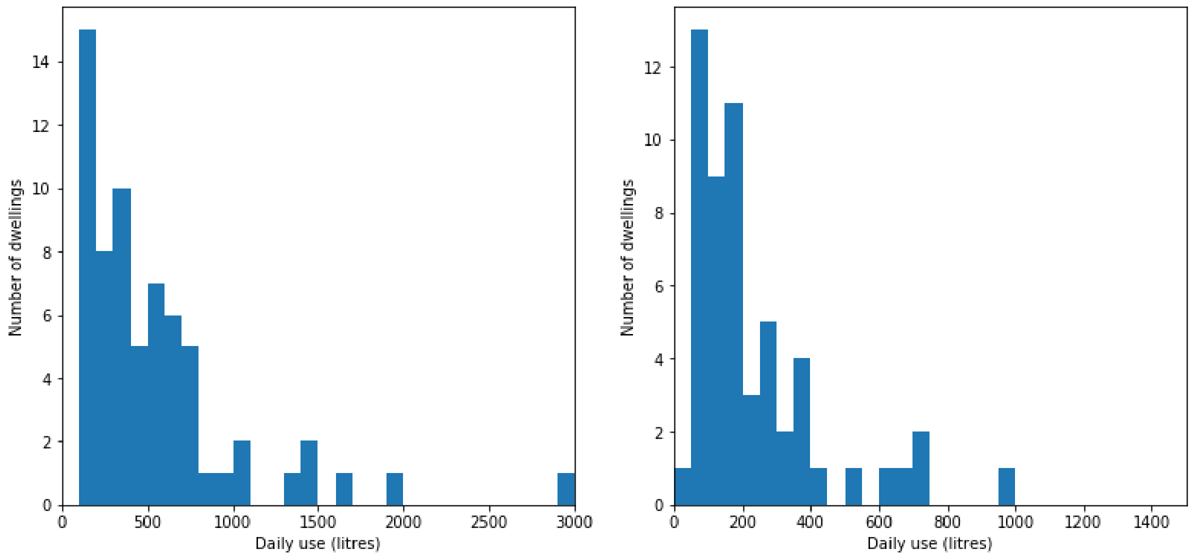

Figure 1 Distribution of average water use in litres per day, considered on a per-dwelling (left figure) and per-person (right figure) basis.

The survey data identified which of the households were metered and volumetrically charged for their water use. As shown in Table 2, although the results were limited by the sample size, the households that were metered and volumetrically charged did use less water. However, this may not be a causal link; dwellings with higher average use may simply be less likely to be metered and volumetrically charged. This result was consistent whether considering the average or median daily volumes; this also applied when considering per-person use. This table also details the differences in seasonal use between houses based on metering and volumetric charging, although some of these are affected by data completeness (refer to Figure 6 of Pollard, 2022, and the associated discussion). Several households recorded the highest water use in Winter. The variability in the data is clear, with very large standard deviations observed (particularly in Summer). Figure 2 also illustrates the distribution of average daily water use for dwellings, differentiated by metering and volumetric charging.

Table 2 Daily use statistics, considering whether dwellings were metered and volumetrically charged.

	Average Daily Volume (litres)	Median Daily Volume (litres)	Average Daily Volume Per Person (litres)	Median Daily Volume Per Person (litres)
Entire record				
All households	543 (σ=482)	397	237 (σ=197)	165
Metered and volumetrically charged	424 (σ=359)	319	198 (σ=198)	127
Not metered and volumetrically charged	682 (σ=570)	551	304 (σ=182)	266
Spring				
All households	464 (σ=498)	277	201 (σ=200)	145
Metered and volumetrically charged	286 (σ=362)	180	138 (σ=195)	68
Not metered and volumetrically charged	655 (σ=555)	513	295 (σ=172)	244
Summer				
All households	614 (σ=690)	412	272 (σ=371)	163
Metered and volumetrically charged	554 (σ=802)	331	250 (σ=411)	119
Not metered and volumetrically charged	687 (σ=533)	574	310 (σ=295)	237
Autumn				
All households	476 (σ=301)	411	209 (σ=144)	156
Metered and volumetrically charged	419 (σ=232)	363	186 (σ=127)	144
Not metered and volumetrically charged	550 (σ=364)	537	256 (σ=167)	250
Winter				
All households	548 (σ=696)	390	226 (σ=202)	163
Metered and volumetrically charged	379 (σ=201)	336	177 (σ=198)	128
Not metered and volumetrically charged	765 (σ=996)	446	322 (σ=274)	260

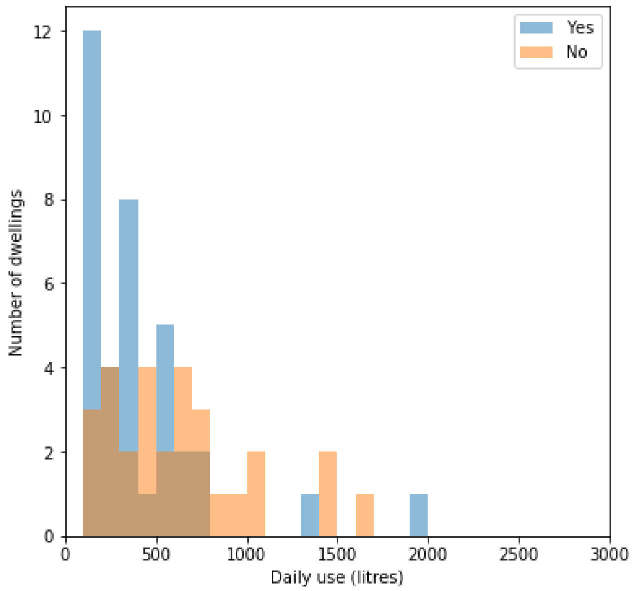

Figure 2 Effects of metering and volumetric charging (identified through the survey data) on the distribution of average water use in litres per day, where "Yes" and "No" denote households that were and were not metered and volumetrically charged, respectively.

The data provided in Figure 1, Figure 2 and Table 2 consider the average daily use from each dwelling. Figure 3 shows the daily use distributions (per property and per person) calculated including all individual days from all dwellings. Figure 4 shows similar distributions, separated by season. Unsurprisingly, the shape of these distributions exhibited seasonal variation, with much broader distributions in Spring and Summer than in Autumn and Winter.

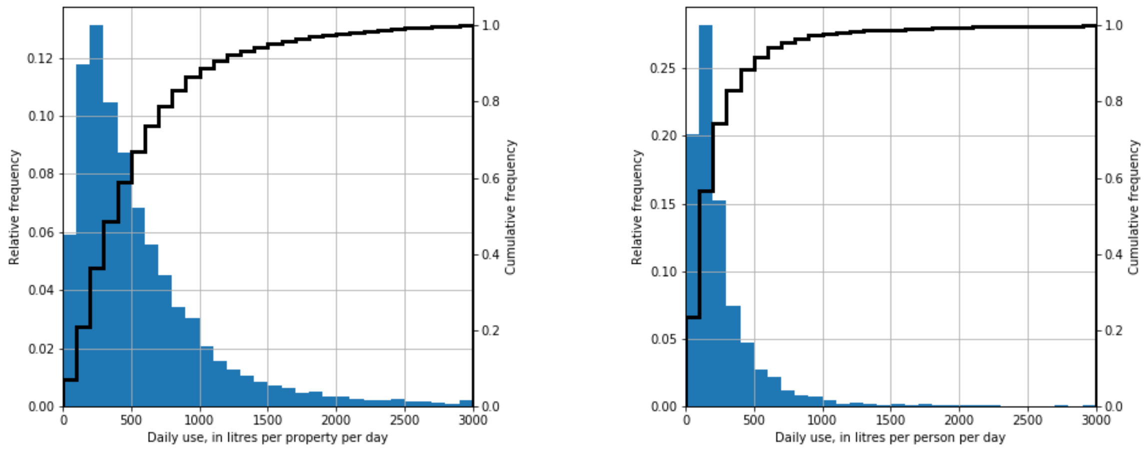

Figure 3 Distributions of daily water use volumes on a per-dwelling (left figure) and per-person (right figure) basis.

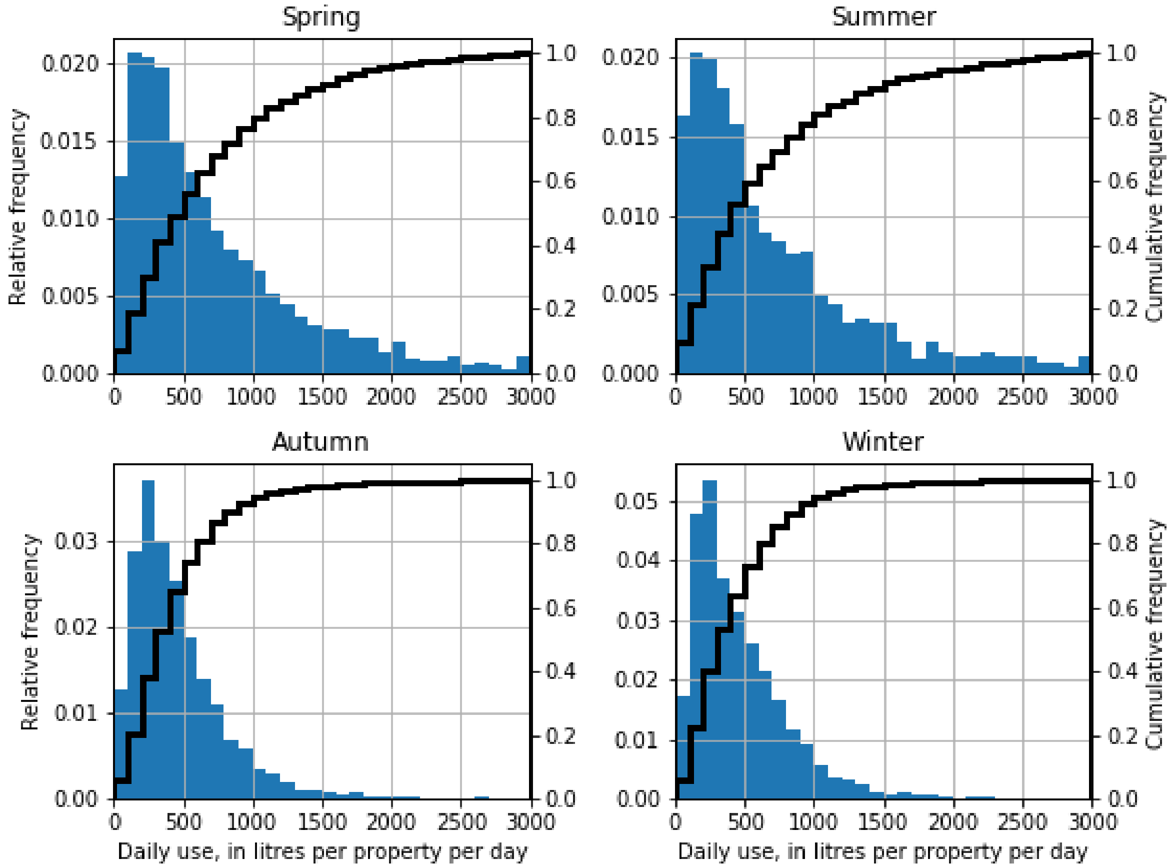

Figure 4 Seasonal distributions of daily water use volumes on a per-dwelling basis. Note that the vertical scales differ between the individual distributions.

The high temporal resolution of the dataset also enabled the calculation of variations in water use during the day. Three types of "diurnal curves" are reported in this study:

- Peaking factors (normalised by average daily use)
- Hourly flowrate per dwelling
- Hourly flowrate per person

These diurnal curves are shown in Figure 5, Figure 6 and Figure 7, respectively. It should be noted that these hourly diurnal curves are plotted from 0000 to 2300. The lack of wraparound explains the slight discontinuities between the start and end points within these figures. Given that the peaking factors are normalised by the average consumption, these exhibit less variability than the flowrates.

These diurnal curves may also be plotted to capture seasonal or weekday/weekend variations. These are not included within the main body of this report; rather, they are included in Appendix A: Diurnal curves. Although these curves can also be further refined into individual months or days of the week, these do not yield much additional information and are not included in this report.

The trends evident in these additional curves are unsurprising (see Appendix A: Diurnal curves). Weekday use is more sharply peaked in the morning and evening, while the weekend use has a later and less pronounced morning peak, with more consistent use throughout the day. The seasonal variations demonstrate that the morning peak occurs at different times depending on the season (earlier in summer) while the afternoon/evening peak is strongest during summer, then Autumn (particularly due to the contribution of March), then Spring and Winter.

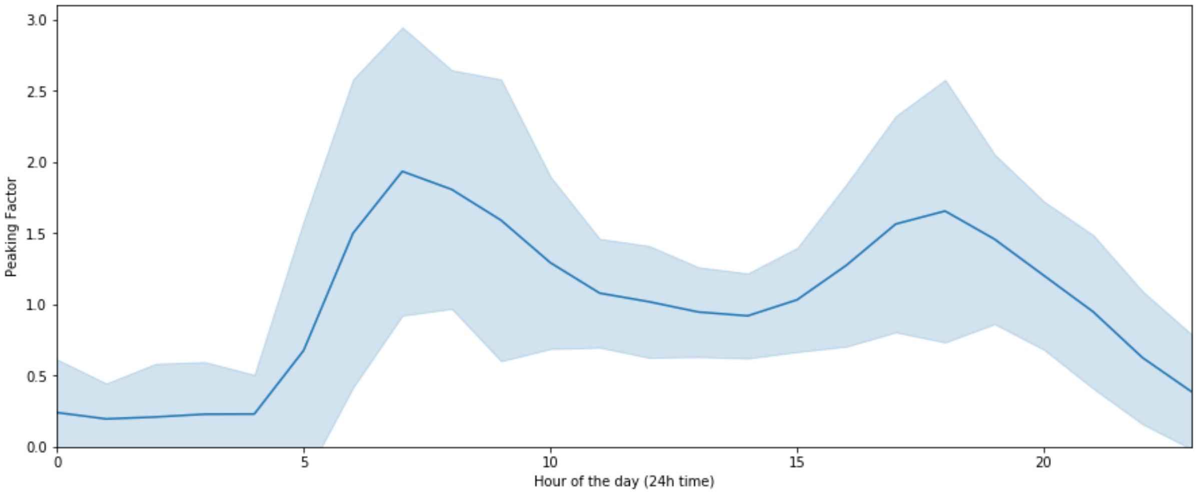

Figure 5 Hourly variation in peaking factors, averaged over all dwellings, where the shaded area represents one standard deviation.

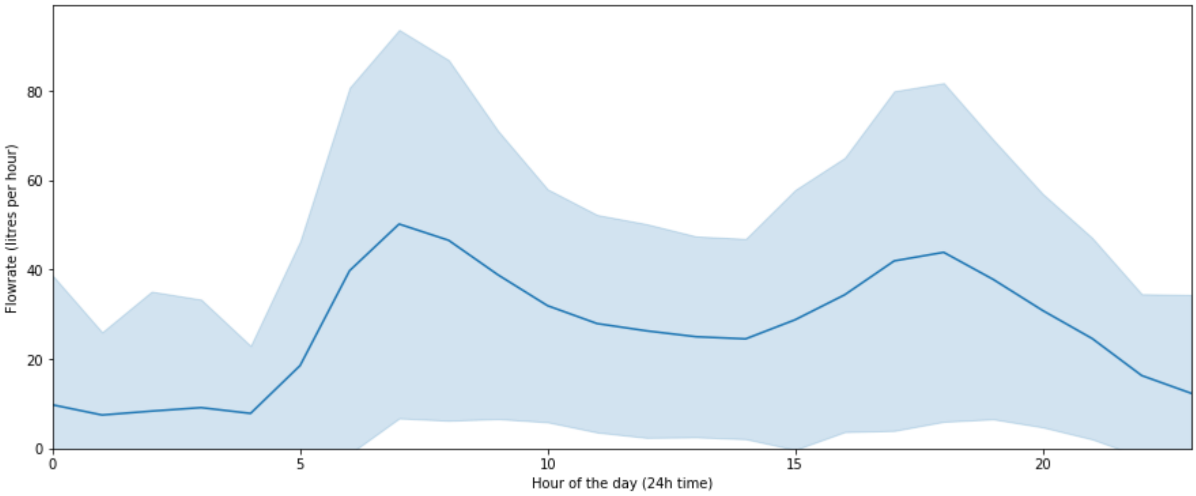

Figure 6 Hourly variation in flowrate, averaged over all dwellings, where the shaded area represents one standard deviation.

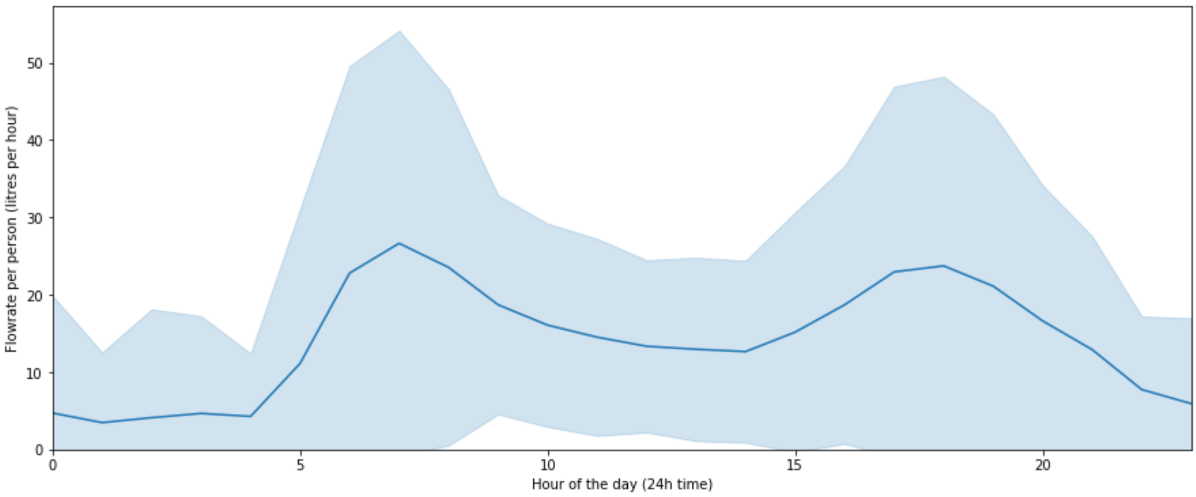

Figure 7 Hourly variation in flowrate per person, averaged over all dwellings, where the shaded area represents one standard deviation.

Error! Not a valid bookmark self-reference. summarises the morning and evening peaking factors (and their timing) for these different cases. Overall, the timing of the peaks within the (hourly resampled) diurnal curves was reasonably consistent across seasons and between weekdays and weekends. Unsurprisingly, the morning peak occurred earlier in Spring and Summer, while it occurred slightly later in Winter. Weekdays on average had a morning peak occurring at 0700, while the morning peak in weekends occurred at 0800. The timing of the evening peak was consistent across all seasons, although the peak occurred slightly earlier in weekends.

The peaking factors exhibited considerable variability across the entire data. The standard deviations listed in Table 3 were very large, in some cases of the same magnitude as the average peaking factors themselves. This demonstrates the diversity of water use profiles across the various dwellings included in the dataset. Although these standard deviations would likely reduce if the data were further broken down by dwelling type or other socio-economic data (the same is true in general of standard deviations reported elsewhere in this report), this more detailed analysis was not undertaken in the current project. It is also noted that the sample size and data quality (see Pollard, 2022) may be insufficient for such an analysis, although the results are likely to be of interest to water service provides in the future.

Table 3 Summary of morning and evening peaking factors, including seasonal and weekday/weekend variations. Peaking factors are reported in terms of average (mean) and median values across all dwellings, and including standard deviations.

	Time of peak	Average peaking factor	Median peaking factor	Standard deviation
Entire record – all days				
Morning peak	0700	1.93	1.69	1.01
Evening peak	1800	1.65	1.49	0.92
Entire record – weekdays				
Morning peak	0700	2.10	1.42	1.91
Evening peak	1800	1.77	1.19	1.61
Entire record – weekends				
Morning peak	0800	1.39	0.91	1.46
Evening peak	1700	1.39	1.04	1.20
Spring – all days				
Morning peak	0600	2.04	1.38	2.07
Evening peak	1800	1.66	1.22	1.41
Summer – all days				
Morning peak	0600	1.72	1.21	1.95
Evening peak	1800	1.67	1.07	1.42
Autumn – all days				
Morning peak	0700	2.01	1.53	1.70
Evening peak	1800	1.83	1.23	1.62
Winter – all days				
Morning peak	0800	2.09	1.28	2.42
Evening peak	1800	1.44	0.84	1.68

3 Summary of analysis approaches

This section summarises the different steps undertaken in the analysis of the meter data for the disaggregation of leakage and end use. The initial disaggregation was undertaken, where possible, for a 2-week period in Winter. This approach ensured a manageable record length, while capturing a snapshot of typical use for each household. Following the Winter analysis, this fortnightly analysis was extended to the other seasons. This enabled testing of the appliance characteristics to account for seasonal variations. Finally, the approach was applied to all seasons within the dataset for that dwelling. The key steps in the analysis undertaken for each dwelling are summarised below:

- Undertake time series disaggregation (for selected dwellings)
- Analyse use for a fortnight in Winter
    - Summarise mean and minimum night flows
    - Identify any potential leakage events
    - Identify any other potential instances of simultaneous use and separate these
    - Disaggregate use based on common event characteristics
    - Check results against time series disaggregation results, if available
- Extend fortnightly analysis to Spring, Autumn, and Summer (in that order)
    - Check event characteristics and iterate if required
- Apply event characteristics to entire dataset for dwelling (all seasons)

3.1 Time series disaggregation

To provide a check on the disaggregation results, a time series analysis of the water use time series data was undertaken. This manual analysis was carried out over a 2-week period. For each water use event, defined as a continuous period of non-zero water use, this analysis involved:

- Calculation of the key event characteristics:
    - Duration
    - Mean flow rate
    - Maximum flow rate
    - Total volume
- Plotting of the flow rate time series

Although appliance signature data was unavailable for all but one of the dwellings considered within this study, the time series analysis provided reliable estimates of the most likely appliances associated with the common water use events. Figure 8 illustrates the results of the times series analysis for one day in Winter from the record of Dwelling 25, where the raw water use is disaggregated into toilet, shower, tap, washing machine and dishwasher events. Given that this analysis was undertaken directly on the water use time series data, it was relatively straight-forward to identify and separate instances of simultaneous water use, such as a toilet flush during a shower.

The time series analysis results provided calibration data to the automated data analysis, which did not involve an examination of the raw time series data, but instead used the key event characteristics to identify the most likely use type for each event. The comparison for a representative Winter fortnight is shown in Section 4.2. Although time constraints prevented the application of the time series analysis to all dwellings for a 2-week period, this analysis was applied to many additional dwellings for shorter time periods (2 days to 1 week) to provide checks on the automated disaggregation. These checks were undertaken particularly when the data contained large “spikes” in water use, or when a particular use type dominated the overall signal.

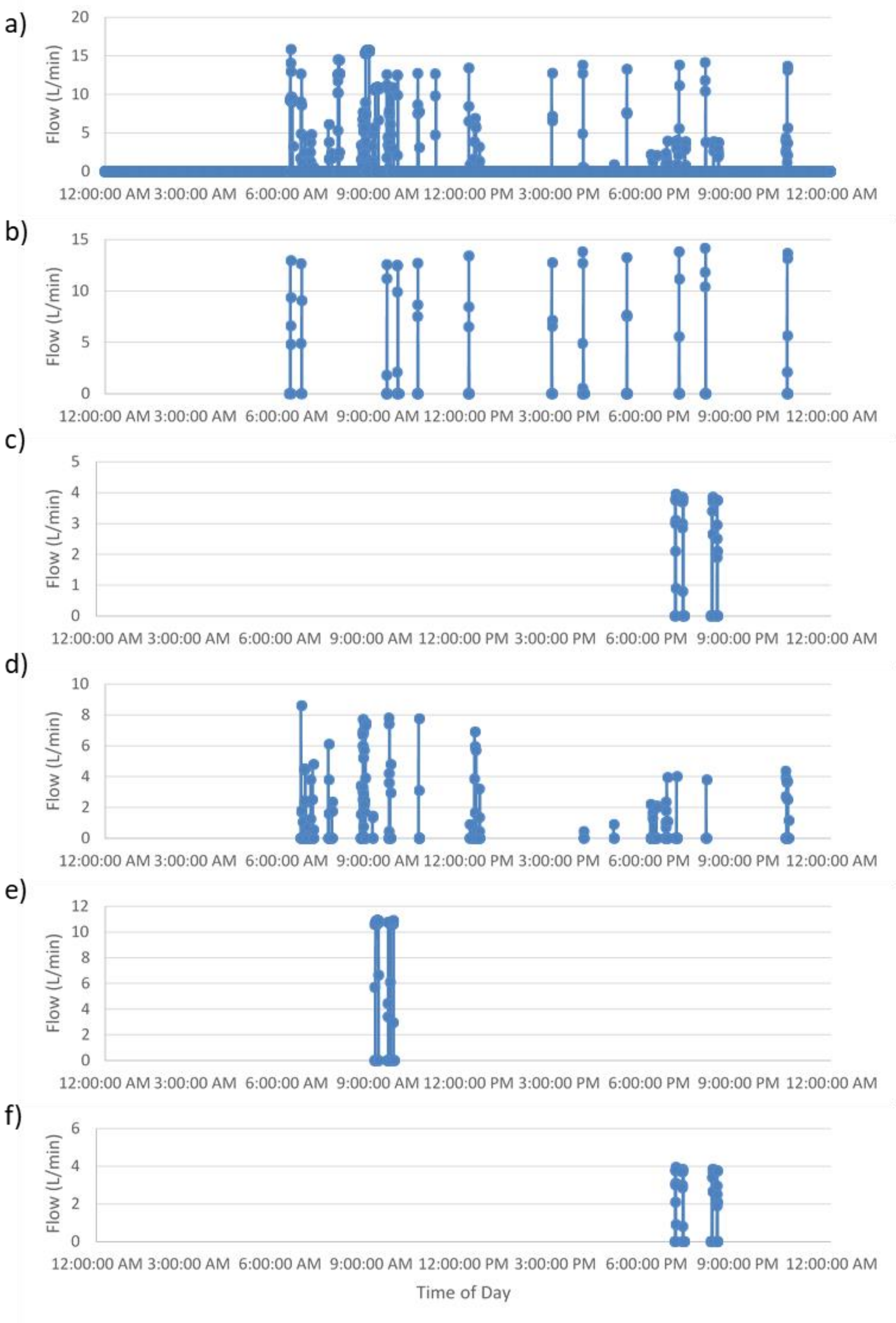

Figure 8 Results of the time series water use disaggregation for one representative Winter day within the overall 2-week period. a) Total water use, including all events; b) Toilet events; c) Shower events; d) Tap events; e) Washing machine events; f) Dishwasher events.

3.2 MNFs and leakage disaggregation

The first step in the overall analysis was to investigate the mean and minimum night flows for a 2-week period in Winter. This season was selected as a baseline for household water use, as it is less likely to include large amounts of irrigation and outdoor use. This facilitates the identification of leaks and appliances within the dataset. The survey also contained questions that allowed households to self-report any instances of leakage or any drips that they were aware of. In their responses, 5 out of the 66 dwellings reported a drip (either a toilet cistern or a tap), while 1 dwelling reported a leak (an outside tap).

Minimum night flows within the fortnight allow a quick check on any potential leaks or drips. Figure 9 illustrates the time series of hourly water use for one of the dwellings. Any long periods with consistently high water use (particularly overnight) may indicate the presence of a leak or drip. Figure 10 plots the daily mean and minimum night flows over the fortnight, where the minimum night flows are calculated from 2am to 4am. Water use during this time period is typically very low, such that any leaks will be more readily identified from MNF data.

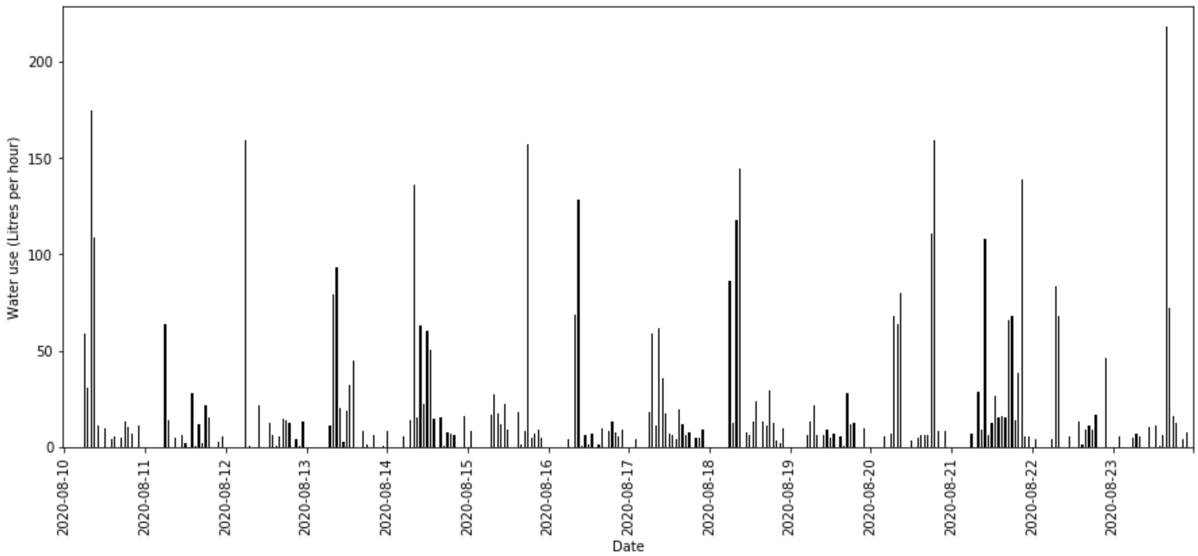

Figure 9 Water use over a fortnight in Winter for Dwelling 25, illustrating the temporal variation of hourly water use.

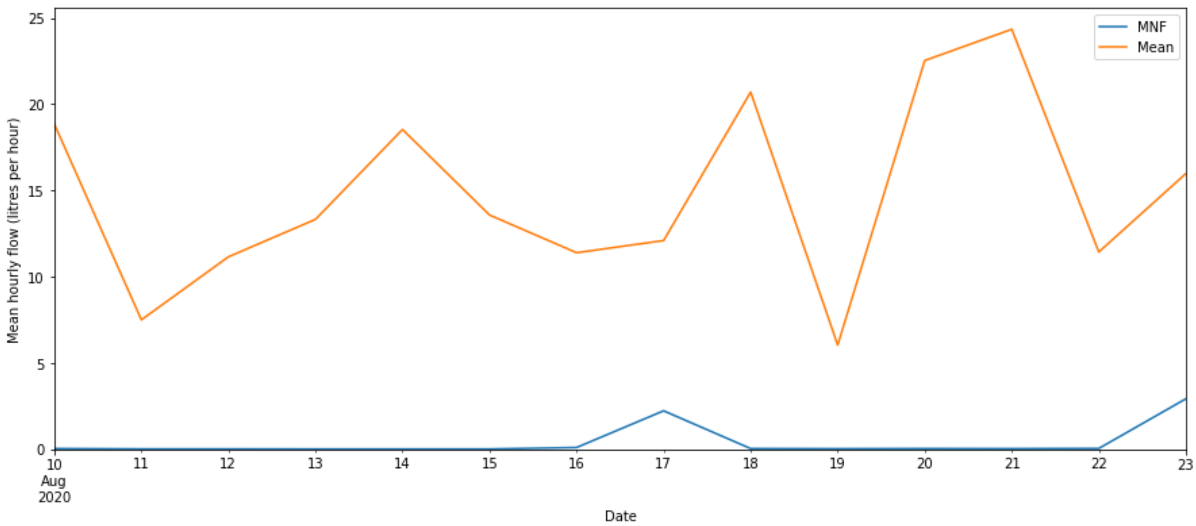

Figure 10 Daily mean and minimum night flows for a fortnight in Winter for Dwelling 25. During this period, the mean flow was 14.8 litres per hour, while the minimum night flow was 0.4 litres per hour.

An additional check on leaks and drips is to check for the presence of very long water use events, particularly those with very low mean flow rates, or to check for the presence of very large numbers of single-timestep events that contain a very small flowrate (and hence, volume). The latter may indicate a slow leak or drip that did not exceed the volume increment (less than 30 mL for these instruments, see Pollard, 2022) during every 10 s timestep. Often such events registered the lowest possible mean flowrate for the instrument, corresponding to the minimum volume change within a single timestep.

Events were classified as leaks or drips if the met one of the following criteria:

- A single event had a very long duration and a very low mean flowrate (e.g. 0.2 litres per minute), and was not consistent with any other use types for that dwelling.
- A large number of single-timestep events had very low flows and occurred at a consistent frequency (e.g. every 4 timesteps) over a long time period (including the MNF period).

Given more detailed information about the instrumentation, the frequency of occurrence for these single-timestep events can be readily converted into an equivalent continuous flow rate for the leak or drip. This was not undertaken in the present study, where only the total volume associated with leaks or drips was recorded for the dwelling and period of interest. Leaks and drips were more challenging to identify for dwellings that contained no data for the Winter period, or dwellings with frequent gaps in the measured data.

3.3 Event-based use disaggregation

Following the disaggregation of any leaks and drips, the remaining events were detected from the water use data. Each event was identified as a continuous period of non-zero water use. In the fortnightly analysis, any extremely long events were flagged for checking, as these could also be instances of leaks with larger flow rates, or other large use events such as the filling of a swimming pool. Given the large number of potential simultaneous events, such as a toilet flush occurring during a shower event, a simple "peaks above threshold" approach was applied to automate the separation of events. Although this approach was not always effective for events with very complicated variations in water use, it worked very well at separating shorter events from long events with a consistent flow rate. Figure 11 and Figure 12 illustrate this automated separation process for a longer event containing one and two shorter events, respectively.

Following the separation of simultaneous events, events were checked for instances of washing machine and dishwasher use. These appliances were more complicated to detect than other appliances, given the multiple cycles associated with a single event. All possible individual "candidate" events were identified based on their duration, mean and maximum flowrate, and total volume. The most common time gaps between these candidate events were identified and checked against standard timeframes for dishwasher and washing machine cycles. Where necessary, these common time gaps were identified for the entire record (not merely for a fortnight), as a fortnight did not typically contain a large enough sample of dishwasher or washing machine events. Finally, any candidate events that were separated by one of the identified time gaps were combined into a single event and classified as a washing machine or dishwasher event, as appropriate. Given the variability in the water use profiles of different types and brands of dishwashers and washing machines (beyond the differences between front and top loaders discussed in Pollard, 2022), this simplistic approach may have under-represented washing machine and dishwasher use. Where possible, the identification of these events was informed by the survey data on the type, size, and frequency of use of these appliances. Reliable signature data could further improve this approach.

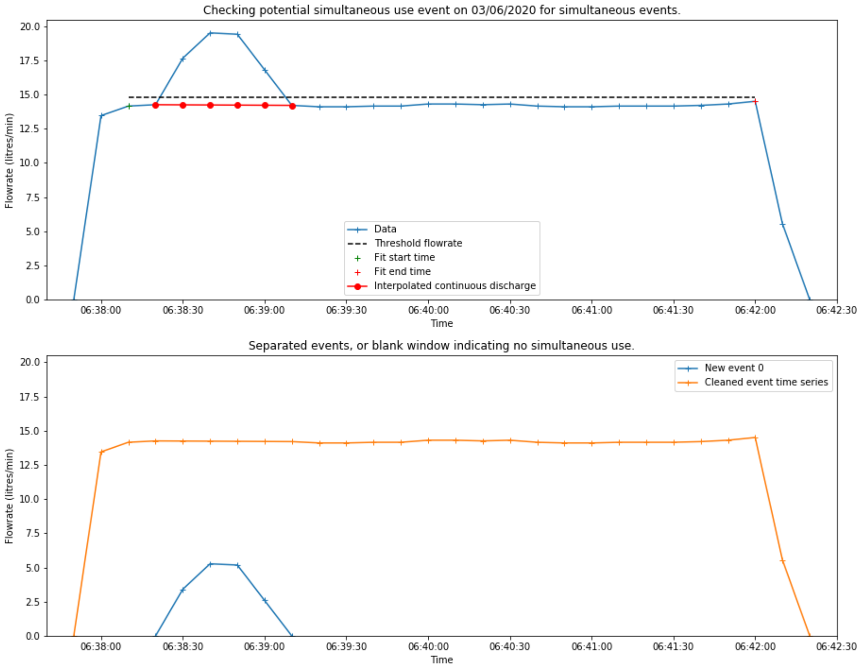

Figure 11 Separation of simultaneous use events: Removal of one short event from a longer event with a consistent discharge.

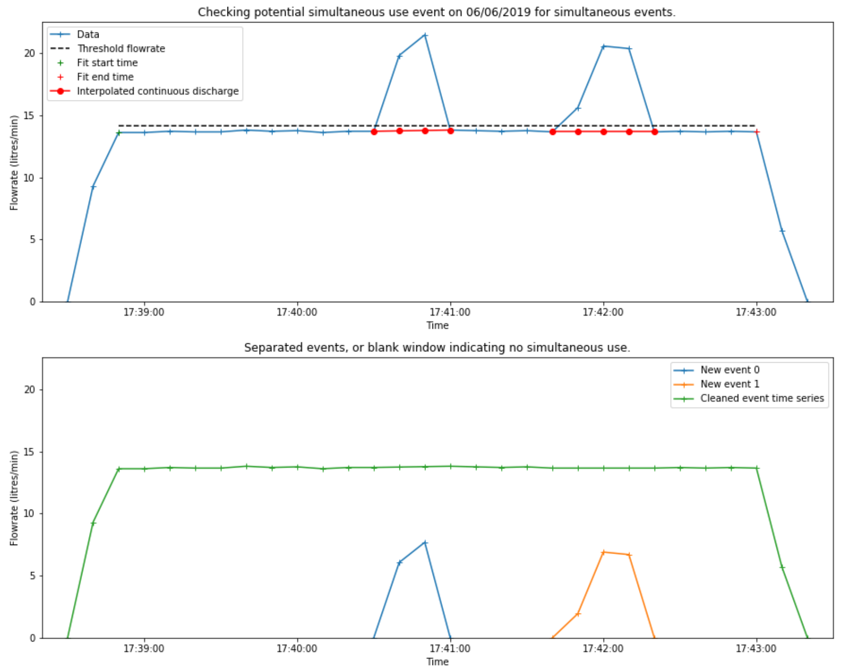

Figure 12 Separation of simultaneous use events: Removal of two short events from a longer event with a consistent discharge.

Following the identification of dishwasher and washing machine events, all events were checked for instances of toilet, shower and tap use. During the initial fortnightly analysis, event characteristics (duration, mean and maximum flowrate, and total volume) were determined using survey data and data visualisation. Figure 13 illustrates this process for the identification of toilet events within the Winter fortnight analysis of Dwelling 25. Survey data indicated that the dwelling contained one small dual-flush toilet. The histograms and "kernel density" contours in Figure 13 clearly show the peaks in duration, mean and maximum flowrates, and total volumes, that are consistent with toilet flush events. In this case, the half-flush and full-flush events can be clearly separated. However, this was not possible in all dwellings; toilet events are not separated into half-flush and full-flush events within this report.

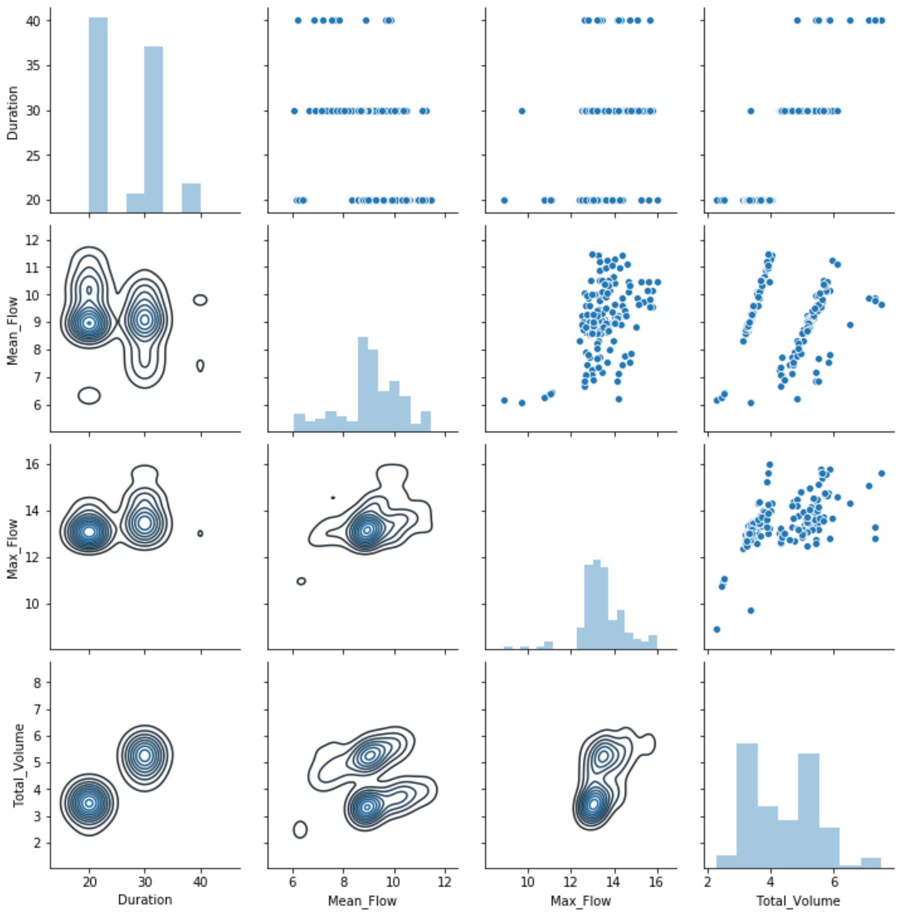

Figure 13 Use of data visualisation to identify the characteristics of dual-flush toilet events.

Although the example in Figure 13 illustrates the use of data visualisation for the identification of toilet events, the same type of "pair plot" was used to identify the most common characteristics for shower, dishwasher, washing machine, and tap events. Where required, further details could be obtained using a time series plot to check the variation of water use within the event(s) of interest. An additional "high flow or outdoor use" category was also created, as it is difficult to differentiate between outdoor use and high-flow taps such as in a laundry tub.

After each subsequent use type was identified, the number of events and contribution to the total water use volume for that fortnight were also determined. The survey data provided valuable information for the use disaggregation, for example self-reported frequency and duration of showers in a typical week. Where possible, these survey results were compared to the disaggregated event data.

Following the disaggregation of the events into their different use components, data were plotted and saved for further analysis. As noted at the start of Section 3, after completing the analysis for a fortnight in winter, the same approach was applied to 2-week periods within Spring, Summer and Autumn (depending on data availability). Where required, the characteristics of each event type were refined at this step, to ensure sensible use profiles across all seasons. Finally, the analysis approach was automated and applied to the full dataset for the dwelling of interest.

3.4 Machine Learning – Support Vector Machines

Support-vector machines (SVMs) are supervised learning models that perform well in classification and regression problems. The supervised learning of SVMs has also been discussed in detail and demonstrated to perform well for end-use classification of water use by Gourmelon et al. (2021), although the authors noted the need for accurate labelling of end use as they used synthetic data from a consumption simulation tool. In the current project, SVM was applied initially to one dwelling within the overall dataset (Dwelling 25), using the end-use labels derived from the event-based analysis. This is an imperfect approach, and ideally a training dataset of this kind would contain independently verified labels for the different end use types.

Often SVM models are trained following exploratory data analysis and visualisation, as illustrated in Figure 14 and Figure 15. These figures show all events with volumes larger and smaller than 5 litres, respectively, for ease of visualisation. Here the use types appear to be well clustered, which is unsurprising given the approach taken in the event-based disaggregation.

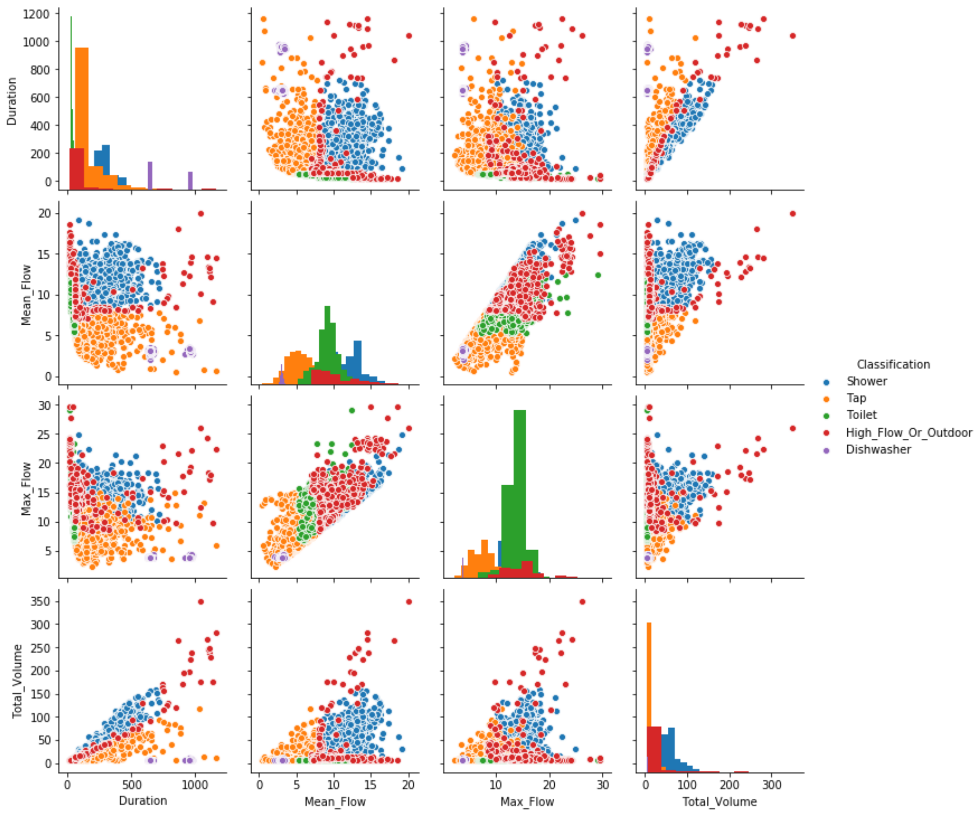

Figure 14 Training data for SVM model, for all events with a total water use volume larger than 5 litres.

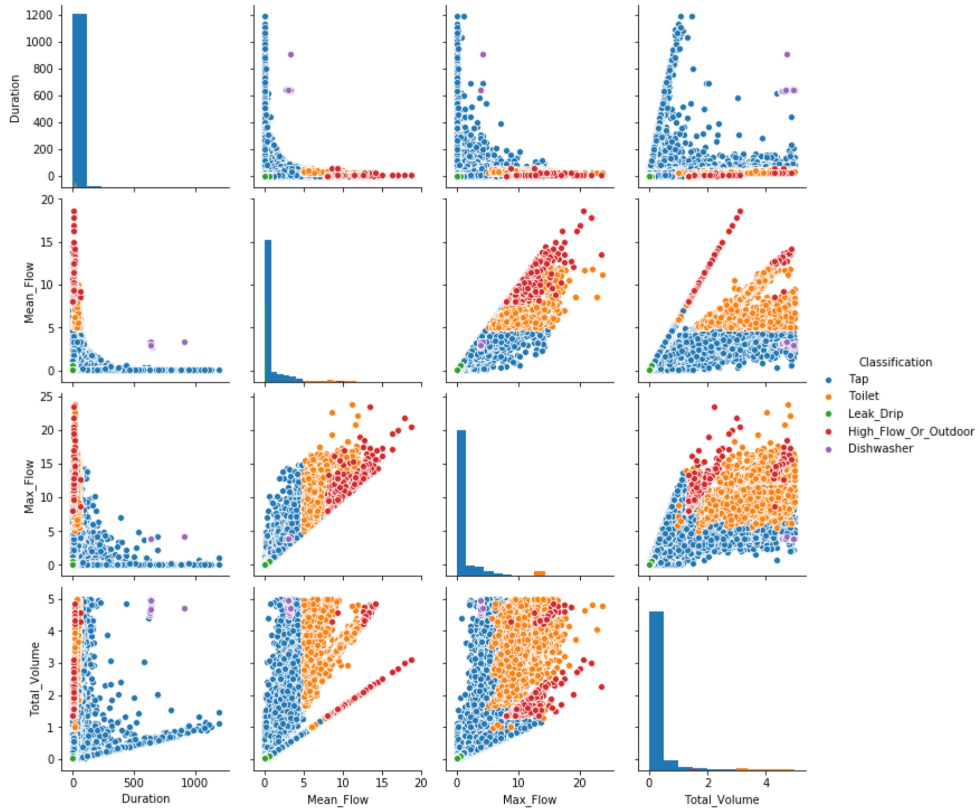

Figure 15 Training data for SVM model, for all events with a total water use volume smaller than 5 litres.

The SVM model was trained as a standard support vector classifier (SVC) with regularisation parameters of 1.0. The model used a Radial Basis Function (RBF) kernel. Subsequently, the classification was enhanced by performing a grid search to determine the optimal regularisation parameters for the classification task. Although time consuming, this provided an improvement in overall performance.

3.5 Limitations and opportunities

The main limitation in applying these end use disaggregation approaches to the residential water use data was the record length and data quality. Although the 10 second data resolution was excellent, many records were missing one or more seasons or had very large gaps in the data. Others also contained more frequent missing values, which complicated the end use disaggregation. Any future projects would benefit from more complete datasets from each dwelling, as well as possibly an increased sample size. Both could help reduce the standard deviations in the results.

Earlier studies on end use disaggregation (Heinrich, 2007) utilised “fingerprint” data or a more sophisticated instrument setup to enable direct matching of different appliances to the recorded water use. The lack of such data in the current project necessitated the approach described in Section 3.3, which did not have an independent “source of truth” for the disaggregated data. Significant improvements would be possible if the signatures of different appliances were known for each dwelling. This would also help reduce the standard deviations in the results.

Finally, the survey questions were designed to provide insights to assist with the end use disaggregation. However, many of the questions were not answered or were answered incompletely in the survey. Pollard (2022) describes the lessons learned from the broader project.

4 Results

4.1 Minimum night flows and leakage disaggregation

Table 3 presents the mean flow and MNF data for all dwellings, again separated according to whether the dwellings were metered and volumetrically charged. Seasonal minimum night flows were consistently higher for dwellings that were not metered and volumetrically charged. This was slightly different for the overall dataset, where the metered and volumetrically charged households had a higher MNF. However, not all dwellings had a sufficient record length to be included in the seasonal analysis, but all were included in the overall MNF analysis.

Table 4 Summary of mean and minimum night flow results, considering all dwellings. Standard deviations are reported beside the average hourly flows and average MNFs.

	Average Hourly Flow (litres/hour)	Median Hourly Flow (litres/hour)	Average MNF (litres/hour)	Median MNF (litres/hour)
Entire record				
All households	25.7 (σ=21.1)	19.6	8.4 (σ=24.0)	2.1
Metered and volumetrically charged	20.8 (σ=15.2)	16.7	9.1 (σ=30.6)	1.6
Not metered and volumetrically charged	31.3 (σ=25.5)	25.2	7.6 (σ=13.1)	3.1
Spring				
All households	25.3 (σ=21.1)	17.4	6.9 (σ=16.7)	1.6
Metered and volumetrically charged	17.6 (σ=9.9)	16.0	2.6 (σ=3.5)	1.4
Not metered and volumetrically charged	30.4 (σ=25.4)	17.5	9.8 (σ=21.6)	2.0
Summer				
All households	33.4 (σ=29.7)	24.1	18.9 (σ=82.4)	2.3
Metered and volumetrically charged	25.0 (σ=14.9)	23.7	3.0 (σ=3.1)	2.0
Not metered and volumetrically charged	43.0 (σ=38.8)	30.5	37.1 (σ=119.2)	3.9
Autumn				
All households	19.9 (σ=10.8)	17.7	2.7 (σ=3.0)	1.8
Metered and volumetrically charged	18.2 (σ=9.1)	16.3	2.6 (σ=2.7)	1.8
Not metered and volumetrically charged	21.9 (σ=12.8)	20.5	2.7 (σ=3.6)	1.8
Winter				
All households	23.9 (σ=29.6)	17.4	6.8 (σ=21.6)	1.7
Metered and volumetrically charged	17.8 (σ=9.4)	15.8	2.7 (σ=3.1)	1.4
Not metered and volumetrically charged	29.2 (σ=41.7)	17.5	10.1 (σ=31.4)	1.7

At least one of these dwellings had a very large MNF, hence distorting the average value. The difference based on metering and volumetric charging reduced considerably (and the trend was more consistent with the seasonal data) when considering the median instead of mean values, implying that the results were skewed by a small number of households with much higher use. The standard deviations in the reported data were generally high, although these reduced when considering the seasonal data separately for dwellings that were both metered and volumetrically charged. The highest standard deviations were obtained for MNFs for households that were not metered and volumetrically charged. This may indicate the use of overnight irrigation or similar in some of these properties.

Figure 16 also shows the distribution of mean flows and MNFs, including all dwellings, while Figure 17 shows the same information differentiated by the metering and volumetric charging.

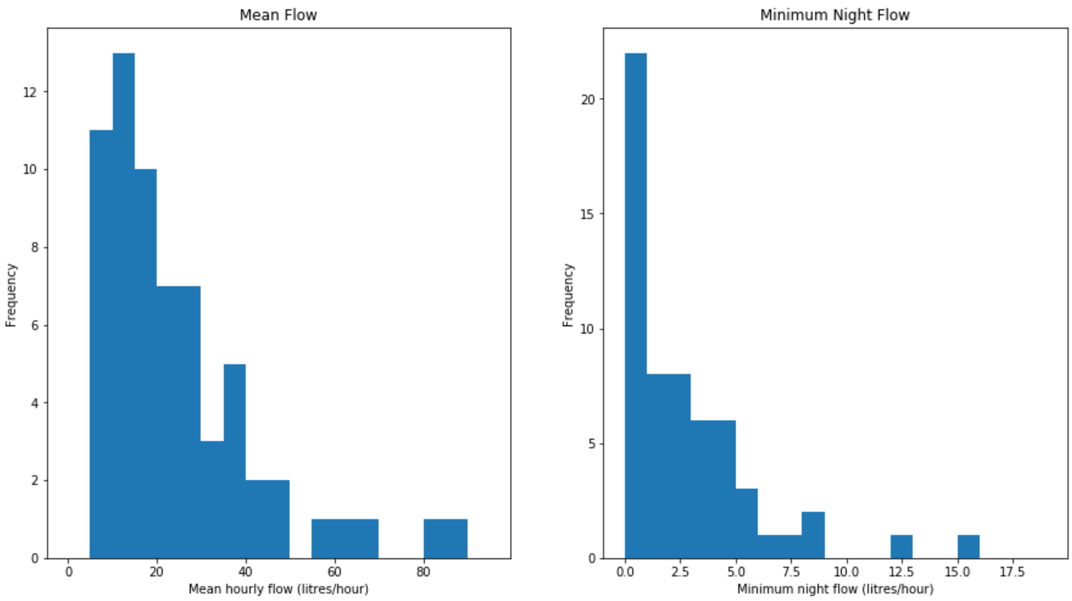

Figure 16 Mean and minimum night flow distributions across all dwellings.

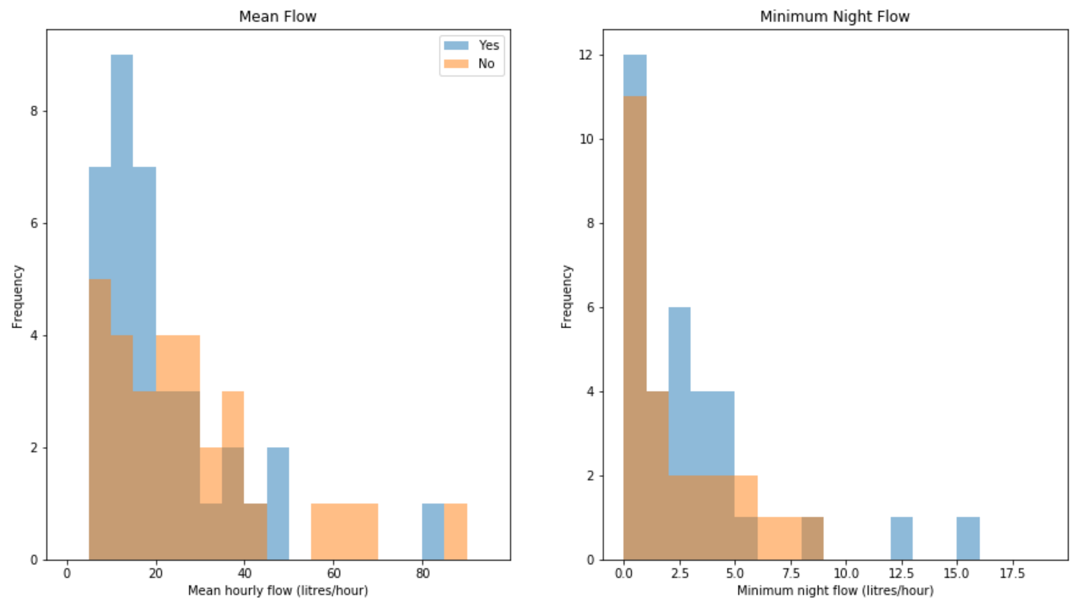

Figure 17 Mean and minimum night flow distributions across all dwellings, considering whether the dwelling was metered and volumetrically charged ("Yes", blue colour) or was not ("No", orange colour).

Periods of high MNF are indicative of potential leaks, particularly during the Winter months when sustained outdoor use and irrigation are less likely. This may be elucidated by plotting the daily mean flow and MNF data (see Figure 10) over the entire record length. Figure 18 shows data for a dwelling where the MNF was elevated during Spring and Summer. More detailed information about the water use during this period would be required to determine whether a leak was present or not. However, the data in Figure 19 appear to show the presence of a progressive leak, with a steady increase in MNF during Autumn and Winter. Methods for determining the stationarity of time series data (e.g. Dickey & Fuller, 1979) could be used to determine whether water use meets the criteria for a progressive leak. However, this relies on long continuous data records that span many seasons; hence, a simple check on mean flows and MNFs is likely to be sufficient for leak checking purposes.

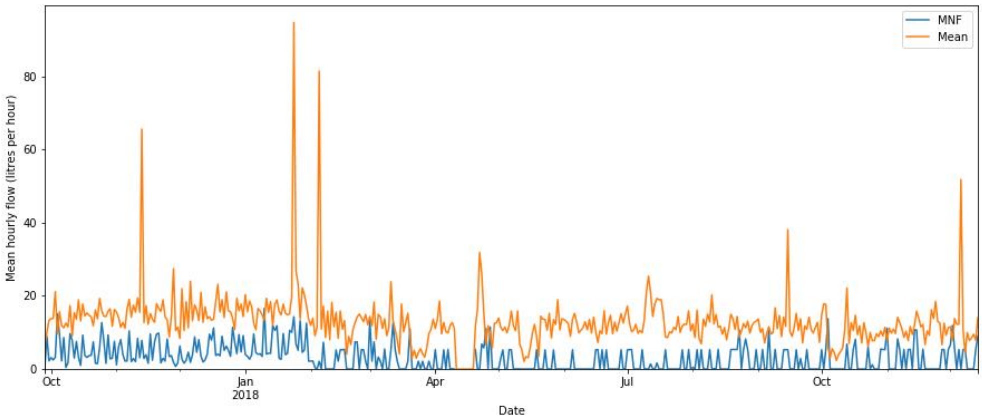

Figure 18 Mean flow and MNF over a long period in time, exhibiting elevated use during Spring and Summer.

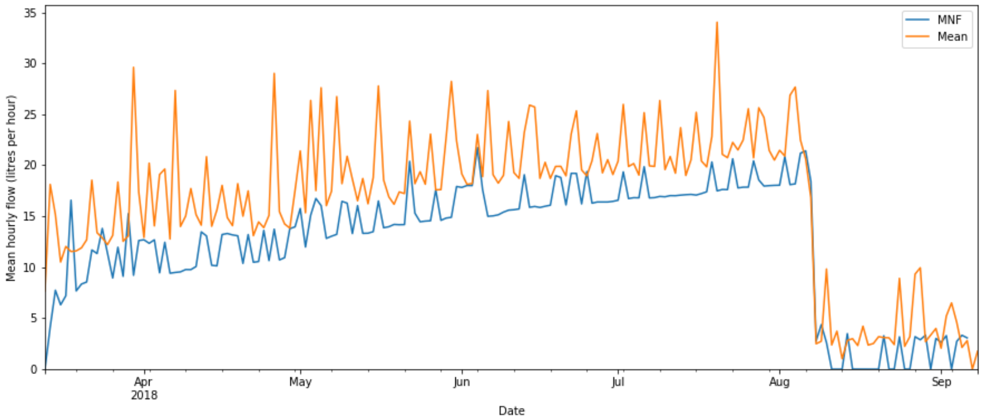

Figure 19 Mean flow and MNF over a long period in time, providing evidence of a possible progressive leak.

The disaggregation of end use types enables an assessment of the most common use types during the MNF period (2am to 4am). Most dwellings only contained contributions from toilet, tap and leaks/drips during this period. However, some dwellings contained significant contributions that were consistent with the characteristics of a shower, dishwasher, outdoor or high-flow event. Given that some dwellings contained very high MNF contributions, the pie chart in Figure 20 considers the relative, rather than absolute, contributions to water use within the MNF period for each dwelling. This avoids the end use types being dominated by a single use associated with the dwellings that had very large MNFs (such as the one illustrated in Figure 19). Figure 20 presents a more "mixed" use profile than would be expected for the MNF period. These results could be refined by implementing an upper limit on the MNF of dwellings included in this analysis, hence only including dwellings with "normal" MNF use. Other improvements relate to the general disaggregation of end use types, as discussed elsewhere.

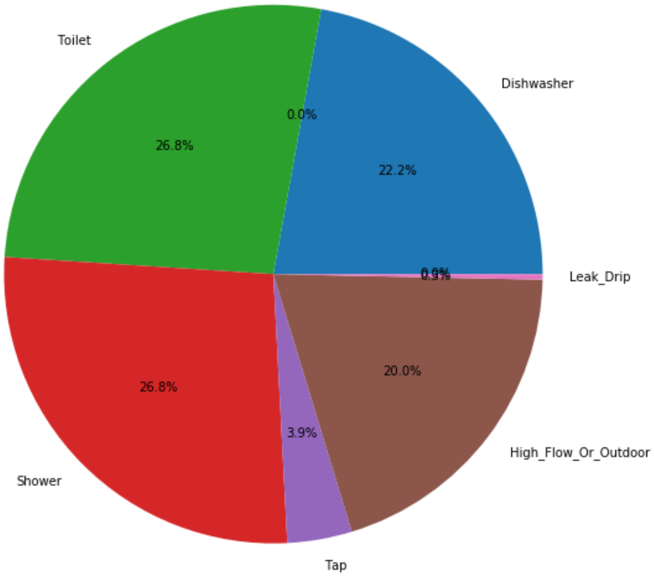

Figure 20 Relative contributions to MNF use, averaged over all dwellings.

4.2 Comparison of time series and event-based use disaggregation methods

The time series analysis described in Section 3.1 is a manual and time-consuming analysis method, hence this was only applied to a fortnight in one dwelling and to isolated days/weeks in other dwellings. The event-based analysis approach initially requires some manual checks on the most common water use patterns in each dwelling but can subsequently be scaled up to a longer period in an automated manner. The more detailed time series analysis therefore provides a good calibration of the event-based analysis. Figure 21 compares the two approaches for a 2-week period in Winter for Dwelling 25. Unsurprisingly, the agreement is very good. Low-flow events that may have been leaks were removed prior to the start of the time series analysis; however, these had a negligible contribution to the total water use within this fortnight. The dishwasher and washing machine contributions determined within the event-based analysis were slightly lower than those of the time series analysis, while the proportion of shower use was slightly higher. However, overall, the agreement between the two analysis methods is excellent, providing confidence in the application of the event-based disaggregation to all dwellings over all seasons. The time series analysis approach was also applied for isolated checks on the disaggregated use of several dwellings within the dataset.

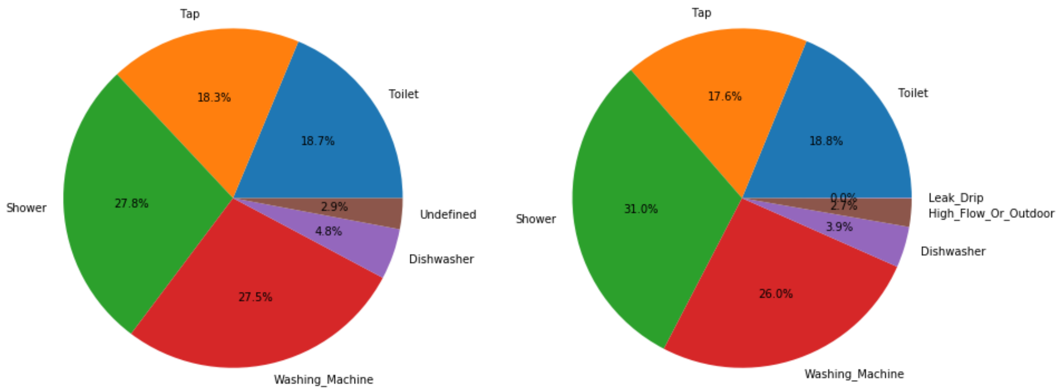

Figure 21 Comparison of time series analysis (left) with event-based analysis (right) in terms of the breakdown of water use within a Winter fortnight for Dwelling 25.

4.3 End use breakdown for event-based analysis

Figure 22 presents the end uses identified from the event-based analysis across all dwellings and seasons, while Figure 23 provides a seasonal breakdown of the end use disaggregation results. Both consider the relative contributions to use for each dwelling.

The results of this analysis are broadly comparable with the data from previous end use studies such as Heinrich (2007) and Roberti (2010), the results of which are illustrated in Figure 24 (these are also discussed in Pollard, 2022). Table 5 summarises the use breakdown for Summer and Winter. For example, leaks and drips accounted for approximately 2% of water use in the current study, which is similar to the 2-4% range from previous studies. However, there are some important differences which merit further investigation. The current study found much lower washing machine use (13%) than previous studies (20-25%). This may be due to increased uptake of front loader washing machines but may also be caused by limitations in the event-based identification process. These limitations could be improved by the provision of robust appliance signature data. As a result of the reduction washing machine use, several other end use types were increased relative to previous studies. These include tap, shower, and toilet events.

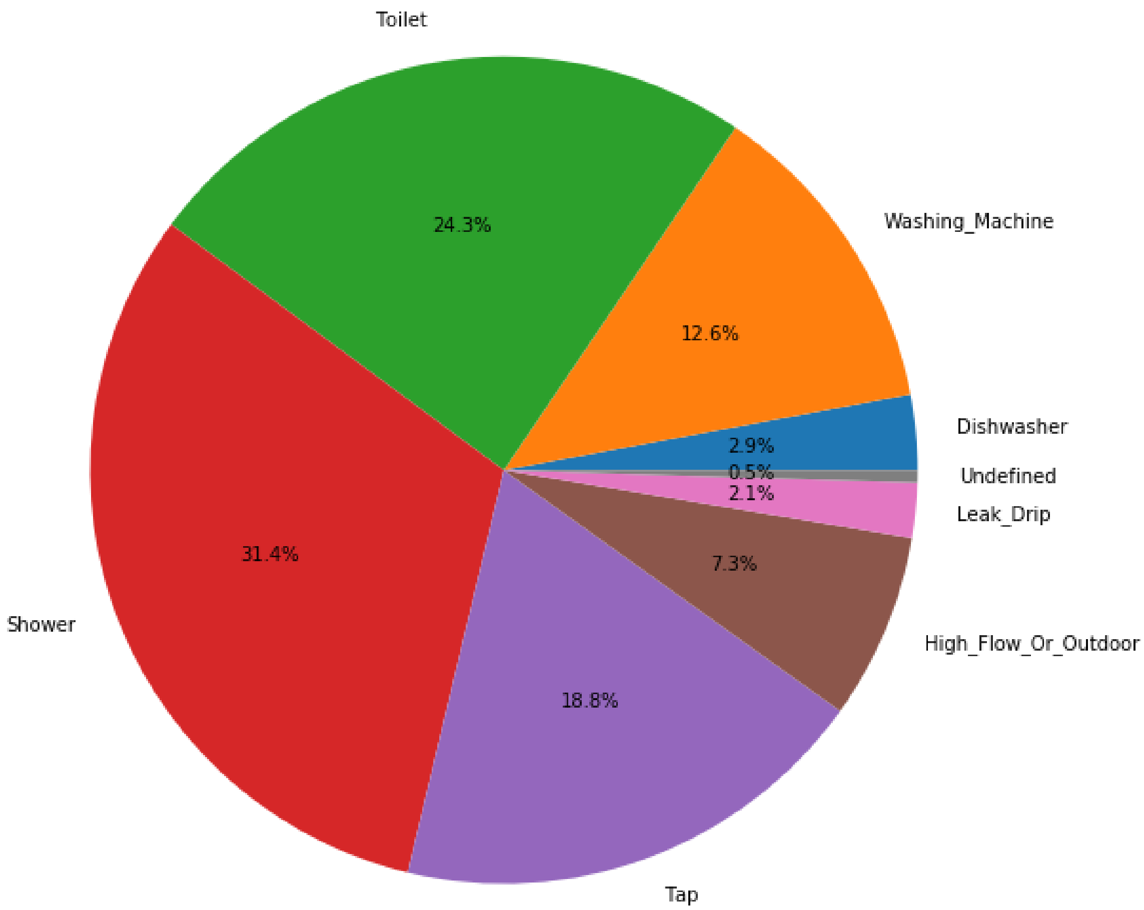

Figure 22 End uses identified from event-based analysis.

Some predictable seasonal trends are well captured by the event-based disaggregation approach, including the increase in outdoor use in Summer. Although leaks and drips were present throughout the dataset, these were largest in a relative sense during Winter. Although previous studies showed an increase in the proportion of shower use during Winter, results from the current study indicate that shower use was relatively consistent across the different seasons.

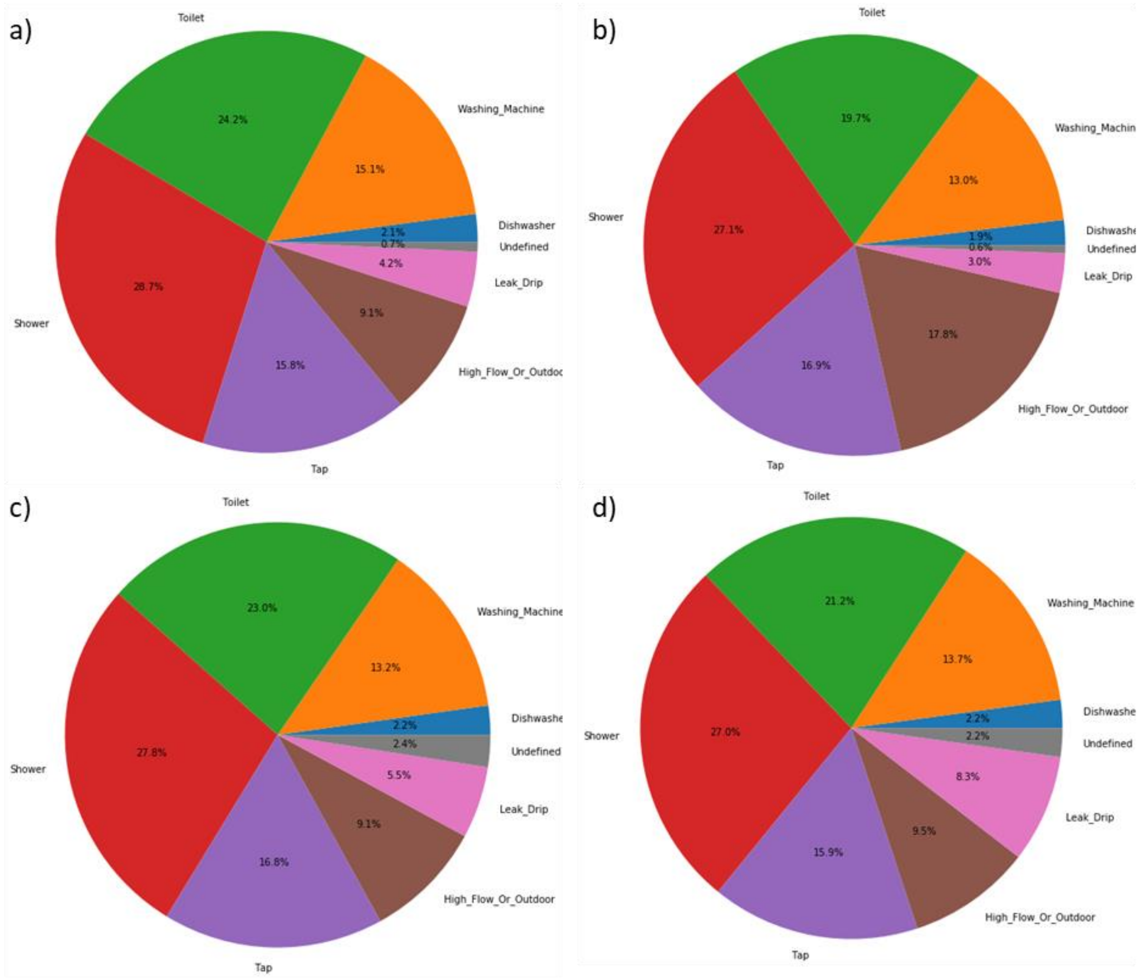

Figure 23 Seasonal variation in end use from the application of the event-based disaggregation approach to all dwellings. The seasons shown are a) Spring, b) Summer, c) Autumn, and d) Winter.

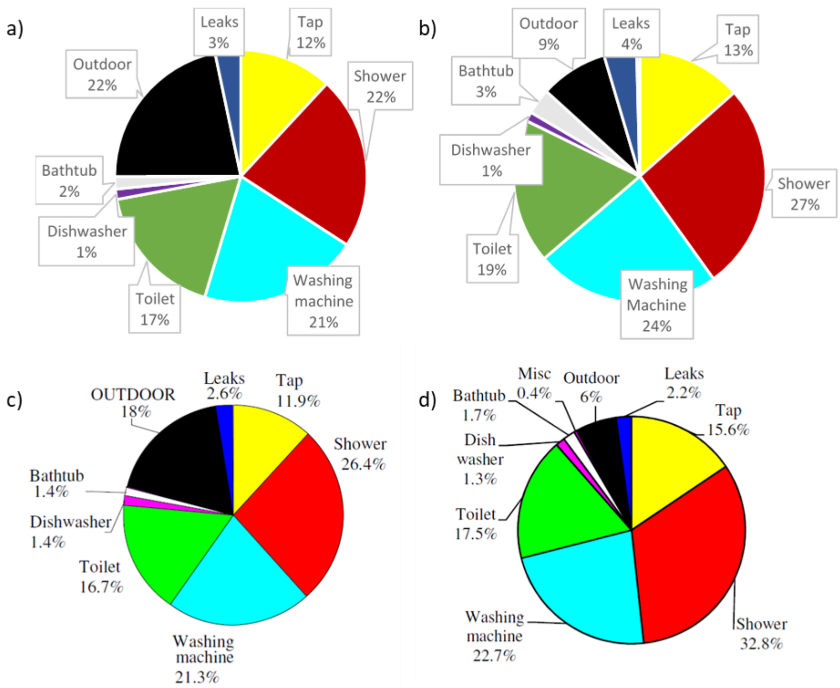

Figure 24 End use breakdown from a) Kapiti during the Summer (Heinrich, 2007), b) Kapiti during the Winter (Heinrich,

2007), c) Auckland during the Summer (Roberti, 2010), d) Auckland during the Winter (Roberti, 2010).

Table 5 Numerical comparison of current results with previous studies.

Use type	**Current study**	**Kapiti study (Heinrich, 2007)**	**Auckland study (Roberti, 2010)**
Winter			
Toilet	16%	19%	18%
Shower	23%	27%	33%
Tap	12%	13%	16%
Outdoor	15%	9%	6%
Leak/drip	10%	4%	2%
Washing machine	13%	24%	23%
Dishwasher	2%	1%	1%
Summer			
Toilet	18%	17%	17%
Shower	26%	22%	26%
Tap	14%	12%	12%
Outdoor	23%	22%	18%
Leak/drip	3%	3%	3%
Washing machine	14%	21%	21%
Dishwasher	1%	1%	1%

Table 6 provides more detailed information relating to the overall use breakdown of Figure 22 and the seasonal use breakdown of Figure 23. This table also provides the standard deviation in addition to the mean value, calculated across all eligible dwellings (not all dwellings had sufficient data in all of the seasons). As demonstrated elsewhere in this report, the relative use by the different dwellings exhibited significant variability. This may partly be explained by some dwellings not using a washing machine or dishwasher, or not having significant outdoor use. Others may have had significant outdoor use as a proportion of their total use, or a larger than average use of one of the other appliances. Although leaks and drips had a relatively small relative contribution to average water use, these also exhibited significant variability, indicating that some dwellings may have had relatively large water losses.

Table 6 Mean, median and standard deviation associated with the water use identified using the event-based analysis.

Use type	Mean value	Standard deviation
Entire dataset		
Dishwasher	2.9%	4.4%
Washing Machine	12.6%	6.0%
Toilet	24.3%	13.0%
Shower	31.4%	12.4%
Tap	18.8%	11.4%
High flow or outdoor	7.3%	7.7%
Leak/Drip	2.1%	4.5%
Undefined	0.5%	0.6%
Spring		
Dishwasher	2.1%	3.5%
Washing Machine	15.1%	7.5%
Toilet	24.2%	11.0%
Shower	28.7%	12.3%
Tap	15.8%	10.8%
High flow or outdoor	9.1%	9.3%
Leak/Drip	4.2%	8.0%
Undefined	0.7%	1.1%
Summer		
Dishwasher	1.9%	3.1%
Washing Machine	13.0%	7.8%
Toilet	19.7%	11.4%
Shower	27.1%	12.3%
Tap	16.9%	13.2%
High flow or outdoor	17.8%	18.5%
Leak/Drip	3.0%	9.9%
Undefined	0.6%	1.3%
Autumn		
Dishwasher	2.2%	3.2%
Washing Machine	13.2%	8.3%
Toilet	23.0%	11.7%
Shower	27.8%	11.3%
Tap	16.8%	11.5%
High flow or outdoor	9.1%	11.6%
Leak/Drip	5.5%	10.9%
Undefined	2.4%	11.9%
Winter		
Dishwasher	2.2%	3.5%
Washing Machine	13.7%	7.9%
Toilet	21.2%	12.9%
Shower	27.0%	13.1%
Tap	15.9%	12.4%
High flow or outdoor	9.5%	14.0%
Dishwasher	8.3%	16.3%
Undefined	2.2%	10.2%

Although the pie charts in Figure 22 and 23 show the relative contributions of the different use types to the overall and seasonal water use, they do not capture the changing total water use during the year or within an average day. Figure 25 shows the variation in water use over the year, averaged over all dwellings. It should be noted that few of the dwellings contained a complete dataset over a full year; hence, the results are averaged over different dwellings' use in different months. A more complete dataset would help to provide more reliable results. However, the general patterns of enhanced consumption during the Spring and particularly Summer, with the lowest consumption in Winter, are well captured. The increase in outdoor use in Spring and Summer is also to be expected, as is the increase in absolute volumes used for showers in these months (even if the relative contribution of shower events remained unchanged).

Table 7 provides a seasonal summary of the data shown in Figure 25, including the average across the entire dataset. It is important to note that the daily resampled data below were calculated after removing the effects of any leaks and drips, such that these are not included. It is also clear that the high average outdoor water use during the summer months was due to very high use in some dwellings, leading to very large standard deviations for this use type. As noted elsewhere in this report, the variability was very high across the dataset, even during Winter when the water use was relatively low.

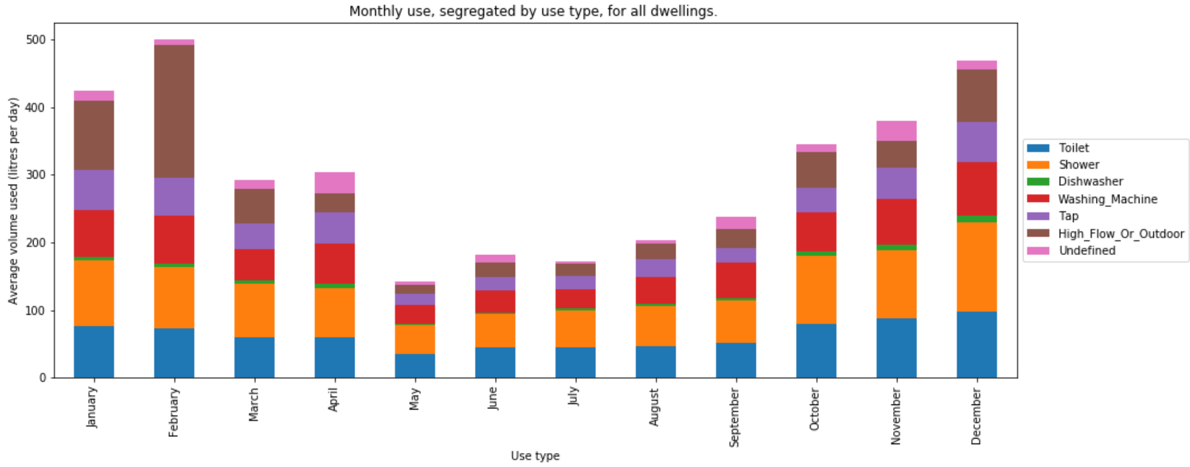

Figure 25 Annual variation in water use, separated by use type.

Table 7 Summary of daily volumes associated with each use type, reporting mean and mean values along with standard deviations for the entire dataset and separated by season.

Use type	Mean daily volume (litres)	Median daily volume (litres)	Standard Deviation (litres)
Entire dataset			
Dishwasher	5.1	3.5	5.8
Washing Machine	52.5	35.5	61.6
Toilet	63.4	39.6	62.8
Shower	78.3	52.6	76.7
Tap	37.0	20.6	48.0
High Flow/Outdoor	54.2	13.5	96.4
Undefined	13.7	5.2	25.7

Table 7 cont. Summary of daily volumes associated with each use type, reporting mean and mean values along with standard deviations for the entire dataset and separated by season.

Spring			
Dishwasher	5.8	3.9	7.2
Washing Machine	59.3	40.7	59.3
Toilet	73.0	51.6	63.1
Shower	88.1	65.2	74.6
Tap	34.9	23.0	40.3
High Flow/Outdoor	39.9	14.7	68.6
Undefined	19.9	8.9	40.9
Summer			
Dishwasher	6.8	5.4	7.1
Washing Machine	73.1	52.8	84.4
Toilet	82.8	60.7	69.9
Shower	106.0	82.9	92.7
Tap	57.7	34.1	77.9
High Flow/Outdoor	125.4	28.5	217.4
Undefined	12.0	5.7	20.4
Autumn			
Dishwasher	4.5	2.6	5.2
Washing Machine	44.4	33.2	50.2
Toilet	51.6	31.9	55.5
Shower	65.0	36.6	70.6
Tap	33.1	15.3	40.7
High Flow/Outdoor	31.1	7.8	60.5
Undefined	16.2	3.8	32.4
Winter			
Dishwasher	3.1	1.9	3.9
Washing Machine	33.0	15.3	52.4
Toilet	46.0	14.2	62.7
Shower	54.0	25.5	69.0
Tap	22.5	10.1	33.1
High Flow/Outdoor	20.2	2.9	39.0
Undefined	6.8	2.7	9.2

Considering the variations in water use over a day, Figure 26 presents the daily water use profiles (diurnal curves) showing the breakdown by use type. Although differences in water use between households tend to “smooth” these results, the results appear to conform to expectations. There is a peak in overall water use, particularly due to showers, in the morning, while any high flow or outdoor use tends to be concentrated in the afternoon. Figure 27 shows the higher peaks (with more prominent shower events) during weekdays, while Figure 28 shows the smaller peaks and more consistent hourly use during weekends.

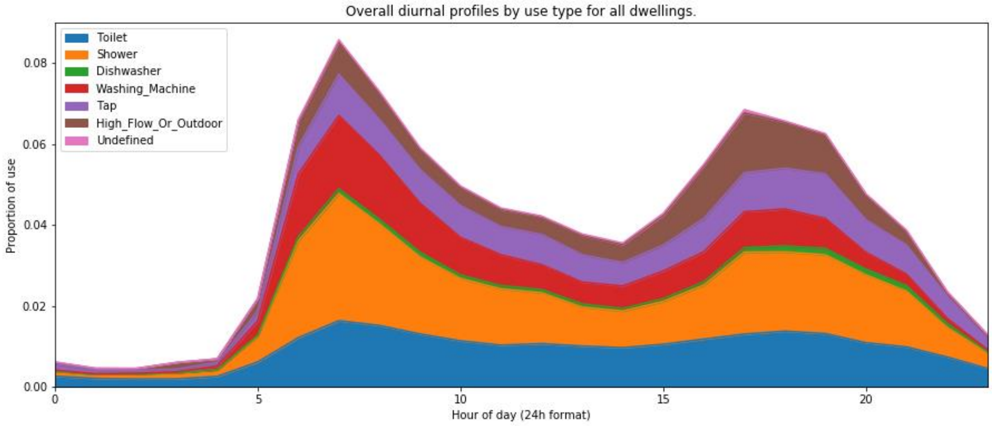

Figure 26 Hourly fluctuations in water use, broken down by use type and averaged across all dwellings.

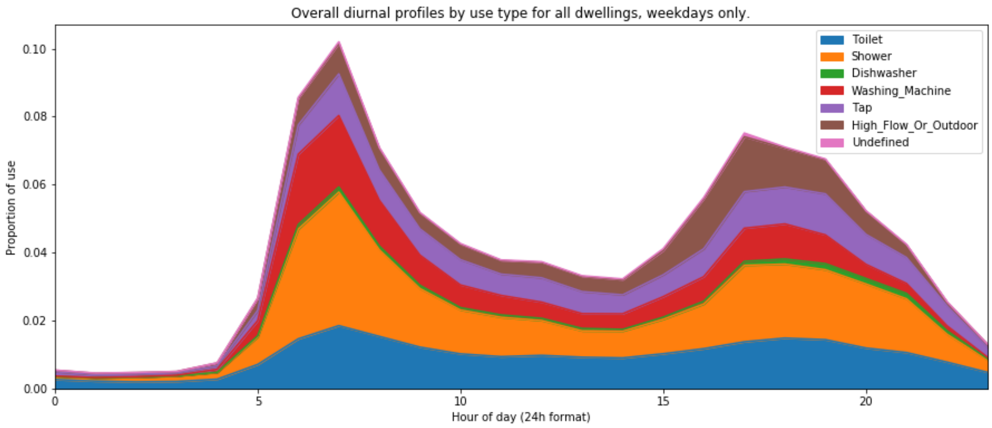

Figure 27 Hourly fluctuations in water use during weekdays, broken down by use type and averaged across all dwellings.

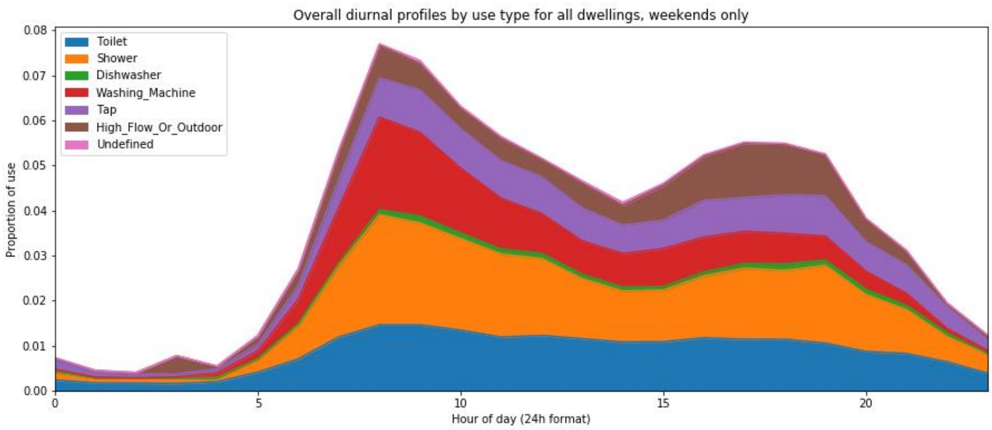

Figure 28 Hourly fluctuations in water use during weekends, broken down by use type and averaged across all dwellings.

4.4 Frequency of use for event-based analysis

The temporal resolution of the dataset allows calculation of the use frequencies associated with the different appliances identified during the disaggregation process. These use frequencies are summarised in Table 8. Although most use frequencies were calculated daily, the washing machine and dishwasher use frequencies were calculated weekly, given the less frequent typical use of these appliances. It should also be noted that all frequencies are reported on a per-dwelling (rather than a per-person) basis. All use types exhibit large standard deviations, again demonstrating the significant variability in the dataset and the likely value in more targeted smart meter investigations in the future. In some cases (shower, washing machine and high flow or outdoor use), large outliers in the dataset contributed to these large standard deviations; these outliers also increased the average use frequency compared to the median use frequency. These may indicate some incorrect classifications arising from the event-based disaggregation. However, given the data quality and quantity issues noted elsewhere in this report and lacking independent data on appliance signatures, a small amount of incorrect classification is somewhat unavoidable. Although these data could also have been broken down by season, for instance to investigate changes in the frequency of (say) showers or outdoor use between seasons (particularly Summer and Winter), this was not undertaken in the current project.

The tap use frequencies reported in Table 8 are very high, in part because many events that did not meet the criteria for leaks and drips (but may not have resulted from tap use) were categorised as tap use by default. Given the overlap in tap and leak/drip signatures, it is challenging to completely separate the two use types.

Table 8 Frequency of use derived from event-based analysis.

	Average use frequency	Median use frequency	Standard deviation
Toilet	14.7 per day	10.7 per day	10.3 per day
Shower	2.0 per day	1.4 per day	2.6 per day
Tap (can include leaks)	118.8 per day	106.4 per day	75.1 per day
High flow or outdoor	2.8 per day	2.9 per day	3.4 per day
Washing machine	3.4 per week	2.2 per week	3.5 per week
Dishwasher	4.2 per week	2.9 per week	3.4 per week

4.5 Support Vector Machines

Confusion matrices and classification reports are standard approaches for assessing the performance of supervised learning techniques. In the confusion matrices shown in Table 4, the diagonals (shaded for ease of interpretation) represent the number of correctly classified events of each kind, while any off-diagonal entries represent incorrectly classified events. Although the standard regularisation parameters led to perfect classification of some event types, they also generated far more incorrect end-use classifications than the optimised parameters determined using a grid search. As the classification report in Table 5 demonstrates, using standard values to assess the effectiveness of supervised learning techniques, the optimised parameters provided a significant improvement in performance for end-use classification (consistent with Gourmelon et al., 2021). Although the results should be interpreted with caution, given that the training and test data were generated themselves using an event-based disaggregation (rather than being supplied directly as part of the dataset), the performance of the SVM model is nonetheless impressive. This supervised learning approach also offers an advantage over unsupervised learning approaches, given the unbalanced nature of the dataset (where some use types dominate in terms of the number of events). Figure 29 compares the end use breakdown for the test dataset as determined by the SVM models to the original use labels.

Table 9 Confusion matrices for SVM end-use disaggregation for Dwelling 25, using standard parameters (1.0 for regularisation parameters), and using parameters optimised through a grid search.

Standard parameters					
200	0	0	0	0	0
1	7	8	108	133	0
1	0	291	46	14	0
53	0	35	28259	7	0
0	0	0	773	1007	0
0	0	0	0	0	32
Optimised parameters through grid search					
199	0	0	1	0	0
0	224	8	17	8	0
0	3	349	0	0	0
0	11	0	28341	2	0
0	2	0	6	1772	0
0	0	0	3	0	29

Table 10 Classification report for SVM end-use disaggregation for Dwelling 25.

Standard parameters				
	precision	recall	f1-score	support
Dishwasher	0.78	1	0.88	200
High_Flow_Or_Outdoor	1	0.03	0.05	257
Shower	0.87	0.83	0.85	352
Tap	0.97	1	0.98	28354
Toilet	0.87	0.57	0.68	1780
Washing_Machine	1	1	1	32
accuracy			0.96	30975
macro avg	0.92	0.74	0.74	30975
weighted avg	0.96	0.96	0.96	30975
Optimised parameters through grid search				
	precision	recall	f1-score	support
Dishwasher	1	0.99	1	200
High_Flow_Or_Outdoor	0.93	0.87	0.9	257
Shower	0.98	0.99	0.98	352
Tap	1	1	1	28354
Toilet	0.99	1	0.99	1780
Washing_Machine	1	0.91	0.95	32
accuracy			1	30975
macro avg	0.98	0.96	0.97	30975
weighted avg	1	1	1	30975

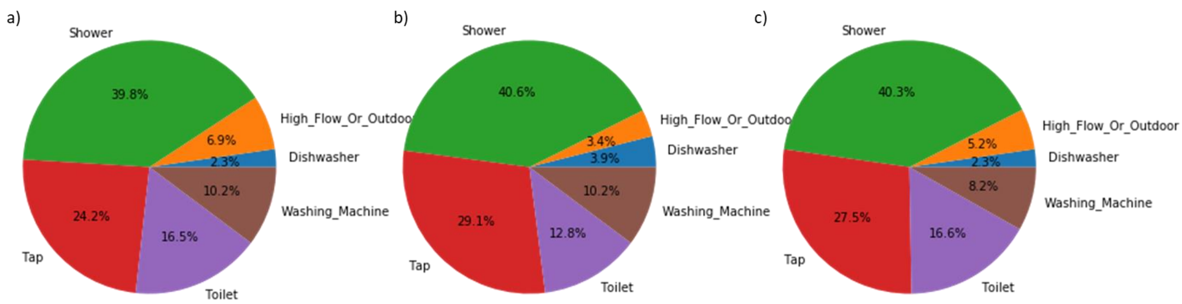

Figure 29 Contributions of different appliances to total residential water use for Dwelling 25, for use types classified by a) the event-based analysis (used to train the SVM model), b) the standard SVM model using regularisation parameters of 1.0, and c) the SVM model using parameters optimised through a grid search.

5 Opportunities for further research

This report has completed the classification of residential water use into standard appliance types using the dataset of Pollard (2022). Three approaches were followed:

- Time series analysis
- Event-based analysis
- Support Vector Machines supervised learning classification

The three approaches were used in a complementary manner within this project, with the time series analysis used to calibrate and check the results of the event-based analysis, which in turn provided training and testing data for the SVM classification model. However, this approach is relatively limited, lacking a "source of truth" for the use types. With the recent proliferation of smart devices, particularly smart water meters, opportunities exist to engage end-users directly in the provision of labelled water use data. The outsourcing of the classification to the end user may be difficult from the perspective of generating reliable data over long periods. However, the significantly larger sample size, as well as the significant reduction in effort in post-labelling of events, may offset this issue.

In addition to assisting in the labelling of events, engaging end users with smart technology may have additional benefits. Many electricity consumers now receive regular personalised summaries of their electricity use, along with suggestions for reducing their energy bill. A similar approach applied to water use would promote conservation through awareness of actual consumption patterns, while also providing rapid data that could determine the presence of a large leak (saving water for the consumer and provider alike).

This report has barely scratched the surface of the opportunities available when applying data analysis to residential water use. It is therefore hoped that this will help to start (or continue) discussions on this topic, rather than being taken as the last word.

References

Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366a), 427-431.
Gourmelon, N., Bayer, S., Mayle, M., Bach, G., Bebber, C., Munck, C., Sosna, C., and Maier, A. (2021). Implications of experiment set-ups for residential water end-use classification. Water, 13(2):236.
Heinrich, M. (2007). Water end use and efficiency project (WEEP) - final report. BRANZ Study Report 159, BRANZ, Judgeford, Porirua.
Pollard, A. 2022. Residential Water Use in New Zealand, BRANZ Study Report 469, BRANZ, Judgeford, Porirua
Roberti, J. (2010). Auckland water use study - monitoring of residential water end uses. In Proceedings of SB10, New Zealand Sustainable Building Conference, number 51. Published as BRANZ Conference Paper 145.

Appendix A: Diurnal curves

Peaking factors

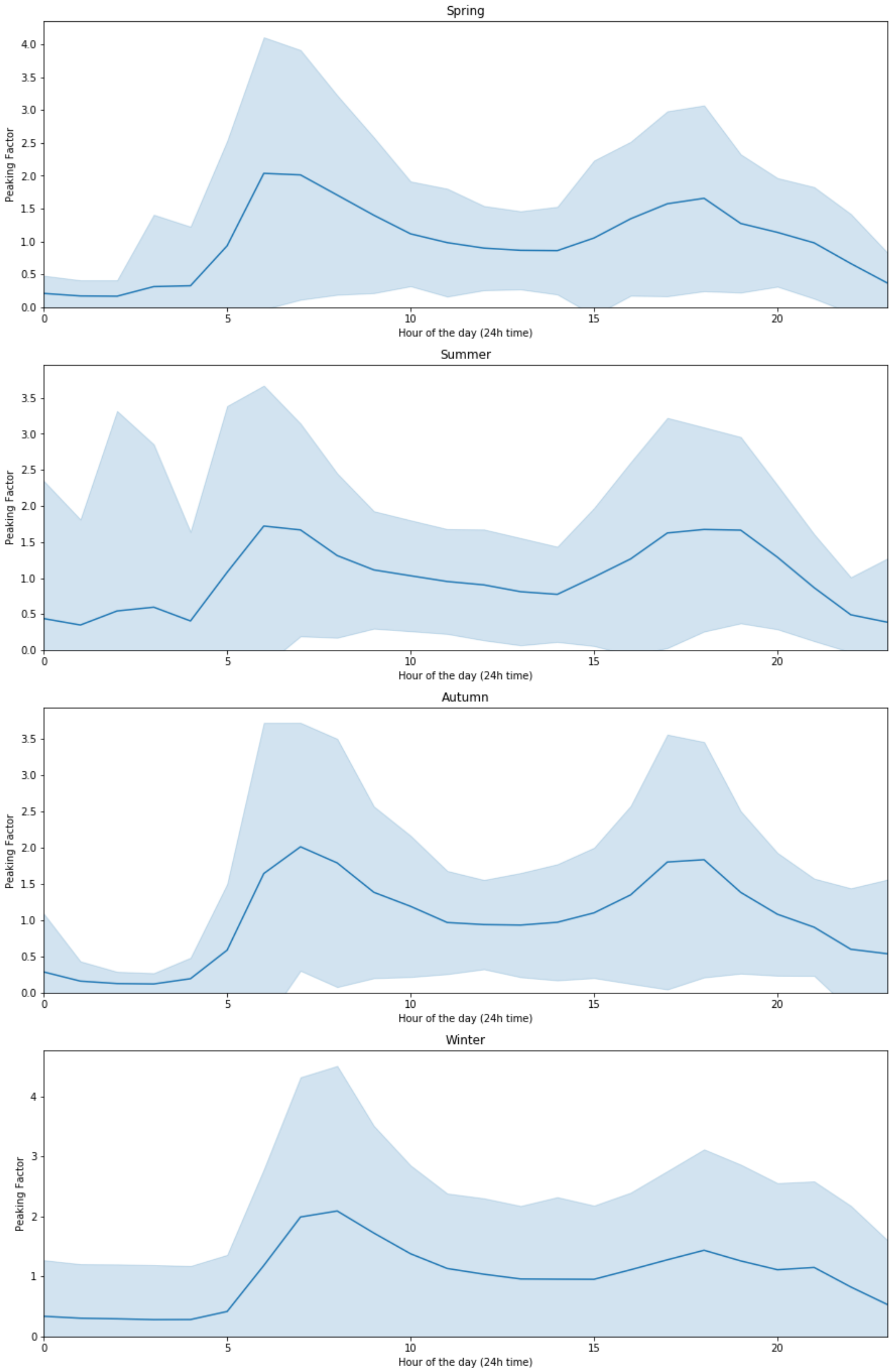

Figure 30 Seasonal variation in peaking factor diurnal curves, where the shaded area represents one standard deviation.

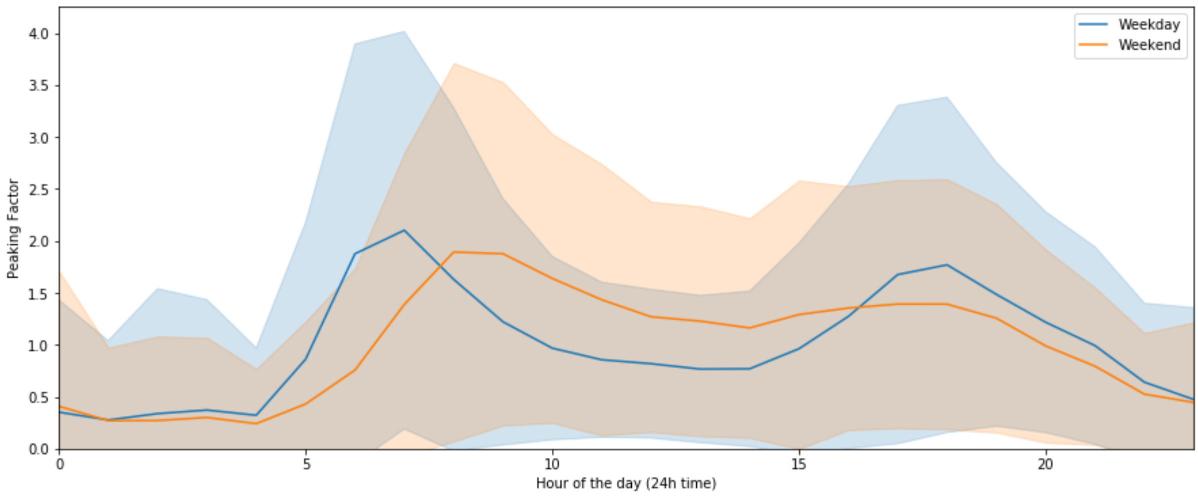

Figure 31 Weekday/weekend variation in peaking factor diurnal curves, where the shaded area represents one standard deviation.

Flowrates per dwelling

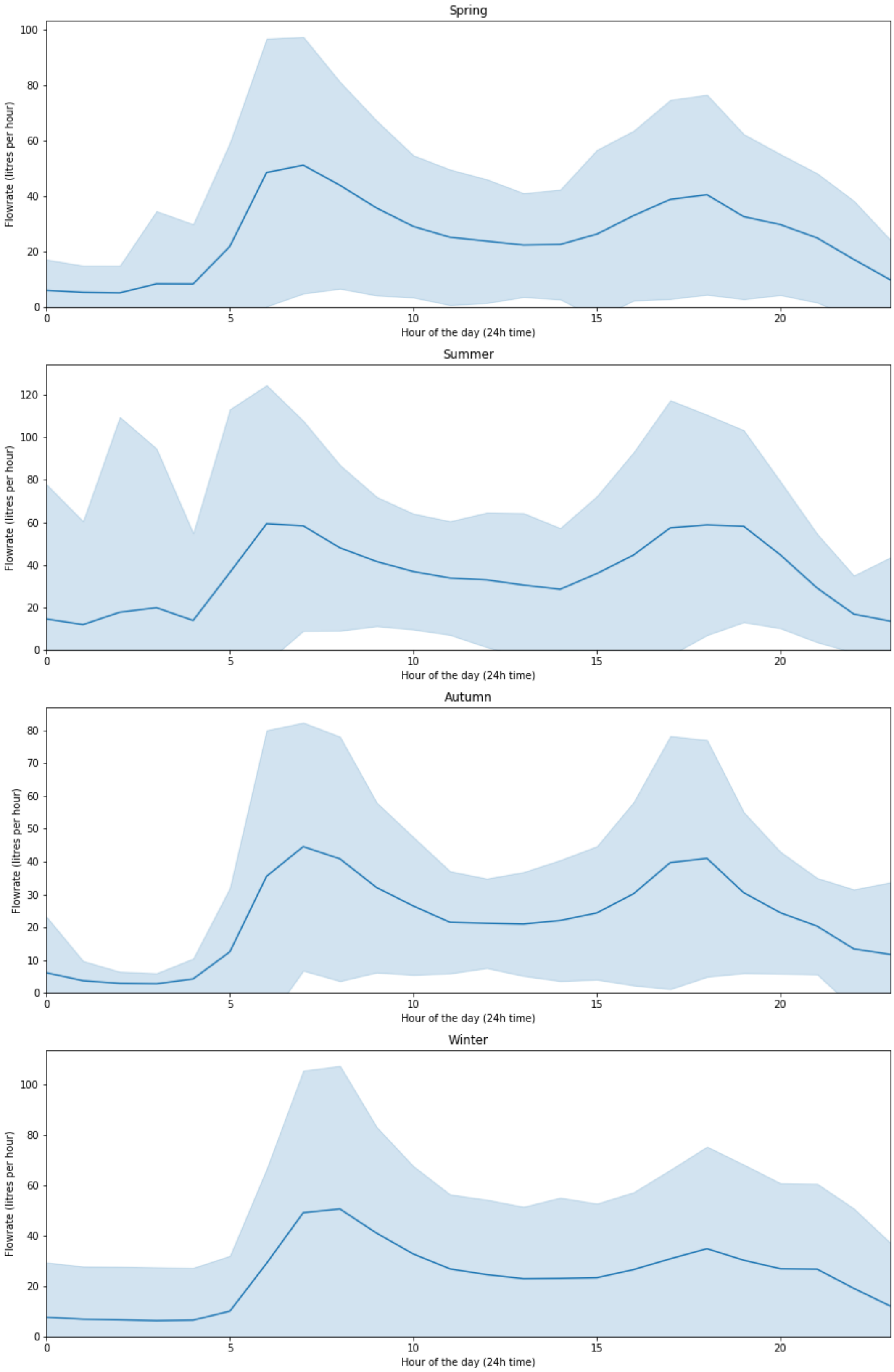

Figure 32 Seasonal variation in hourly volume diurnal curves, where the shaded area represents one standard deviation.

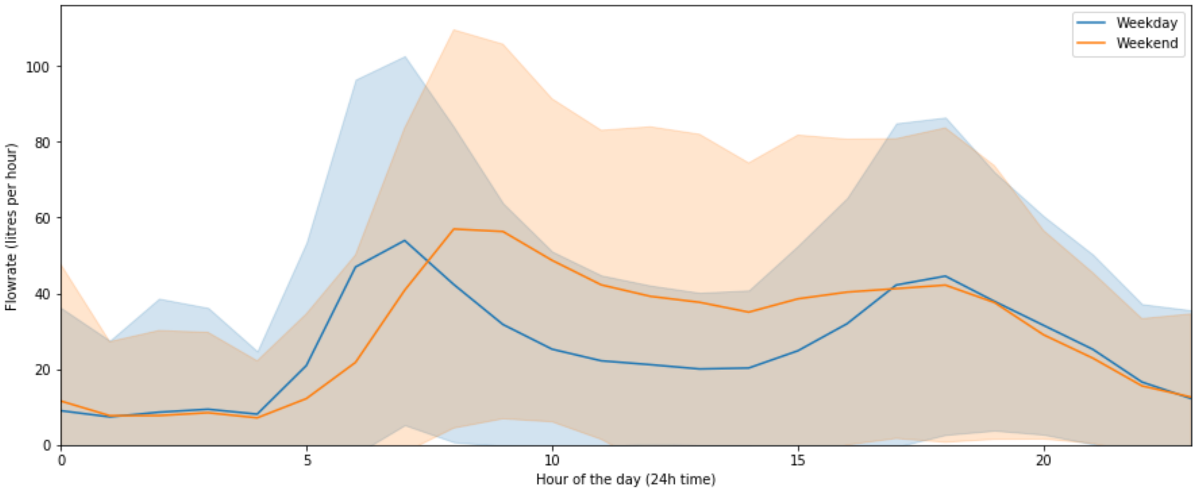

Figure 33 Weekday/weekend variation in hourly volume diurnal curves, where the shaded area represents one standard deviation.

Flowrates per person

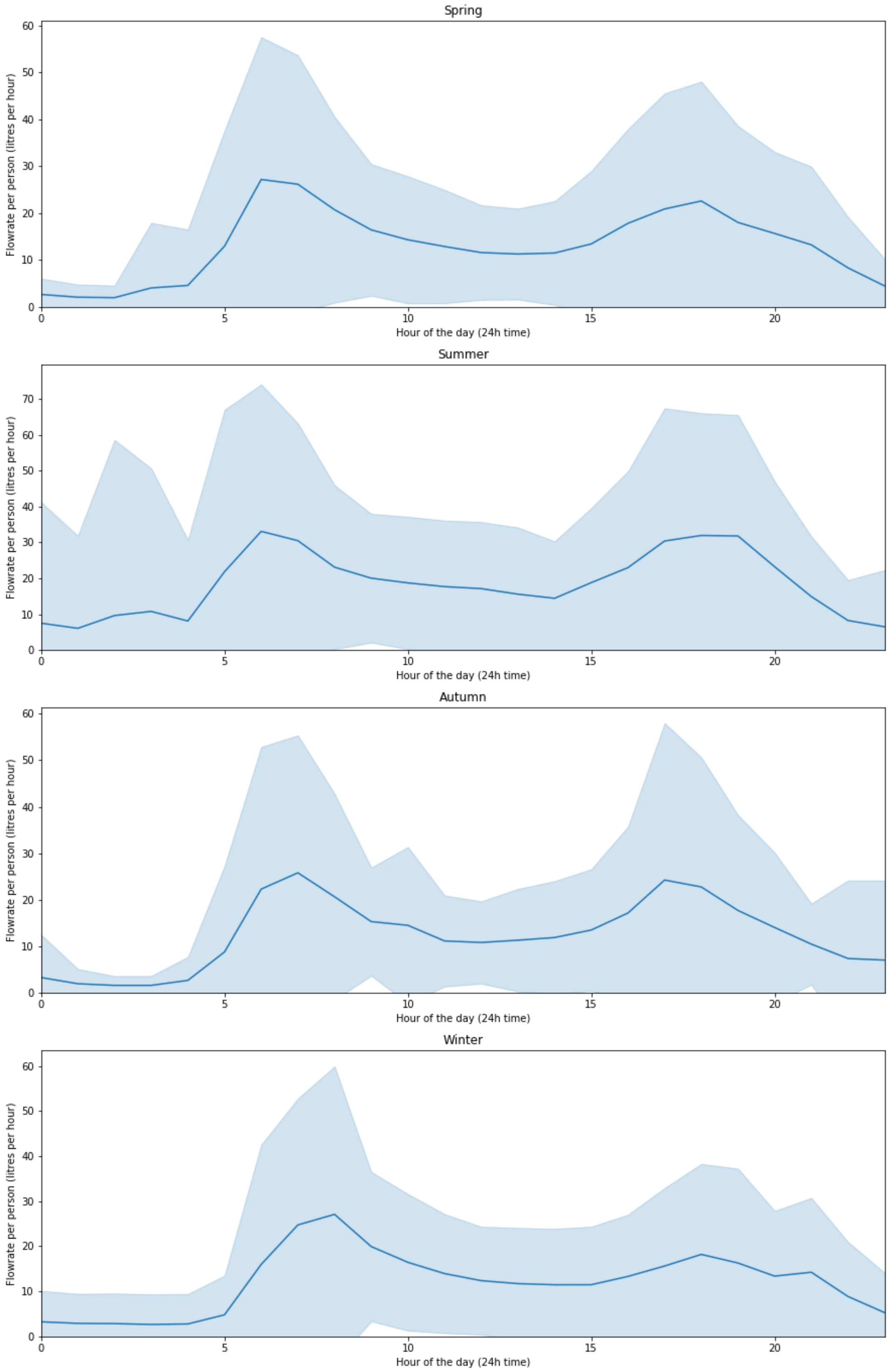

Figure 34 Seasonal variation in per-person hourly volume diurnal curves, where the shaded area represents one standard deviation.

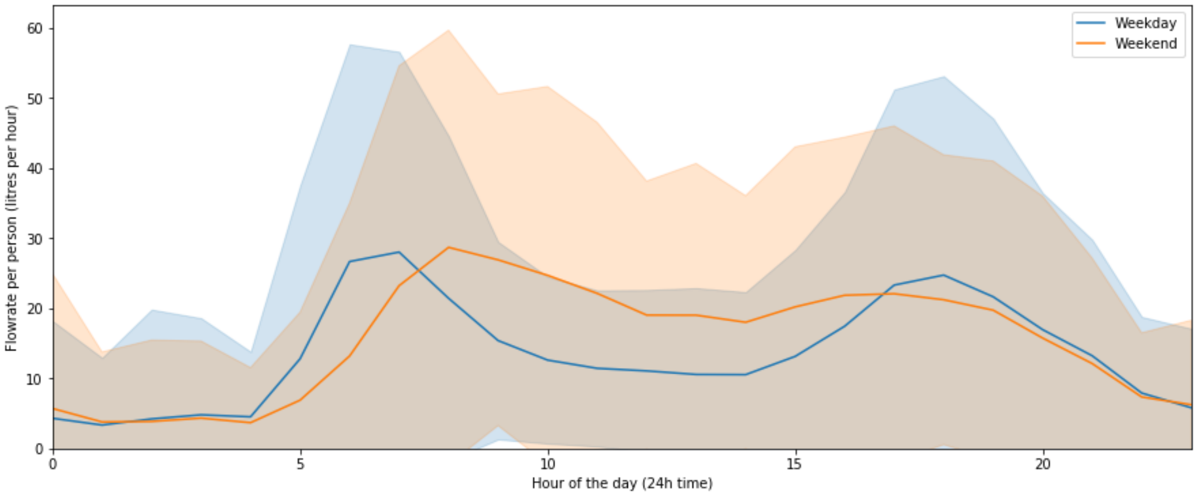

Figure 35 Weekday/weekend variation in per-person hourly volume diurnal curves, where the shaded area represents one standard deviation.

Appendix B: Leak visualisation using heatmap

Although not included in the preceding report, another way to enable leak detection (courtesy of Andrew Pollard) is a "heat map" plot of water use over time, where the data are plotted against the hour of the day. Figure 36 provides an example of this for Dwelling 35. This example would enable the identification of relative large leaks, although very small drips would be unlikely to be detected.

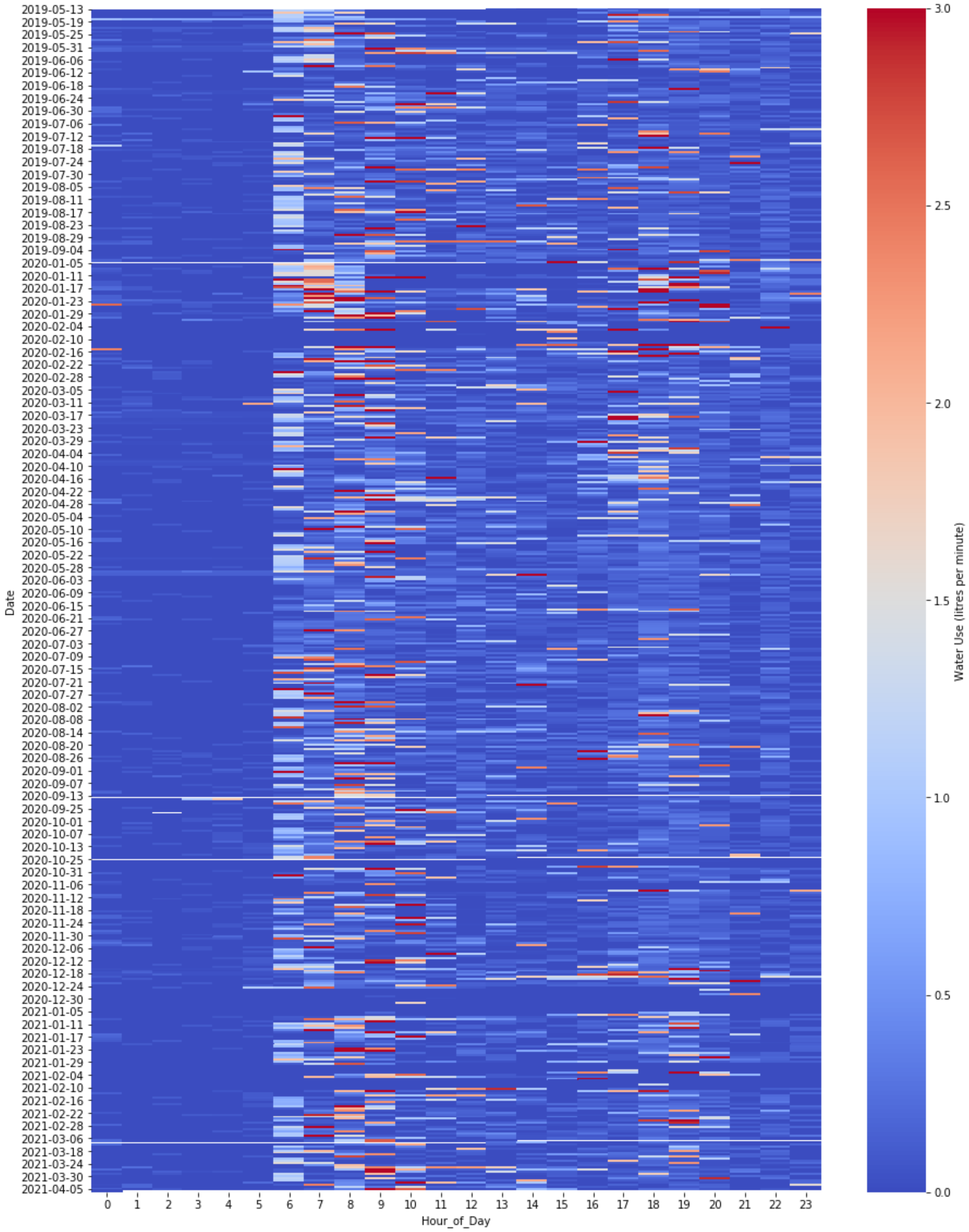

Figure 36 Heatmap plot to enable the identification of leaks within the water use dataset, where missing values are white.

Although the colour scale used in Figure 36 could be modified to include a reduced upper limit for the identification of very small leaks or drips, another approach could be to plot the heatmap data using a logarithmic colour scale. Figure 37 shows an example of a logarithmic colour scale, with the same upper limit as used in Figure 36. This improves the resolution of very small flow events, although any “zero use” values must be replaced with a small but non-zero value or removed entirely. Zero values have been filled with white (the same as missing values) in Figure 37.

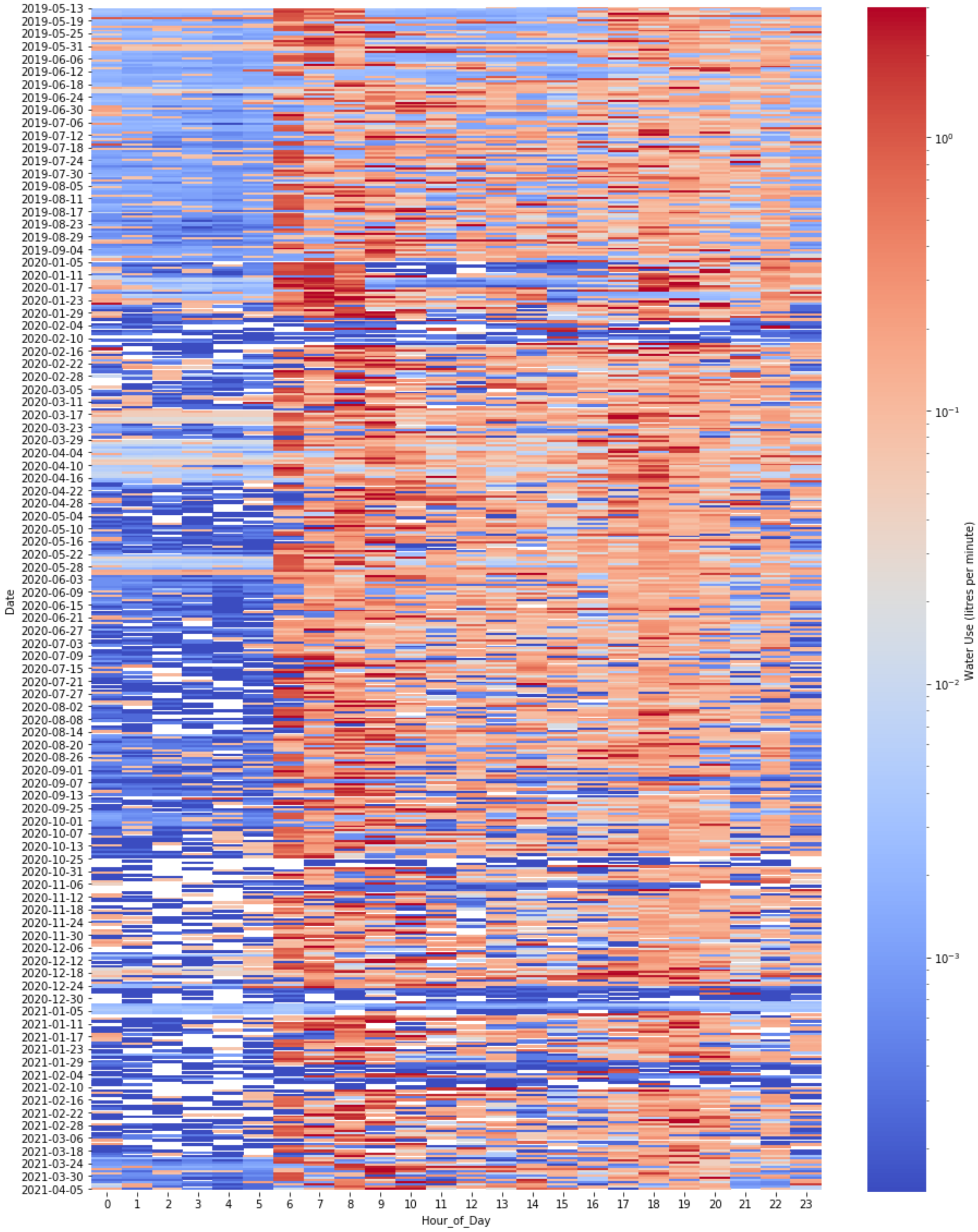

Figure 37 Heatmap plot to enable the identification of leaks within the water use dataset, using a logarithmic colour scale to facilitate easier identification of small leaks or drips. All zero or missing values are denoted by a white colour.