

# A probabilistic hot water usage model and simulator for use in residential energy management

M.J. Ritchie <sup>a</sup>, J.A.A. Engelbrecht <sup>a</sup>, M.J. Booysen <sup>a,\*</sup>

<sup>a</sup>*Department of E&E Engineering, Stellenbosch University, South Africa*

---

## Abstract

Water heating contributes up to 40% of a household's total electricity usage and places a substantial burden on the electricity grid due to high power ratings and users' largely simultaneous hot water usage. The main determinants of its electricity draw are physical properties such as set temperature, insulation, and plumbing configuration; environmental conditions such as ambient temperature and inlet temperature; and the hot water usage profiles. These profiles include the usage volumes, the times of usage and the outlet temperatures. The efficacy of energy management techniques that model water heaters and the accuracy of their simulation results therefore rely on representative hot water usage profiles. Existing models for household hot water usage neglect differences between users, and temporal variations such as the season and the day of the week, and are not fully autonomous. We propose a probabilistic data-driven model for modelling individualised hot water profiles and an accompanying hot water usage simulator that includes all these factors. We gathered data from 77 residential households over a period of one year to train and evaluate the model for all four seasons. The results show that the simulated hot water usage profiles match the statistical properties of the measured data. Moreover, the individual hot water usage modelling and the resulting aggregated energy load on the grid closely match the measured data, improving on the existing hot water usage by halving the modelling error.

*Keywords:* Residential electricity demand modelling; Domestic hot water; Hot water usage modelling; Water heater simulation; Demand-side management; Hot water usage model; Probabilistic modelling

---

## 1. Introduction

Electric water heaters contribute up to 40% of a household's total energy usage (Hohne et al., 2019). In South Africa, which relies almost exclusively on electricity for energy, the approximately five million water heaters contribute an estimated 30% to 50% of the grid load at peak times, and 7% of the country's daily energy requirements (Hohne et al., 2019; Skinner et al., 2012). Moreover, since 88% of the country's electricity is generated from burning coal, the environmental water heating footprint is substantial (Mathu, 2017).

Fortunately water heaters with tanks are reasonably good at storing thermal energy, which makes them well-suited to demand-side management – an approach that reduces the impact of heavy loads on a constrained grid. However, besides managing demand from water heaters, effective demand-side management strategies should balance two additional, competing requirements: managing the total energy used and ensuring user satisfaction (Roux et al., 2018). To achieve this balance requires large-scale simulation of water heaters using accurate thermal models. The main determinants of such a model are environmental (e.g. inlet water temperature and ambient temperature), the physical properties of the water heater (e.g. thermal losses specification, element rating, volume), installation and configuration parameters (e.g. plumbing setup, retrofitted insulation blankets, orientation), and the hot water usage profile (also called the 'draw profile' or simply the 'water profile') (Matos et al., 2019; Nel et al., 2017; Kepplinger et al., 2015; Diao et al., 2012).

After a demand-side-management-ready smart water heater is installed, the main determinant of variable use is the user's water profile, which directly affects the energy profile (Nel et al., 2017). The accuracy of

---

\*Corresponding author

Email address: [mjbooyesen@sun.ac.za](mailto:mjbooyesen@sun.ac.za) (M.J. Booysen )

a water usage profile has a substantial and direct influence on the accuracy of modelling a tanked water heater (Dongellini et al., 2015; Spur et al., 2006). For example, Wang et al. (2007) modelled and simulated a gas-powered water storage tank and found deviations of between 8% and 15% for different water profiles. Domestic hot water and the resulting energy profiles typically have two peaks daily, one in the morning and one in the afternoon (Jack et al., 2018; Roux and Booyesen, 2017; Forlee, 1998; ASHRAE, HVAC, 2003).

The hot water draw model that is possibly the most frequently used, the ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers) standard, has two versions. The first, and also the older and more widely used of the two, comes from the ASHRAE Standard 90.2 (ASHRAE, 1993) which does not specify the source of data. The second, from the HVAC Applications Handbook, 2003, has data collected from Canadian residences and provides a profile for a typical household with two adults, two children, a clothes washer and a dishwasher (Fairey and Parker, 2004). The ASHRAE standard has been used to perform large simulations of grids comprising water heaters, but it does not differentiate between users. Moreover, it uses a coarse time segmentation of one hour and does not consider weekly and seasonal variations.

Individual differentiation and characterisation of water profiles reduces the uncertainties for predictive analyses and removes assumptions based on over-sizing and generalisation (Iglesias and Palensky, 2013). Forecasting household water usage patterns for individual heaters in demand-side management strategies could improve the reliability of the response to energy surpluses and shortages (Gelažanskas and Gamage, 2015). Almost all of the existing research uses individual measurements of point-of-use devices in a household to determine the end-use cases, which are used to develop the household’s total water profile. Such models are generally complex and expensive; a cheaper alternative is load modelling (Widén et al., 2009).

These models provide individual differentiation for all water used, but a household’s usage varies over time: weekly patterns vary for days of the week and annual patterns vary for seasons, regions and climate (Booyesen et al., 2019; Roux and Booyesen, 2017). The day of the week and the season have distinct effects on the water demand, the significant factors being temperature and precipitation (Gerin et al., 2014; Gato et al., 2007; Zhou et al., 2002). Other factors to consider are holidays and once-off special events (Zhou et al., 2002). Accordingly, when assessing energy-saving initiatives it is important to characterise heater profiles accurately (Bertrand et al., 2017).

### 1.1. Modelling water profiles

Given the importance of getting accurate usage profiles, various models have been proposed in the literature. The characteristics of the main models are discussed here and summarised in Table 1.

Table 1: Work and models for simulated water usage profiles. An S in the Day column indicates that days were differentiated only by workdays and weekends and a tick indicates that the differentiation extended to the day of the week. A dash indicates that the source didn’t specify the information. The Unique column indicates if the model accounted for household individuality.

Ref	Category	Hot water	Season	Day	Unique	Res.
ASHRAE (1993)	Technical standards	✓	-	-	✗	1 hr
ASHRAE, HVAC (2003)	Technical standards	✓	-	-	✗	1 hr
Edwards et al. (2015)	Data collected	✓	-	-	✓	5 min
Widén et al. (2009)	Stochastic	✓	-	S	✓	5 min
Hendron et al. (2010)	Stochastic	✓	✓	S	✗	<1 min
Jordan and Vajen (2001)	Stochastic	✓	✓	✓	✗	1 min
Buchberger and Wells (1996)	Stochastic	✓	✓	S	✓	1 min
Blokker et al. (2009)	Stochastic	✓	-	S	-	1 sec
Scheepers and Jacobs (2014)	Stochastic	✓	✗	-	-	-
Richard (2016)	Stochastic	✓	✓	✗	-	1 min
Yao and Steemers (2005)	Stochastic	✗	✓	-	-	1 min
Negnevitsky and Wong (2014)	Stochastic	✓	✗	✗	✗	5 min
Aki et al. (2016)	Stochastic	✓	✗	✗	✓	10 min
Bakker et al. (2008)	Data-learning	✗	✗	✓	✓	1 hr
Pflugradt (2016)	Behavioral	✗	✓	✓	✓	-
Popescu and Serban (2008)	Time-series forecasting	✓	-	✓	✓	-
Gelažanskas and Gamage (2015)	Time-series forecasting	✓	-	✓	✓	1 hr
Roux (2018)	Time-series forecasting	✓	-	✓	✓	1 min
Jack et al. (2018)	Data-driven	✗	✗	✗	✓	1 min
This paper	Data-driven	✓	✓	✓	✓	1 min

### *Technical standards*

Existing standards provide typical daily hot water usage profiles that are available for design purposes. [Marini et al. \(2015\)](#) evaluated simulation tools based on various standards to compare simulated results with monitored results for modelling hot water usage in buildings. They found that the simulated total hot water demand was underestimated by up to 30% and overestimated by up to 40%. The energy efficiency of the simulated model was also overestimated by between 14% and 22% compared to the actual hot water heater. [Gholizadeh and Aravinthan \(2016\)](#) proposed a model to reduce the electricity costs of energy usage by adjusting the heaters' set-point temperatures and simulating water profiles based on the ASHRAE Standard 90.2. However, like other technical standards, the ASHRAE standard does not take into account seasonal or individual variation, so water profiles based on technical standards are typically generalised.

### *Data-collected models*

Data-collected models stray from using probabilistic models and advance towards a technique that focuses on numerous participants and large amounts of data. The technique uses long-term databases of water measurements and simulations to produce a water profile based on the average. [Edwards et al. \(2015\)](#) used data measurements from 73 Canadian households to produce annual water draw profiles. Of the total households, 12 were selected to produce five minute resolution profiles based on four usage levels and three discrete temporal usage patterns for each level. These usage levels corresponded to the mean, median, 20% and 80% percentiles of usage for the sample of households, and the temporal usage patterns were made up of predominantly morning usage, predominantly evening usage, and usage constant throughout the day. The research revealed that different water usage patterns can lead to significant changes when it comes to predicting water usage profiles.

### *Stochastic models*

Stochastic methods model the probability distribution of household water usage using time-of-use data. [Widén et al. \(2009\)](#) used time-of-use data to generate electricity and hot water profiles. Individuals in Sweden used diaries to track the sequence and time of electrical or hot water activities during one weekday and one weekend day. They then tried to match the activities to end-use categories and used information from surveys and product tests to estimate appliance parameters. The results showed a promising correlation between simulated and measured data. [Hendron et al. \(2010\)](#) determined hot water events for different activities using clustering and probability distributions. The event characteristics were approximated by fitting normal and exponential functions. [Jordan and Vajen \(2001\)](#) used a similar approach with the exclusion of clustering. [Yao and Steemers \(2005\)](#) produced domestic energy demand profiles from aggregated random appliance profiles. They clustered the energy profiles according to the factors that influence a household's profile, such as the number of occupants, the times of the first and last appliance usage of each day, and the occupancy patterns, determined by the periods of the day when the house was unoccupied ([Negnevitsky and Wong, 2014](#); [Aki et al., 2016](#)).

[Buchberger and Wells \(1996\)](#) based their event probability distributions on the assumption that water demand follows a homogeneous Poisson rectangular pulse process, which means that events can be approximated by rectangular shapes with the rate of occurrences being time dependent. However, the results showed that this process cannot model the variance in occurrence well enough and the data came from a total of only four households. [Scheepers and Jacobs \(2014\)](#) devised a method for generating a residential indoor cold water profile based on probabilistic end-use models. The profiles consisted of high resolutions with the temporal sample difference being as small as one second. Residential water demand profiles were simulated using SIMDEUM, which was developed by [Blokker et al. \(2009\)](#). As in the model by [Buchberger and Wells \(1996\)](#), water events were approximated by rectangular pulses for six end-use cases and aggregated water demands were obtained using a Monte Carlo method. However, they used information on total water usage and could not distinguish between hot and cold water. [Richard \(2016\)](#) improved on this by using volume balancing techniques to approximate the portion of hot water from the total. The resulting daily water demand profiles were similar to popular existing patterns such as the ASHRAE Standard 90.2.

### *Behavioural models*

Behavioural models are based on factors that influence load profiles such as work days and holidays. [Pflugradt \(2016\)](#) produced a psychological desire model to avoid calculating probability distributions. He

used a program called a behaviour-based load profile generator which makes it possible to simulate an individual’s behaviour, to obtain energy and water usage data. He concluded that it works well for individual households; however, the quality of the results is strongly determined by the input. It is difficult to gather information about residents and more assumptions have to be made for larger areas of households.

#### *Data-learning models*

Neural networks are computational models consisting of neurons that can learn, generalise and categorise data. [Bakker et al. \(2008\)](#) obtained the heat demand of a household to calculate the electricity demand and used them in a multi-layer feed-forward neural network. The heat demand for each following day was predicted with knowledge of the expected temperature and the previous day’s heat profile. The output data was generated with an hourly resolution, as not enough training data was available to make more accurate predictions. The results were good, despite an hourly mean absolute deviation of approximately 36 kWh. The authors suggested that the results could be improved if more factors were considered in the model.

#### *Time series models*

Time series models forecast a system’s future behaviour by analysing historical data to make statistical conclusions based on time correlation. The original data is decomposed into trend, seasonality and residual components. Autocorrelation functions and partial autocorrelation functions are used to compare a time series with a time-displaced version of itself. A residual component showing zero correlation means that the components were successfully separated. A strict requirement for time series modelling is that the data must be stationary, meaning that it has a time-invariant mean, variance and autocorrelation. Many tests are available to check if time series data is stationary. A variety of forecasting models can then use this information to predict future data. More information on time series forecasting can be found in [Shumway and Stoffer \(2000\)](#).

To forecast the hot water demand for residential households, [Gelažanskas and Gamage \(2015\)](#) tested various time series forecasting techniques and used different forecasting models. As the data showed strong daily and weekly correlation, the best results were obtained using a Loess (locally estimated scatterplot smoothing) seasonal decomposition of a time series with an ARIMA (autoregressive integrated moving average) model and an STL (seasonal and trend decomposition) exponential smoothing model. The study did not, however, account for the time of year, which contributes to both seasonality and holidays. [Popescu and Serban \(2008\)](#) explored a similar approach using an ARMA (autoregressive moving average) model to simulate the water demand for a block of apartments.

The existing time series forecasting models are suitable for forecasting the profile for an individual household, but they fail to capture seasonal effects. Time series is closely related to machine learning, where a great deal of data is required.

#### *Data-driven models*

These models forecast water usage by analysing the relationships and pattern of measured data without matching water usages to specific household devices. [Jack et al. \(2018\)](#) proposed a minimal simulation model (MSM) that simulates smart control of a water heater to assess its impact on service. This model used the monitored electricity demand of a domestic water heater to generate water profiles that are simulated from a hot water tank model. The parameters for generating water profiles were adjusted manually until the simulated and monitored electricity demand profiles matched – a process that had to be repeated for each individual heater.

### *1.2. Contributions*

We present a novel data-driven probabilistic model that statistically models hot water profiles and a simulator that uses the model to generate hot water usage profiles, which can be used to support demand-side energy management of domestic water heaters. To deal with the limitations we identified in the studies reviewed above, our model features fine time resolution, differentiation according to the season of the year and the day of the week, individual household differences, and unsupervised automation. Our proposed model uses clustering and statistical analysis to model domestic hot water usage behaviour from historically measured usage profiles. Hot water usage data from 77 South African households was used. The measured hot water usage data from the households, the source code that fits the probabilistic model to the measured

data, the source code for the hot water usage simulator, and the simulated hot water usage profiles that were generated by the simulator for this paper can all be found in the supplementary material.

## 2. Hot water usage model and simulator

The probabilistic model presented here statistically models the daily hot water usage profile for individual electric water heaters (EWHs) based on historically measured flow rate data. This model can be used to model hot water demand and the resulting residential energy demand. We also created a hot water usage simulator that implements the probabilistic model. This simulator generates simulated hot water usage profiles (flow rate versus time) based on the probabilistic model that was fitted to the historically measured usage data. The hot water usage simulator can be used as a component of a larger EWH simulator. We first describe the probabilistic model in Section 2.1 and then the usage simulator in Section 2.2.

We used the data set gathered from 77 domestic water heaters in South Africa. Each heater was equipped with a smart controller that captured the flow rate every minute with a resolution of 0.5L using an in-line flow meter attached to the inlet pipe. After cleaning the data, each heater provided four months of data, one month for each season, with the starting dates being 5 February 2018, 9 April 2018, 21 August 2018 and 2 October 2018. The chosen periods do not include major holiday periods, but the model does make provision for their inclusion. The full data set is available with the model and the simulator code in the supplementary material.

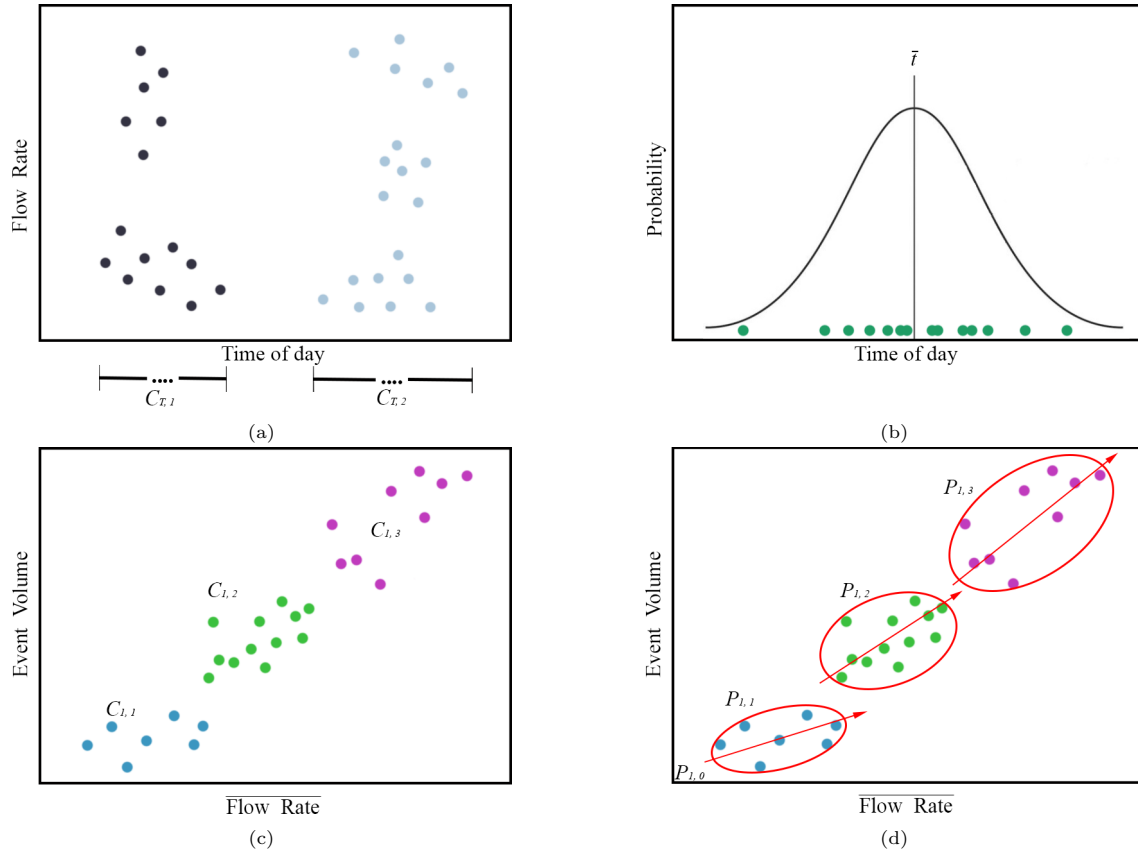


Figure 1: (a) The flow rate data points are clustered according to the time of day to produce time clusters  $C_{T,1}$  and  $C_{T,2}$ . (b) Water event parameters recorded for each time cluster are clustered according to volume. (c) Volume clusters are fitted with 2D Gaussian distributions (red ovals) with correlation of volume and flow rate shown with a red arrow. Each cluster is associated with its own probability of occurrence,  $P_{n,m}$ . (d) Start times measured for a time cluster (green) are fitted with a Gaussian distribution where  $t'$  represents the mean middle time of the time cluster.

### 2.1. Probabilistic hot water usage model

The hot water usage model statistically models the hot water usage behaviour for a given EWH by identifying repeating daily hot water usage events, and determining the statistical distributions of each usage event’s time of day, volume of water used, and average flow rate. To do this, the model groups the measured flow rate data into clusters of time, volume and flow rates, to which it fits distributions and determines their probabilities of occurrence, as shown in Figure 1.

The model handles hot water usage variations due to seasonality by separating the measured data for a given EWH into four seasonal sets, and determining different model parameters for each season (summer, autumn, winter, spring). The model handles variations in the usage pattern due to the day of the week by further separating the seasonal data set into three subsets: the undifferentiated data set, the semi-differentiated set, and the differentiated set. The undifferentiated set assumes that the usage behaviour does not differ based on the day of the week, and that there is only one usage pattern which is the same for all seven days of the week.

The semi-differentiated set assumes that the usage behaviour differs between weekdays and weekends, and that there are two different usage patterns, one for weekdays and one for weekends. The differentiated set assumes that the usage behaviour differs based on the day of the week, and that there are seven different usage patterns, one for each day of the week (Mondays, Tuesdays, ..., Sundays). For each season, we therefore derive three variations of our probabilistic hot water usage model, by applying the model to each of the three different data subsets: undifferentiated, semi-differentiated, and differentiated. We now present the procedure for deriving the model from the measured flow rate data. The same procedure is followed to derive the model parameters for the four seasons, and for the three model variations; the only difference is the data set that is used to derive the model parameters.

To identify repeating usage events, the daily measured flow rate data for all the days in a given data set are superimposed on a single day cycle, and the flow rate data are grouped into time clusters according to time of the day using k-means clustering (Jain et al., 1999), as shown in Figure 1a. We determine the optimal number of time clusters using the elbow method (Kodinariya and Makwana, 2013). Time clusters are represented by  $C_{T,p}$ , where the subscript  $p$  is the index of the time cluster. Each time cluster now represents a repeating usage event, and the flow rate data points in the same time cluster on the same day are considered to represent a single usage event. We determine the probabilistic distribution of the time-of-day of each repeating event by fitting a Gaussian probability density function to the time-of-day component of all the flow rate data points that belong to each time cluster, as shown in Figure 1b. The Gaussian probability density function is determined for time cluster  $p$  by calculating the time-of-day mean  $\bar{t}_p$  and variance  $\sigma_{t_p}^2$  for the flow rate data points within the time cluster.

Next, we analyse the distribution of the volume and the average flow rate of the daily usage events within each time cluster. For a given time cluster, the volume and the average flow rate of the usage event on each day are calculated and plotted on a scatter plot, as shown in Figure 1b. To identify different types of usage events that may occur within the same time slot, the (volume, flow rate) data points are subclustered according to volume, as shown in Figure 1c. The volume subclustering within each time cluster is again performed using k-means clustering and the elbow method. Volume subclusters are represented by  $C_{p,q}$ , where subscript  $q$  is the index of the volume subcluster within time cluster  $p$ . Each event volume is captured with its associated average flow rate to retain correlation between the two variables.

The volume subclusters within a time cluster represent different hot water usage events that may occur in the same time slot. However, on a given day, only one hot water usage event can occur in each time slot. The volume subclusters within the same time cluster therefore represent different possible realisations of a single hot water usage event within the given time slot. Every day, each possible realisation of the hot water usage event has a certain probability of occurring. For each time cluster, we therefore need to calculate the probability of each volume subcluster being realised. The probability  $P_{p,q}$  of volume subcluster  $q$  being realised in time cluster  $p$ , is calculated by dividing the number of days that the given subcluster occurs by the total number of days in the data set. (Note that this approach also accommodates the probability that no usage event occurs in the time slot on certain days, since there will also be a volume cluster for zero volume events).

Finally, we characterise the probabilistic distribution of the volume and average flow rate for each volume subcluster. Since the volume and the average flow rate of a usage event may be statistically dependent, we fit a Gaussian joint probability density function to the (volume, flow rate) data points in each volume subcluster,

as shown in Figure 1d. Each volume subcluster is therefore characterised by a joint (volume, average flow rate) mean represented by  $(\bar{V}_{p,q}, \bar{f}_{p,q})$  and a joint (volume, average flow rate) variance represented by the volume variance  $\sigma_{vp,q}^2$ , the average flow rate variance  $\sigma_{fp,q}^2$ , and the correlation coefficient  $\rho_{p,q}$ .

To summarise, the probabilistic hot water usage model models the number of expected hot water usage events each day (time clusters), the probabilistic distribution of the time of day for each usage event (Gaussian probability density function), a number of different types of usage events that may occur in the same time slot (volume subclusters), the probability of each type of usage event occurring, and the probabilistic distribution of the volume and average flow rate of each type of usage event (using the Gaussian joint probability distribution function for each volume subcluster).

## 2.2. Hot water usage simulator

Algorithm 1 shows the procedure used to generate simulated usage profiles from a water heater's probabilistic usage model. Assume that we wish to generate the hot water usage profile for a number of days. To do this, we iterate through all the days, and for each day, we iterate through all the time clusters and generate a random hot water usage event for each time cluster. To generate the hot water usage event for a specific time cluster, we first determine the time of the event by drawing a random sample from its Gaussian time of day distribution, as shown in Figure 1b. Next, we determine which volume subcluster is realised within the time cluster by randomly sampling the subcluster index  $q$  based on the probability distribution of all the volume subclusters within time cluster  $p$ . Once the volume subcluster has been determined, we determine the volume and the average flow rate of the hot water usage event by drawing a random sample from the Gaussian joint (volume, flow rate) distribution for the given volume subcluster, as shown in Figure 1d. Given the randomly sampled time, volume, and average flow rate of the event, the simulated flow rate versus time is constructed. First, the duration of the usage event is calculated by dividing the volume by the average flow rate, using

$$T_{\text{dur}} = \frac{V}{\bar{f}} \quad (1)$$

where  $T_{\text{dur}}$  is the duration of the event, and  $V$  and  $\bar{f}$  are the randomly sampled volume and average flow

---

### Algorithm 1 Hot water usage algorithm

---

```

1: procedure GENERATE SIMULATED USAGE DATA
2:   for each day do
3:     for each time cluster  $p$  do
4:        $x \leftarrow$  Roll number between 0 and 1 // where  $x$  is a random variable
5:        $q, i \leftarrow 0$ 
6:       while  $i < Q_p$  and  $q = 0$  do // where  $Q_p$  is the optimal number of volume subclusters
7:          $i \leftarrow i + 1$ 
8:         if  $x < \sum_{n=1}^i P_{p,i}$  then // where  $P$  is the volume subcluster occurrence probability
9:            $q \leftarrow i$  // assign picked volume subcluster to  $q$ 
10:        end if
11:      end while
12:      if  $q > 0$  then // event occurred
13:         $V_{p,q}, f_{p,q} \leftarrow \mathcal{N}(\bar{V}_{p,q}, \bar{f}_{p,q}, \sigma_{vp,q}^2, \sigma_{fp,q}^2, \rho_{p,q})$  // generate one sample
14:         $T_{\text{dur},p,q} \leftarrow \frac{V_{p,q}}{f_{p,q}}$ 
15:         $t_{mp,q} \leftarrow \mathcal{N}(\bar{t}_p, \sigma_{tp}^2)$  // generate one sample
16:         $t_{\text{start},p,q} \leftarrow t_{mp,q} - \frac{1}{2} \times T_{\text{dur},p,q}$ 
17:      end if
18:    end for
19:    for each water usage event do // construct events
20:       $\text{flow rate}(t) \leftarrow f$     $t_{\text{start}} \leq t \leq t_{\text{start}} + T_{\text{dur}}$ 
21:       $\leftarrow 0$     $t < t_{\text{start}}$  and  $t > t_{\text{start}} + T_{\text{dur}}$ 
22:    end for
23:    Superimpose all water usage events to obtain simulated hot water usage profile for day
24:  end for
25: end procedure

```

---

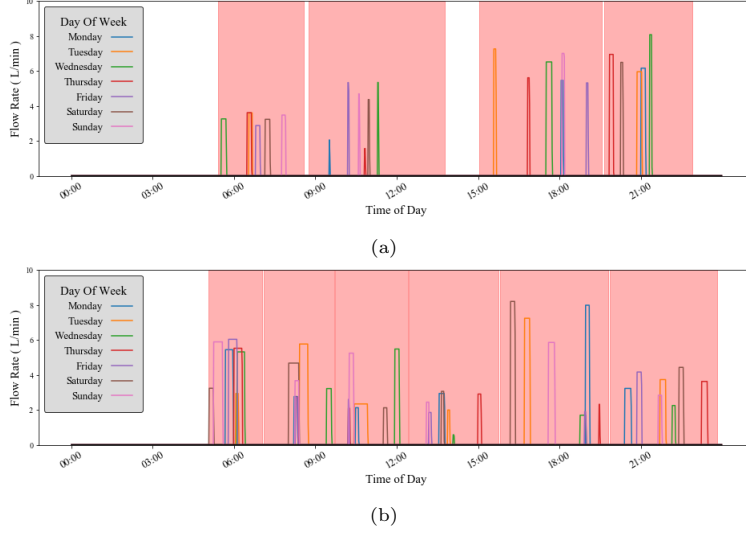


Figure 2: Simulated hot water profiles generated in the identified time clusters (shaded red region) for (a) a light usage EWH and (b) a heavy usage EWH. The undifferentiated model was used.

rate. The start time of the event is then calculated using

$$t_{\text{start}} = t_m - \frac{T_{\text{dur}}}{2} \quad (2)$$

where  $t_{\text{start}}$  is the start time of the event,  $t_m$  is the randomly sampled time of the event, and  $T_{\text{dur}}$  is the duration of the event. (The time  $t_m$  is the time in the middle of the event, and therefore half of the duration  $T_{\text{dur}}$  must be subtracted to obtain the start time of the event  $t_{\text{start}}$ . The simulated usage event is then constructed using the following equation

$$f(t) = \begin{cases} \bar{f}, & t_{\text{start}} \leq t \leq t_{\text{start}} + T_{\text{dur}} \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where  $f(t)$  is the instantaneous hot water flow rate as a function of time. A hot water usage event is therefore simulated as a constant hot water flow rate between the start time and the end time of the event, and has a zero flow rate otherwise. The constant flow rate equals the randomly sampled average flow rate of the event, and the end time equals the randomly sampled start time plus the duration of the event.

The simulated hot water usage profile for the entire day is constructed by superimposing the simulated hot water usage events of all the time clusters. This hot water usage model can generate simulated profiles for a specific day (e.g. Monday, Tuesday, etc.) over a desired period of time. Using the differentiated model, we generated a week's profile by producing a one-day water profile from each of seven hot water usage models, each representing a day of the week. Figure 2 shows the simulated profiles generated by the undifferentiated model for seven days against the backdrop of the identified time clusters in red. It shows the profile for one week on a 24 hour cycle, for a heater with light usage and one with heavy usage.

### 3. Model and simulator verification

The simulator was verified by comparing the statistical distribution of the hot water usage events generated by the simulator with the statistical distribution of the measured usage events identified from the measured data. The verification consisted of four evaluations:

1. The statistical distributions of the simulated hot water usage events were compared with the statistical distributions of the real hot water usage events that were identified in the measured flow rate data. The comparison was done for two example EWHs, one with a light usage profile, and one with a heavy usage profile, and with their probabilistic models derived using the undifferentiated data set.

2. A large number of simulated usage profiles were generated for an EWH with an average usage profile. Histograms of the time of day of usage events were plotted and compared against the Gaussian probability density functions for the time of day identified from the measured flow rate data.
3. The frequency of the simulated usage events that belong to each volume subcluster within the same time cluster was calculated and compared with the probability of the volume subcluster occurring, as identified from the measured flow rate data.
4. Scatter plots of the volume and average flow rates of the daily usage event were plotted and compared with the confidence ellipses of the Gaussian joint probability functions for the volume and flow rate identified from the measured flow rate data.

The generation of volume clusters for a profile consisting of five time clusters for the undifferentiated model is shown in Figure 3a. The water events were obtained for 200 weeks worth of Monte Carlo realisations of a simulated hot water profile and were plotted against the volumes of each water event. The time clusters are colour-coded and the volume clusters are differentiated by symbols to make it clear to which clusters the water events belong. This profile specifically shows that the user has larger water events occurring in the morning and evening hours and smaller water events occurring in-between. Table 2 shows the expected probabilities for each volume cluster and the percentage of volume clusters that actually occurred in the following columns. The smallness of the differences between the two columns shows that the model generates water events very close to the expected probabilities.

Table 2: Table of event occurrences of each volume cluster within each time cluster for the 200 weeks of measured and simulated water events taken from Figure 3a.

Time cluster	Volume cluster 1		Volume cluster 2	
	Measured	Simulated	Measured	Simulated
1	50 %	49 %	18 %	15 %
2	25 %	24 %	36 %	34 %
3	25 %	22 %	11 %	7 %
4	61 %	57 %	18 %	16 %
5	57 %	54 %	11 %	12 %

The rest of Figure 3 shows the modelled distributions (based on measured data) and the simulated statistical distributions for an EWH that has three different time clusters. Approximately 2000 water events were generated for each time cluster to produce these distributions of the data that accurately reflect the cluster’s statistical properties. Figures 3b, 3c and 3d show the modelled normal distribution (orange) of the time of day based on the measured data and a histogram of the simulated even times (green) generated by the usage simulator. The legend of each figure shows the mean and standard deviations of the measured data and the simulated events. The statistical properties of the simulated start times agree with those of the statistical properties of the measured data.

Figures 3e, 3f and 3g show scatter plots of the volume and flow rate parameters for simulated water events for the same three time clusters as the start times. From each time cluster three volume clusters were identified and colour-coded accordingly. A confidence ellipse is also shown for the measured data that was used to identify each volume cluster. Each ellipse has a confidence interval of 95% to show that 95% of the simulated profile’s water events fall within these regions. These figures show that the simulated profiles retain the mean and standard deviation for both volume and flow rate as well as the correlation between them.

#### 4. Comparison with existing models

We firstly compared the ability of the three variations of our model to model hot water demand against that of two popular models from the literature. We also used a validated heater model to compare the heaters’ simulated energy profiles with their modelled profiles. Finally, we compared the results with the “notch test”, an experiment that was done on the South African national electricity grid using an estimated 120 000 heaters, to determine the aggregate electrical profiles of residential households’ water heaters (Forlee, 1998).

For comparison, hot water usage profiles were generated from the minimal simulation model (MSM) devised by Jack et al. (2018) and the ASHRAE 90.2 implementation by Gholizadeh and Aravinthan (2016).

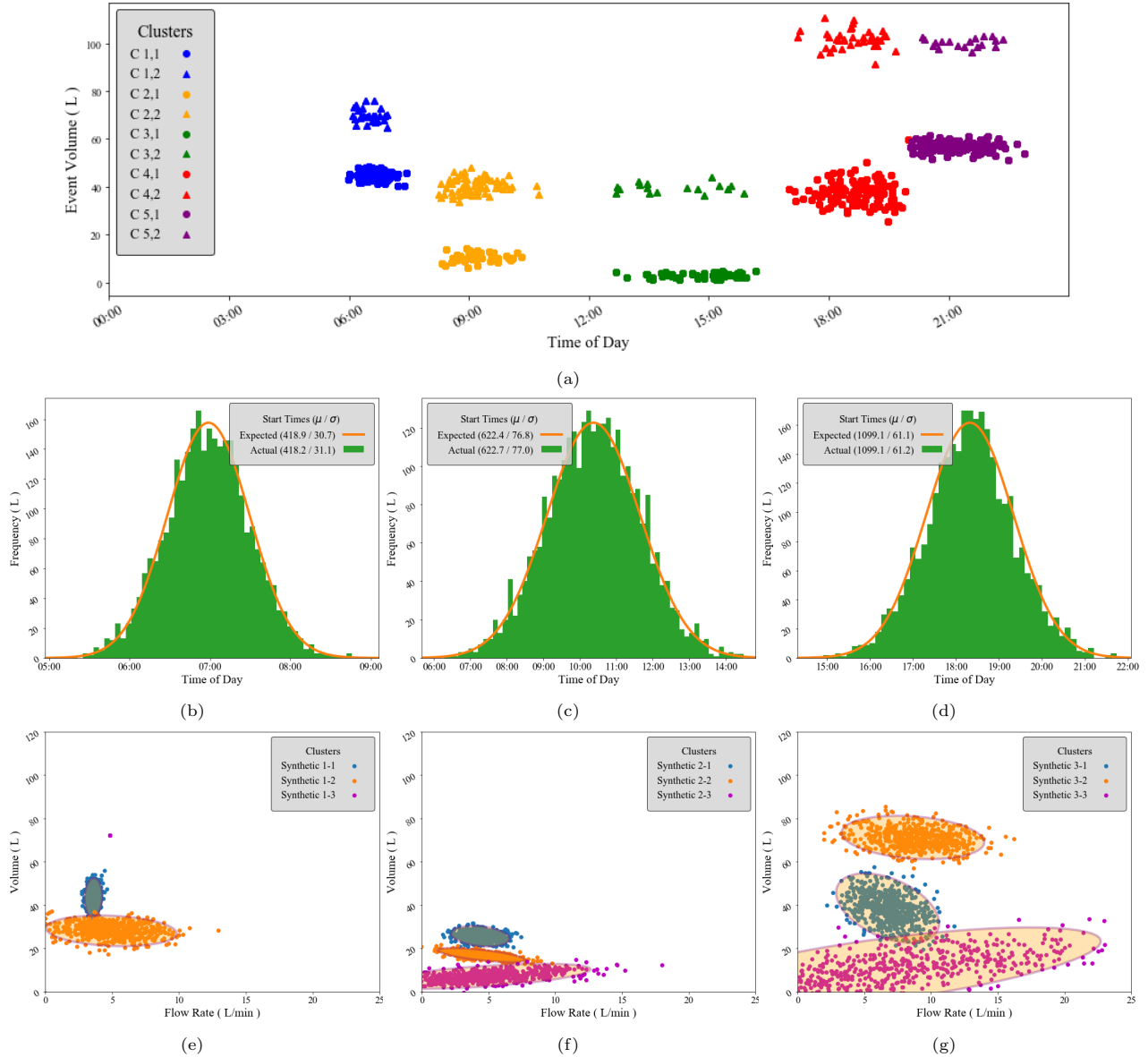


Figure 3: a) Scatter plot of water events generated for many realisations of a single EWH’s simulated hot water profile for the undifferentiated model. The occurrence frequencies are shown in Table 2. b-d) Start times of measured data distributions (orange) and generated simulated water events (green) for three time clusters. Statistical properties of each are shown in the legend. e-g) Volume and flow rate scatter plots of the simulated water events generated for each time cluster. Also shown are 95% confidence ellipses produced by the measured data for each volume cluster.

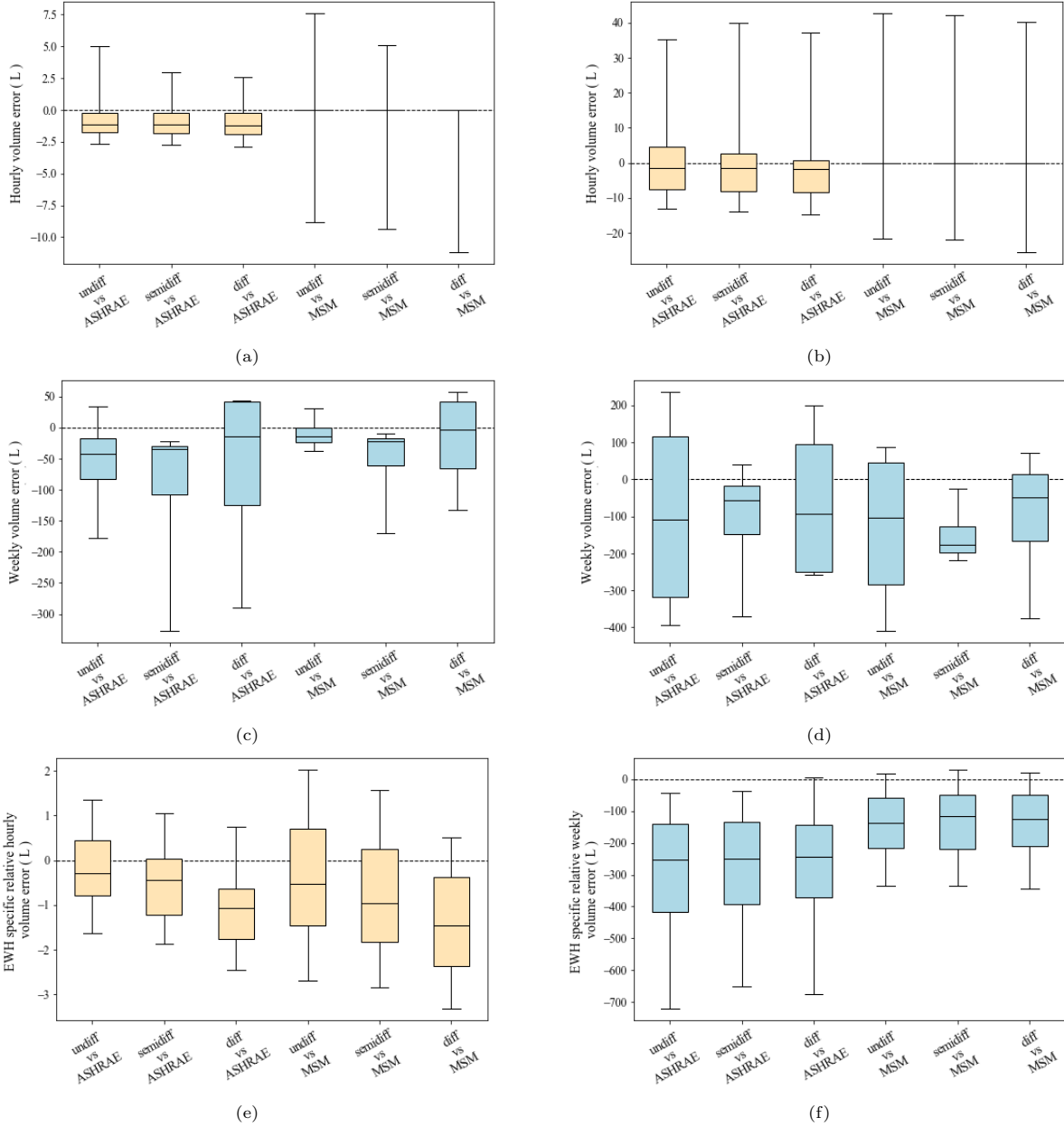


Figure 4: a-d) Comparison of the water volume modelling errors for the nominal simulated hot water profiles of our model for all data sets and other existing models. Results are shown for hourly volume modelling errors (Equation 7) for a) a light water usage household and b) a heavy water usage household, and results are also shown for weekly volume modelling errors (Equation 10) for c) a light water usage EWH and d) a heavy water usage EWH. A negative value indicates that our model had a smaller modelling error, compared to the measured data, than the model against which it is compared. e-f) Comparison of the EWH-specific e) hourly and f) weekly water volume modelling errors (Equations 11 and 12) for the nominal simulated hot water profiles of our model for all data sets and other existing models for 77 households.

The original MSM model was inconvenient for a large number of profiles because the parameters would have to be manually adjusted to fit each unique profile, which would take many hours to do. We therefore adapted this model to fit the data by autonomously iterating the parameters to allow the model to work with larger data-sets. Setting up the model requires manual intervention to assess the quality of the match between the measured and modelled profiles. We automated this assessment using a normalised coefficient of determination,  $R^2_{\text{diff}}$ .

The ASHRAE model is a deterministic profile with no random variations. To implement the model we added random noise to the ASHRAE 90.2 profile, similar to [Gholizadeh and Aravinthan \(2016\)](#). We used a normal Gaussian distribution with mean  $\mu = 0$  and variance  $\sigma^2 = 0.06$ . To ensure a fair comparison, the ASHRAE profiles were scaled to match the aggregated median weekly water usage volume of the measured data. Each water usage event was spread to occur over the hour with a per-minute water flow rate, so that it fitted a normal distribution with deviations at the hour boundaries of

$$T^{\text{hour}} = \text{Average Usage} + U(-500 \text{ sec}, 500 \text{ sec}) \quad (4)$$

#### 4.1. Evaluation of water usage modelling performance and comparison with other models

We compared the ability of our hot water usage model to model hot water usage demand (differentiated, semi-differentiated and undifferentiated versions) against two prominent existing models for all 77 household water heaters. We evaluated the models' hourly and weekly volume modelling and then compared the error levels of the modelled profiles and the measured profiles. The hourly modelling error is the difference between the modelled volume of water usage over the next hour and the measured volume of water usage over the same hour and is calculated as follows:

$$e_{\text{model|h}}[n] = |V_{\text{meas}}[n] - V_{\text{model}}[n]| \quad (5)$$

where  $n$  refers to the minute time sample and

$$V_{\text{meas}}[n] = \sum_n^{n+60} f[n] \quad (6)$$

The hourly modelling error was calculated using a moving window that calculates the hourly volume modelling error for each minute of the day. For a given EWH, we can compare the hourly modelling error of our model with the hourly modelling errors of the other models (MSM and ASHRAE) by comparing their box and whisker plots. The differences between the resulting hourly modelling errors of one of our models, such as the differentiated model, and other model, in this case ASHRAE, are calculated as follows:

$$e_{\text{res|h}}[n] = e_{\text{diff|h}}[n] - e_{\text{ASH|h}}[n] \quad (7)$$

The weekly volume modelling is calculated by finding the difference between the volumes of water used for each week between the measured and the other model's generated profiles, as follows:

$$e_{\text{model|w}}[m] = |V_{\text{meas}}[m] - V_{\text{model}}[m]| \quad (8)$$

where  $m$  refers to the week of a profile and

$$V_{\text{meas}}[m] = \sum_{m \times 60 \times 24 \times 7}^{(m+1) \times 60 \times 24 \times 7} f[m] \quad (9)$$

The resulting weekly modelling error is calculated as follows:

$$e_{\text{res|w}}[m] = e_{\text{diff|w}}[m] - e_{\text{ASH|w}}[m] \quad (10)$$

The EWH-specific relative volume modelling error is the average volume modelling error of our model for a specific EWH minus the average volume modelling error of the other model for the same heater. To compare the general modelling performance of our model (differentiated, semi-differentiated and undifferentiated data sets) against the general modelling performance of the other model for all 77 heaters, we plot the distribution

Table 3: Table showing the mean hourly volume modelling error (measured in liters) for each model and are differentiated by day of the week. The last column shows the mean error of a full week.

Model	Mon	Tue	Wed	Thu	Fri	Sat	Sun	Mean
ASHRAE	5.9	6.0	6.1	5.9	5.8	5.7	5.9	5.9
MSM	6.2	5.8	6.2	5.8	5.7	6.0	6.2	6.0
undiff	5.4	5.1	5.4	5.2	5.4	6.4	6.4	5.6
semidiff	5.3	4.9	5.2	5.0	5.2	5.8	5.7	5.3
diff	4.7	4.1	4.6	4.3	4.5	5.3	5.0	4.7

of the relative heater-specific modelling error for all the heaters. The hourly and weekly heater-specific relative modelling error for EWH  $i$  for our model, such as the differentiated model, and other models, such as ASHRAE, is calculated as follows:

$$\bar{e}_{\text{rel|h}}[i] = \bar{e}_{\text{diff|h}}[i] - \bar{e}_{\text{ASH|h}}[i] \quad (11)$$

$$\bar{e}_{\text{rel|w}}[i] = \bar{e}_{\text{diff|w}}[i] - \bar{e}_{\text{ASH|w}}[i] \quad (12)$$

The results also compare the three versions of our model for the three different data sets used. The following notations are given for our model: *diff*, the differentiated data set; *semidiff*, the semi-differentiated data set; and *undiff*, the undifferentiated data set.

Nominal water profiles were generated for a heater with a light usage (EWH A) and one with a heavy usage (EWH B). Figures 4a and 4b show box and whisker plots of the hourly volume modelling errors (calculated using Equation 7) of the different models for EWH A and B respectively. Looking at the subtraction order of this equation, we see that a negative value indicates that our model had a smaller modelling error when compared to the measured data than the two other models did, and a positive value indicates that our models modelling error was larger. Figure 4a shows that our model’s hourly modelling errors were mostly smaller than those of the ASHRAE model for the light profile. Comparing our model to the MSM model, we see the that performance of our model improves as we move from the undifferentiated to the differentiated version. The absence of a box for the MSM plots is because the flow is mostly zero during the day, a phenomenon captured in the measured data and both our and the MSM models, whereas the ASHRAE model has a water event for every hour. The outcome of this is that most hourly volume errors remain zero for both our and the MSM models. For the heavy profile of EWH B in Figure 4b our model does not perform as well as the ASHRAE model. However, its performance improves as we move from the undifferentiated to the differentiated version. Figures 4c and 4d show box and whisker plots of the weekly volume modelling errors (calculated using Equation 10) of the various models for EWH A and EWH B respectively. These figures also show that all three versions of our model have a smaller modelling errors than the other models in the majority of cases.

We generated a weekly nominal simulated water profile of 77 households to produce the results in this section. Table 3 shows the mean of the hourly volume modelling errors in litres for each model compared with those of the measured data for the seven days of the week and the mean of all the days. The mean errors for the ASHRAE and MSM models are 5.9 and 6.0 L.

Our model dramatically improves on these figures, from the undifferentiated version’s 5.6 L error, to the semi-differentiated version’s 5.3 L error, to the differentiated version’s modest 4.7 L error.

The table not only shows that our model performs better; it also highlights the importance of differentiating between days of the week to forecast water usages with only small errors.

Figures 4e and 4f show box and whisker plots similar to those in Figure 4, but instead show the heater-specific hourly and weekly volume modelling errors calculated according to Equations 11 and 12 respectively. The negative bias of each plot in both figures shows that our model will on average produce smaller modelling errors than the ASHRAE and MSM models. In Figure 4e, it is clear that the modelling errors of our model become even smaller than those of the ASHRAE and MSM models when we use the more differentiated data sets. Figure 4f does not show any significant changes when we use a different data set for our model, showing that the weekly volume variations average out over the week.

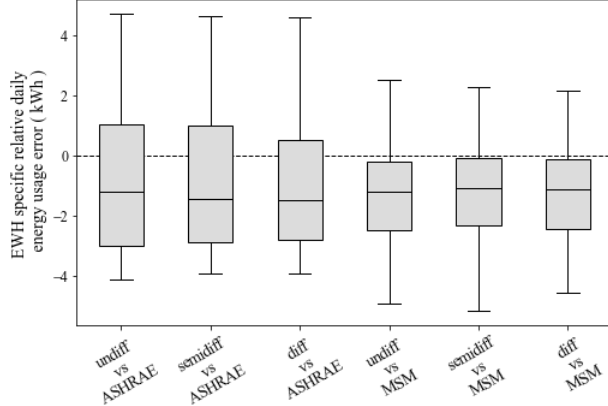


Figure 5: Comparison of EWH-specific mean daily energy modelling errors (Equation 14) for the nominal simulated hot water profiles of our model for all data sets and other existing models for 77 households.

#### 4.2. Individual EWH energy usage modelling performance

The energy profiles for all heaters were simulated using a validated thermal model of an EWH, and using both the measured hot water usage data and the simulated hot water usage events. The heater-specific energy usage modelling error was determined by comparing the simulated energy usage when using the simulated hot water usage events to the simulated energy usage when using the measured hot water usage data. The daily energy usage modelling error (in kWh) for an individual EWH is calculated as follows:

$$e_{\text{model|kWh}}[d] = |E_{\text{meas}}[d] - E_{\text{model}}[d]| \quad (13)$$

where  $d$  refers to the day in the profile and  $E$  refers to the total daily energy usage in kWh.

We compared the energy modelling performance of our probabilistic model against the energy modelling performance of the other two models by comparing our average daily energy modelling error against their average daily energy modelling error for each individual EWH, and then plotting the distribution of the EWH-specific *relative* daily energy modelling error over all 77 EWHs. The relative daily energy modelling error of your model relative to another model, e.g. the ASHRAE model, is calculated by taking the difference between our average daily energy modelling error and their modelling average daily energy modelling error, as follows:

$$\bar{e}_{\text{res|kWh}}[i] = \bar{e}_{\text{diff|kWh}}[i] - \bar{e}_{\text{ASH|kWh}}[i] \quad (14)$$

where  $i$  is the EWH index,  $\bar{e}_{\text{rel|kWh}}$  is our relative daily energy modelling error,  $\bar{e}_{\text{diff|kWh}}$  is the average daily energy modelling error of our probabilistic model using the differentiated data set, and  $\bar{e}_{\text{ASH|kWh}}$  is the average daily energy modelling error of the ASHRAE model.

The measured and simulated water profiles for all the heaters in autumn were simulated on an EWH model (Nel et al., 2016) to produce the energy profile for each household. Figure 5 shows box and whisker plots comparing the daily modelling errors of the ASHRAE and MSM models with all three versions of our model (calculated using Equation 14). The figure shows that our model will on average model the daily energy load 1kWh/day better than the other models, as shown by the negative median of each box and whisker plot, which is 17% of the median heater’s 5.9 kWh/day.

#### 4.3. Aggregated EWH energy usage modelling performance compared to Eskom load data

Finally, we compared a normalised 24-hour energy profile of our 77-household grid, using each heater’s simulated water profile, with the “notch test” from Eskom (the South African national electricity grid) (Forlee, 1998). The test considered only workdays in autumn, to match the data from Eskom.

Figure 6 shows the normalised electricity demand over a 24-hour cycle for the measured data, comparing all the models as well as the results obtained from “notch test” for all 77 heaters. The ASHRAE model deviates the most from the measured profile, performing worst at the morning peak, and has early morning

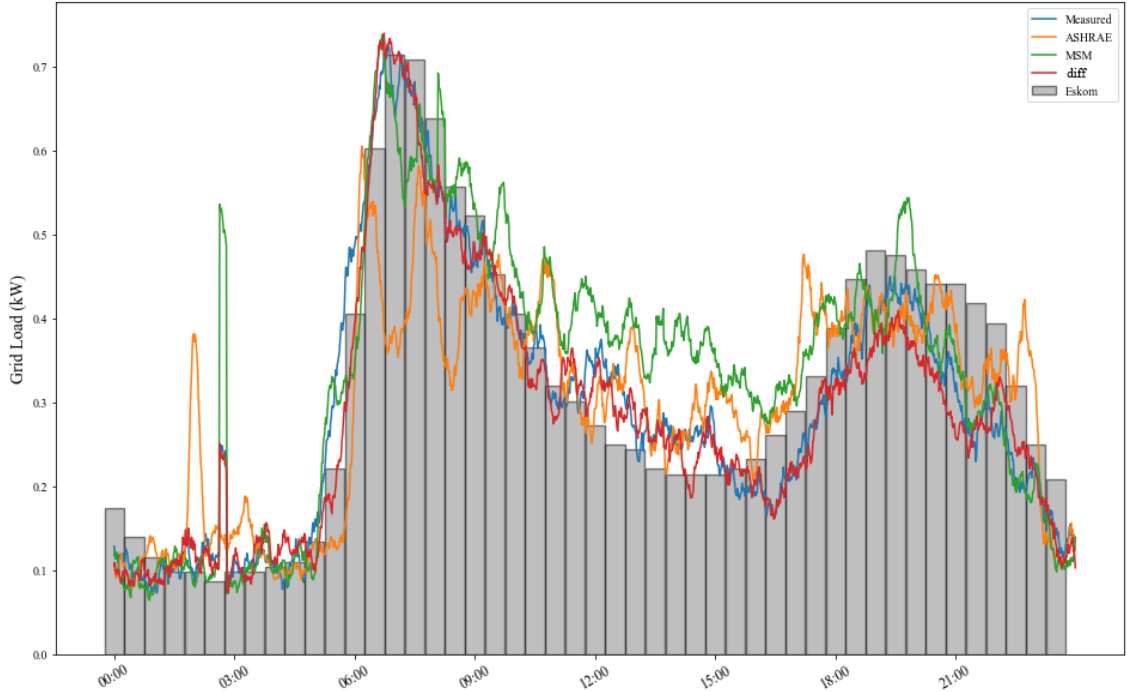


Figure 6: Normalised electrical demand plot over 24 hours for 77 EWH's for weekdays in autumn. The Eskom “notch test” electrical demand is represented by a bar plot (Forlee, 1998).

demand spikes misaligned with the measured profile. The MSM model and our model both model the early morning spike, but the MSM model overestimates it substantially. The MSM models’ profile generally follows the trend of the measured profile, but again overestimates the demand in the period between the morning and evening peaks. Our model’s profile accurately follows the measured profile, except for the evening peak where the load is slightly lower than the measured one. This could be because our sample has regional and external variations from the Eskom sample. The root mean squared (RMS) errors of the models’ demand profiles are 0.10 for the ASHRAE model, 0.08 for the MSM model, and 0.04 for the differentiated version of our model. Expressed as a percentage of the median heater, these are 16.9% for the ASHRAE model, 13.6% for the MSM model, and 6.8% for our differentiated model.

## 5. Conclusion

The probabilistic model we present in this paper statistically models the daily hot water usage profile for individual electric water heaters (EWHs) based on historically measured flow rate data. This model can be used to model hot water demand and the resulting residential energy demand. We also present a hot water usage simulator that uses the probabilistic model to generate simulated hot water usage profiles (flow rate versus time). Our model can generate high-resolution water profiles for specific applications and is fully automated. It excels in its ability to characterise a household’s water profile, taking into consideration differences between users and temporal variations such as the season of the year and the day of week. Our results confirm that the model produces simulated hot water profiles that closely match the measured data and are representative enough to be used in energy management strategies for modelling electric water heaters. Furthermore, the model’s results improve on those of existing hot water usage models for representing the energy drawn from the grid.

The hot water usage simulator can be used as a component of a larger EWH simulator to improve the efficiency of energy systems such as water heaters and strategies such as demand-side management. The efficiencies can be further improved if more measured water flow data is available to increase the accuracy of forecasting. The availability of more measured data could also provide enough information to fit more

accurate probability density functions to the clustered water usages. The model's unique visualisation of the data offers ideas that could be productive for further research in the field.

## Acknowledgements

The authors thank MTN South Africa for funding through grant S003061; the Water Research Commission (South Africa) for funding the water heater project through grant K1-7163, and Eskom for funding through the tertiary educational support programme.

## References

- Aki, H., Wakui, T., Yokoyama, R., 2016. Development of a domestic hot water demand prediction model based on a bottom-up approach for residential energy management systems. *Applied Thermal Engineering* 108, 697–708. doi:[10.1016/j.applthermaleng.2016.07.094](https://doi.org/10.1016/j.applthermaleng.2016.07.094).
- ASHRAE, 1993. ANSI/ASHRAE Standard 90.2, 1993. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. .
- ASHRAE, HVAC, 2003. Applications handbook, 2003. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., USA .
- Bakker, V., Molderink, A., Hurink, J.L., Smit, G.J., 2008. Domestic heat demand prediction using neural networks, in: 2008 19th International Conference on Systems Engineering, IEEE. pp. 189–194. doi:[10.1109/ICSEng.2008.51](https://doi.org/10.1109/ICSEng.2008.51).
- Bertrand, A., Mastrucci, A., Schüler, N., Aggoune, R., Maréchal, F., 2017. Characterisation of domestic hot water end-uses for integrated urban thermal energy assessment and optimisation. *Applied energy* 186, 152–166. doi:[10.1016/j.apenergy.2016.02.107](https://doi.org/10.1016/j.apenergy.2016.02.107).
- Blokker, E., Vreeburg, J., Van Dijk, J., 2009. Simulating residential water demand with a stochastic end-use model. *Journal of Water Resources Planning and Management* 136, 19–26. doi:[10.1061/\(ASCE\)WR.1943-5452.0000002](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000002).
- Booyesen, M., Visser, M., Burger, R., 2019. Temporal case study of household behavioural response to cape town's "day zero" using smart meter data. *Water research* 149, 414–420. doi:[10.1016/j.watres.2018.11.035](https://doi.org/10.1016/j.watres.2018.11.035).
- Buchberger, S.G., Wells, G.J., 1996. Intensity, duration, and frequency of residential water demands. *Journal of Water Resources Planning and Management* 122, 11–19. doi:[10.1061/\(ASCE\)0733-9496\(1996\)122:1\(11\)](https://doi.org/10.1061/(ASCE)0733-9496(1996)122:1(11)).
- Diao, R., Lu, S., Elizondo, M., Mayhorn, E., Zhang, Y., Samaan, N., 2012. Electric water heater modeling and control strategies for demand response, in: 2012 IEEE power and energy society general meeting, IEEE. pp. 1–8. doi:[10.1109/PESGM.2012.6345632](https://doi.org/10.1109/PESGM.2012.6345632).
- Dongellini, M., Falcioni, S., Morini, G.L., 2015. Dynamic simulation of solar thermal collectors for domestic hot water production. *Energy Procedia* 82, 630–636. doi:[10.1016/j.egypro.2015.12.012](https://doi.org/10.1016/j.egypro.2015.12.012).
- Edwards, S., Beausoleil-Morrison, I., Laperrière, A., 2015. Representative hot water draw profiles at high temporal resolution for simulating the performance of solar thermal systems. *Solar Energy* 111, 43–52. doi:[10.1016/j.solener.2014.10.026](https://doi.org/10.1016/j.solener.2014.10.026).
- Fairey, P., Parker, D., 2004. A review of hot water draw profiles used in performance analysis of residential domestic hot water systems. Florida Solar Energy Center URL: <http://www.fsec.ucf.edu/en/publications/pdf/FSEC-RR-56-04.pdf>.
- Forlee, C., 1998. Water heating ins South Africa facts and figures from the 1997 "notch testing" program. Presented at 5th International Domestic Use of Electrical Energy Conference, Cape Town. Available at [https://drive.google.com/open?id=0B6M3SFQa2Dy\\_Q3ZnZk1RdTJMa00](https://drive.google.com/open?id=0B6M3SFQa2Dy_Q3ZnZk1RdTJMa00).

- Gato, S., Jayasuriya, N., Roberts, P., 2007. Forecasting residential water demand: Case study. *Journal of Water Resources Planning and Management* 133, 309–319. doi:[10.1061/\(ASCE\)0733-9496\(2007\)133:4\(309\)](https://doi.org/10.1061/(ASCE)0733-9496(2007)133:4(309)).
- Gelažanskas, L., Gamage, K., 2015. Forecasting hot water consumption in residential houses. *Energies* 8, 12702–12717. doi:[10.3390/en81112336](https://doi.org/10.3390/en81112336).
- Gerin, O., Bleys, B., De Cuyper, K., 2014. Seasonal variation of hot and cold water consumption in apartment buildings, in: *Proceedings of CIB W062, 40th International Symposium on Water Supply and Drainage for Building (Sao Paulo, Brazil)*, pp. 1–9. URL: [http://www.siscone.com.br/Uploads/CIB2014/Trab0128000015232014070\\_000000.docx](http://www.siscone.com.br/Uploads/CIB2014/Trab0128000015232014070_000000.docx).
- Gholizadeh, A., Aravinthan, V., 2016. Benefit assessment of water-heater management on residential demand response: An event driven approach, in: *2016 North American Power Symposium (NAPS)*, IEEE. pp. 1–6. doi:[10.1109/NAPS.2016.7747831](https://doi.org/10.1109/NAPS.2016.7747831).
- Hendron, B., Burch, J., Barker, G., 2010. Tool for generating realistic residential hot water event schedules. Technical Report. National Renewable Energy Lab.(NREL), Golden, CO (United States). URL: <https://www.osti.gov/biblio/989020>.
- Hohne, P., Kusakana, K., Numbi, B., 2019. A review of water heating technologies: An application to the South African context. *Energy Reports* 5, 1–19. doi:[10.1016/j.egy.2018.10.013](https://doi.org/10.1016/j.egy.2018.10.013).
- Iglesias, F., Palensky, P., 2013. Profile-based control for central domestic hot water distribution. *IEEE Transactions on Industrial Informatics* 10, 697–705. doi:[10.1109/TII.2013.2275032](https://doi.org/10.1109/TII.2013.2275032).
- Jack, M., Suomalainen, K., Dew, J., Evers, D., 2018. A minimal simulation of the electricity demand of a domestic hot water cylinder for smart control. *Applied Energy* 211, 104–112. doi:[10.1016/j.apenergy.2017.11.044](https://doi.org/10.1016/j.apenergy.2017.11.044).
- Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: A review. *ACM Comput. Surv.* 31, 264–323. doi:[10.1145/331499.331504](https://doi.org/10.1145/331499.331504).
- Jordan, U., Vajen, K., 2001. Realistic domestic hot-water profiles in different time scales. Report for IEA-SHC Task 26. URL: <https://sel.me.wisc.edu/trnsys/trnlib/iea-shc-task26/iea-shc-task26-load-profiles-description-jordan.pdf>.
- Kepplinger, P., Huber, G., Petrasch, J., 2015. Autonomous optimal control for demand side management with resistive domestic hot water heaters using linear optimization. *Energy and Buildings* 100, 50–55. doi:[10.1016/j.enbuild.2014.12.016](https://doi.org/10.1016/j.enbuild.2014.12.016).
- Kodinariya, T.M., Makwana, P.R., 2013. Review on determining number of cluster in k-means clustering. *International Journal* 1, 90–95. URL: <https://bit.ly/2VJsxOE>.
- Marini, D., Buswell, R.A., Hopfe, C.J., 2015. A critical software review-how is hot water modelled in current building simulation. URL: <https://dspace.lboro.ac.uk/2134/19285>.
- Mathu, K., 2017. Cleaning south africa’s coal supply chain. *Journal of Business Diversity* 17. doi:[10.33423/jbd.v17i3.1236](https://doi.org/10.33423/jbd.v17i3.1236).
- Matos, C., Bentes, I., Pereira, S., Faria, D., Briga-Sa, A., 2019. Energy consumption, CO2 emissions and costs related to baths water consumption depending on the temperature and the use of flow reducing valves. *Science of The Total Environment* 646, 280–289. doi:[10.1016/j.scitotenv.2018.07.290](https://doi.org/10.1016/j.scitotenv.2018.07.290).
- Negnevitsky, M., Wong, K., 2014. Demand-side management evaluation tool. *IEEE Transactions on Power Systems* 30, 212–222. doi:[10.1109/TPWRS.2014.2329323](https://doi.org/10.1109/TPWRS.2014.2329323).
- Nel, P., Booysen, M.J., van der Merwe, B., 2016. A computationally inexpensive energy model for horizontal electric water heaters with scheduling. *IEEE Transactions on Smart Grid* 9, 48–56. doi:[10.1109/TSG.2016.2544882](https://doi.org/10.1109/TSG.2016.2544882).

- Nel, P., Booysen, M.J., Van der Merwe, B., 2017. Saving on household electric water heating: What works best and by how much?, in: 2017 IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia), IEEE. pp. 1–6. doi:[10.1109/ISGT-Asia.2017.8378439](https://doi.org/10.1109/ISGT-Asia.2017.8378439).
- Pflugradt, N.D., 2016. Modellierung von wasser und energieverbräuchen in haushalten. English: “Modeling water and energy consumption in households.” URL: <https://nbn-resolving.org/urn:nbn:de:bsz:ch1-qucosa-209036>.
- Popescu, D., Serban, E., 2008. Simulation of domestic hot-water consumption using time-series models, in: Proceedings of the 6th IASME/WSEAS International Conference on Heat Transfer, Thermal Engineering and Environment, Rhodes, Greece, pp. 20–22. URL: [https://www.researchgate.net/publication/267414381\\_Simulation\\_of\\_Domestic\\_Hot-Water\\_Consumption\\_Using\\_Time-Series\\_Models](https://www.researchgate.net/publication/267414381_Simulation_of_Domestic_Hot-Water_Consumption_Using_Time-Series_Models).
- Richard, M., 2016. Simulating domestic hot water demand by means of a stochastic end-use model. Ph.D. thesis. Stellenbosch: Stellenbosch University. URL: <https://scholar.sun.ac.za/handle/10019.1/98434>.
- Roux, M., 2018. Load management of electric water heaters in a smart grid through forecasting and intelligent centralised control. Master’s thesis. Stellenbosch: Stellenbosch University. URL: <http://scholar.sun.ac.za/handle/10019.1/103591>.
- Roux, M., Apperley, M., Booysen, M., 2018. Comfort, peak load and energy: Centralised control of water heaters for demand-driven prioritisation. *Energy for Sustainable Development* 44, 78–86. doi:[10.1016/j.esd.2018.03.006](https://doi.org/10.1016/j.esd.2018.03.006).
- Roux, M., Booysen, M., 2017. Use of smart grid technology to compare regions and days of the week in household water heating, in: 2017 International Conference on the Domestic Use of Energy (DUE), IEEE. pp. 276–283. doi:[10.23919/DUE.2017.7931855](https://doi.org/10.23919/DUE.2017.7931855).
- Scheepers, H., Jacobs, H., 2014. Simulating residential indoor water demand by means of a probability based end-use model. *Journal of Water Supply: Research and Technology-AQUA* 63, 476–488. doi:[10.2166/aqua.2014.100](https://doi.org/10.2166/aqua.2014.100).
- Shumway, R.H., Stoffer, D.S., 2000. Time series analysis and its applications. *Studies In Informatics And Control* 9, 375–376. doi:[10.1007/978-3-319-52452-8](https://doi.org/10.1007/978-3-319-52452-8).
- Skinner, T., et al., 2012. An overview of energy efficiency and demand side management in south africa, in: Presentation to the world bank/IFC workshop on appropriate incentives to deploy renewable energy and energy efficiency, Washington, DC. Available online: [www.eskom.co.za](http://www.eskom.co.za). URL: [https://www.esi-africa.com/wp-content/uploads/Tom\\_Skinner.pdf](https://www.esi-africa.com/wp-content/uploads/Tom_Skinner.pdf).
- Spur, R., Fiala, D., Nevrala, D., Probert, D., 2006. Influence of the domestic hot-water daily draw-off profile on the performance of a hot-water store. *Applied Energy* 83, 749–773. doi:[10.1016/j.apenergy.2005.07.001](https://doi.org/10.1016/j.apenergy.2005.07.001).
- Wang, W., Beausoleil-Morrison, I., Thomas, M., Ferguson, A., 2007. Validation of a fully mixed model for simulating gas fired water storage tanks, in: Proc. Building Simulation. URL: [https://www.aivc.org/sites/default/files/p240\\_final.pdf](https://www.aivc.org/sites/default/files/p240_final.pdf).
- Widén, J., Lundh, M., Vassileva, I., Dahlquist, E., Ellegård, K., Wäckelgård, E., 2009. Constructing load profiles for household electricity and hot water from time-use data—modelling approach and validation. *Energy and Buildings* 41, 753–768. doi:[10.1016/j.enbuild.2009.02.013](https://doi.org/10.1016/j.enbuild.2009.02.013).
- Yao, R., Steemers, K., 2005. A method of formulating energy load profile for domestic buildings in the uk. *Energy and buildings* 37, 663–671. doi:[10.1016/j.enbuild.2004.09.007](https://doi.org/10.1016/j.enbuild.2004.09.007).
- Zhou, S., McMahon, T., Walton, A., Lewis, J., 2002. Forecasting operational demand for an urban water supply zone. *Journal of hydrology* 259, 189–202. doi:[10.1016/S0022-1694\(01\)00582-0](https://doi.org/10.1016/S0022-1694(01)00582-0).

## Supplementary material

- The model and simulator source code, measured water usage data, and simulated output are available at <https://bit.ly/SyntheticHotWater>.