Nour A. Attallah[1,2] and Jeffery S. Horsburgh[1,2]

[1]Department of Civil and Environmental Engineering, Utah State University, 4110 Old Main Hill, Logan, UT 84322-4110, USA

[2]Utah Water Research Laboratory, Utah State University, 8200 Old Main Hill, Logan UT 84322-8200

Corresponding author: Nour Attallah (nour.atallah@usu.edu)

Key Points:

- We developed an end use water demand model that simulates detailed household water use and potential savings using Monte Carlo techniques.
- The model is generally applicable and can be modified to simulate the detailed water end uses for different cities.
- Results can assist water utilities in identifying opportunities for incentive programs and to encourage water conservation.

Abstract

We present a model of indoor residential water use that estimates water demand and conservation potential by end use for a target community by simulating indoor water end use events at a household level. The model uses end use event data from a set of representative residential households to simulate a larger community and advances existing end use models by: 1) accounting for an expanded set of indoor water end uses; 2) considering the variability in flowrates, durations, and volumes for end use events over different days of the week; and 3) providing a generalized approach for simulating indoor water usage and potential conservation at the city level. The model simulates residential water use behavior in individual households by randomly sampling water end use events for different end use types for each day of the week and then aggregating the sampled water end use events per day to estimate the daily water use per household. We used the model to evaluate a set of technological and behavioral conservation actions to quantify the conservation potential in each simulated household as well as aggregated to the city level. We evaluated the performance of the model in predicting the observed average daily water use of households in Logan City, Utah, USA and compared against other common water demand models to demonstrate the model's reliability. The results of this paper are reproducible using openly available code and data, representing an accessible platform for advancing water demand modeling using detailed water end use data.

1 Introduction

With rapid growth of urban populations and limited resources, improving the short and long term planning and management of urban water supply has created a persistent need to develop and adopt alternative management schemes

(Gaudin, 2006). In the last decade, several water demand forecasting and simulation strategies have been proposed to promote water conservation and water demand management (Koutiva and Makropoulos, 2019). Residential water demand modeling aims to simulate the water demand behavior of households and how it is influenced by management strategies and external factors (e.g., environmental, social, etc.). Since the 1960s, many residential water demand modeling-oriented studies have been published, where monthly water use data have been frequently used for management programs. For example, in 2008, Aurora Water tracked and analyzed residential monthly water use records for the city of Aurora, Colorado, USA for a period of one year and investigated the impacts of different demand management programs enacted for different months (e.g., price, water restrictions, and rebate programs) (Kenney et al., 2008). Despite their dissimilar contexts and techniques, residential water demand modeling studies have mostly shared the same procedure in simulating water demand that first determines a set of independent variables to be used in the model for predicting water use (e.g., number of residents, age of the property, etc.) and second determines an estimation method or model formulation (Worthington and Hoffman, 2008).

The major determinants of water use included in most existing demand modeling studies have been the number of residents in a household, the existence of swimming pools, precipitation rates, price of water, and the outdoor lot size (Wentz et al., 2013; Kenney et al., 2008; Haley et al., 2007; Arbués et al., 2003; Dalhuisen et al., 2003; Gaudin, 2006; Mayer et al., 1999; Espey et al., 1997). Regression models have been the prime estimation method adopted in several studies to simulate and predict residential water use, including ordinary least squares regression (Agthe and Billings, 1980; Carver and Boland, 1980; Schefter and David, 1985), two-stage least squares (Chicoine et al., 1986; Renwick et al., 2019; Stevens et al., 1992), three-stage least squares (Chicoine et al., 1986), instrumental variable approach (Higgs and Worthington, 2001; Martinez-Espiñeira, 2002; Renwick et al., 2019), maximum likelihood approach (Hajispyrou et al., 2002), generalized least squares approach (Gaudin et al., 2006; Höglund, 1999), and generalized method of moments approach (Garcia and Reynaud, 2004; Nauges and Thomas, 2003).

With respect to spatial scale, residential water demand models have been developed at district levels (Mamade et al., 2014), household levels (Kontokosta and Jain, 2015), and water end use levels (Cahill et al., 2013). At district levels, water demand models have used a spatial scale consisting of a group of residential households in one or more cities. Such a spatial scale is typically relevant for infrastructure planning and long term water demand forecasting (di Mauro et al., 2020). At the household level, water demand models have been primarily used to estimate peak water demand and timing with output estimates for a single household (di Mauro et al., 2020). At the end use scale, water demand models have been used to better understand residential water use behavior, the consumption rate of each water end use inside household units, and to develop targeted water end use conservation actions. Given the variability of models at different spatial and temporal scales, the required input data and model output also varies. The temporal scale of district-level water demand modeling varies from hourly to monthly and annual intervals (di Mauro et al., 2020), whereas the temporal scale of household and end use models can vary from minutes to one day. The majority of household scale models use data inputs collected with a time resolution of 15 minutes to one day. End use scale models use data inputs gathered at seconds to one minute resolutions.

Achieving an appropriate balance between water supply capacity expansion and water conservation requires more mechanistic and detailed modeling approaches that allow water managers to control for demographic, behavioral, and social variation in water use across households (Jorgensen et al., 2009). This can be vital for utilities where water is scarce and developing more water supply is expensive or even impossible. In addition, given new standards and technologies in water end uses and demographic and behavioral heterogeneities of water consumers, growth in water demand is unlikely to be homogenous. Thus, detailed water modeling and targeted conservation actions may be necessary planning tools for water supplying agencies. Over the last two decades, models have started to include behavioral factors (e.g., shower duration) (Matos et al., 2013; Romano and Kapelan, 2014; Talebpour et al., 2014) and geospatial factors (e.g., climate) (Maeda et al., 2011; Praskievicz and Chang, 2009; Kuski et al., 2020). The emergence of smart metering technology and the high temporal and spatial resolutions of recent water end use monitoring studies has enhanced the development of residential water use models that account for economic, behavioral, and geospatial factors (Cominola et al., 2015; Makki et al., 2015). Some of these more advanced models integrate end use data to simulate the water demands of individual water end uses such as faucets, showers, toilets, etc. and then aggregate end uses to estimate consumption at the household level (Cominola et al., 2018). Coupling such an event level water demand model with demographic surveys about households and their residents including number of residents, age distribution, age of household, and characteristics of water-using fixtures can lead to more realistic simulations of water demand patterns that compare well to those that have been observed.

For example, Blokker et al., (2009) developed a water demand model to predict water use from end use measurements. Statistical data from a survey conducted across 46 households in the city of Amsterdam, Netherlands, including census data and the average age in each household, were incorporated into the model along with water use data obtained from different end uses. The frequency of water use for each event type was simulated using a Poisson distribution, water use volume for each individual event of different end use types was assumed to be constant, and the flowrate of water use for each event was simulated as a lognormal distribution. Williamson et al., (2002) modified the model developed by Blokker et al., (2009) to develop an enhanced water end use model. Modifications included changing fixed volumes of water use for different end uses to probability distributions, which allowed them to account for the water use variability present in each end use type. However, Williamson et al., (2002) used water end use data collected from only 20 residential households in South East Queensland, Australia and generated probability distribution curves for sampling using end use data and statistical data from a survey conducted across those households. End use probability curves were used to sample water consumption, while statistical probability curves were used to sample demographic variation of residents, including the number of residents of a simulated household and technical performance of its water end uses.

Table 1 lists characteristics of several approaches for modeling residential end uses, including: indoor end uses incorporated, whether the model can simulate conservation potential, software used, whether the model uses an open source software license, whether the model accounts for daily variation in water use, and whether the model is generalizable to other communities. Despite the recent improvements in residential water use modeling established by these models, some important variables have been left out or not adequately integrated into the models. This includes not accounting for all different types of water end uses, assuming constant flow rate and/or constant water use volume for all end uses of the same type, and not having a realistic probability for occurrence of water use events over different days of the week. In 2011, a team of researchers conducted a review study of the existing residential urban water end use models and concluded that the ability of existing models to simulate water end use demands especially at a city scale is limited (House-Peters and Chang, 2011).

In this paper, we present an end use water demand model that addresses these gaps in prior modeling efforts reported in Table 1 and that is aimed at improving understanding of residential water use behavior and promoting water management and conservation strategies for water utilities. The model described in this paper simulates a more complete set of indoor water end uses than other models and uses realistic probability of occurrence for all events and their associated features (frequency, volume, duration, and flowrate) instead of assuming average values for these features. The model accounts for heterogeneity of water use behavior amongst different residential households by using an event dataset drawn from a representative set of households. The model is also open source for further testing and reuse.

Table 1. Approaches used for different end use modeling studies

Study	This paper
Incorporated end uses	Faucet, toilet, shower, clothes washer, dishwasher, b
Water conservation prediction	Yes
Software used	Python
Open source software license	Yes
Daily variation per end use type	Yes
Representative set of households used in the simulation	Yes
Handling unclassified events	Yes
Variable water use per end use type	Yes

Study	This paper
Model propagation at city scale	Yes

The simulation process results in estimates of water use for a group of households that reflect realistic variability in water end use technologies and residents with diverse water use behavior. We utilized this detailed technical and behavioral information to investigate a set of water conservation actions and quantified their associated water saving potential. Technological practices included those designed to reduce water irrespective of the residents' behavior (e.g., retrofitting an inefficient showerhead). Behavioral practices focused on changing residents' habits irrespective of the technology being used (e.g., fewer showers). Water use savings associated with these actions was calculated as the difference in water use before and after conservation actions were implemented. This study was focused on answering the following research questions: a) How can improving the representation of water end uses at a detailed level within a water demand model improve our ability to predict residential water use and the effects of conservation actions? b) What is the water saving potential for individual homes as well as aggregated to a city level associated with different technological and behavioral conservation actions designed to reduce current indoor water use?

The case study presented demonstrates how detailed water end use records from an existing study can be used to simulate the water use behavior of residential households for which there is no detailed water end use data available. In the case study application, we used the simulation results to analyze the variability of water use in terms of timing and distribution of end uses, efficiency of end uses, and water conservation potential of residential households in the city of Logan, Utah, USA. We demonstrate how the model is generally applicable and can be modified to simulate the detailed water end uses of other cities. Applying the model requires availability of monthly water use records for the simulated households and the existence of a sample of households from a detailed water end use dataset such that the water use behavior of the sample households is representative of the water use behavior of the simulated households.

2 Materials and Methods

2.1 Water End Uses

We identified seven indoor water end uses to be incorporated in the water end use demand model, including faucet, toilet, shower, bathtub, clothes washer, dishwasher, and unclassified (events not associated to any end use type, e.g., leaks). In the U.S., these are the main water end uses expected in single-family residential households. To evaluate the efficiency of these end uses, we used specifications from the current federal standard defined by the U.S. Energy Policy Act of 1992 (DOE, 1992), the Environmental Protection Agency's (EPA) Energy Star Program (EPA, 2021a), and the U.S. EPA WaterSense efficient fixtures (EPA, 2021b). The Energy Policy Act of 1992, which became a law in

1994, mandates a maximum water use volume or flowrate for different end use fixtures manufactured and installed in the U.S. after 1994 and was designed to encourage manufacturing of high performing, water efficient fixtures. Based on these specifications, we divided faucet, toilet, and shower events into three categories: inefficient events, typical events, and efficient events. Inefficient events are those that have water use volumes or flowrates higher than the maximum water use volume or flowrate mandated by the U.S. Energy Policy Act of 1992. Typical events are those that have water use volumes or flowrates less than the maximum mandated standard by the U.S. Energy Policy Act of 1992 and higher than the EPA WaterSense program specifications. Efficient events are those that have water use volumes or flowrates less than or equal to the EPA WaterSense program specifications. For clothes washer and dishwasher events, we used the specifications defined by the EPA EnergyStar Program to classify events as efficient, typical, or inefficient (Table 2). We assessed the efficiency of bathtub filling events using the size of a bathtub. Standard bathtubs can hold up to 300 liters (L) of water. Smaller bathtubs can hold up to 150 L of water. However, since bathtub filling events do not use the full capacity of the bathtub, we assumed that a bathtub filling event will use approximately two thirds of its capacity. Based on that, we identified efficient bathtub filling events as those that use less than or equal to 100 L of water, typical events as those that use between 100 and 200 L of water, and inefficient events as those that use more than 200 L of water. Table 2 summarizes the technical performance of each end use type according to the Federal Standard and EPA specifications.

Table 2. Technical performance by end use type.

End use type	Inefficient event	Typical event	Efficient event
Toilet	Volume > 6.1 LPF[a]	4.8 LPF < Volume < 6.1 LPF	Volume < 4.8 LPF
Faucet	8.3 > LPM[b]	5.7 LPM < Flowrate < 8.3 LPM	Flowrate < 5.7 LPM
Shower	Flowrate > 9.5 LPM	7.6 LPM < Flowrate < 9.5 LPM	Flowrate < 7.6 LPM
Clothes washer	Volume > 110 L/load	70 L/load < Volume < 110 L/load	Volume < 70 L/load
Dishwasher	Volume > 13 L/cycle	6 L/cycle < Volume < 13 L/cycle	Volume < 6 L/cycle
Bathtub	Volume > 200 L/filling	100 L/filling < Volume < 200 L/filling	Volume < 100 L/filling

[a]L per flush

[b]L per minute

2.2 End Use-Level Water Demand Model Formulation

An end use water demand model can be formulated based on the premise that total water use for a household is the sum of all of the end uses of water. Given that, the total water use volume for an individual simulated household for a given day can be calculated as:

$$V_{T,D} = \left(\sum_{i=1}^{n} B_i V_{i,D} \right) \tag{1}$$

where $V_{T,D}$ is the total water use volume for a household (L) on day of the week D, B_i is a coefficient indicating the absence (0) or presence (1) of an end use i, $V_{i,D}$ is the volume of water used by end use type i during day of the week D (L), and n is the number of end uses within the simulated household. The volume of water consumed during the day by unclassified events that cannot be prescribed to a particular end use (e.g., leaks) (L), is modeled as a separate end use.

Water end use technical performance, number of residents in a household, water use behaviors, variation in occupancy of the household on different days of the week, and demographic factors that vary across households and individual water use events within a household all affect the total volume of water used by each end use during a day ($V_{i,D}$). In order to account for this variability, the volume of each individual end use event is simulated in the model. To do so, the model accounts for: 1) the number of individual water use events from each end use type that occurs during a day of the week D, or frequency ($f_{i,D}$), and 2) the volume of each end use event j of type i ($v_{i,j}$) (Eqn. 2):

$$V_{T,D} = \left(\sum_{i=1}^{n} B_i \sum_{j=1}^{f_{i,D}} v_{i,j} \right) \tag{2}$$

This enables the model to estimate the total daily water use for each of the different end uses while accounting for variation in volumes of each individual water use event across each of the different end uses. Instead of assuming average volume and frequency estimates for events of each end use type, the model simulates the frequency of event occurrence for each household and day along with the volume of each individual end use event using a Monte Carlo sampling approach. We chose a Monte Carlo sampling approach rather than assuming average volume and frequency because we have observed that volume and frequency are not consistent across homes or days of the week (Rosenberg et al., 2007), and we were interested in the conservation potential associated with different event types, which depends on variability in event volumes. Another approach that could be used consists of choosing individual homes with detailed end use data and using the events for those homes without manipulation. However, given the relatively small number of homes with detailed end use data, this would result in reusing the same events over and over which may not be representative of the distribution of events from the much larger set of homes to be simulated. A Monte Carlo simulation approach provides a wider variety of water use events to sample from and results in a smoother, more realistic distribution of events for simulated homes, providing an opportunity to simulate a wider variety of water use behaviors reflective of a broader group of residential water users.

Event frequency values for each day (Monday - Sunday) and a volume for each simulated event are drawn from cumulative distribution functions (CDFs) for frequency and volume derived from detailed event data for a subset of the households from a detailed water use monitoring study. By doing so, the model is able to simulate variability in water use across different households and across the different end use types.

To satisfy the input requirements of the model, detailed, disaggregated water end use event data obtained from smart metering studies are needed. Detailed water end use event data consist of individual water use events for a household and additional information about each event, including the date, start time, volume, duration, and flowrate. To simulate residential homes within a city using the model formulation above, a representative number of households with their detailed water end use data can be scaled up to the level of all residential homes within a city. However, most cities lack detailed water end use datasets. Existing end use studies have necessarily focused on a small group of households within a municipality boundary, and there have been few large scale studies to date (Boyle et al., 2013). Furthermore, existing studies may include bias associated with their spatial distribution, with most of them having been conducted in Queensland, Australia and scattered cities across the U.S. (e.g., Jorgensen et al., 2009; Makki et al., 2015; Willis et al., 2013). These limitations have restricted existing end use demand models to places where detailed end uses of water studies were conducted.

For cities where no detailed water end use datasets are available, an alternative is to draw a sample of households with their detailed water end use data from one of the existing end use studies such that the water use behavior of the drawn sample is representative of the water use behavior of the households to be modeled (see Section 2.5 for how we did this for our case study). Similarity in water use can be quantified using data that are widely available for different cities (e.g., monthly billing data). The monthly water use data for households to be simulated can be used to calculate the overall water use probability distribution for those households, where the distribution shows the probabilities of occurrence of different monthly water use volumes for all households within the city. Then, monthly water use volumes for households with detailed water end use data are used to draw a representative sample of households from existing end use studies.

The input to the model is a comma-separated values (CSV) file that contains the water end use event data for the representative sample of households and the number of households to be simulated. The output of the model is a CSV file that contains the simulated water end use events for the number of residential households in the input (e.g., all single family residential homes in a modeled city). In the following sections, we first describe the Monte Carlo sampling procedure used in executing the model. We then describe in more detail how the model inputs were developed for our case study application in Logan, including how the representative sample of households with detailed water end use data was selected and how CDFs input to the Monte Carlo sampling procedure were constructed. Following that, we describe how we validated the “existing conditions” model simulation results and then implemented the ability to simulate water conservation strategies.

2.3 Model Execuion Procedure

The model initiates the sampling procedure from CDFs for event frequency so that water use behavioral factors for a simulated household are related. For example, a simulated household with a high toilet flush frequency on one day is expected to have high flush frequency across all days. We used the cumulative distribution function that characterizes frequency of water use to rank households as having low (< 33rd percentile), medium (33rd - 66th percentile), or high (>66th percentile) frequency of water use, depending on their percentile ranking of number of events per day. We then devised a Monte Carlo sampling procedure to ensure that the frequency of end use events of the same type within the same simulated household for different days of the week were drawn randomly, but from the same group of frequencies. For example, if the frequency of the first end use type is sampled from the low frequency group (< 33rd percentile), the frequency values for all simulated days for that end use type for the same household are sampled from the low frequency group. The model assumes that once a high, medium, or low frequency has been set for an end use type for a simulated household, that end use type for the simulated household remains in that category to preserve the same frequency behavior for end uses of the same type throughout different days of the week.

In order to account for the variation of technical performance of end uses across different households, we constrained the sampling process for events of the same type to choose only events within the same level of technical performance (i.e., inefficient, typical, or efficient). For example, if the first event of one type is sampled from efficient events, all subsequent events of the same type are randomly sampled from the group of efficient events. This was implemented in our sampling procedure by sub-setting flowrates and/or volumes from different end use types into different groups based on their technical performance. We then devised a Monte Carlo sampling procedure to ensure that the simulated events of the same type share the same technical performance, but still capture observed variability across water use events. By doing this, we ensured a realistic water use behavior for each simulated household.

Using these Monte Carlo methods, we sampled from the distributions of event frequencies and event volumes/flowrates for each day of the week to generate a simulated set of events that when summed provide a water use estimate for each simulated household over a one-week period. For sampling purposes, we used the CDF for each input (event frequency and event volume/flowrate for different days of the week). The CDF's x-axis encloses the range of possible values of an input, while its y-axis holds the non-exceedance probability values, which vary from 0 to 1. After generating CDFs for frequency and individual water use event volumes/flowrates for each day of the week using the event data input to the model, Equation 2 was evaluated as follows for each individual simulated residence:

- Select day of the week (D) for the simulated household.
- For each end use type i:
  - Determine the frequency of end use event type i (e.g., shower) for the simulated household for the selected day of the week by randomly sampling from the CDF of frequency values for that day – generate a random number between 0 and 1 representing a non-exceedance probability and select the corresponding frequency from the x-axis (Figure 1). In the example below for the shower end use, the randomly generated non-exceedance probability value of 0.4 indicates that the simulated household is drawn from the medium water use frequency group. For other days of the week, narrow the randomly selected non-exceedance probability value for the same end use type to be within the range of medium water use frequency group (0.33-0.66).

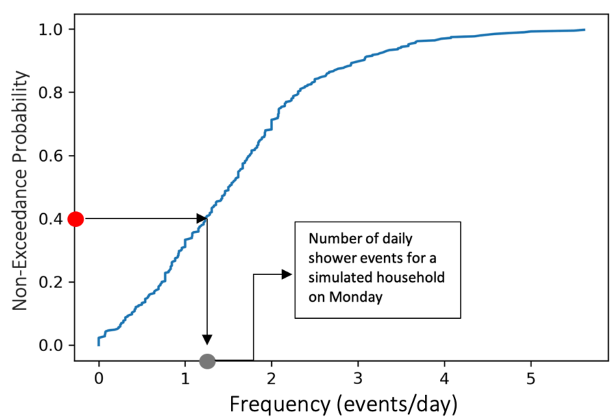

Figure 1. Sampling process from the CDF of shower frequency.

- If the selected frequency is zero, the simulated household does not have end use i. Set the value of B_i to 0 and the end use volume $v_{i,j}$ to zero. If $B_i = 1$, proceed to the next step. The shape of the generated CDF curve of frequencies is influenced by the frequency values in the original data. For example, if 50% of all households in the representative sample do not have bathtub filling events, the CDF curve of frequencies will have a steeper slope segment at the beginning of the curve (Figure 2) indicating that many of the frequency values used to generate the distribution have a value of zero. This implies that the likelihood of a sampled household having a B_i value that equals to zero (no bathtub filling events) is high assuming that the sampling is random and unbiased.

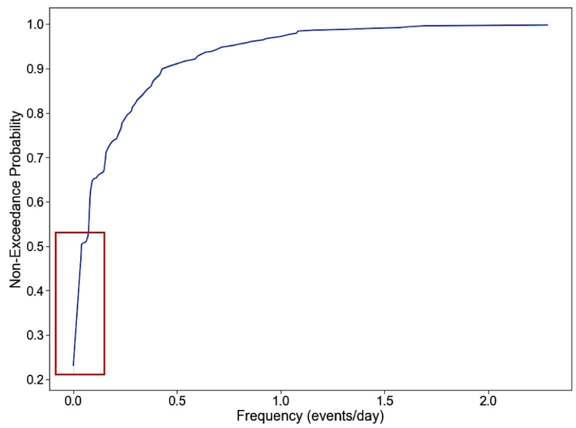

Figure 2. Example CDF for bathtub filling events.

- Based on end use type, generate three CDFs of water volumes or flowrates, one for efficient events, one for typical events, and one for inefficient events. For faucet and shower end uses, flow rates are used to reflect their technical performance, while user behavior is captured in the duration for each event. For other end use types, including toilet, bathtub, clothes washer, and dishwasher end uses, only volume is considered since it is more relevant than the flow rate in terms of technical performance. An example of the CDFs used for shower event sampling is shown in Figure 3 with both shower event flowrate and duration CDFs used for sampling.
  - For each event j in the set of end use events of type i defined by frequency $f_{i,D}$:
  - For the first event of toilet, bathtub, clothes washer, and dishwasher ($j = 1$), randomly pick a CDF curve of volumes from the inefficient, typical, and efficient distributions generated in the previous step for events of type i. Determine an end use volume, $v_{i,1}$, by randomly picking a volume from the selected CDF of volumes – generate a random number between 0 and 1 representing a non-exceedance probability and select the corresponding volume from the x-axis.
  - For the first event of faucet and shower end uses ($j = 1$), Determine an end use flowrate, $\mathrm{FR}_{i,1}$, by randomly picking a flowrate from the selected CDF of flowrates – generate a random number between 0 and 1 representing a non-exceedance probability and select the corresponding flowrate from the x-axis. Determine an end use duration $D_{i,1}$ by randomly picking a duration from the CDF of durations – generate a random number between 0 and 1 representing a non-exceedance probability and select the corresponding duration from

the x-axis. Calculate the water use volume of the first event of faucet and shower end uses by multiplying its flowrate by its duration. In the example below, the first shower event was picked from the CDF for efficient showerheads, and its duration was randomly picked from the CDF of shower durations (Figure 4).

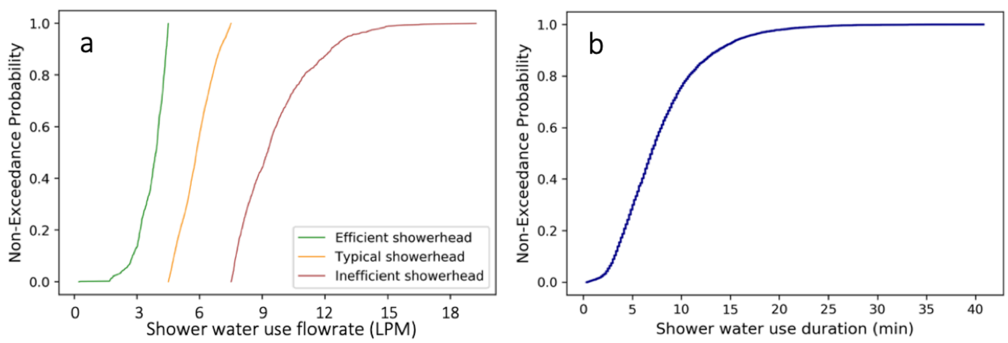

Figure 3. Panel a: flowrate CDFs for showers based on their technical performance, and Panel b: duration CDFs for showers based on their behavioral performance.

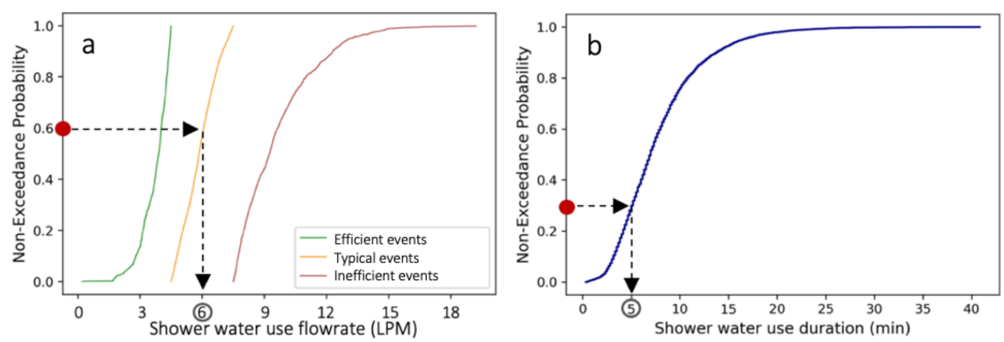

Figure 4. Panel a: sampling from flowrate CDF for shower events based on their technical performance, and Panel b: sampling from duration CDF for shower events based on their behavioral performance.

- For succeeding events ($j > 1$) of types toilet, bathtub, clothes washer, and dishwasher, determine an end use volume by randomly sampling from the CDF for event volume after narrowing the sampling range to a set of event volumes that matches the technical performance of the first selected event – generate a random number between 0 and 1 representing a non-exceedance probability and select the corresponding volume from the x-axis. For succeeding events ($j > 1$) of types faucet and shower, determine an end use flowrate by randomly sampling from the flowrate CDFs after narrowing the sampling range to a set of event flowrate values that match the technical performance of the first selected event. By doing this, we ensure a consistent technical performance of water use events of the same type in the same household for different days of the week. Determine an end use duration by randomly sampling from the CDF of shower durations. Calculate the water use volume of each event by multiplying the flow rate of each event by its duration. In this example, succeeding shower events are sampled from the typical event CDF (Figure 4, Panel a) while their duration can be any value within the CDF curve of durations (Figure 4, Panel b).

- Add the volume of the current event j to a total volume tally for event type i for the current day D.
- For water use events that are not prescribed to a particular end use type, generate a single CDF of event volumes. Determine an event volume, $v_{i,1}$, by randomly picking a volume from the generated CDF of volumes – generate a random number between 0 and 1 representing a non-exceedance probability and select the corresponding volume from the x-axis. For succeeding events ($j > 1$) determine an end use volume by randomly sampling from the CDF of volumes. By not constraining the sampling procedure of water use events that are not prescribed to a particular end use type (e.g., leaks), we ensure realistic behavior of these events given that they have been observed to vary drastically from one day to another within the same home.
- Add the total volume tally for events of type i to the total daily volume tally for the current day D.
- Repeat the steps described above for each day of the week for each simulated residence until the number of residences in the input has been simulated.

2.4 Simulating Water Conservation Stratigies

The total volume of water savings (L) is calculated in the model as the difference between water use before and after conservation actions are applied (e.g., installing a low-flow showerhead for a certain household will reduce overall water use by reducing water used by showers). The savings associated with conservation actions depend on the initial state of a household. For example, a household that already has efficient shower heads will not realize water savings by installing low-flow showerheads. The expected amount of water saved by making end uses more efficient can be calculated as:

$$V_{S,D,i} = \left(\sum_{i=1}^{n} B_i \sum_{j=1}^{f_{i,D}} (v_{i,j} - v'_{i,j}) \right) \tag{3}$$

where $V_{S,D,i}$ (L) is the water savings from retrofitting end use i and/or changing water use behavior for the household on day of the week D, B_i is a coefficient indicating the absence (0) or presence (1) of an end use i, $v_{i,j}$ is the volume of water used by end use type i for an individual event j during day of the week D (L) before retrofitting, n is the number of end uses within the simulated household, and $v'_{i,j}$ is the volume of water used by end use type i for an individual event j during day of the week D (L) after retrofitting.

This expression enables the model to investigate technological and behavioral conservation actions at the household level. While households that already have efficient fixtures will not save water for conservation actions that involve retrofitting fixtures, conservation actions that involve behavioral change of water use can still be considered for those households. The amount of water savings is assumed by the model to be a simple superposition of the effectiveness of each independent action. For example, if a household chooses to reduce shower lengths and reduce clothes washer use frequencies, the total effectiveness of those actions together is modeled as the sum of the effectiveness of each of those actions when implemented independently. The total water savings from adopting multiple conservation actions is then estimated as the sum of water savings associated with each implemented action, which can be denoted as:

$$V_{S,D,\ T} = \sum_{i=1}^{n} V_{S,D,i} \tag{4}$$

where $V_{S,D,\ T}$ is the total water savings (L) for day of the week D, and $V_{S,D,i}$ is as described above.

2.5 Case Study Application

The water end use model we developed can be applied to simulate any set of residential households where the following conditions are met: 1) monthly or more frequent water use data for the residential households to be simulated is available, and 2) there is a set of households with detailed water end use event data that are representative of the households to be simulated. As a demonstration case, we picked the city of Logan, Utah, USA as a medium sized municipality to demonstrate the capability of the model to simulate the indoor residential water use of all households in a city. Logan City is the hub of a growing metropolitan area in northern Utah's Cache Valley (Figure 5) and relies entirely on springs and groundwater wells to supply municipal water needs. Logan's drinking water is drawn from groundwater in DeWitt Spring located in Logan Canyon to the east of the city. Although the spring generally provides a sufficient amount of water to supply the City, it is supplemented by four culinary wells that assist the supply, primarily in the summer. More than 70% of the total supplied fresh water is consumed by the residential sector in Logan City. The majority of residential buildings in Logan are classified as single-family household (SFH), with 7,500 SFH connections reported in the city's monthly water records. SFH connections account for 90% of residential users in Logan.

The Logan City water utility provided us with monthly water use records collected from 2012 to 2018 for all SFH connections within the city. The provided billing dataset contains the total monthly water use volume in gallons per household along with other secondary attributes, including the billed days, square footage of the home, property number, account number, and bill date (Attallah et al., 2022). To select a representative sample of households with water end use event data for input to the model, we used data from the 2016 Residential End Uses of Water Study (REUWS) collected by AquaCraft, Inc. (DeOreo et al., 2016). The 2016 REUWS dataset provides information about individual water use events derived from high temporal resolution smart metering data. AquaCraft monitored 762 single-family households across 11 cities in the U.S. and Canada between 2000 and 2016 for a period of two weeks. They used their TraceWizard software (DeOreo et al., 1996) to disaggregate the high resolution flow trace from each household's water meter to identify and classify individual water use events. The resulting dataset contains individual water use events along with several event attributes, including the date, start time, volume, and peak flow rate. In addition to the detailed end use dataset, the 2016 REUWS recorded daily water use for each participating household. Table 3 summarizes the geographical coverage and other parameters collected in the 2016 REUWS.

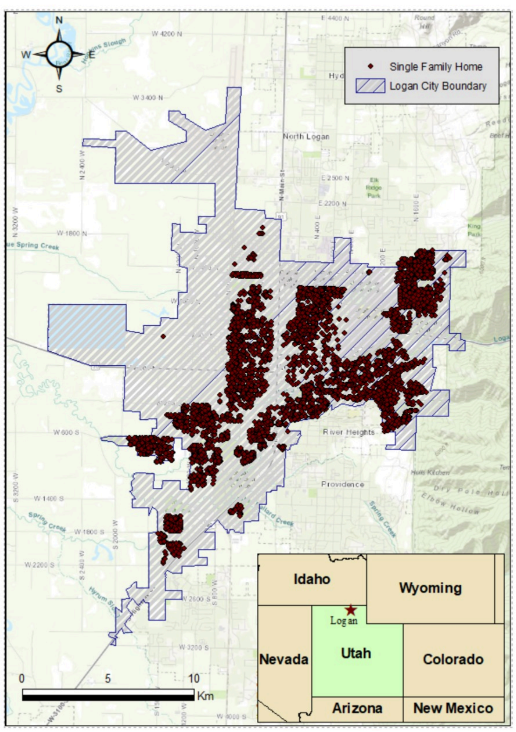

Figure 5. Distribution of single-family households in Logan.

To draw a sample of 2016 REUWS households that is representative of the Logan households, we used average daily water use as the metric for comparison. The monthly water use data provided by Logan City and the 2016 REUWS daily water use dataset were collected at different temporal aggregations (monthly versus daily), and over different time periods (2010-2016 for REUWS versus 2012-2018 for Logan). To enable comparison across the datasets, we arranged both into a similar temporal aggregation. We used the years of 2014-2018 for Logan as the most recent five years of data. We downscaled the monthly billing data for all households in Logan to average daily water use by dividing the monthly water use volumes of each year by the number of billed days for each month. To ensure we were only accounting for indoor water use, we excluded summer months from the dataset where outdoor water use is anticipated and considered winter months only (January to March and November to December). We estimated four values of average daily water use for each household in the Logan dataset for each year, one value for each winter month, then averaged them together to get one estimate of daily water use for each household. For the 2016 REUWS dataset, we estimated average daily water use volumes for all households across all days by excluding irrigation events where they existed for all years available.

Table 3. Data collected in the 2016 Residential End Uses of Water study.

Data	Description
Geographic Coverage	Clayton County, GA; Denver, CO; Fort Collins, CO; Peel, Ontario; San Antonio, TX;
Temporal Coverage	AquaCraft recorded water flow through each individual customer's water meter every
Demographics	Number of residents, rent versus own, highest level of education in the household, ann
End uses	Toilet, bathtub, faucet, shower, clothes washer, dishwasher, evaporative/swamp cooler
Fixture information	Presence of low-flush, ultra-low-flush, dual-flush toilets, number of showerheads in sho

After calculating the average daily water use for each household in the Logan dataset, we used a weighted random sampling approach to identify a set of households from the 2016 REUWS dataset that would generate a probability distribution of average daily water use representative of the one generated from Logan households. Weighted random sampling utilizes PDF curves to randomly sample data points from a distribution (in this case the 2016 REUWS households) based on weights assigned to each data point in the sampling dataset based on the PDF of another dataset (in this case the Logan households). The sampling weights effectively set the likelihood with which households in the 2016 REUWS dataset will be selected so that the sampling procedure generates a set of households having a distribution of average daily water use that represents the distribution of average daily water use for Logan households as closely as reasonably possible. The following steps summarize the weighted random sampling procedure:

- Identify the range of values of average daily water use volume for households in both Logan and the 2016 REUWS datasets. Remove households from the 2016 REUWS dataset with daily water use volumes beyond the range of water use volumes of Logan dataset.

- Generate a PDF curve of average daily water use volumes for households in the Logan dataset (Figure 6). The x-axis of the PDF represents the range of average daily water use volumes for Logan households. The y-axis represents the probability density, or the likelihood of the corresponding value on the x-axis occurring. Since a PDF is a graphical representation of a numerical distribution where the outcomes are continuous, for each household in the 2016 REUWS dataset with average daily water use within the range of average daily water use values from the Logan dataset, there is a probability density value on the Logan dataset's PDF curve.

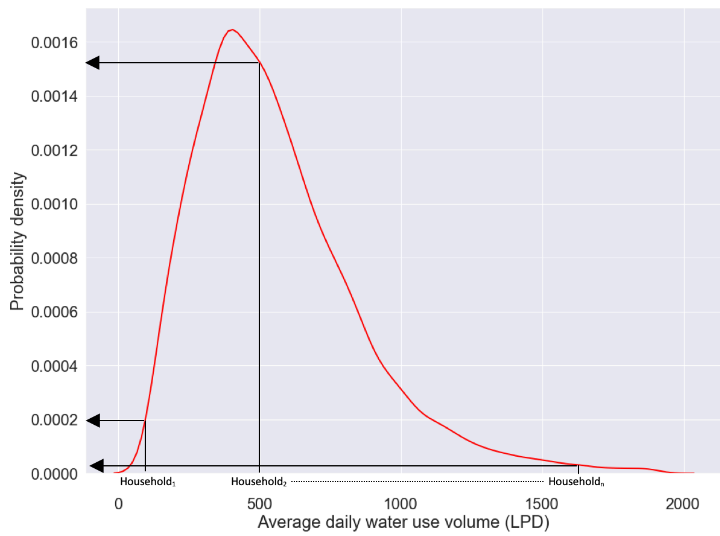

Figure 6. Weight assigning to each 2016 REUWS dataset household. The average daily water use volume for each 2016 REUWS household is intersected with the PDF curve for the Logan dataset to obtain a probability density value for each 2016 REUWS household. These probability density values are used as the weights for 2016 REUWS households in the sampling procedure.

- Calculate the probability density value for the average daily water use volume for each 2016 REUWS household using the PDF curve of the Logan dataset (Figure 6). The calculated probability density value for a 2016 REUWS household is called the sampling weight. The sampling weight sets the importance of each household in the 2016 REUWS dataset such that the likelihood of a household being selected is equal to the probability density of that point from the Logan PDF.

- Normalize the sampling weight of each 2016 REUWS household by dividing weights by the summation of weights for all 2016 REUWS households.

The summation of normalized weights from all 2016 REUWS households should equal to 1.

- Use the random.choice function from the NumPy Python package to randomly select a subset of 2016 REUWS households using the normalized weights for the 2016 REUWS households as input to the function. We also set the replace parameter of the function to be true to sample with replacement. We chose to sample with replacement given the small number of 2016 REUWS households (less than 400 households) compared to 7,500 households in the Logan dataset.

- The sampling function requires predefining the number of 2016 REUWS households to be selected. To identify the optimal number of 2016 REUWS households to select, we used a statistical test of equality metric to evaluate different sample sizes. Many statistical tests can be used to test the equality of continuous, one-dimensional probability distributions. The most common ones include the Chi-square test (Looney, 2008), the Anderson-Darling test (Nelson, 1998), and the Kolmogorov-Smirnov (KS) test (Massey et al., 1951). A one-sample KS test can be used to compare a sample (i.e., daily water use volumes drawn from the 2016 REUWS dataset) with a reference probability distribution (i.e., daily water use volumes obtained from Logan dataset) to determine whether they are the same. An attractive feature of the KS test is that it does not depend on the underlying CDF being tested. The KS test uses the p-value significance level to examine whether two distributions are equivalent. The KS test returns a D statistic and a p-value corresponding to the D statistic. The D statistic is the absolute max distance between the CDFs of the two samples. The closer this number is to 0, the more likely it is that the two distributions are equivalent. The p-value returned by the KS test has the same interpretation as other p-values. If the p-value is lower than some significance level (e.g., =0.05), then the null hypothesis is rejected, signifying the modeled and observed results are not from the same distribution. If the p-value is greater than the =0.05 significance level, then both datasets were drawn the same distribution. The KS test was implemented using the SciPy 1.7.2 Python Package.

- For the KS test configuration, we used an initial sample size of 50 2016 REUWS households, and then increased the sample size by one household on each iteration, and stopped when the population size of 7,500 households was reached. We estimated the D statistic and a p-value for each sample size and selected the sample size that produced the least D static value with a p-value greater than the =0.05 significance level.

Utilizing the procedure described above, a total of 92 households that generated a probability distribution of average daily water use representative of Logan households was drawn from the 2016 REUWS dataset. The selected households resulted in minimum D static value, indicating that the daily water use volumes of the drawn households most closely represent the overall daily water use volumes of households in Logan dataset. The selected sample of 92 households included 69 unique households and 23 replicated households. We used the detailed end use event data for all 92 households in this set to simulate the detailed water end use events for Logan residents. The PDF of average daily indoor water use of Logan households during winter months of the years of 2014-2018 versus the sample of households drawn from the 2016 REUWS dataset is shown in Figure 7.

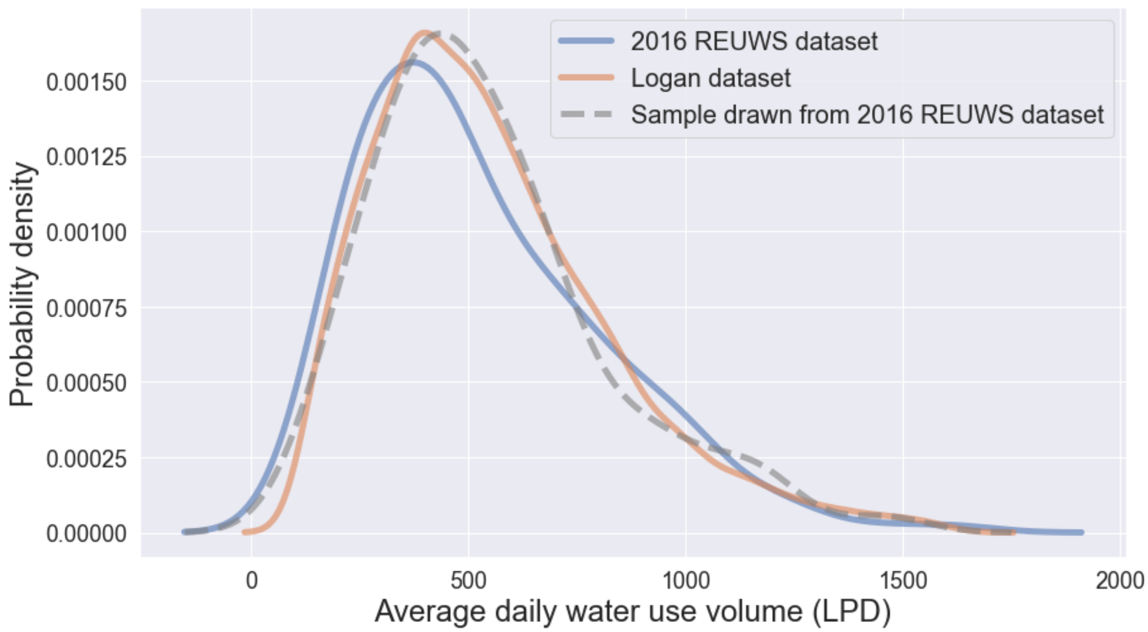

Figure 7. PDF for average daily indoor water use of Logan's households versus the sample of households from the 2016 REUWS.

2.6 Model Validation and Comparison

The simulation process resulted in estimates of daily water use volumes for 7,500 residential households located in Logan City for a period of one week. To confirm that the water use volumes from the simulation model accurately represent the water use behavior of residents in Logan City, we compared the simulated water use volumes to observed water use volumes retrieved from the monthly billing dataset. Given that simulated water use volumes were generated using Monte Carlo simulation, comparing them directly with observed data is not possible. Instead, we compared the distribution and characteristics of simulation results to the distribution and characteristics of the observed data to ensure that they match. Since the simulated water use volumes and the observed water use volumes from the monthly billing dataset have different temporal scales (simulated daily water use volume for one week period versus observed monthly water use volume), we first arranged both datasets into a common temporal aggregation. For the monthly water use records, we calculated average daily water use volumes for each household by dividing the total monthly water use volume for that household by the number of billed days using winter months data for the years between 2014 and 2018, then we averaged across winter months to get one value of average daily water use volume for each household in the dataset. For the model results, we calculated the average daily water use volume for each household by summing the daily water use volume for the whole one week simulation period and dividing the total by seven days. To evaluate whether the actual average daily water use volumes and the simulated average daily water use volumes were drawn from the same distribution we used the KS test.

To evaluate how improving the representation of water uses at a detailed level within a water demand model can improve our ability to predict the water uses – our first research question – we evaluated the performance of the developed model in predicting the actual average daily water uses of households in Logan City dataset against other urban water demand simulation models. In our review of existing urban water demand models, we found that code is not openly available. In most cases, access was restricted or can only be obtained by contacting the authors. We could not replicate other end use models because source code was not available and their formulations/equations were not well enough described in the papers that we could re-implement them. Moreover, other end use modeling studies were restricted to communities where water end use data are available, which inhibits their ability to predict the detailed water use of other residential communities.

In response to these issues, we compiled a list of theoretical and empirical methodologies reported in urban water demand simulation papers published over the past two decades. We searched on different web search engines and scientific databases including Google Scholar, Zotero, and Mendeley for the following combination of words: "urban water demand model", "water demand simulation", and "residential water demand model". We then compiled a list with the methods and related publications retrieved with the above search, we reviewed and classified the list according to model replicability, equation availability, and directions to replicate the method presented in the paper. From this list, we selected a subset of models that meet the following criteria: 1) can simulate current water use conditions, 2) commonly used and recognized (e.g., regression), 3) we have input data for (e.g., landscaped area, census count), 4) well enough described in the paper that we could replicate them, and 5) the specific model selected is representative of a class of models reported in the literature. We then implemented those models to simulate current residential water use in Logan by generating 7,500 daily water use volumes that represent the number of residential water connections of Logan City.

Based on the aforementioned criteria, we replicated three different water demand models including an Ordinary Least Squares (OLS) model (Polebitski and Palmer, 2010), Piecewise Regression model (Chang et al., 2014), and Multiple Regression model (Arbués et al., 2010). Independent variables implemented to predict indoor water use in these models included socio-demographic variables (e.g., number of residents) and meteorological variables (e.g., precipitation). Demographic variables used as inputs in each model were retrieved from the Cache County GIS Parcel data website (https://www.cachecounty.org/gis/). These variables and the interaction between them were implemented differently in each model; however, where possible, we used the same variables to simulate current residential water use in Logan City as did the authors of the prior modeling studies – e.g., like Polebitski and Palmer, (2010), we used the building area (ft^2), number of residents, income, property age, and household value to predict indoor water use using an OLS model.

To evaluate the reliability of each model in predicting current water use, we compared the cumulative distribution of average daily water use volumes estimated from the Logan dataset versus the cumulative distributions of average daily water use volumes obtained from different water demand simulation models including the model developed in this paper. Quantitatively, we utilized the KS statistical test on the CDFs output from the different models tested against the CDF of actual water use data from Logan dataset. We estimated the D statistic and a p-value for each model and presumed that the model that produced the smallest D static value with a p-value greater than the =0.05 significance level is the best performing model.

2.7 Water Conservation Actions

To quantify the water saving potential associated with different technological and behavioral conservation actions – our second research question – we examined the efficiency of water end uses of different types across all simulated households to identify households and end use types with water conservation potential. We then quantified the water saving potential associated with a set of potential technological and behavioral conservation actions (Table 4) for water use in the model. Technological actions include actions associated with the technical performance and water use efficiency of different end use types inside a household (e.g., retrofitting an inefficient showerhead). Behavioral actions include actions associated with the water use behavior of a household's residents (e.g., reduce shower length).

Based on the end use type, we used either the volume or flowrate of the simulated water use events to investigate the technical performance of the existing end uses and compared them with typical and efficient end uses. Volume was used to reflect the technical performance of toilet, bathtub, clothes washer, and dishwasher end uses. Retrofitting actions on these end uses were applied on events with volumes exceeding efficient volumes, and the expected water use after retrofitting was calculated as the volume of events from retrofitted fixtures. For faucet and shower end uses, flowrate was used to reflect their technical performance. Retrofitting actions on these end uses were applied on events with flowrates exceeding efficient flowrates, and the expected water use after retrofitting was calculated as the flowrate of retrofitted fixtures multiplied by the duration of their corresponding events. For all retrofitting actions, we assumed that a retrofit would change an end use's technical performance, but not user behavior.

Table 4. Proposed water conservation actions and their associated characteristics in terms of water use*.

Technological conservation actions	Characteristic flows/volumes	Behavioral conservation a
Retrofit showerheads	7.6 LPM	Fix leaks
Retrofit faucets	5.7 LPM	Reduce faucet use duratio
Retrofit toilets with low flush toilets (LFT)	6.1 LPF	Reduce shower duration
Retrofit toilets with highly efficient toilets (HET)	4.8 LPF	Reduce clothes washer use
Retrofit top load washers with front load washers	~ 100 L/load	

* Values reported in this table were retrieved from the EPA WaterSense and EnergyStar Websites (EPA, 2021a, EPA, 2021b).

Besides retrofitting actions, we used the duration, volume, and the number of simulated events per household per day to account for behavioral change in water use for those end uses that are associated with the behavior of residents (e.g., reduce shower duration). Four different actions were examined (Table 4). For the fixing leaks action, we assumed that 50% of total unclassified events are leaks and thus residents of a household can reduce unclassified water use inside their home by 50% by fixing leaks. Thus, the amount of water saved by fixing leaks was calculated as the volume of unclassified events divided by two. For the reducing shower length action, we first identified long shower events as events that last longer than the 80th percentile of all shower events in a simulated household and assumed that residents of that household can reduce their long shower events down to the 80th percentile of all shower durations. The amount of water saved by reducing shower duration per household was calculated for simulated shower events that exceeded the 80th percentile shower length as the difference in shower duration before and after the duration reduction of each event multiplied by the flowrate of the event. The same procedure was used for the reducing faucet duration action. To reduce clothes washer event frequency, we assumed that residents can reduce their current frequency of laundry events by 10%, although other frequencies could easily be simulated.

3 Results and Discussion

The end use water demand model simulated 367,500 water use events for 7,500 households in the City of Logan over the period of one week. The average execution time for the water demand model, which simulates one week of water use for both existing conditions and the water conservation scenarios for all households was approximately six hours on a 2017 MacBook Pro laptop computer with a 3.1 GHz quad-core Intel i7 processor and 16 GB of RAM.

3.1 Model Comparision and Applicability

All of the models we tested resulted in a cumulative distribution curve relatively similar to the Logan data (Figure 8), but the end use model we developed most closely matched the distribution of the Logan City data. Moreover, it provides detailed end use results that could assist water suppliers in identifying opportunities for incentive programs to encourage water conservation and monitoring effectiveness of those programs where the other models do not. The disparity between our model and other models in simulating current water use indicates that using water end use events to predict total daily water use volumes instead of using regression approaches will likely generate more realistic results.

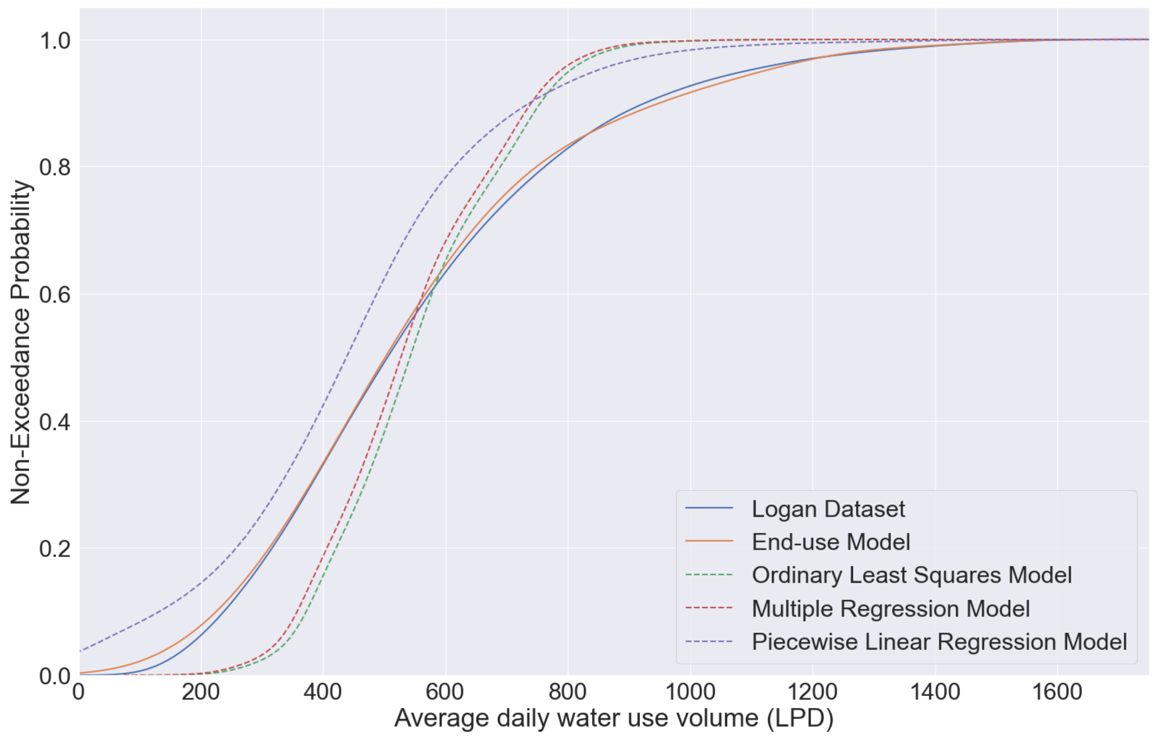

Figure 8. Cumulative distributions for observed average daily water use volume and simulated average daily water use volume of different simulation methods for all residential connections in Logan.

Using the KS test, the resulting p-value of the end use model was 0.84, which is higher than the 0.05 significance level. The D statistic value for our model was 0.049, which is less than the D statistic value of the other models we tested (Table 5). Thus, both the observed average daily water use volume records calculated from monthly billing data for winter months during 2014-2018 and the simulated average daily water use volumes obtained from the end use model have very similar distributions. P-values for the other models we tested were not significant, indicating that their resulting distributions are different than that of the monthly billing data.

Table 5. Estimated D statistic and a p-value for each model.

Model	D statistic	P-value
End use model	0.049	0.84
OLS model	0.204	$2.1\ X\ 10^{-15}$
Multiple regression model	0.180	$2.1\ X\ 10^{-15}$
Piecewise linear regression model	0.160	$2.1\ X\ 10^{-15}$

To assess the applicability of the end use model in predicting current conditions given that it was based on data from 2014 – 2018, we explored the variability in indoor water use of households in the Logan dataset for those years (Figure 9). The white dots in the figure represent the medians of the distributions, the thick grey bar in the center represents the interquartile range, the thin grey line represents the whole distribution, except for outlier data points, wider sections of the violin represent a higher approximate frequency of data points in that section, and thinner sections represent a lower approximate frequency of data points in that section. As illustrated in Figure 9, the overall distribution of indoor water use for Logan City households was fairly stable between 2014 and 2018.

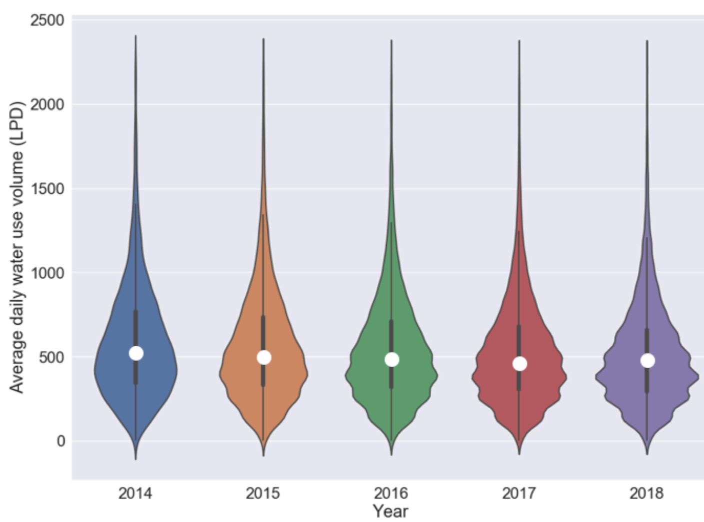

Figure 9. Distributions of average daily indoor water use for Logan City households between 2014-2018.

With respect to water end uses simulated by the model, toilet flushing accounted for the largest volume of indoor water use, followed by showers, faucets, clothes washers, and bathtubs, which matches the relative contribution of indoor water use type reported by the 2016 REUWS (Table 6). While each simulated household had a unique behavioral pattern, bulk behavior across all simulated households, matched that of the 2016 REUWS with the biggest difference between the two studies being 3%.

Table 6. Relative contribution of indoor water use type.

End use type	Simulation results	2016 REUWS	Difference
Bathtub	4%	4%	0%
Clothes washer	15%	18%	3%
Faucet	22%	22%	0%
Dishwasher	0.5%	2%	1.5%
Shower	26%	23%	3%
Toilet	31%	28%	3%
Unclassified	1%	3%	2%

3.2. End Use Efficiency and Water Conservation Potential

The maximum amount of water savings is expected when all retrofitting actions are implemented at the same time. However, toilets have two retrofit options that are mutually exclusive. To maximize water savings, we assumed that typical and inefficient toilets are retrofitted to highly efficient toilets since they save more water than low flush toilets (Table 4). Based on that, the expected proportion of water saved if all technological conservation actions are implemented together at the same time ranged from 0% to 50% for individual households and totaled approximately 23% of total water use across all households.

Generally, technological conservation actions are effective and more likely to persist since they involve changing fixtures. Adoption of behavioral conservation actions may vary from one household to another and even in the same household from one day to another since they are associated with phycological, social, and behavioral changes of household's residents (Addo et al., 2018). In the matter of durability of technological actions compared to behavioral actions, technological actions can perform up to 20 years (EPA, 2021a), while behavioral actions have been shown to be effective for six months at most (Schultz et al., 2019). Retrofitting toilets to 4.8 LPF toilets had the most water saving potential (assuming behavioral water use does not change) at a total water savings of approximately 10.5 million LPD for Logan City. On the behavioral side, reducing shower lengths can save over 1.9 million LPD for Logan City. Figure 10 summarizes the household-level water savings rates for both technological and behavioral conservation actions. The box plots in both figures show the distribution of daily water savings across all 7,500 Logan households, with each box showing a different implemented water conservation action.

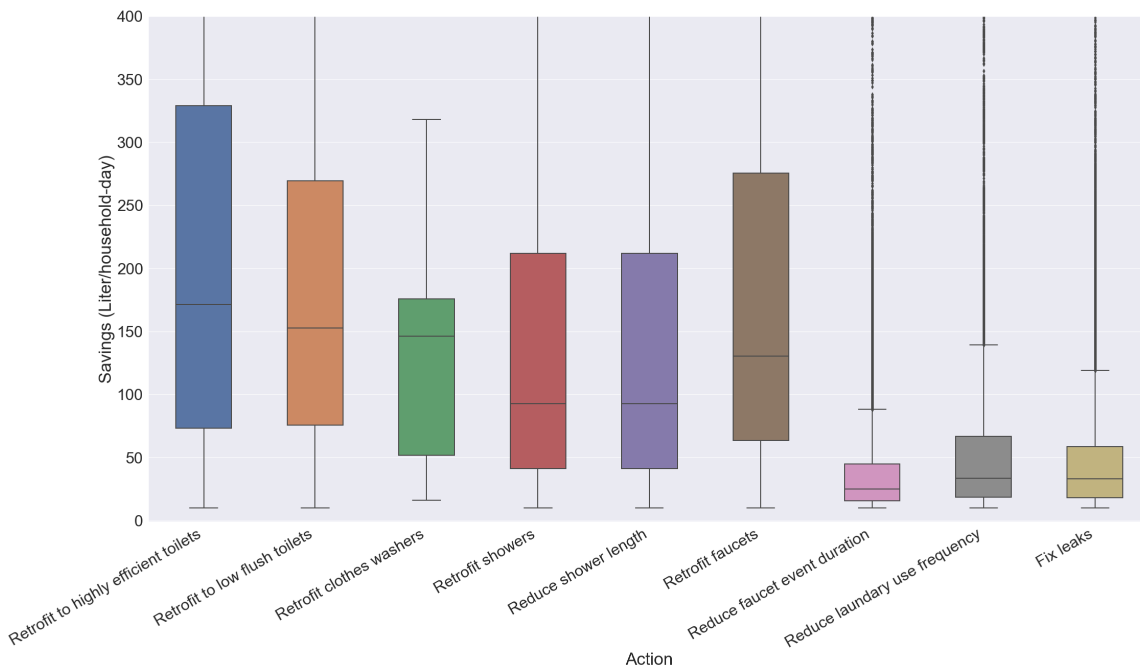

Figure 10. Ranges of potential water savings for technological and behavioral conservation actions. Some outliers were removed to enhance the readability of the figure.

4 Conclusions

We developed an end use water demand model that simulates detailed household water use using Monte Carlo techniques. The model advances existing end use modeling studies that used similar techniques by accounting for differences in event frequency among different days of the week, simulating variabilities in event volume or flowrate and duration for different end use types for different days of the week while constraining the technical performance of different end uses, incorporating all expected indoor water use events in the simulation process, providing estimates of baseline use and maximum conservation potential at the individual home and city levels, and developing a generic model that can be scaled to any number of single family residential homes.

The model uses event data from a sample of households in the 2016 REUWS dataset as input to simulate water use behavior of Logan residents. The input dataset consists of detailed end use event data for a sample of households that are representative of the households to be simulated. The model is generally applicable and can be modified to simulate the detailed water end uses of other cities with the following constraints: 1) the city to be simulated must have available water use records (e.g., monthly or more frequent billing data records), and 2) there must be a sample of households in the 2016 REUWS (or another dataset) such that the water use of the sample households is representative of the water use of the households to be simulated (e.g., similar daily water use distribution). Since the 2016 REUWS dataset collected data across 11 different cities in the U.S. and considered monthly data in selecting households with different water use behaviors, we anticipate that the likelihood of extracting a sample of households from the 2016 REUWS dataset with an overall water use representative to other cities is high, although selection of households for simulation would be enhanced by the availability of more households with detailed end use data.

In our case study application, we demonstrated how existing water use event data can be used to predict the detailed water use of other residential communities, with the only required data from the city to be simulated being their monthly water use billing records. Since we used data from 2014-2018 only, we acknowledge that the water demand model quantifies the detailed water use and evaluates potential conservation in the context of the years of 2014-2018. However, Figure 9 shows that indoor water use was stable for Logan over this period, and we anticipate similar water use behavior from many other communities across the U.S. Thus, the model should reflect current conditions and conservation potential, but may need to be adjusted in the future to reflect changes in indoor water use behavior.

The retrofitting and behavioral conservation actions for selected end uses showed high potential for water conservation across the 7,500 residential households we simulated. The expected upper band of total water savings at the household level is 2,700 L/household-day and the expected total water savings at the city level is more than 20 million L/day, representing approximately 23% of all water currently used indoors by residential users in Logan City.

The type of detailed water end use simulation produced by the model, including practical water conservation actions and the ability to simulate their savings at the city level, could assist water utilities in identifying opportunities for incentive programs that will have the greatest impact and to encourage water conservation. Effectiveness of these efforts could be monitored using new methods for collection of high resolution water use data or through more conventional comparison of pre- and post-retrofit monthly data, although effectiveness of multiple, simultaneous programs would be difficult to separate using only monthly data. Furthermore, this type of modeling can be used for forecasting demand and determining how water use patterns may change over time in response to population growth, demographic shifts, behavioral change, and improvements in technology. It may also be useful in better characterizing how and when water is being used inside of households and in the design of improvements to the residential water distribution infrastructure. Supplying this type of information to water users can also be a tool for impacting water use behavior and managing demand.

Acknowledgments

This research was funded by the United States National Science Foundation under grant number 1552444. Any opinions, findings, and conclusions or recommendations expressed are those of the authors and do not necessarily reflect the views of the National Science Foundation. Additional financial support was provided by the Utah Water Research Laboratory at Utah State University. We acknowledge Logan City for their cooperation in this study and for providing the monthly water use data for residential connections in the city.

Open Research

The water demand model was designed and developed using Version 3.7 of the Python programming language. It was implemented as a single script that can be executed using any Python programming environment and was developed using the SciPy 1.7.2, Pandas 1.3.4, NumPy 1.21.4, and scikit-learn 1.0.1 packages for Python.

Code for the water demand model is open-source, released under the Creative Commons Attribution CC BY license, and available in the HydroShare repository (Attallah et al., 2022). Documentation of software requirements, Python Jupyter notebooks with examples of workflows implementing each part of the code, and instructions for running the code are provided in the HydroShare resource.

References

Attallah, N., J. S. Horsburgh, C. J. Bastidas Pacheco (2022). Supporting data and tools for "Simulating Community Water Use Behavior And Potential Water Conservation Using Detailed End Use Event Data", HydroShare, http://www.hydroshare.org/resource/6cbedaad1c19414d9015ed10c88fb761

Abdallah, A. M., & Rosenberg, D. E. (2014). Heterogeneous Residential Water and Energy Linkages and Implications for Conservation and Management. *Journal of Water Resources Planning and Management, 140*(3), 288–297. doi:10.1061/(asce)wr.1943-5452.0000340

Agthe, D. E., & Billings, R. B. (1980). Dynamic models of residential water demand. *Water Resources Research*, *16*(3), 476–480. doi:10.1029/wr016i003p00476

Arbués, F., Villanúa, I., & Barberán, R. (2010). Household size and residential water demand: an empirical approach. *Australian Journal of Agricultural and Resource Economics*, *54*(1), 61–80. doi:10.1111/j.1467-8489.2009.00479.x

Arbués, F., García-Valiñas, M. N., & Martínez-Espiñeira, R. (2003). Estimation of residential water demand: a state-of-the-art review. *The Journal of Socio-Economics*, *32*(1), 81–102. doi:10.1016/s1053-5357(03)00005-2

Addo, I. B., Thoms, M. C., & Parsons, M. (2018). Household Water Use and Conservation Behavior: A Meta-Analysis. *Water Resources Research, 54*(10), 8381–8400. doi:10.1029/2018wr023306

Blokker, E. J. M., Buchberger, S. G., Vreeburg, J. H. G., & van Dijk, J. C.

(2009). Comparison of Water Demand Models: PRP and SIMDEUM Applied to Milford, Ohio, Data. *Water Distribution Systems Analysis 2008*. Published. doi:10.1061/41024(340)17

Boyle, T., Giurco, D., Mukheibir, P., Liu, A., Moy, C., White, S., & Stewart, R. (2013). Intelligent Metering for Urban Water: A Review. *Water*, *5*(3), 1052–1081. doi:10.3390/w5031052

Cahill, R., R. Lund, J., DeOreo, B., & Medellín-Azuara, J. (2013). Household water use and conservation models using Monte Carlo techniques. *Hydrology and Earth System Sciences*, *17*(10), 3957–3967. doi:10.5194/hess-17-3957-2013

Carver, P. H., & Boland, J. J. (1980). Short- and long-run effects of price on municipal water use. *Water Resources Research*, *16*(4), 609–616. doi:10.1029/wr016i004p00609

Chang, J. S., Jung, D., Kim, J., & Kang, T. (2014). Comparative analysis of trip generation models: results using home-based work trips in the Seoul metropolitan area. *Transportation Letters*, *6*(2), 78–88. doi:10.1179/1942787514y.0000000011

Chicoine, D. L., Deller, S. C., & Ramamurthy, G. (1986). Water Demand Estimation Under Block Rate Pricing: A Simultaneous Equation Approach. *Water Resources Research*, *22*(6), 859–863. doi:10.1029/wr022i006p00859

Cominola, A., Giuliani, M., Castelletti, A., Rosenberg, D., & Abdallah, A. (2018). Implications of data sampling resolution on water use simulation, end-use disaggregation, and demand management. *Environmental Modelling & Software*, *102*, 199–212. doi:10.1016/j.envsoft.2017.11.022

Cominola, A., Giuliani, M., Piga, D., Castelletti, A., & Rizzoli, A. (2015). Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. *Environmental Modelling & Software*, *72*, 198–214. doi:10.1016/j.envsoft.2015.07.012

Dalhuisen, J. M., Florax, R. J. G. M., de Groot, H. L. F., & Nijkamp, P. (2003). Price and Income Elasticities of Residential Water Demand: A Meta-Analysis. *Land Economics*, *79*(2), 292–308. doi:10.2307/3146872

DeOreo, W. B., Mayer, P. W., Dziegielewski, B., Kiefer, J. C. (2016). Residential End Uses of Water, Version 2. Water Research Foundation. https://www.waterrf.org/research/projects/residential-end-uses-water-version-2

DeOreo, W. B., Heaney, J. P., Mayer, P. W. (1996). Flow trace analysis to access water use, *Journal - American Water Works Association*, 88(1), 79–90. doi:10.1002/j.1551-8833.1996.tb06487.x

di Mauro, A., Cominola, A., Castelletti, A., & di Nardo, A. (2020). Urban Water Consumption at Multiple Spatial and Temporal Scales. A Review of Existing Datasets. *Water*, *13*(1), 36. doi:10.3390/w13010036

DOE, 1992. Energy Policy Act.

Escriva-Bou, A., Lund, J. R., & Pulido-Velazquez, M. (2015). Modeling residential water and related energy, carbon footprint and costs in California. *Environmental Science & Policy, 50*, 270–281. doi:10.1016/j.envsci.2015.03.005

EPA, 2021a. "WaterSense Products". Accessed May 7, 2021. https://www.epa.gov/watersense/watersense-products

EPA, 2021b. "WaterSense Products". Accessed December 25, 2021. https://www.energystar.gov/productfinder

Espey, M., Espey, J., & Shaw, W. D. (1997). Price elasticity of residential demand for water: A meta-analysis. *Water Resources Research, 33*(6), 1369–1374. doi:10.1029/97wr00571

Garcia, S., & Reynaud, A. (2004). Estimating the benefits of efficient water pricing in France. *Resource and Energy Economics*, *26*(1), 1–25. doi:10.1016/j.reseneeco.2003.05.001

Gaudin, S. (2006). Effect of price information on residential water demand. *Applied Economics*, *38*(4), 383–393. doi:10.1080/00036840500397499

Hajispyrou, S., Koundouri, P., & Pashardes, P. (2002). Household demand and welfare: implications of water pricing in Cyprus. *Environment and Development Economics*, *7*(04). doi:10.1017/s1355770x02000402

Haley, M. B., Dukes, M. D., & Miller, G. L. (2007, October). Residential Irrigation Water Use in Central Florida. *Journal of Irrigation and Drainage Engineering*, 427–434. doi:10.1061/(asce)0733-9437(2007)133:5(427)

Higgs, H., & Worthington, A. (2001). Consumer Preferences and Water Charging Options in a Large Urban Municipality. *Public Works Management & Policy*, *5*(3), 209–217. doi:10.1177/1087724x0153003

Höglund, L. (1999). Household demand for water in sweden with implications of a potential tax on water use. *Water Resources Research*, *35*(12), 3853–3863. doi:10.1029/1999wr900219

House-Peters, L. A., & Chang, H. (2011). Urban water demand modeling: Review of concepts, methods, and organizing principles. *Water Resources Research*, *47*(5). doi:10.1029/2010wr009624

James, R., & Rosenberg, D. E. (2022). Agent-Based Model to Manage Household Water Use Through Social-Environmental Strategies of Encouragement and Peer Pressure. *Earth's Future*, *10*(2). doi:10.1029/2020ef001883

Jorgensen, B., Graymore, M., & O'Toole, K. (2009). Household water use behavior: An integrated model. *Journal of Environmental Management*, *91*(1), 227–236. doi:10.1016/j.jenvman.2009.08.009

Kenney, D. S., Goemans, C., Klein, R., Lowrey, J., & Reidy, K. (2008). Residential Water Demand Management: Lessons from Aurora, Colorado1. *JAWRA*

Journal of the American Water Resources Association, *44*(1), 192–207. doi:10.1111/j.1752-1688.2007.00147.x

Kontokosta, C. E., & Jain, R. K. (2015). Modeling the determinants of large-scale building water use: Implications for data-driven urban sustainability policy. *Sustainable Cities and Society*, *18*, 44–55. doi:10.1016/j.scs.2015.05.007

Koutiva, I., & Makropoulos, C. (2019). Exploring the Effects of Alternative Water Demand Management Strategies Using an Agent-Based Model. *Water*, *11*(11), 2216. doi:10.3390/w11112216

Kuski, L., Maia, E., Moura, P., Caetano, N., & Felgueiras, C. (2020). Development of a decentralized monitoring system of domestic water consumption. *Energy Reports*, *6*, 856–861. doi:10.1016/j.egyr.2019.11.019

Looney, S. W. (2008). A Review of: "An Introduction to Categorical Data Analysis, 2nd ed., by A. Agresti." *Journal of Biopharmaceutical Statistics*, *18*(3), 585–587. doi:10.1080/10543400802001664

Maeda, S., Nagamochi, T., Kawachi, T., & Takeuchi, J. (2011). Regional allocation of irrigation water in a rice paddy area with water-saving practices. *Irrigation and Drainage Systems*, *25*(2), 81–96. doi:10.1007/s10795-011-9113-2

Makki, A. A., Stewart, R. A., Beal, C. D., & Panuwatwanich, K. (2015). Novel bottom-up urban water demand forecasting model: Revealing the determinants, drivers and predictors of residential indoor end-use consumption. *Resources, Conservation and Recycling*, *95*, 15–37. doi:10.1016/j.resconrec.2014.11.009

Mamade, A., Loureiro, D., Covas, D., Coelho, S., & Amado, C. (2014). Spatial and Temporal Forecasting of Water Consumption at the DMA Level Using Extensive Measurements. *Procedia Engineering*, *70*, 1063–1073. doi:10.1016/j.proeng.2014.02.118

Martinez-Espiñeira, R. (2002). Residential Water Demand in the Northwest of Spain. *Environmental and Resource Economics*, *21*(2), 161–187. doi:10.1023/a:1014547616408

Massey, F. J. (1951). The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association*, *46*(253), 68–78. doi:10.1080/01621459.1951.10500769

Matos, C., Teixeira, C. A., Duarte, A., & Bentes, I. (2013). Domestic water uses: Characterization of daily cycles in the north region of Portugal. *Science of The Total Environment*, *458–460*, 444–450. doi:10.1016/j.scitotenv.2013.04.018

Mayer, P. W., DeOreo, W. B. (1999). Residential End Uses of Water. American Water Works Association Research Foundation, Denver, Colorado. https://energy.mo.gov/sites/energy/files/deoreo-residential-end-uses-of-water-study-update-2013.pdf

Nauges, C., & Thomas, A. (2003). Long-Run Study of Residential Water Consumption. *Environmental and Resource Economics*, *26*(1), 25–43.

doi:10.1023/a:1025673318692

Nelson, L. S. (1998). The Anderson-Darling Test for Normality. *Journal of Quality Technology*, *30*(3), 298–299. doi:10.1080/00224065.1998.11979858

Praskievicz, S., & Chang, H. (2009). Identifying the Relationships Between Urban Water Consumption and Weather Variables in Seoul, Korea. *Physical Geography*, *30*(4), 324–337. doi:10.2747/0272-3646.30.4.324

Polebitski, A. S., & Palmer, R. N. (2010). Seasonal Residential Water Demand Forecasting for Census Tracts. *Journal of Water Resources Planning and Management*, *136*(1), 27–36. https://doi.org/10.1061/(asce)wr.1943-5452.0000003

Renwick, A., Dynes, R., Johnstone, P., King, W., Holt, L., & Penelope, J. (2019). Challenges and Opportunities for Land Use Transformation: Insights from the Central Plains Water Scheme in New Zealand. *Sustainability*, *11*(18), 4912. doi:10.3390/su11184912

Romano, M., & Kapelan, Z. (2014). Adaptive water demand forecasting for near real-time management of smart water distribution systems. *Environmental Modelling & Software*, *60*, 265–276. doi:10.1016/j.envsoft.2014.06.016

Rosenberg, D. E., Tarawneh, T., Abdel-Khaleq, R., & Lund, J. R. (2007). Modeling integrated water user decisions in intermittent supply systems. *Water Resources Research*, *43*(7). doi:10.1029/2006wr005340

Schefter, J. E., & David, E. L. (1985). Estimating Residential Water Demand under Multi-Part Tariffs Using Aggregate Data. *Land Economics*, *61*(3), 272. doi:10.2307/3145842

Schultz, W., Javey, S., & Sorokina, A. (2019). Social Comparison as a Tool to Promote Residential Water Conservation. *Frontiers in Water, 1*. doi:10.3389/frwa.2019.00002

Stevens, T. H., Miller, J., & Willis, C. (1992). Effect of Price Structure on Residential Water Demand. *Journal of the American Water Resources Association*, *28*(4), 681–685. doi:10.1111/j.1752-1688.1992.tb01489.x

Suero, F. J., Mayer, P. W., & Rosenberg, D. E. (2012). Estimating and Verifying United States Households' Potential to Conserve Water. *Journal of Water Resources Planning and Management*, *138*(3), 299–306. doi:10.1061/(asce)wr.1943-5452.0000182

Talebpour, M., Sahin, O., Siems, R., & Stewart, R. (2014). Water and energy nexus of residential rainwater tanks at an end use level: Case of Australia. *Energy and Buildings*, *80*, 195–207. doi:10.1016/j.enbuild.2014.05.018

Wentz, E. A., Wills, A. J., Kim, W. K., Myint, S. W., Gober, P., & Balling, R. C. (2013). Factors Influencing Water Consumption in Multifamily Housing in Tempe, Arizona. *The Professional Geographer*, *66*(3), 501–510. doi:10.1080/00330124.2013.805627

Williamson, P., Mitchell, G., & McDonald, A. T. (2002). Domestic Water Demand Forecasting: A Static Microsimulation Approach. *Water and Environment Journal*, *16*(4), 243–248. doi:10.1111/j.1747-6593.2002.tb00410.x

Willis, R. M., Stewart, R. A., Giurco, D. P., Talebpour, M. R., & Mousavinejad, A. (2013). End use water consumption in households: impact of socio-demographic factors and efficient devices. *Journal of Cleaner Production*, *60*, 107–115. doi:10.1016/j.jclepro.2011.08.006

Worthington, A. C., & Hoffman, M. (2008). An Empirical Survey of Residential Water Demand Modelling. *Journal of Economic Surveys*, *22*(5), 842–871. doi:10.1111/j.1467-6419.2008.00551.x