# A. Stochastic Modelling of Water End Use

## A.1. Mathematical Representation of the Community Water end-use

For purposes of predicting water demand or water use in a community, Creaco et al. (2016) stated that there is a high volume of scientific studies that has been carried out in the development of water use or water demand models . These models can be classified into those that break the water use into micro-components (or water end-uses) and those that lump the water use as one. This study focuses on the former. Models that break the water use into micro-components, spatially and temporally cascade the water component pulses (across different households) through a bottom-up approach to generate the overall demand.

### A.1.1 The community

In this study, the community was schematised spatially and temporally to build its water end-use model in shown in figure A.1. The residential mains water users in the community were categorised into two groups - the house connected users (group 1) and the yard-tap connected users (group 2). It was noted that there were significant differences in the water end-use behaviours between these two groups, as explained in section 2.2 of this paper, that warranted their separation. And within each group, the constituent households were considered to be homogeneous.

Rathnayaka et al. (2011) observed that the concept of representing the spatial heterogeneity of a community across different household types is common in various water end-use models. This concept has been applied in Haarhoff (2006); Blokker et al. (2010). This study extended this concept to a community of a low-income country - hence introducing new set of heterogeneous household groups.

### A.1.2 The Household

At household level, the general principles used in the generation of the total household water use are similar in principle to works by Buchberger and Wu (1995); Buchberger and Wells (1996); Rauch et al. (2003); Memon et al. (2005); Haarhoff (2006); Blokker et al. (2010); Creaco et al. (2016) but closer in the modelled water end-use details to works by Haarhoff (2006); Blokker et al. (2010). In all these works, each water end-use pulse was generated as a product of two independent pulses: the intensity of the water end-use and its frequency. Though, Creaco et al. (2015, 2016) recommend that the duration and its corresponding intensity of an end-use should be modelled as dependent variables, in this study they were maintained as independent variables.

The intensity of the water end-use represents the volume of water used for that end-use event per unit time (which was days or weeks in this study) while the frequency of the end-use represents the number of end-use events per unit time. A product of these two quantities yielded the amount of end-use water used per person in a given time unit (day). The per-capita water quantity was then further multiplied by the household population (no of adults per household) to yield the total end-use volume of water used per day in that household. Water end-use events showering and dish washing of the house-connected users group had their intensity broken down further as a product of the event duration (in minutes) and the event flow rate (litres per min).

In Memon et al. (2005); Jacobs and Haarhoff (2006); Haarhoff (2006), the end-use volume parameters of intensity and frequency were generated from community averaged intensities and frequencies of end-uses of all the households in the community. Statistically, the use of community averaged position and spread parameters presumes that the household population water uses follow a symmetric distribution like normal distribution. In this community however, this presumption would be a misrepresentation of the community water use, since the socially collected data of this community showed that most of the water use distributions across the community were skewed, and hence not symmetric. Therefore, the approach employed in Blokker et al. (2010); Rauch et al. (2003) that involves the probabilistic distribution of each water end-use parameter was also applied in this study. This approach was preferred for this study because it does not prescribe any pre-requisite distributions to any parameters, and is also well-suited for studies involving relatively small sample sizes.

All in all, mathematically the overall water use in a household (and then across all households of a group) was represented as:

$$Q = \sum_{j}\sum_{i}[a_{i,j} * b_{i,j} * n_j] \quad (1)$$

where $a_{i,j}$ represents the volume (or intensity) of water consumed per end-use($i$) per person in a household ($j$), $b_{i,j}$ represents the frequency of use events of end-use ($i$) per person in a day in household ($j$), and $n_j$ is the number of adults in household $j$. The summation is done for all end uses considered, and for all the households in the group.

The parameters ($a_{i,j}$ , $b_{i,j}$, $n_j$) are instances of probability density distribution models. For each consumer group and for each constituent water end-use parameter (that is its intensity and its frequency), their statistical position and statistical spread values were estimated from the socially surveyed data of the community. The histogram of each water end-use parameter was then fitted to each of the native probability density distributions (like Normal, Lognormal, Gamma, Weibull,Exponential, Rayleigh, and Chi-Squared), from which the probability distribution with the closest fit (as measured by a correlation coefficient) was taken as the most appropriate probability distribution model of that end-use parameter. The details of the appropriate probability distributions chosen for each water end-use parameter is represented in figure A.3, figure A.4 and figure A.5.

## A.2. Temporal Resolution

The product ($a_{i,j} * b_{i,j} * n_j$) in equation 1 represented the total water volume consumed in a single day (for toilet use, showering, hand washing) or in a single week (for laundry, for general cleaning). To simulate the water consumed over a month, the daily flow water end-uses were multiplied by 30, while the weekly flows were multiplied by 4.3. In other words, the time space was modelled as a deterministic variable in this study.

In Blokker et al. (2010), a methodology that originated from Buchberger and Wells (1996) and extended by Creaco et al. (2016) is used to describe the time space. This method considers each time ordinate as an output of a Poisson probability distribution model. Therefore, in a modelling system with small time steps for example one second simulating up to one hour (or to a day), this would create a large enough sample space to allow for statistical consistency in applying this probability distribution model.

In this study, the time evolves from a single day to one month (30 time ordinates) for toilet, showering and handwashing water end-uses and from a single week to a month (4.3 ordinates) for laundry and general cleaning water end-uses. These time space sample sizes are not adequate to ensure statistical consistency of the outputs, justifying the need for a simplified representation of the time space.

## A.3. Making the water end-use model more realistic

In general, the purpose of the water end-use model development was to provide a platform upon which the surveyed or measured water end-use volumes can be justified as representative of the water used in the community. Therefore, the the water end-use model also had to be modified further to make it more representative of the water consumed in the community, and also more representative of the monthly billing measurements.

In the measurement of household water use (for the billing process), household metering errors are experienced. Two, it was revealed from the analysis of the socially collected data that a significant proportion of the community mains water users employ rain water harvesting. The model was therefore updated to include: a parameter for household metering errors, a parameter for the effect of rainwater harvesting on the household, and finally a parameter to allow for temporal summation of the end-use volumes for all the days of the month (that is 30 for daily end-uses, and 4.3 for weekly end-uses). The temporal evolution parameter was introduced because it is expected there could be a temporal evolution trend with the water consumption that was not considered in the model development.

The final updated stochastic water end-use for the community model that estimates the total amount of mains water consumed by a group of households in a month was represented as:

$$Q_s = \sum_{i=1}^{i=h} \sum_{j=1}^{j=e} [m_j * rw_j] * [a_{i,j} * b_{i,j} * n_j] * [ts_i] \quad (2)$$

where $h$ and $e$ represent the total number of households and end-uses considered respectively. The parameters are still represented the same way as in equation 1, except for the rain water harvesting parameter $rw_j$, the temporal frequency parameter $ts_i$, and the metering error parameter $m_j$ that were added.

In addition, the monthly water consumption data for the community shows that the the number of households is not a constant from month to month, it varies randomly. Because of this, the water end-use model was therefore modified to compute the total monthly consumption of each month independently for each month, in that the number of households billed in that month was used as an input into the stochastic model (for the number of households simulated).

Further, since this was a stochastic simulation, each month was simulated for a thousand 1000 times (under a Monte Carlo simulation) to provide a distribution of monthly consumptions, from which a monthly consumption distribution was generated.

## A.4. Calibration and Validation

In summary, the water end-use model in equation 2 generates instances of the monthly consumption of the community - which should statistically and numerically be equivalent to the measured monthly (and billed) water consumption of the community. And if the two datasets (the computed and the measured water consumptions) are both statistically and numerically equivalent, then their statistical centeralised values (mean or median) and their cumulative density functions (a representation of their statistical frequency distribution) should converge. Therefore, the comparison of these datasets is not one of comparing statistical moments like mean or median, but also of their cumulative frequency distributions.

This calibration approach for stochastic models has been applied in similar works by Blokker et al. (2010); Creaco et al. (2016). In both works, the authors agree that total monthly water consumptions can be generated by summing up the individual ordinate consumptions, confirming the procedures applied in this model. Again, both authors employ cumulative density functions are a calibration procedure to confirm the statistical consistency between the simulated and measured observations. Therefore, the objective of the calibration was to ensure that the measured (billed) monthly total water consumption falls within the 95% confidence interval bounds of the monthly total water consumption simulations, and that the both cumulative density functions are similar in shape. Mathematically, this is illustrated as:

$$Minimise \ \ (1) \quad |Q_s - Q_m| \tag{3}$$

$$Minimise \ \ (2) \quad |CDF_s - CDF_m| \tag{4}$$

BY: *adjusting various water intensity end-use and household size parameters*

$Q_s$ represents the median of the simulated total monthly water end-use volumes, while $Q_m$ represents the measured or billed total monthly water consumption in the community. $CDF_s$ is the cumulative distribution function of the simulated median total water use volume, while $CDF_m$ is the cumulative distribution function of the measured or billed water volumes. Objective one (1) was not necessarily focussed on ensuring that the median simulated water volume is comparable to the measured volumes, but on ensuring that the measured volumes like within the simulated 95% confidence interval boundaries.

One of the major weaknesses of the social surveying of water end-use investigation was the estimation of the water use volumes, that is, the water end-use intensities. The calibration procedure therefore focused on the adjustment of the water end-use intensities. Each water use intensity parameter (the position and or spread) of interest was increased or decreased incrementally with a factor, till both calibration objectives above were satisfied. The final factors employed in the calibration are shown in table A.1. In addition, the household size was another parameter that was calibrated. According to the recently released national census details of household size, the survey data collected in this community showed minor variations with those shown in the national census. This difference warranted a modification of the household size as a calibration parameter. The calibration results are shown in figure A.6.

The calibration procedure was carried out on the months during the dry season, and the wet season months were employed to validate the calibration. In this, it was assumed that the only difference between the wet and dry seasons, was the rate of rain water harvesting taking place between both periods. So, in the validation period, only the rain water harvesting parameter was adjusted all the other parameters were not adjusted any further. The validation comparisons between the measured and the simulated total monthly consumptions were comparable, as shown in figure A.8.

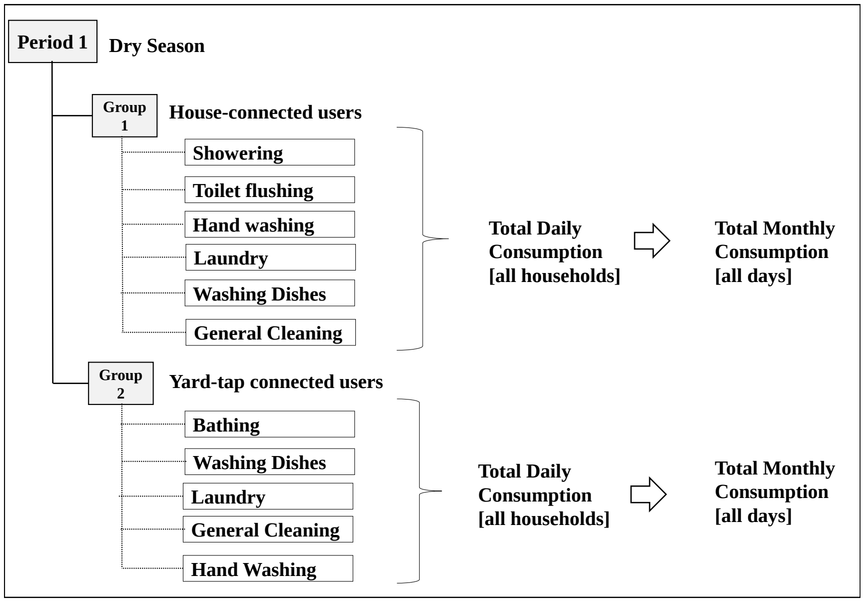



**Figure A.1:** *A schematisation of the water user groups in the community*

## REFERENCES


Blokker, E. J. M., Vreeburg, J. H. G., van Dijk, J. C., 2010. Simulating Residential Water Demand with a Stochastic End-Use Model. Journal of Water Resources Planning and Management 136 (1), 19–26.

Buchberger, S. G., Wells, G. J., 1996. Intensity, Duration, and Frequency of Residential Water Demands. Journal of Water Resources Management 122 (1), 11 – 19.

Buchberger, S. G., Wu, L., 1995. Model for Instantaneous Residential Water Demands. Journal of Hydraulic Engineering 121 (3), 232–246.

Creaco, E., Farmani, R., Kapelan, Z., Vamvakeridou-Lyroudia, L., Savic, D., 2015. A model for generating residential water demand pulses with correlated duration and intensity. In: e-proceedings of the 36th IAHR World Congress. IAHR, The Hague, The Netherlands.

Creaco, E., Kossieris, P., Vamvakeridou-Lyroudia, L., Makropoulos, C., Kapelan, Z., Savic, D., 2016. Parameterizing residential water demand pulse models through smart meter readings. Environmental Modelling & Software 80, 33–40.

Haarhoff, J., 2006. Structure and data requirements of an end-use model for residential water demand and return flow. Water SA 30 (3), 293 – 304.

Jacobs, H., Haarhoff, J., 2006. End-Use Modelling as a Means to Predict the Effects of Water Demand Management. In: Biennial Conference of the Water Institute of Southern Africa(WISA). No. May 2002. Water Research Commision(WRC), Durban, South Africa, pp. 1 – 9.

Memon, F. A., Butler, D., Han, W., 2005. Economic assessment tool for greywater recycling systems. In: Institution of Civil Engineers - Engineering Sustainability. No. September. pp. 155–161.

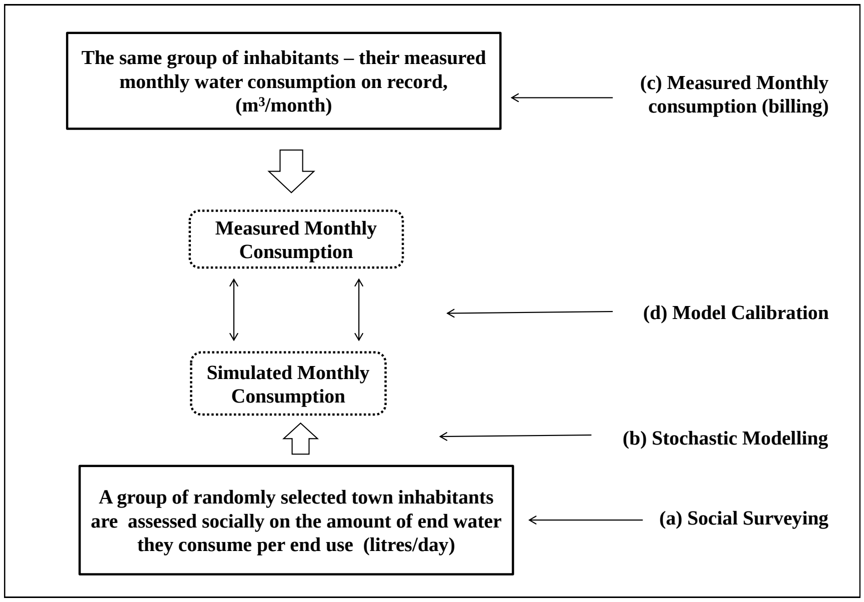



**Figure A.2:** *The Modelling methodology approach applied*

Rathnayaka, K., Malano, H., Maheepala, S., Nawarathna, B., George, B., Arora, M., 2011. Review of residential urban water end-use modelling. In: 19th International Congress on Modelling and Simulation. No. December. Perth, Australia, pp. 3321 – 3327.

Rauch, W., Brockmann, D., Peters, I., Larsen, T. a., Gujer, W., Feb. 2003. Combining urine separation with waste design: an analysis using a stochastic model for urine production. Water research 37 (3), 681–9.

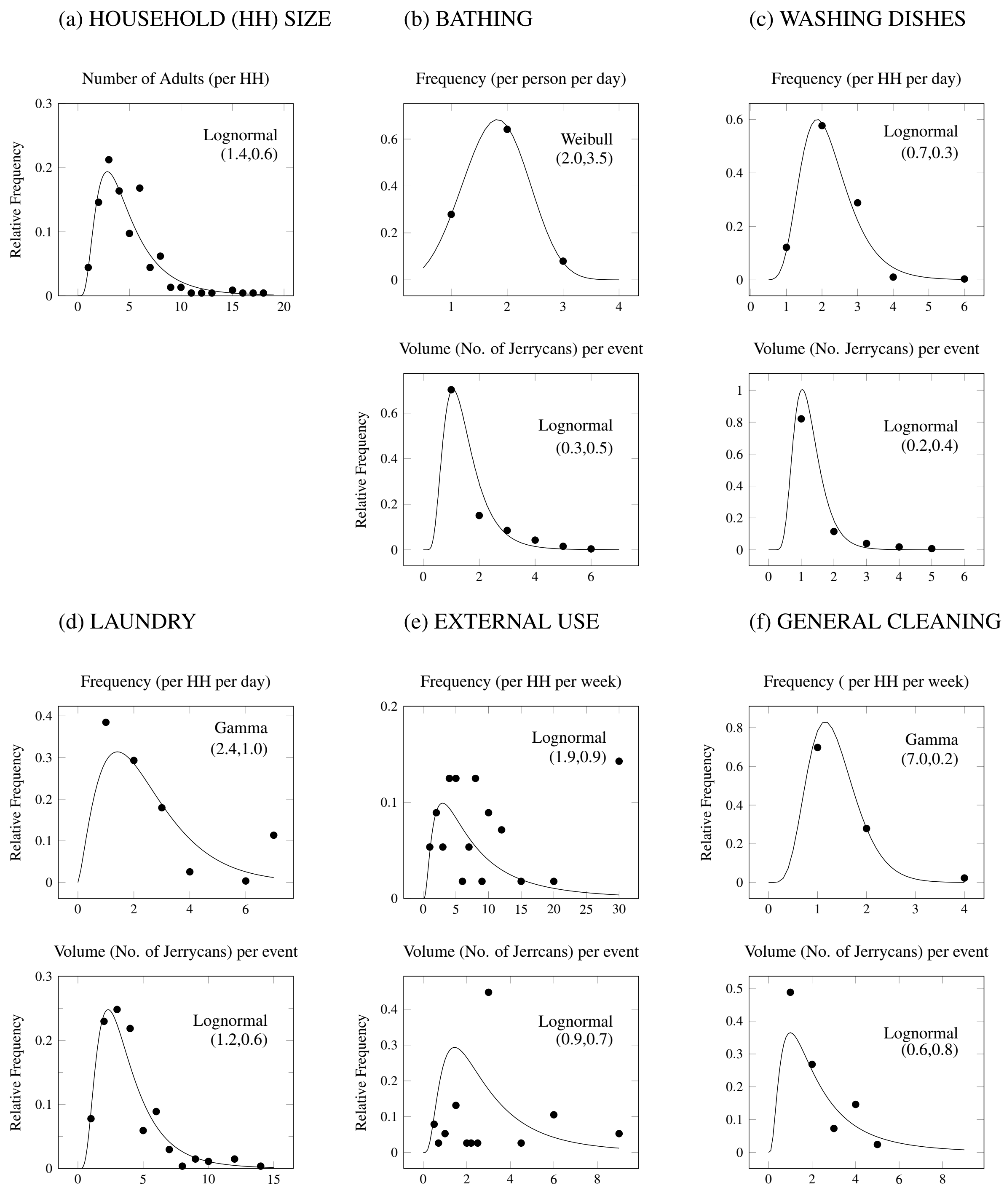



**Figure A.3:** *Probability fit models (continuous line) used to represent frequency distribution data (black dots) of (a) household size and water uses ( of (b) bathing, (c) dish washing, (d) laundry, (e) external use, and (f) general cleaning) across* ***yard-tap water users***

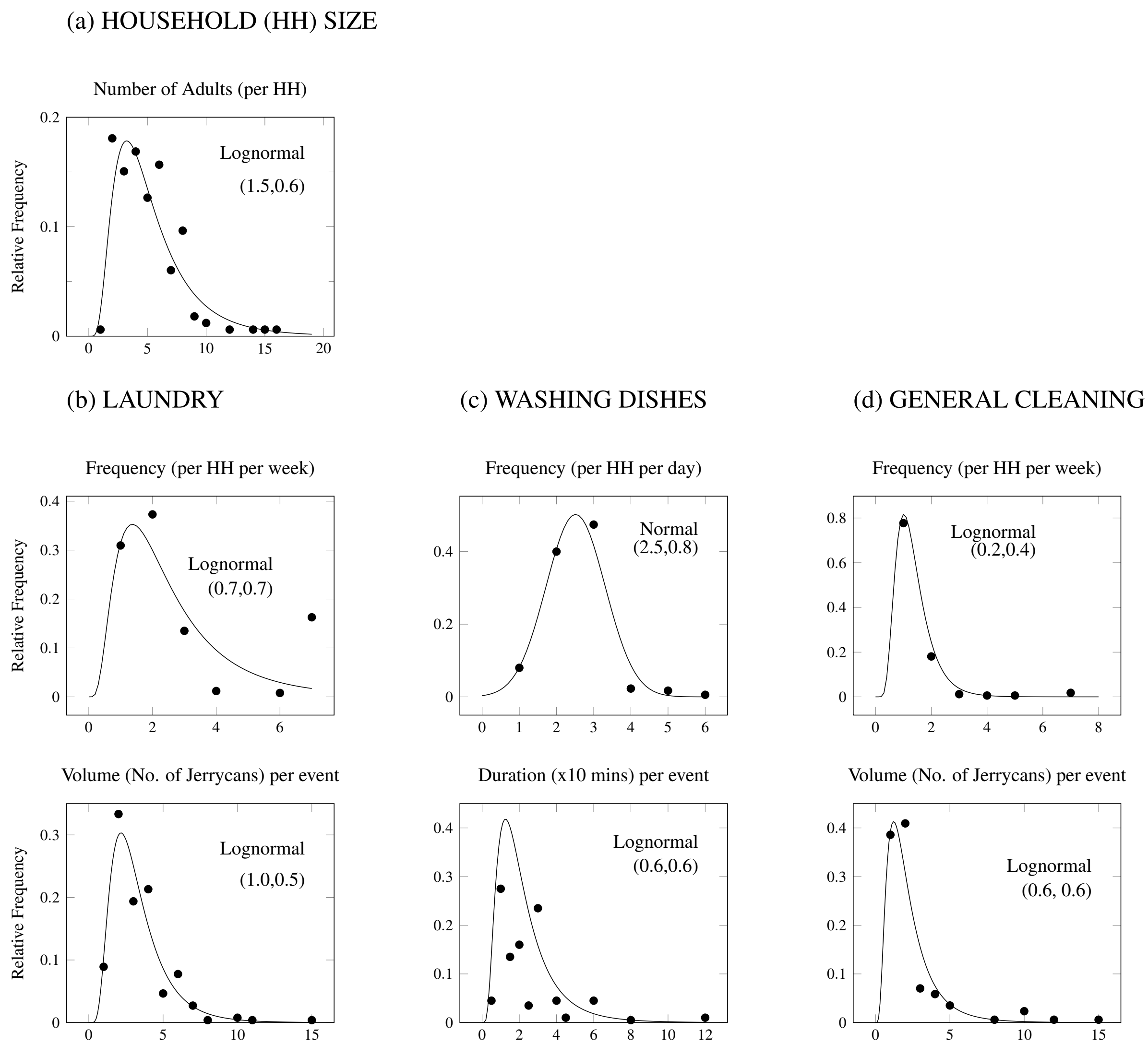
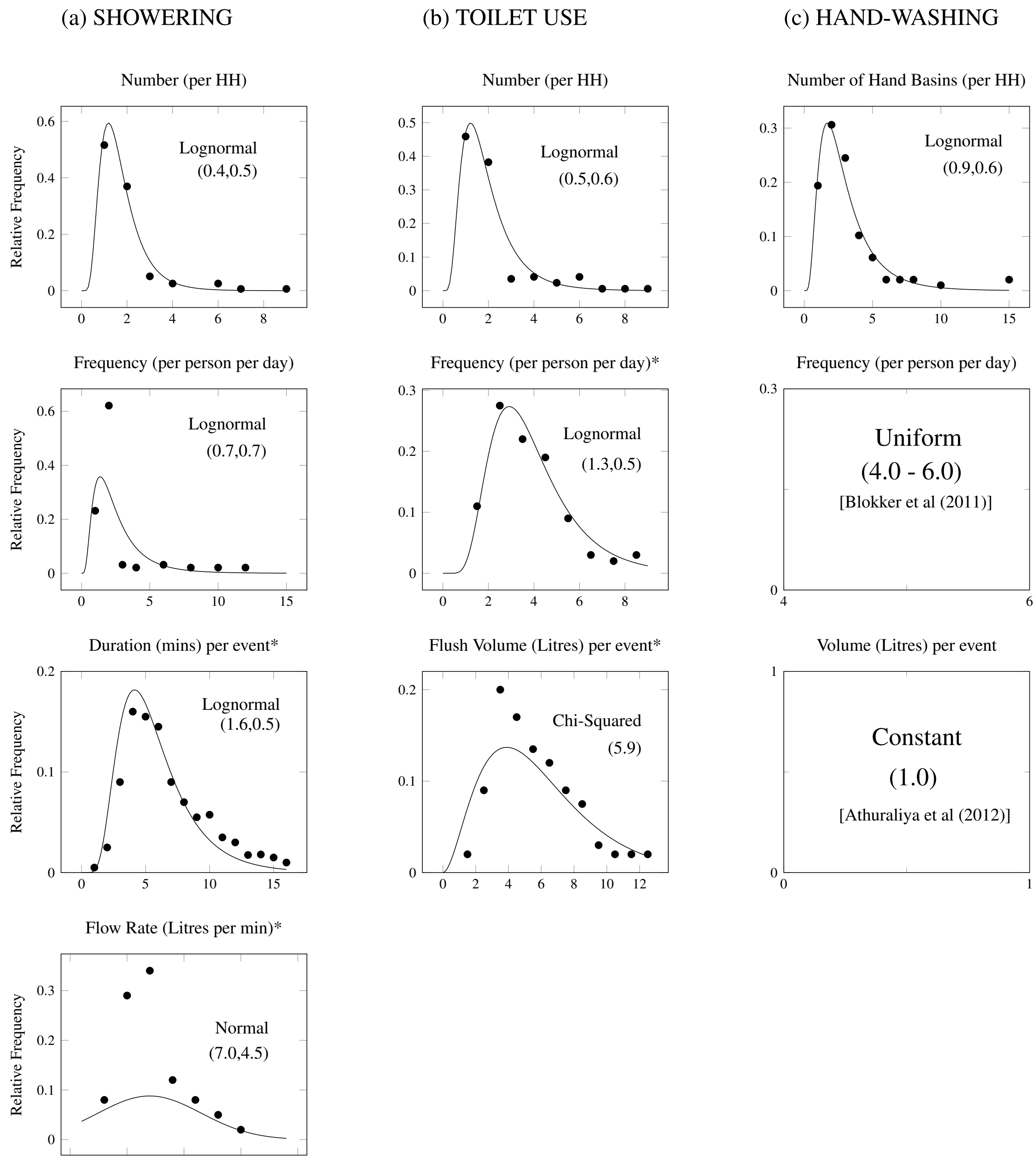



**Figure A.4:** *Probability fit models (continuous line) used to represent frequency distribution data (black dots) of (a) household size and water uses (of (b) Laundry, (c) dish washing, and (c) general cleaning ) across* ***house-connected water users***

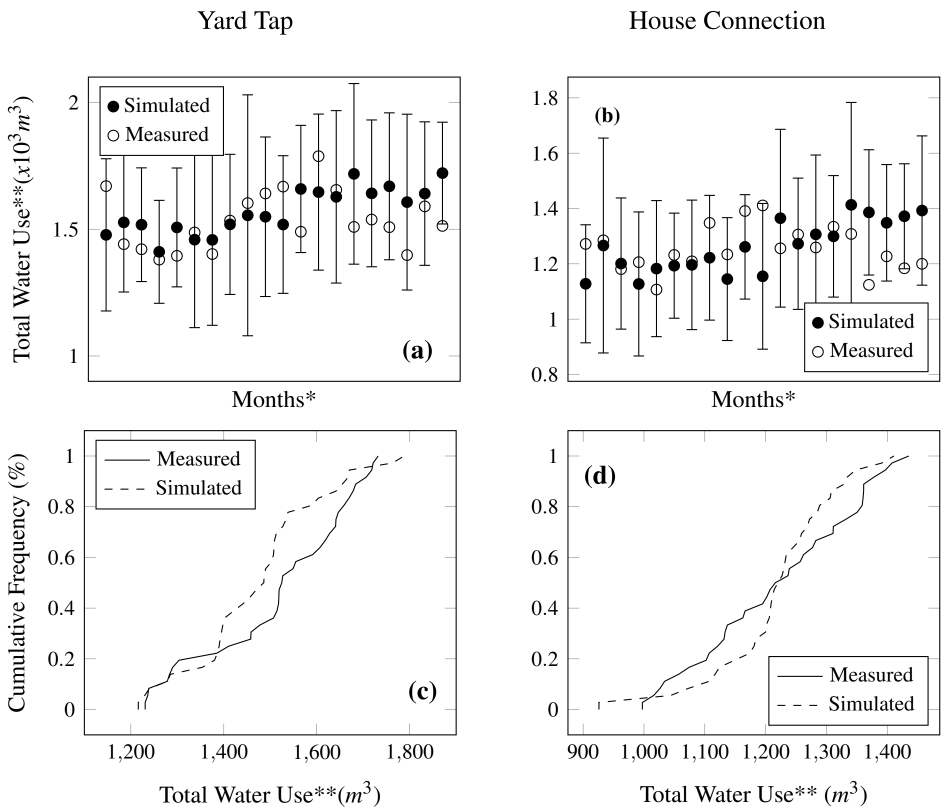



*Employed data from Yarra Valley Residential Water Use Study (Australia) in Athuraliya et al (2012)

**Figure A.5:** *Probability fit models (continuous line) used to represent frequency distribution data (black dots) of more water uses ((a) showering, (b) toilet use, and (c) hand-washing) across **house-connected water users** in the community.*

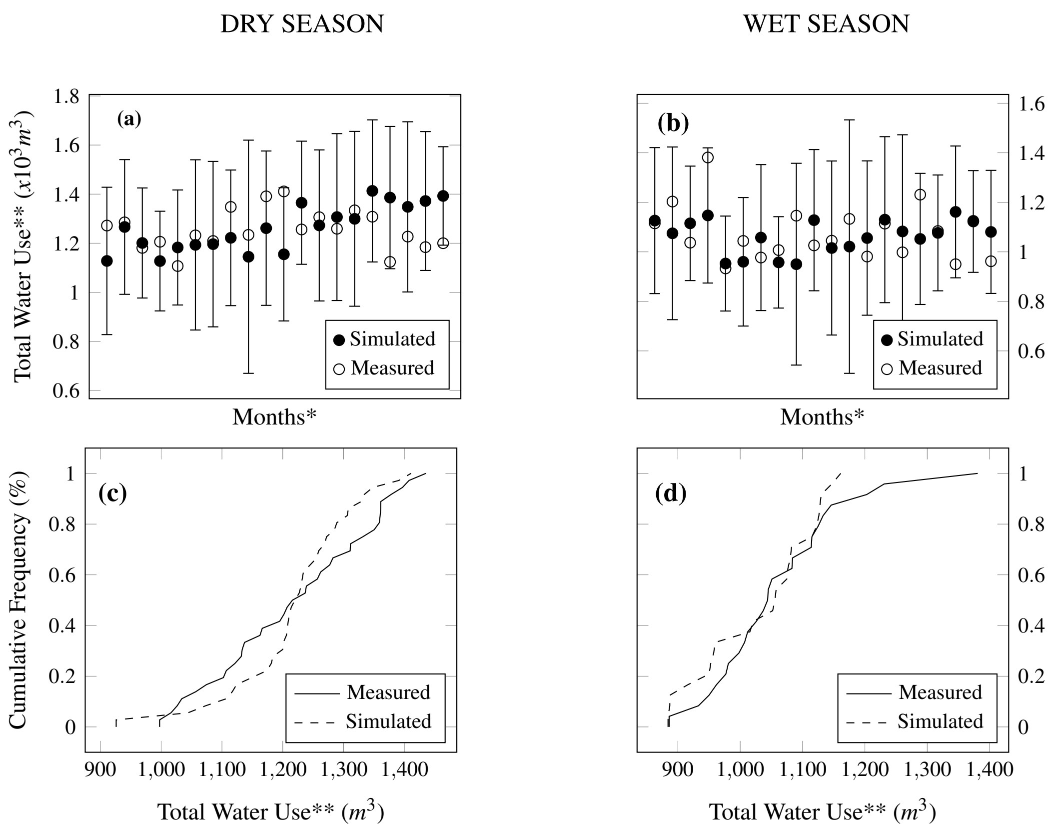
**Table A.1:** *Calibration Parameters*

| Category | Water end-use | Factor | Remark [Target Parameter] |
|---|---|---|---|
| Yard Taps | Bathing | 0.7 | No. of Jerry cans |
| | Washing Dishes | 1.4 | No. of Jerry cans |
| | Household Size | 1.18 | General Increase |
| | Metering Error | 0 | – |
| House Connections | Toilet Use | 0.5 | No. of flushes |
| | Showering | 0.5 | Intensity Volume |
| | Washing Dishes | 0.5 | Flow Rate |
| | Household Size | 1.35 | General Increase |
| | Metering Error | 0.1 - 0.2 | Uniform Distribution |
| General | Rain Water Use [Dry] | 0.15 - 0.25 | Uniform Distribution |
| | Rain Water Use [Wet] | 0.23 - 0.50 | Uniform Distribution |
| | Temporal Evolution factor | 1.0 | – |



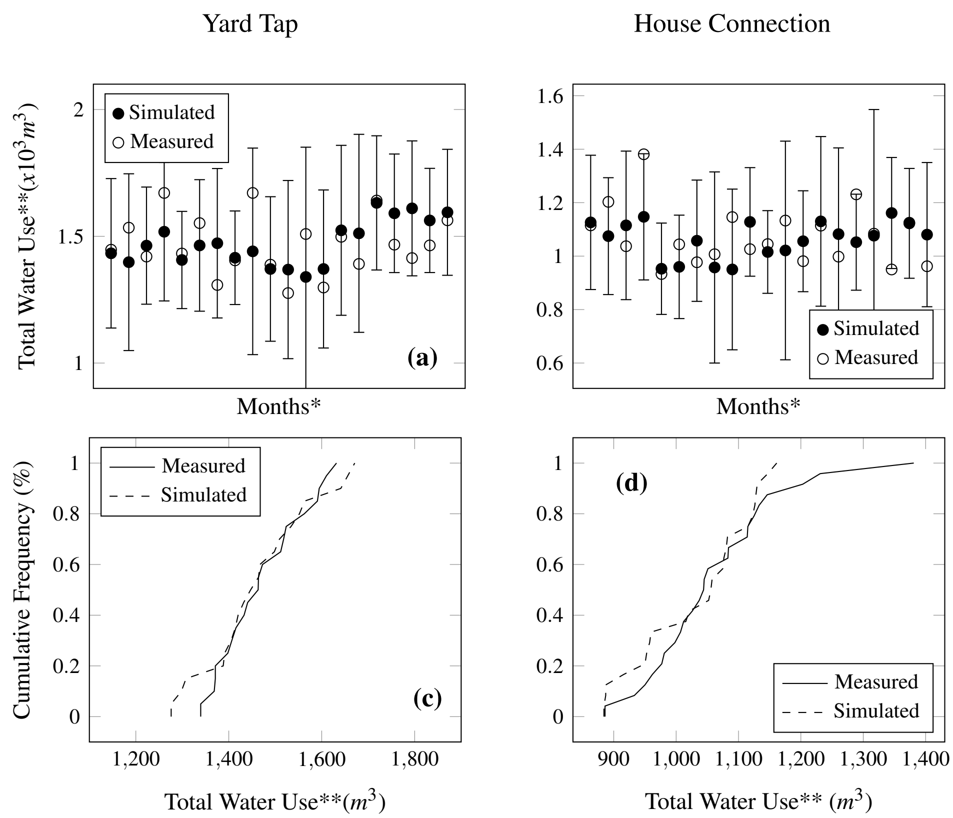

[*] The order or nature of the month was not a subject of the research apart from the number of households measured in that month, each month was simulated independently. That is why the months are not labelled on the x-axis. [**] Total Water Use - represents the total water used up by all the households in a given group in a given month.

**Figure A.6:** *Overall calibration process results for total monthly water consumption of* ***for yard-tap water users*** *shown in (a) and (c), and also for* ***house-connected water users*** *shown in (c) and (d).*




[*] The order or nature of the month was not a subject of the research apart from the number of households measured in that month, each month was simulated independently. That is why the months are not labelled on the x-axis. [**] Total Water Use - represents the total water used up by all the households in a given group in a given month.

**Figure A.7:** *Overall calibration process results for total monthly water consumption* ***for house-connected water users*** *during (a) dry season and (b) wet season, and also for their respective cumulative distribution functions in (c) and (d).*




[*] The order or nature of the month was not a subject of the research apart from the number of households measured in that month, each month was simulated independently. That is why the months are not labelled on the x-axis. [**] Total Water Use - represents the total water used up by all the households in a given group in a given month.

**Figure A.8:** *Validation process results for total monthly water consumption of* ***for yard-tap water users*** *shown in (a) and (c), and also for* ***house-connected water users*** *shown in (c) and (d).*