Implications of data sampling resolution on water use simulation, end-use disaggregation, and demand management

A. Cominola[a,*], M. Giuliani[a], A. Castelletti[a,c], D.E. Rosenberg[b], A.M. Abdallah[b]

[a]*Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Milan, Italy*
[b]*Department of Civil and Environmental Engineering and Utah Water Research Laboratory, Utah State University, Old Main Hill, 4110, UT 84322-4110, Logan, USA*
[c]*Institute of Environmental Engineering, ETH, Zurich, Switzerland*

Abstract

Understanding the tradeoff between the information of high-resolution water use data and the costs of smart meters to collect data with sub-minute resolution is crucial to inform smart meter networks. To explore this tradeoff, we first present STREaM, a STochastic Residential water End-use Model that generates synthetic water end-use time series with 10-second and progressively coarser sampling resolutions. Second, we apply a comparative framework to STREaM output and assess the impact of data sampling resolution on end-use disaggregation, leak detection, peak demand estimation, data storage, and availability. Our findings show that increased sampling resolution allows more accurate end-use disaggregation, prompt water leakage detection, and accurate and timely estimates of peak demand. Simultaneously, data storage requirements and limited product availability mean most

*Corresponding author. Tel.:+39-02-2399-9040
Email address: `andrea.cominola@polimi.it` (A. Cominola)

large-scale, commercial smart metering deployments sense data with hourly, daily, or coarser sampling frequencies. Overall, this work provides insights for further research and commercial deployment of smart water meters.
Keywords: smart meter, sampling resolution, water demand management, STREaM, synthetic end-use model

---

Software availability

- Name of software: STREaM - STochastic Residential water End-use Model - tested on Matlab R2016a
- Developers: Andrea Cominola, Matteo Giuliani, Andrea Castelletti, David Ezechiel Rosenberg, Adel Abdallah
- Contact email: andrea.cominola@polimi.it
- Year first available: 2017
- Available from: GitHub repository - https://github.com/acominola/STREaM

1. Introduction

Over the last two decades, technological advances in the field of urban water demand metering have fostered the development of smart metering technologies that can sense water use with fine sub-daily sampling resolutions, down to a few seconds (Mayer and DeOreo, 1999). Scientific literature on water demand modelling and management reports an increasing number of successful studies and use cases (for a review, see Cominola et al., 2015, and references therein) demonstrating the benefits of smart metering technologies to support demand-side management strategies that can complement traditional water supply development (Gleick et al., 2003). Recent applications showed that effective demand management strategies are a result of understanding users' typical behaviours and the associated consumption patterns at different spatial and temporal resolutions (Jorgensen et al., 2009, 2013). Yet, the adoption of smart metering technologies is still limited in utility and commercial applications because utilities are conservative, reluctant to change (Stewart et al., 2010), and the costs, benefits, and tradeoffs for investing in smart meters are unclear.

At coarse temporal resolutions, water use data are usually collected on a quarterly or monthly basis focusing on the urban or suburban scale to inform strategic regional planning with predictions of the aggregated water demand at the municipal or district level (House-Peters and Chang, 2011). Moving towards higher temporal resolutions, the advent of smart meters in the late 1990s opened up a new potential to better characterize water demand patterns on the basis of water consumption data at very high spatial and temporal resolution, for instance enabling end-use disaggregation (Nguyen et al., 2013) and better estimates of demand peaks (Beal et al., 2016). Depending on the technology exploited in the meter, we can distinguish four types of sensors: (i) Accelerometers (e.g., Evans et al., 2004), which analyze vibrations in a pipe induced by the turbulence of the water flow; (ii) Ultrasonic sensors (e.g., Mori et al., 2004), which estimate the flow velocity by measuring the difference in time between ultrasonic beams generated by piezoelectric devices and transmitted within the water flow; (iii) Pressure sensors (e.g., Froehlich et al., 2011), which estimate the flow rate as a function of the pressure change generated by the opening/close of the water devices valves via Poiseuille's Law; (iv) Mechanical or magnetic flow meters (e.g., Mayer and DeOreo, 1999; Kowalski and Marshallsay, 2003), which correlate the number of revolutions or nutations of a piston, magnet, or disk to the water volume passing through the meter. These sensors offer theoretical resolutions finer than 0.02 liters, but cost, staff time, privacy, and regulations strongly constrain the actual resolutions that can be guaranteed by large scale Advanced Metering Infrastructure (AMI) (Boyle et al., 2013). Understanding the tradeoff between the value of the information provided by high-resolution data and metering economic and operational costs is crucial to inform the design of smart metering networks as well as to discover and guard against unintended consequences of deployment options.

At one extreme of this tradeoff curve, the availability of high-resolution smart metered data generates several opportunities for advancing water demand management. Sub-minute sampling resolution is needed to run most water end-use disaggregation algorithms and provide a reliable breakdown household level water use into different categories (e.g., shower, toilet, clothes washing machine) (Nguyen et al., 2013b, 2015). The knowledge of timings, peak-hours, and frequencies of use of the different consumption devices is key to understand consumer behaviours, identify consumption anomalies, and, ultimately, design targeted personalized demand management strategies, including economic incentives to upgrade inefficient appliances (e.g., Mayer et al., 2004; Suero et al., 2012) or awareness campaigns targeting specific end uses (e.g., Willis et al., 2010; Abdallah and Rosenberg, 2014).

Yet, this metering strategy inevitably increases the amount of data the water utility must collect and handle. Sampling at one-minute resolution, for instance, implies replacing the four annual readings per user with 525,600 data readings. This increase may challenge business hardware and software performance due to existing issues with respect to power source, battery life, telemetry network capacity, and black spots, i.e., data gaps, and billing software (Stewart et al., 2010). In addition, there is still no consensus about the best architecture to store consumption data. A centralized system facilitates checking the accuracy of the collected data, while a distributed one would significantly reduce transmission costs (Oracle, 2009).

Intermediate metering strategies attempt to balance these competing interests by sampling at resolutions of a few minutes to 1 hour. Although this choice prevents an accurate characterization of end-use consumption profiles from aggregate signals with time spacing larger than a minute (e.g., toilet flushing or tap usage usually last a few seconds, showering a few minutes, thus it is hard to unpack end-use information from aggregate signals at coarser resolutions), these data still provide valuable information to water utilities and agencies for designing and managing the water supply system. In fact, sub-daily sampling resolutions allow extracting consumption patterns and accurately estimating the total water demand that the water supply system should be able to deliver to a group of users (e.g., Cardell-Oliver, 2013). This can be seen by looking at the sample water use data reported in Figure 1, which shows how the variability of water use patterns is gradually masked as data are sampled at progressively longer time intervals. Moreover, medium-resolution data can also support the identification of anomalous events occurring on the network or downstream the household meter (e.g., leakage, empty houses, or frauds). This is a major interest for water utilities because post meter leakages account for up to 10% of total residential water use. Reducing the amount of water wasted through leakages also generates secondary benefits in terms of reduced water-related energy consumption and treatment costs (see, for instance, Britton et al. (2013) study in Australia).

This tradeoff between metering cost and accuracy can influence the type of demand management operations and strategies available to utility managers, program costs, and corresponding benefits for water consumers and utilities. In this paper we quantitatively assess how different temporal resolutions to read residential water meters impact information retrieval and demand management by answering the following research questions: which aspects of water demand modelling and management can be accurately, feasibly, and cost-effectively informed by different data resolutions? Are there resolution thresholds discriminating on these aspects?

To answer these questions, we contribute a comparative framework to explore the tradeoffs between data sampling resolution and accuracy in end use disaggregation, time to detect leaks, errors in estimating the volume and timing of peak flows, data storage requirements, and commercial availability. Given the low availability of residential water use data at different resolutions, we first developed a stochastic simulation model named STochastic Residential water End-use Model (STREaM). STREaM relies on a large dataset including observed and disaggregated water end-uses from over 300 single-family households in nine U.S. cities (DeOreo, 2011). STREaM generates synthetic time series of water end use with diverse sampling resolutions. Second, we applied the comparative framework on STREaM output. STREaM allows the generation of residential water demand traces at the end-use level up to a 10-second resolution. Each water end-use fixture in our model is characterized by its signature (i.e., typical consumption pattern), as well as its probability distributions of number of uses per day, single use durations, water demand contribution and time of use during the day. STREaM was used to generate a set of annual consumption traces for 500 heterogeneous households in terms of both number of occupants and efficiency of the end-use fixtures. The implications of adopting different data sampling resolutions are then explored by aggregating the generated 10-second water consumption trajectories up to the 1-day resolution and by evaluating a set of performance metrics including end-use disaggregation accuracy, costs due to leakage detection delay, precision in reproducing volume and timing of water demand peaks, data storage requirements, and commercial availability of metering systems. We use the framework to explore which temporal data resolutions might enable water demand management actions, utilities operations, and communication of customized information to water consumers. Findings from our multi-resolution assessment can support further research and commercial development in water meters and deployment of AMI, as well as assist utilities in trading off benefits from second-to-minute data sampling resolution and cost of adopting and maintaining high-resolution metering infrastructures.

The paper is organized as follows: the next section introduces the proposed comparative framework for multi-resolution assessment and formalises the set of performance metrics used in this study. Section 3 illustrates the synthetic generation of residential water demand traces via STREaM. Numerical results are then reported and discussed in terms of their policy implications. The last section conclude with final remarks and directions for further research.

2. Comparative framework for multi-resolution assessment

To assess the implications to record water consumption data at different temporal frequencies on water demand modelling and management, we introduce a comparative framework composed of seven performance metrics (Table 1). Each metric quantifies the impact of temporal data resolution on a specific aspect of water demand modelling and management, i.e., end-use disaggregation, leakage detection, peak demand estimation, data storage, and commercial availability of water meters. These components and related metrics are important because managers and researchers want to know how well data can be used to disaggregate end-uses, inform customized feedback, detect and respond to fix leaks, avoid related water waste and costs, and estimate peak water demands. Managers are also interested in feasibility aspects, such as the volume of data generated and commercial availability of metering systems for purchase.

2.1. End-use disaggregation

The literature inconsistently defines performance metrics to assess the suitability of end-use disaggregation methods (Makonin and Popowich, 2015). In this work, we select two performance metrics among those available in the literature to assess disaggregation at different temporal resolutions both in terms of accuracy in assigning water consumption to the contributing fixtures, and capability to properly reproduce water end-use time series (i.e., their pattern, with time of use and peaks). The first metric is the Appliance Contribution Accuracy, formulated as the average of the Water Contribution Accuracy (WCA) across all households and fixtures. We derived its formulation adapting similar metrics measuring the power contribution accuracy/error in the electricity field (Cominola et al., 2017):

$$\text{Appliance Contribution Accuracy} = \frac{1}{N} \times \sum_{i=1}^{N} \frac{\sum_{k=1}^{M_i} \text{WCA}_i^k}{M_i},$$
$$\text{WCA}_i^k = 1 - \frac{\left| \sum_{t=1}^{H} y_{i,t}^k - \sum_{t=1}^{H} \hat{y}_{i,t}^k \right|}{\sum_{t=1}^{H} \bar{Y}_{i,t}} \tag{1}$$

where N is the total number of households metered, M_i the total number of water fixtures in each house i, H is the length of the monitoring period, $\bar{Y}_{i,t}$ the total observed water use of house i at time t, and $y_{i,t}^k$ and $\hat{y}_{i,t}^k$ are, respectively, the observed and estimated water consumption for appliance k of house i at time t (t is a discrete-time index). The above metric measures the accuracy of end-use model in assigning the water contribution share to each fixture. Water Contribution Accuracy reflects cases when the disaggregation algorithm correctly assigns positive water use to an appliance when the appliance was actually used plus cases when the algorithm assigns zero water use to an appliance that was not used. The closer accuracy is to 1, the better the algorithm disaggregates water use by appliance, and vice versa for accuracy values close to 0. Accurate estimations of the contribution of each end-use to total demand allow water managers to tailor water demand management strategies to users and provide customized feedback (Sønderlund et al., 2016). As a second metric to assess the performance of end-use disaggregation, we selected the Appliance Root-Mean Square Error (Appliance RMSE), formulated as:

$$
\begin{aligned}
\text{Appliance RMSE} &= \frac{1}{N} \times \sum_{i=1}^{N} \frac{\sum_{k=1}^{Mi} \text{NRMSE}_i^k}{M_i}, \\
\text{NRMSE}_i^k &= \frac{\sqrt{\frac{1}{H}\sum_{t=1}^{H} (y_{i,t}^k - \hat{y}_{i,t}^k)^2}}{\max(y_{i,t}^k) - \min(y_{i,t}^k)}
\end{aligned}
\tag{2}
$$

where N, M_i, H, $y_{i,t}^k$ and $\hat{y}_{i,t}^k$ are as previously and NRMSE is the normalized root-mean square error for appliance k in house i. Performance metrics based on square error or RMSE have been widely used in the field of end-use disaggregation (e.g., Figueiredo et al., 2014; Piga et al., 2016; Rahimpour et al., 2017). This second metric is complementary to the first because Appliance Contribution Accuracy assesses end-use accuracy at the level of aggregate end-use contribution, while Appliance RMSE quantifies model over- and under-estimation of water use time series, thus allowing for a more detailed evaluation the capabilities of an end-use algorithm to reproduce end-use time series patterns. This is key for demand modelling and management because low RMSE values allow retrieving accurate information on peak water use, end-use frequencies, time of use for the major end-uses, and to monitor changes in demand patterns overtime. In the above formulation, we normalized RMSE to account for the different flow range of each appliance. We divide by the flow range rather than the average flow value because water datasets are highly unbalanced with numerous zero readings. Dividing by a mean close to zero would give high errors independent of the appliance type. Dividing by the range balances estimation error with the maximum error that can potentially occur at each time step.

The main limits to use the Appliance Contribution Accuracy and Appliance RMSE to assess end-use disaggregation performances are related to the formulation of the first metric. Overall, if two or more appliances flow in similar ranges (as can happen with indoor household water fixtures) and an algorithm incorrectly disaggregates the end uses, terms in the numerator of Eq. 1 will be large and cause the WCA to be close to 0. Dividing by the total observed water use $\bar{Y}_{i,t}$ in the evaluation of WCA maintains the relative importance of appliances but can mask small inaccuracies for individual appliances. If an appliance is used only occasionally (i.e., water use is often 0) a disaggregation algorithm might classify all estimated use as zero and achieve a WCA close to 1 even though it missed a few infrequent events for the appliance. Finally, WCA represents an aggregate performance of end-use disaggregation and can provide useful information to utilities that use smart meter data to communicate a breakdown of water use by appliance to their customers. Considering the above limitations, care should be taken to use the Appliance Contribution Accuracy with unbalanced datasets. Yet, a coupled analysis of Appliance Contribution Accuracy with other, less aggregated, performance metrics such as Appliance RMSE can help interpret results.

2.2. Leakage detection

Leakage detection represents a major challenge for utilities because of direct and indirect costs of leakages (Britton et al., 2013). To assess the potential to correctly detect leaks, we define the Average Water Loss performance metric that is based on the average water volume lost for all end uses (in liters) before the leakage is detected:

$$\text{Average Water Loss} = \frac{\sum_{i=1}^{N}\sum_{t=LS_i}^{LD_i}(\bar{Y}_{i,t} - \sum_{k=1}^{M_i} y_{i,t})}{N} \tag{3}$$

where $\bar{Y}_{i,t}$ is the total observed water use of house i at time t, $\sum_{k=1}^{M} y_{i,t}$ is the legitimate water use of house i at time t over its M_i appliances, LS_i the starting time of a leakage in house i, LD_i the time step when the leakage is detected in house i, and N the total number of households metered. Lower Average Water Loss indicates faster leak detection. This formulation assumes that only one leak episode occurs along the whole time series of water use of each house. In this research, we do not consider the subsequent time after detection to respond, locate, and fix the leak. Thus, $LD = LS + r$, where r represents the time between the start of the leakage and its detection and is equal to $r = u - (\frac{LS}{u} - \lfloor\frac{LS}{u}\rfloor)$ (u is the considered sampling interval, e.g., 1 minute, 1 hour). This treatment allows isolating the sole effect of data sampling resolution on leak detection without including errors and impacts deriving from the application of a given leakage detection algorithm (e.g., Minimum Night Flow (Britton et al., 2008)). This treatment also ignores how promptly the utility can respond to fix the leak and time to complete the repair. In reality, the time to detect a leak is likely shorter than the subsequent time to respond and fix the leak. Thus, here the volume of water loss depends only on the sampling time frequency and the size of the leak.

2.3. Peak demand estimation

Data sampling resolution affects the estimation of water demand peaks at the various scales (i.e., household, district, and utility), which is key to design water distribution systems and support management strategies to reduce or shift peak demand (Beal et al., 2016). In order to assess the impact of data sampling resolution on the accurate estimation of water demand peaks, we formulate the Peak Estimation Error:

$$\text{Peak Estimation Error} = \frac{1}{H_{\text{day}}} \sum_{d=1}^{H_{\text{day}}} \left| \frac{\bar{Y}_{d,u_{benchmark}}^{\text{TOT,PEAK}} - \bar{Y}_{d,u}^{\text{TOT,PEAK}}}{\bar{Y}_{d,u_{\text{benchmark}}}^{\text{TOT,PEAK}}} \right| \tag{4}$$

where $\bar{Y}_{d,u_{\text{benchmark}}}^{\text{TOT,PEAK}}$ is the observed peak water use for day d, aggregated over all metered households, and metered with the finest available resolution $u_{\text{benchmark}}$; $\bar{Y}_{d,u}^{\text{TOT,PEAK}}$ is the observed peak water use for day d, aggregated over all metered households, and metered with sampling resolution u; and H_{day} is the number of monitored days. It follows that, at the 1-day sampling frequency, the reported flow is the average flow per day. The Peak Estimation Error measures the percentage of under- or over-estimation of peak demand, against the best available peak observation (i.e., the one observed at the finest available resolution).

Data sampling resolution affects also the ability to identify the times of the day when demand peaks occur, and coarse resolutions can mask peaks with short duration and high magnitude. Accurate peak time estimates can help schedule supply operations and pumping, as well as inform programs to shift peak demands. To complement the Peak Estimation Error metric with information on time of the peak, we define the Peak Estimation Time Gap:

$$\text{Peak Estimation Time Gap} = \frac{1}{H_{\text{day}}} \sum_{d=1}^{H_{\text{day}}} \left| t_{d,u_{benchmark}}^{\text{TOT,PEAK}} - t_{d,u}^{\text{TOT,PEAK}} \right| \tag{5}$$

where $t^{\text{TOT,PEAK}}_{d,u_{\text{benchmark}}}$ is the time step when the observed peak water use for day d for all households occurs measured using the finest available resolution data $u_{\text{benchmark}}$; $t^{\text{TOT,PEAK}}_{d,u}$ is time step when the observed maximum value of water use for day d for all households occurs measured with data of sampling resolution u; and H_{day} is, as before, the number of monitored days. The Peak Estimation Time Gap measures, in minutes, the average time lag between the peak demand measured from a time-series at a specified temporal resolution and the finest temporal resolution.

Metrics for the magnitude and timing of peak demand are readily modified to include other metrics of interest to utilties such as minimum and average demands. To keep the set of metrics compact, we only consider peak demand in this work.

2.4. Data storage

While providing more detailed data on water use, high-frequency smart metering inevitably increases the size of datasets to transfer, store, and analyze, plus related costs (Oracle, 2009). Here, we define a Data Size metric that quantifies the amount of memory needed to store water use data at a given resolution:

$$\text{Data Size} = 4 \times 2 \times R_{\text{year}} \tag{6}$$

where R_{year} is the number of water use readings collected for a single household over a year. R_{year} depends on the sampling frequency (e.g., it is equal to 365 with daily sampling frequency, 8760 with hourly sampling frequency, etc.). In the definition of the Data Size metric, we assume that each monitored water consumption data point can be stored as a record of 2 floating-point variables, i.e., date/time stamp and corresponding water consumption reading, using 4 bytes of memory each (Zuras et al., 2008), thus Data Size is measured in bytes/(household $\times$ year). This storage assumption is conservative and provides an upper bound reference metric. In practice, there are smarter ways to transmit and store data such send one starting date/time stamp then follow with the list of regularly-spaced readings (this would redue the storage requirement indicated by the metric by roughly half). Smarter meters may do more initial processing on the meter itself before transmitting more aggregated data.

2.5. Commercial availability of water meters

Numerous commercial water metering systems exist and have been used both in experimental trials, as well as real-world deployments (Boyle et al., 2013). Their cost, storage capability, frequency of data collection and transmission depend on the meter, the register, associated hardware and accessories, and available power. In order to assess the actual capabilities of commercial meters based on state-of-the art experiences, we define the Availability as a binary metric. This metric assumes a value of 1 if a metering system is commercially available and can sample water use with a given resolution. Otherwise, the metric takes a value of 0 (i.e., no commercial metering systems exist or water use data can only be sampled at the specific sampling frequency with *ad hoc*, non-commercial systems).

3. STREaM STochastic Residential water End-use Model

As real world residential water use data with different temporal resolutions were not available, we synthetically generated them with a stochastic water end use generator. STREaM (STochastic Residential water End-use Model) synthetically generates time series of residential water use at the end-use level with time resolutions spanning from 10 seconds to one day.

3.1. Model structure

The structure of STREaM is built upon the prototype synthetic water consumption generator presented in Cominola et al. (2016). In short, given a user-defined house with specified number of occupants, available water consuming fixtures, fixture efficiency, time horizon, and sampling resolution, STREaM simulates time series of water use for individual appliances and their sum as total household water demand. STREaM relies on the assumption that the water use time series of the j-th water end-use fixture (e.g., toilet, faucet, shower, etc.) in the d-th day of the simulation horizon can be characterized by the following elements: (i) number of times the j-th fixture is used during the day (we will refer to each usage as *consumption event* hereafter); (ii) starting time of use during the day for each consumption event; and (iii) duration and volume of water used for each consumption event. In addition, we assume that the pattern of each end-use consumption event is characterized by a specific *signature*, i.e., the characteristic water use flow pattern over time of a single consumption event for a specific end-use.

According to the model structure illustrated in Figure 2, the inputs required by STREaM are (i) sample size N, i.e., number of households for which STREaM will simulate end-use time series of water use; (ii) house demography, i.e., number of occupants for each house in the sample $O = \{o_1, o_2, ..., o_N\},\ o_i > 0\, \forall i \in [1, N]$; (iii) fixture presence $P = \{p_1, p_2, ..., p_M\},\ p_k \in \{0, 1\}\ \forall k \in [1, M]$, i.e., a binary index specifying the presence (absence) of the k-th fixture in the i-th household; (iv) fixture efficiency level $E = \{e_1, e_2, ..., e_M\},\ e_k \in \{0, 1\}\ \forall k \in [1, M]$, i.e., a binary index specifying the efficiency level (*standard* or *high*) of each fixture in each household; (v) length of the simulation horizon H; (vi) time sampling resolution $u,\ u > 0$ for the output water use time series. The finest temporal resolution allowed by STREaM is 10 seconds. As output, STREaM returns the end-use time series of water use y_i^k for each house i and its fixtures k, as well as each household's total water use time series $\bar{Y}_i = \sum_{k=1}^{M} y_i^k$.

The core of STREaM is the generation of end-use water use time series. Let' s consider the i-th house, characterized by o_i occupants, fixture presence P_i and fixture efficiencies E_i. STREaM generates the end-use time series y_i^k according to the following procedure:

- **Sample Daily Consumption Events.** The number of consumption events for each fixture k and each day d of the simulation horizon H is Monte-Carlo sampled from its probability distribution as $NCE_{i,d,k} \sim \mathcal{P}(NCE_k|o_i, e_{i,k})$, where $\mathcal{P}(NCE_k|o_i, e_{i,k})$ is the probability distribution of the number of usages per day for appliance k, conditioned to the number of house occupants (o_i) and fixture efficiencies ($e_{i,k}$).

- **Sample Event Characteristics.** For each consumption event $l \in [0, NCE_{i,d,k}]$, duration (D) and water volume (V) are Monte-Carlo sampled from the joint duration-volume probability distribution of the k-th fixture, conditioned to o_i and appliance efficiency $e_{i,k}$ as $(D_{i,d,k,l}, V_{i,d,k,l}) \sim \mathcal{P}(D_k V_k | o_i, e_{i,k})$. The joint probability is considered, as volume of water used and event durations are generally correlated. Also, the time of use of each consumption event l is sampled from its conditioned probability distribution $T_{i,d,k,l} \sim \mathcal{P}(T_k | o_i, e_{i,k})$.

- **Scale Event Signatures and Generate Event Time Series.** The time series of water use of each water consumption event is generated by uniformly selecting one of the specific signatures of the considered fixture k and scaling it in duration and magnitude to match the sampled values of duration and water volume $(D_{i,d,k,l}, V_{i,d,k,l})$. As the number of signatures available for each water end-use can vary in the input dataset, STREaM randomly selects one, among the available signatures, and then scales it in duration and magnitude. In order to do so, first randomly chosen points of the selected signature are iteratively removed/replicated, in order to match the desired event duration $D_{i,d,k,l}$. Then, the magnitude of each point of the signature is scaled proportionally to its original value, so that the integral under the signature matches the desired water volume $V_{i,d,k,l}$. Finally, the scaled signature is positioned over the end-use time series y_i^k according to its time of use $T_{i,d,k,l}$.

The above procedure is iterated from step 1 to step 3 until the simulation is completed, for all the M fixtures and the days of the simulation period H. Finally, end-use time series of water use y_i^k for each house i and its fixtures k, as well as its total water use time series $\bar{Y}_i = \sum_{k=1}^{M} y_i^k$ are returned, scaled to the chosen sampling resolution u. It is worth noticing that the procedure adopted in STREaM allows generating multiple simultaneous end-use events, in order to reproduce potentially overlapping water uses as they occur in any home in reality. Thus, as STREAM allows potentially concurrent events, end-use disaggregation should aim at decomposing the aggregate signal into its components, rather than classifying purely isolated end-use events.

Optionally, STREaM can include the superimposition of a randomly sampled end-use leakage on the total household water use time series w_i^k, simulating the partial break or total burst of one end use. We synthetically generated each leak by uniform sampling of four parameters, i.e., the *leaking end-use* k (uniformly sampled among the available end uses), *starting time* t_{start} (uniformly sampled over the length of the time series), *rise length* r_{length} (uniformly sampled between the leakage starting time and the length of the time series), *rate of rise* r_{rate} (uniformly sampled as one of four categories defined in Britton et al. (2009), i.e., constant leak, linear, polynomial, and exponential rate of rise). We assumed that the maximum flow reached by the leakage only depends on the leak end-use, and is equal to the maximum value assumed by that end-use over the whole time series.

3.2. Data source and STREaM calibration

We use a large dataset for single-family households observed and disaggregated water end-uses in nine U.S. cities between 2007-2009 collected by Aquacraft Inc. (DeOreo, 2011). Water use was measured over two weeks at 10 seconds resolution for 288 houses. The houses were built after 2001 and have appliances and fixtures that comply with the standards set forth by the Energy Policy Act of 1992 (United States, 1992) (Standard-efficiency houses, hereafter). The study also measured water use for 25 houses that were built after 2007 and comply with the WaterSense high efficiency standards (High-efficiency houses, hereafter).

The number of occupants was reported for each house in the Standard Houses dataset. 11% of the households have 1 occupant, 45% have 2 occupants, 15% have 3, 18.4% have 4, 7.6% have 5, and 3% have more than 5 occupants. Aquacraft Inc. disaggregated water use events for each end use using their FlowTrace Wizard software (DeOreo et al., 1996), reporting the start time, duration, and volume of each event for all the major indoor water end uses, namely shower, toilet, faucet, bathtub, clothes washer, and dishwasher. This version of STREaM focuses on and includes indoor use because available appliances and their operation are consistent across households in the nine cities. We exclude outdoor uses because they differ across households and cities in seasonality use, types of outdoor irrigation systems, landscape type, and area. Future work could expand STREaM to include outdoor use. In total, Aquacraft disaggregated 240,443 separate water use events for 313 houses over 3,731 days (Table 2). Dishwasher and clothes washer events cover the entire appliance cycle and include intermediary wash, rinse, etc. cycles.

We used event volume, duration, time of use, and number of occupants statistics from the above dataset to estimate corresponding probability distributions required by STREaM. After fitting multiple distributions to the data, we found that the number of events per day is best modelled with a negative binomial distribution in 70% of the cases, and Poisson distribution in the remaining cases. Event start time is always modelled with a Kernel distribution. Finally, we jointly modelled event durations and volumes with two-component Gaussian Mixtures.

We noted that the dataset of High-efficiency houses only included duration and volume data for end-use events. Thus, we assumed distributions of start time and number of uses per day identical to those of Standard-efficiency households. The rationale behind this hypothesis is that technological efficiency mostly influences flows (thus volume) rather than user's behaviours such as starting time, duration, or frequency (Abdallah and Rosenberg, 2014). Moreover, given the reduced data for High-efficiency houses, we were unable to estimate duration and volume statistics as a function of number of house occupants. As a last step, we built the dataset of water fixture signatures by using GetData Graph Digitizer software (GetData Graph Digitizer [Computer software], 2017) to visually extract signature patterns from Acquacraft reports (DeOreo, 2011). The number of signatures available for each end-use in STREaM varies between 1 and 15.

3.3. STREaM validation

To validate the STREaM output, we evaluated the observed total average household daily water use by summing the volume of observed water use for all end-uses across the reported day. We validated STREaM according to the following procedure. First, we generated a 1-year long water use time series at 10-second resolution for a sample of 250 standard efficiency households. We included all available end uses, i.e., toilet, shower, faucet, clothes washer, dishwasher, and bathtub, and set the household demography coherently with the occurrences we found in the data used for STREaM calibration (Section 3.2). Second, we summed the generated end-use time series for each household into time series of total household water consumption, and aggregated these to the daily scale (see Section 3.1). Finally, we cross-compared the distribution of simulated and observed total daily water use (Figure 3), which a non-parametric Mann-Whitney U test (McKnight and Najab, 2010) showed were similar (significance = 1%, p-value = 0.012) if values above 20.5 liters/(household*day) were considered for both samples. While the figure shows that the distribution of STREaM output well fits observations, STREaM slightly overestes low daily water use. This overestimation is likely due to the cumulative error resulting from STREaM calibration, when fitting the lower tails of end-use distributions, and specifically those regarding statistics on the number of events per day. As a further test, we computed the average household daily water use and obtained values of 454 and 464 liters/(household*day) for the synthetic and actual datasets.

As further validation, we also performed independent non-parametric Mann-Whitney U test for each end use. These tests compare simulated and observed distributions of number of usages per day, event volumes, durations, and times of use at 10-second sampling resolution for the same 250 standard efficiency households. The outcomes of the Mann-Whitney U tests performed with 1% significance level suggest to accept the null hypothesis of similar distributions for most cases Table 3. Two exceptions were for toilet and faucet end-uses, which are often characterized by short and small-volume water consumption events. Thus, small estimation errors can highly impact on the outcome of statistical tests. Since the time-of-use data were fitted with non-parametric Kernel distributions, we do not report the results of the Mann-Whitney U tests as the sampled timings from them are unlikely to line up with observed times at 10-second sampling resolution. Rather, we visually compare (Figure 4) the time of use of STREaM end uses against observations. The visual comparison shows that the distribution of STREaM output satisfactorily match observations, with small overall timing underestimations.

Overall, the validation demonstrates that STREaM statistically well reproduces the variability of observed data.

4. Application

4.1. Experimental settings

To assess the value of data sampling resolution, we use the performance metrics detailed in Section 2 to evaluate water use time series generated via STREaM for a sample of 500 heterogeneous households. These 500 households differ in terms of demography and efficiency of end-use fixtures. We set the number of occupants to the same proportions adopted for model validation (see Section 3.3), and equipped all houses in the model with toilet, shower, faucet, clothes washer, dishwasher, and bathtub end-uses. We set 50% of appliances as Standard-efficiency households and 50% as High-efficiency households.

Given the above settings, we generated 1-year long water end-use time series for each household, with 10 second sampling frequency. We then aggregated the time series to resolutions of 1 min, 5 min, 15 min, 1 hour, and 1 day to perform multi-resolution assessment. The 1-min resolution has been recognized to be a critical threshold for certain end-use data analytics also in the electricity sector (Armel et al., 2013). We chose 5 min because more than 95% of consumption events in the original dataset used by STREaM has a duration shorter than 5 minutes. Also, 5 min resolution has been adopted in utility metering programs (Mohassel et al., 2014). Finally, 15 min, 1 hour, and 1 day are commonly adopted resolutions in most real-world smart metering deployments (Cardell-Oliver, 2013; Cominola et al., 2015; Thames Water Utilities Limited, 2017).

Additional experimental settings were required to evaluate the performances metrics on end-use disaggregation. We adopted the supervised version of HSID (Hybrid Signature-based Iterative Disaggregation) algorithm for end-use disaggregation (Cominola et al., 2017) and finely tuned it (i.e., calibrated by trial and error the parameters of its Factorial Hidden Markov Models and Iterative Dynamic Time Warping components) to perform end-use disaggregation of water consumption data on a set of 6 generated households with 1 to more-than-5 occupants to account for different frequencies of use due to increasing number of occupants. For each selected household, we calibrated HSID using 2-months data and evaluated the Appliance Contribution Accuracy and Appliance RMSE metrics (Section 2.1) by averaging the outcomes of 1-month end-use disaggregation per household.

4.2. Multi-Resolution Assessment: Numerical Results

A summery of results of metric performance (Figure 5, rows) of each sampling frequency (Figure 5, columns) shows a tradeoff between the top four performance metrics and the bottom two. The value of information for demand modelling and management increases with data sampling resolution (Figure 5, darker colors to left and higher sampling frequency). Accuracy of leakage detection, end-use disaggregation, and peak demand estimation, increase when using data at resolutions of 1 minute or a few seconds. At coarser resolutions, leakage volume dramatically increases, water demand peaks are underestimated by at least 20%, and average RMSE in end-use disaggregation exceeds 5%. At the same time finer resolution data imply larger data size and limited or no commercial products available for utilities to deploy. Most smart metering trials and experiments from the state-of-the-art literature (Cominola et al., 2015) exploit custom metering systems developed with *ad hoc* settings to collect data with minute or finer time intervals. Conversely, due to technical issues related, for instance, to preserving meter battery, most water utilities currently adopting smart meters are collecting water consumption data with hourly, or at most 15-minute, data sampling resolution.

4.2.1. End-use disaggregation

Appliance Contribution Accuracy exhibits a u-shaped pattern where accuracy is high for 1-day resolution data, lowers for intermediary frequencies, and increases again at 1-second resolution. Overall, ACA ranges between 89% and 95% and follows prior studies that demonstrated to achieve disaggregation accuracies in the order of 80-90% with an intrusive calibration process and data sampled at sub-minute resolution (e.g., Nguyen et al., 2013; Froehlich et al., 2011). The large ACA value of 95% for 1 day sampling resolution is counterintuitive. However, we can explain this finding because the water use contribution of major end-uses can also be approximated by their average proportion of total use. An average proportion coupled with a long simulation horizon (1 month) relative to the 1-day sample frequency means the model estimated ACA will closely approximate the actual appliance contribution. Yet, ACA does not quantify model over- and underestimation in reproducing the patterns of water use time series. For this reason we assess end-use disaggregation performance via a coupled analysis of ACA and NRMSE.

The average 25% and 75% confidence limit on appliance RMSE grows substantially with coarser sampling resolutions (Figure 6). Taken together with ACA, three findings emerge. First, the aggregate contribution of each end-use is well estimated even at medium-low resolutions. Second, time series patterns are well estimated only for finer resolutions. And third, water use by each major end-use can be fairly well approximated by their average value. An in-depth analysis breaking down these aggregate results for each appliance (Figure 7) confirms the above comments. In the figure, Water Contribution Accuracy does not present a well-defined pattern across resolutions. Moreover, it can achieve high performance values even at low sampling resolutions, and it generally high for the frequently used appliances such as the toilet. Conversely, Normalized RMSE monotonically decreases with coarse data sampling resolutions, suggesting that fine sampling resolutions are needed to achieve high disaggregation accuracy.

These findings can only be identified by controlled experiments like the one carried out in this work, where data are synthetically generated. However, experiments can miss changing trends of real-world data over time due to user behavioural changes between weekdays and weekends, attitudes, and climatic factors, e.g., seasonality and drought conditions that would emerge if outdoor uses were included (Kenney et al., 2008). Our results show large Appliance RMSE for course data sampling resolutions (RMSE gets up to above 30% for daily data sampling resolutions, meaning end-use estimates are not reliable at this resolution for management applications). Appliance RMSE values would very likely be worse if disaggregating real-world data collected at minute to hourly frequency, as (i) demand patterns would be affected by heterogeneous, irregular, water use behaviours, (ii) water use signatures would be much more diverse than the signatures embedded in STREaM, and (iii) there would be limited calibration data. Considering these several factors, we find that only resolutions of few seconds or, at most, 1 minute can be used to perform accurate end-use disaggregation, provide customized information about consumption of each end-use, peak magnitude, and time of use when multiple and potentially overlapping fixtures are active. These results are also consistent with the analysis by Armel et al. (2013) in the electricity field. Rather than an a priori expectation, it is worth mentioning about this consistency between water and electricity to inform potential integrated water-energy approaches and solutions.

Finally, the results may also depend on the HSID algorithm chosen for disaggregation (Cominola et al., 2017), thus the application of different disaggregation algorithms might change the numerical values obtained for the two performance metrics.

4.2.2. Leakage detection

Results for Average Water Loss demonstrate that data resolution strongly impacts the volume of water that can be saved by more prompt leak detection. Fine resolutions of 10 seconds to 1 minute allow prompt detection of small leaks that otherwise would easily blend with signal noise. Also, the amount of water lost significantly increases at a 5-15 minute resolution. These results do not include leakage after a leak is detected and before it is fixed. Current leakage detection systems typically act on longer detection time intervals, which depend on the leak flow rate (Puust et al., 2010). Moreover, methods based on Minimum Night Flow (Britton et al., 2008) usually detect leakages with above daily delays and their accuracy and rate of false alarms can be affected by signal noise on consumption time series. Thus, there is need for research to improve leakage detection systems (use high frequency data, reduce false alarms) in real case studies. For example, even with a medium resolution of 5 minutes, more than 20 liters are wasted on average, i.e., approximately the amount of water used for a 2.5-minute shower with a flow of 9.5 liters/minute (equal to approximately 2.5 gallons per minute) (DeOreo et al., 2016). At a daily resolution the water loss increases to more than 6 cubic meters, i.e., about the same amount of water an average Italian consumer would use in more than 1 month (approximately 35 days) — the average per-capita daily water use in Italy is approximately 175 liters/(person $\times$ day) (Italian National Institute of Statistics, 2013). At an average price of 2.03 $\$/m^3$ (Intelligence, Global Water, 2011), the leakage would cost the customer 25\$/day. There are also indirect costs for water-related energy use and waste water treatment. Thus, both public and private water suppliers should be interested to use high frequency data collection to improve leak detection.

4.2.3. Peak demand estimation

Peak demand estimation error increases dramatically as the resolution becomes coarser, growing to 60% error with a daily sampling resolution. This increasing estimation error derives from aggregation and averaging of data as the sampling resolution decreases. Consequently, peaks (minimums and maximums) associated with high frequency measurements are dampened and flattened at the measurement resolution becomes coarser Figure 8. At the extreme, the single daily reading is a flat line that shows no variability. Similarly, the Peak Estimation Time Gap grows steadily from 24 min at 1-minute sampling frequency to more than 15 hours (9.4e+02 minutes) at daily sampling frequency. This values can be acceptable for scheduling hourly supply operations, and still allow discriminating between time windows in the day (e.g., morning, afternoon, evening, night) to design time-dependent demand management strategies (e.g., pricing schemes). Yet in real cases with more noisy data, higher number of users, and more asynchronous behaviours, such performance might degrade and hamper the capabilities of utilities to optimize hourly operations and design effective hourly pricing schemes.

These results suggest the benefit to undertake demand management programs using high-resolution data. Indeed, Peak Estimation Error is above 20% when the data resolution is coarser than 5 minutes. For water utilities, underestimation of aggregate water demands across the whole community of consumers would limit knowledge about the actual usage of the network. Further, underestimation of peak demands of single-users would hide the variability of demand patterns across different segments of users, thus limiting the proper design and customization of demand management strategies based on pursuing peak shifting or penalizing high peaks of water demand and intense water consumption levels, e.g., block tariffs and dynamic pricing schemes based on time of use (Cole and Stewart, 2013). In this regard, relevant underestimation or incorrect time estimation of demand peaks would also likely limit the capabilities of detecting anomalous behaviours and leakage events based on water use threshold criteria. Finally, inaccurate estimation of demand peaks prohibits advanced data analysis aimed at cross-correlating peak demand with candidate demand drivers (e.g., presence of swimming pools or outdoor end-uses).

4.2.4. Data storage

Data size depends on the sampling resolution. For example, only 3 kbytes are needed to store the 365 daily data points for a single household in a 1-year time series. The storage needed would increase to over 25 Mbytes if the same data were collected at 10-second sampling resolution. Even though storing 25 Mbytes of data per year is low cost for a single household (for instance, the price of Amazon S3 Standard Storage cloud system in the United States is 0.023 $/GB), the cost increases when projected to the utility scale, with increasing costs for cloud infrastructures, as well as database design and maintenance. Data can become a burdensome asset, especially for those utilities that provide water, electricity, and gas. There is also the need to develop techniques to extract relevant information for decision making. We acknowledge that utilities often analyze aggregate water use data, rather than the raw data. In principle, this can relieve them from data storage costs. Yet, data storage is a proxy measure for the computational burden of big data in terms of data analytics and database design. Therefore, utilities should balance the marginal information value given by high-resolution data to their operations and demand management programs, against costs to acquire and maintain hardware, cloud storage, analyze data, maintain databases, and transmit data (e.g., duration of meter battery). Such costs should also consider the frequency of data transmission: systems can use different frequencies to collect and transmit data.

4.2.5. Commercial availability

The results discussed so far rely on the end-use trajectories generated via STREaM under the assumption that we could potentially meter 500 households at sampling resolutions ranging from seconds to one day. In this section, we provide a few examples to describe the ranges of capabilities of existing commercial and customized metering systems to support the sampling resolutions shown in Figure 5. Metering products are numerous, rapidly changing, and there are many ways to combine meters, registers, and data transmission services into a metering system. Meter system accuracy depends on the meter type, service line size, flow rate, water meter age, and whether the meter complies with accuracy recommendations put forward by the American Water Works Association (Barfuss et al., 2011). Below, we discuss similarities and differences between commercially available systems that can provide sampling resolutions down to about 5 minutes. We also review customized systems deployed in recent end use studies that recorded water use at 1 minute or more frequent intervals (Table 4).

A commercial water meter with a commercial analogue register continuously reads total water use, has no power requirements, but has no ability to store readings. Total water use can only be read when a person visits the meter. The same meter configured with a register and radio transmitter allows a person to read the total water use from near the vicinity of the meter (e.g., from a passing vehicle). Many U.S. water providers use this type of system to pass by the meter once per month to record customers water use and bill customers. More advanced registers, such as the Neptune E-

CODER®)R900i (Neptune Technology Group, 2017) can record total water use every 15 minutes for up to 96 days and use Advanced Meter Reading (AMR) technology to transmit readings via a mobile phone network, fixed radio network, or optical sensor to a person standing in the vicinity of the meter. The MetronFarnier Innov8-VN register offers similar capabilities but can record water use every 5 minutes and transmit data once per day via a mobile phone network to a website where a user can view data (Metron-Farnier, 2017).

Water utilities read commercial registers every five minutes to daily to help monitor or detect leaks or reduce non-revenue water. Similarly, AMI systems connect meters and registers to a line-of-sight, fixed radio frequency network that generally operates at 30 MHz or higher (Hawkins and Berthold, 2015). With AMI, a water utility can automatically read meters over the network at daily, hourly, or even 15 minute intervals.

Currently, reading more frequently than about every 5 minutes requires adding customized hardware and software to the meter or register. For example, Mayer and DeOreo (1999); Beal and Stewart (2013); DeOreo et al. (2016) installed a Halls effect magnetic sensor between the meter and register and data logger to record water use every 10 seconds for up to 2 weeks. Horsburgh et al. (2017) improved the system to collect data every 5 seconds, use low-cost, off-the-shelf hardware components, make the software open source, and transmit data via WiFi. And where the commercial meter or register has pulse (2-wire) or AMR (3-wire) outputs — such as the Innov8-VNadditional devices — pulse counters or data loggers can be connected to outputs and programmed to read as frequently as desired for as long as storage memory allows (e.g., every 5 seconds for ~1 month or every 1 second for ~10 days with MadgeTech (2017)).

5. Conclusions

In this research, we assess the tradeoffs between the value of information provided by water use data sampled at different temporal resolutions and economic, operational, and feasibility issues. We answer the questions: (i) which aspects of water demand modelling and management can be accurately, feasibly, and cost-effectively informed with different data resolutions? and (ii) are there resolution thresholds discriminating on these aspects?

We developed the STREaM tool, to synthetically generate residential water demands for individual end-uses of water, estimate total water use, and develop demand scenarios that consider the number of households and heterogeneity/homogeneity in household demographic characteristics and water use appliances. The tool also generates time-series of water demands at varying temporal intervals ranging from days to seconds. We used these features to identify the effects of increasing the temporal frequency at which water use data are generated and sampled on end-use disaggregation, leak detection, peak demand estimation, data storage, and product availability.

We found that increasing sampling frequency to minutes or seconds increases the average accuracy of end-use disaggregation and decreases disaggregation errors. Increased sampling frequency also decreases the volume of leaked water that goes undetected and decreases the error on estimates of instantaneous peak demand. At the same time, more frequent sampling increases required data storage and the need to develop and deploy custom systems. Currently, commercially available water metering systems sample water use down to about 5 minute intervals.

Several benefits of increased sampling frequency will likely spur further commercial development in water meters, registers, and AMR systems that can sample more frequently than every 5 minutes. Increased frequency will permit more accurate estimation of peak demand which is a key parameter to design and size water distribution systems. Increased frequency will also reduce the time it takes to detect leaks, decrease the corresponding volume of leaked water, and reduce non-revenue water. Non-revenue water is an important metric by with water utilities are evaluated. Additionally, sampling at higher temporal frequency will also allow managers to more accurately estimate the water volumes of individual customer end uses (toilets, faucets, showers, etc.) and reduce error. More accurately resolving water end uses can help managers better understand customer water use and component end-uses. It can also help identify appliances, water use behaviours, and customized conservation programs (e.g., rebates for retrofits, technical assistance, and other incentives) that allow customers to save more water with minimal effort and cost. Resolving water end uses can also help utilities determine which customers to target with conservation programs and efforts. Despite these benefits, smart meters are not fully exploited by water utilities because of costs, concerns related to meter battery life, amount of data to transfer and store, and product availability.

The STREaM tool also opens important opportunities for research. STREaM extends the state-of-the-art literature of stochastic models to simulate residential water use (e.g., Blokker et al., 2009; Aksela and Aksela, 2010; Makropoulos and Rozos, 2011; Koutiva and Makropoulos, 2016). First, STREaM can generate end-use data both at the fine spatial scale (household) and time scale (seconds), while other state-of-the-art models either only reproduce the aggregate water use time series of single household (e.g., Aksela and Aksela, 2010), or generate end-use water use data with daily or coarser resolution (e.g., Makropoulos and Rozos, 2011; Koutiva and Makropoulos, 2016). Second, STREaM is built on a uniquely big and consistent dataset of end-use data metered at sub-minute sampling frequency. In contrast, other models from the literature (e.g., Blokker et al., 2009) are usually calibrated using census data and statistic information on fixture and fixture use from heterogeneous sources. Moreover, STREaM allows to generate water use under different demographic and water efficiency conditions, and its output end-use time series represent an actual trajectory with event signatures, rather than simplified pulses. Finally, STREaM is an open-source project, so that the it can collaboratively grow as new data become available.

STREaM can be used to reproduce and benchmark water demand and disaggregation algorithms. For example, other researchers can use generated water demand traces to test and compare new disaggregation algorithms to existing algorithms. Scenario features (number of households and heterogeneity/homogeneity in household demographic and water use appliances) allow researchers to test and compare disaggregation algorithms under a variety of conditions that are typically difficult to measure or observe or may not occur yet in existing water systems. Further, end-use disaggregation experiments can include (i) randomized combinations of types of considered appliances and (ii) randomized number of appliances per type. Outdoor irrigation would enhance the comparative analysis of end-use disaggregation performance for different appliances. At the same time, managers can compare observations from their existing water system to benchmarks or estimate fluctuations in water system demands at higher temporal frequencies than what they can currently measure. Features of the STREaM tool help show implications of measuring water use at higher temporal frequencies. Similarly, managers can use higher frequency estimates to better manage their water systems. Finally, STREaM is provided as an open source software (available at https://github.com/acominola/STREaM/), therefore we wish more end uses and data from different locations will be made available in the future to make it more usable and represent better consumptions of different communities worldwide.

Acknowledgement

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 619172 (SmartH2O: an ICT Platform to leverage on Social Computing for the efficient management of Water Consumption).

References

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, 288–297.

Aksela, K., Aksela, M., 2010. Demand estimation with automated meter reading in a distribution network. Journal of Water Resources Planning and Management 137, 456–467.

Armel, K.C., Gupta, A., Shrimali, G., Albert, A., 2013. Is disaggregation the holy grail of energy efficiency? the case of electricity. Energy Policy 52, 213–234.

Barfuss, S., Johnson, M.C., Neilson, M.A., 2011. Accuracy of in-service water meters at low and high flow rates .

Beal, C.D., Gurung, T.R., Stewart, R.A., 2016. Demand-side management for supply-side efficiency: Modeling tailored strategies for reducing peak residential water demand. Sustainable Production and Consumption 6, 1–11.

Beal, C.D., Stewart, R.A., 2013. Identifying residential water end uses underpinning peak day and peak hour demand. Journal of Water Resources Planning and Management 140, 04014008.

Blokker, E., Vreeburg, J., Van Dijk, J., 2009. Simulating residential water demand with a stochastic end-use model. Journal of Water Resources Planning and Management 136, 19–26.

Boyle, T., Giurco, D., Mukheibir, P., Liu, A., Moy, C., White, S., Stewart, R., 2013. Intelligent metering for urban water: A review. Water 5, 1052–1081.

Britton, T., Cole, G., Stewart, R., Wiskar, D., 2008. Remote diagnosis of leakage in residential households. Journal of Australian Water Association 35, 89–93.

Britton, T., Stewart, R.A., O'Halloran, K., 2009. Smart metering: Providing the foundation for post meter leakage management .

Britton, T.C., Stewart, R.A., O'Halloran, K.R., 2013. Smart metering: enabler for rapid and effective post meter leakage identification and water loss management. Journal of Cleaner Production 54, 166–176.

Cardell-Oliver, R., 2013. Water use signature patterns for analyzing household consumption using medium resolution meter data. Water Resources Research 49, 8589–8599.

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

Cominola, A., Giuliani, M., Castelletti, A., Abdallah, A., Rosenberg, D., 2016. Developing a stochastic simulation model for the generation of residential water end-use demand time series, in: Proceedings of the 8th International Congress on Environmental Modeling and Software (iEMSs 2016), Toulouse (France).

Cominola, A., Giuliani, M., Piga, D., Castelletti, A., Rizzoli, A., 2017. A Hybrid Signature-based Iterative Disaggregation algorithm for Non-Intrusive Load Monitoring. Applied Energy 185, 331–344.

Cominola, A., Giuliani, M., Piga, D., Castelletti, A., Rizzoli, A.E., 2015. Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. Environmental Modelling & Software 72, 198–214.

DeOreo, W., Mayer, P., Dziegielwski, B., Kiefer, J., 2016. Residential Uses of Water 2016. Water Research Foundation, Denver, CO.

DeOreo, W.B., 2011. Analysis of water use in new single family homes. Salt Lake City Corp. and US EPA, Boulder, CO .

DeOreo, W.B., Heaney, J.P., Mayer, P.W., 1996. Flow trace analysis to assess water use. American Water Works Association. Journal 88, 79.

Evans, R.P., Blotter, J.D., Stephens, A.G., 2004. Flow rate measurements using flow-induced pipe vibration. Journal of fluids engineering 126, 280–285.

Figueiredo, M., Ribeiro, B., de Almeida, A., 2014. Electrical signal source separation via nonnegative tensor factorization using on site measurements in a smart home. Instrumentation and Measurement, IEEE Transactions on 63, 364–373.

Froehlich, J., Larson, E., Saba, E., Campbell, T., Atlas, L., Fogarty, J., Patel, S., 2011. A longitudinal study of pressure sensing to infer real-world water usage events in the home, in: Pervasive Computing. Springer, pp. 50–69.

GetData Graph Digitizer [Computer software], 2017. Available at: http://getdata-graph-digitizer.com/index.php.

Gleick, P.H., Haasz, D., Henges-Jeck, C., Srinivas, V., Wolff, G., Cushing, K.K., Mann, A., 2003. Waste not, want not: The potential for urban water conservation in California. Pacific Institute for Studies in Development, Environment, and Security Oakland, CA.

Hawkins, C., Berthold, A., 2015. Considerations for adopting AMI and AMR: A comprehensive guide for water utilities. Technical Report. Texas A&M Research Extension, College Station, Texas.

Horsburgh, J.S., Leonardo, M.E., Abdallah, A.M., Rosenberg, D.E., 2017. Measuring water use, conservation, and differences by gender using an inexpensive, high frequency metering system. Environmental Modelling & Software 96, 83–94.

House-Peters, L.A., Chang, H., 2011. Urban water demand modeling: Review of concepts, methods, and organizing principles. Water Resources Research 47.

Intelligence, Global Water, 2011. Global water tariffs continue upward trend. Global Water Intelligence 12.

Italian National Institute of Statistics, 2013. Noi italia. 100 statistiche per capire il paese in cui viviamo.

Jorgensen, B., Graymore, M., O'Toole, K., 2009. Household water use behavior: An integrated model. Journal of Environmental Management 91, 227 – 236.

Jorgensen, B.S., Martin, J.F., Pearce, M., Willis, E., 2013. Some difficulties and inconsistencies when using habit strength and reasoned action variables in models of metered household water conservation. Journal of Environmental Management 115, 124 – 135. doi:http://dx.doi.org/10.1016/j.jenvman.2012.11.008.

Kenney, D.S., Goemans, C., Klein, R., Lowrey, J., Reidy, K., 2008. Residential water demand management: lessons from Aurora, Colorado.

Koutiva, I., Makropoulos, C., 2016. Modelling domestic water demand: An agent based approach. Environmental Modelling & Software 79, 35–54.

Kowalski, M., Marshallsay, D., 2003. A system for improved assessment of domestic water use components, International Water Association.

MadgeTech, 2017. Pulse101a data logger. URL: http://www.madgetech.com/pulse101a.html.

Makonin, S., Popowich, F., 2015. Nonintrusive load monitoring (NILM) performance evaluation. Energy Efficiency 8, 809–814.

Makropoulos, C., Rozos, E., 2011. Managing the complete urban water cycle: the urban water optioneering tool, in: SWITCH Conference: The Future of Urban Water: Solutions For Livable and Resilient Cities, Paris (France).

Mayer, P.W., DeOreo, W.B., 1999. Residential end uses of water. American Water Works Association.

Mayer, P.W., DeOreo, W.B., Towler, E., Martien, L., Lewis, D., 2004. Tampa water department residential water conservation study: the impacts of high efficiency plumbing fixture retrofits in single-family homes. A Report Prepared for Tampa Water Department and the United States Environmental Protection Agency .

McKnight, P.E., Najab, J., 2010. Mann-whitney u test. Corsini Encyclopedia of Psychology .

MeterWorks, 2017. Pers. comm.

MetronFarnier, 2017. Innov8-vn register technical reference. URL: http://metronfarnier.com/innov8-registers-remotes/.

Mohassel, R.R., Fung, A., Mohammadi, F., Raahemifar, K., 2014. A survey on advanced metering infrastructure. International Journal of Electrical Power & Energy Systems 63, 473–484.

Mori, M., Tezuka, K., Tezuki, H., 2004. Ultrasonic flow meter, flow measurement method, and computer program. US Patent App. 10/584,318.

Neptune Technology Group, 2017. E-coder®)r900i product sheet. URL: https://www.neptunetg.com/globalassets/products/literature/15-ntg-481-ps-e-coderr900i-10.15.pdf.

Nguyen, K., Stewart, R., Zhang, H., 2013b. An intelligent pattern recognition model to automate the categorisation of residential water end-use events. Environmental Modelling & Software 47, 108–127. doi:10.1016/j.envsoft.2013.05.002.

Nguyen, K.A., Stewart, R.A., Zhang, H., Jones, C., 2015. Intelligent autonomous system for residential water end use classification: Autoflow. Applied Soft Computing 31, 118–131.

Nguyen, K.A., Zhang, H., Stewart, R.A., 2013. Development of an intelligent model to categorise residential water end use events. Journal of Hydro-environment Research 7, 182 – 201.

Oracle, 2009. Smart Metering for Water Utilities. Technical Report.

Piga, D., Cominola, A., Giuliani, M., Castelletti, A., Rizzoli, A.E., 2016. Sparse optimization for automated energy end use disaggregation. IEEE Transactions on Control Systems Technology 24, 1044–1051.

Puust, R., Kapelan, Z., Savic, D., Koppel, T., 2010. A review of methods for leakage management in pipe networks. Urban Water Journal 7, 25–45.

Rahimpour, A., Qi, H., Fugate, D., Kuruganti, T., 2017. Non-intrusive energy disaggregation using non-negative matrix factorization with sum-to-k constraint. IEEE Transactions on Power Systems .

Sønderlund, A.L., Smith, J.R., Hutton, C.J., Kapelan, Z., Savic, D., 2016. Effectiveness of smart meter-based consumption feedback in curbing household water use: Knowns and unknowns. Journal of Water Resources Planning and Management 142, 04016060.

Stewart, R.A., Willis, R., Giurco, D., Panuwatwanich, K., Capati, G., 2010. Web-based knowledge management system: linking smart metering to the future of urban water planning. Australian Planner 47, 66–74. doi:10.1080/07293681003767769.

Suero, F., Mayer, P., Rosenberg, D., 2012. Estimating and verifying united states households' potential to conserve water. Journal of Water Resources Planning and Management 138, 299–306.

Thames Water Utilities Limited, 2017. Metering policy. URL: https://wholesale.thameswater.co.uk/-/media/Site-Content/Wholesale/Wholesale-Uploads/Metering-policy-V42.pdf.

United States, 1992. Energy policy act of 1992 : Public law 102-486-oct. 24, 1992 .

Willis, R.M., Stewart, R.A., Panuwatwanich, K., Jones, S., Kyriakides, A., 2010. Alarming visual display monitors affecting shower end use water and energy conservation in australian residential households. Resources, Conservation and Recycling 54, 1117 – 1127.

Zuras, D., Cowlishaw, M., Aiken, A., Applegate, M., Bailey, D., Bass, S., Bhandarkar, D., Bhat, M., Bindel, D., Boldo, S., et al., 2008. Ieee standard for floating-point arithmetic. IEEE Std 754-2008 , 1–70.

Table 1: Summary of performance metrics for multi-resolution comparative assessment.

Metric ID	Framework component	Metric	Unit
1	End-use disaggregation	Appliance Contribution Accuracy	%
2	End-use disaggregation	Appliance Root-Mean Square Error	-
3	Leakage detection	Average Water Loss	Liters/(household $\times$ year)
4	Peak demand estimation	Peak Estimation Error	%
5	Peak demand estimation	Peak Estimation Time Gap	Minutes
6	Data storage	Data Size	Mbytes/(household $\times$ year)
7	Commercial product(s) available for purchase	Availability	Yes/No

Table 2: Summary of total water use events extracted from a training dataset of 313 households over 3,731 days.

	Standard-efficiency houses	High-efficiency houses
End-use/summary item	Total count of events	
Shower	6,571	688
Toilet	45,167	3,641
Faucet	168,612	10,568
Bathtub	585	65
Clothes washer	3,067	258
Dishwasher	1,111	110
Number of days monitored (measuring water)	3,413	318

Table 3: P-value statistics obtained via Mann-Whitney U testing comparing the distribution of water end-use statistics for STREaM simulated use data at 10-second sampling resolution against the distribution of statistics for observed water use data. Test dataset includes water end-use events for 250 STREaM simulated households over one year (91,250 household-days) and observed data (3,413 household-days). Significance level: 1%. P-value is not reported when the test rejects the null hypothesis of similar distributions.

	Mann-Whitney U test p-value		
Appliance name	Number of usages/day	Consumption event volumes	Consumption event durations
Shower	0.796	0.740	0.526
Toilet	0.499	-	-
Faucet	-	-	-
Bathtub	0.596	0.474	0.685
Clothes washer	0.775	0.368	0.996
Dishwasher	0.569	0.869	0.849

Table 4: Comparison of commercially available systems that can provide sampling resolutions down to about 5 minutes and customized systems deployed in recent end use studies that recorded water use at 1 minute or more frequent intervals.

Measuring frequency	Example Technology	Cost ($)	Availability	At-meter Data Storage	Data trasmission	Reference
1 month	Analogue register	~ 100	Commercial	None	Manual	-
15 min	Neptune E-CODER®)R900i	208	Commercial	96 days	AMR/AMI, Cell network, fixed radio network, optical sensor	Neptune Technology Group (2017); MeterWorks (2017)
1 day, 1 hour, 15 min	Advanced Meter Infrastructure	Site specific	Commercial	Hours to day	Fixed radio network	Hawkins and Berthold (2015)
5 min	MetronFannier Innov8-VN	~ 300	Commercial	Days	Cell network	MetronFarnier (2017)
10 sec	Aquacraft Halls effect sensor; data logger	~ 2400	Custom	2 weeks	Manual	Mayer and DeOreo (1999); Beal and Stewart (2013); DeOreo et al. (2016)
5 sec	Halls effect sensor; RasberryPi	< 200	Custom	Month	Wifi	Horsburgh et al. (2017)

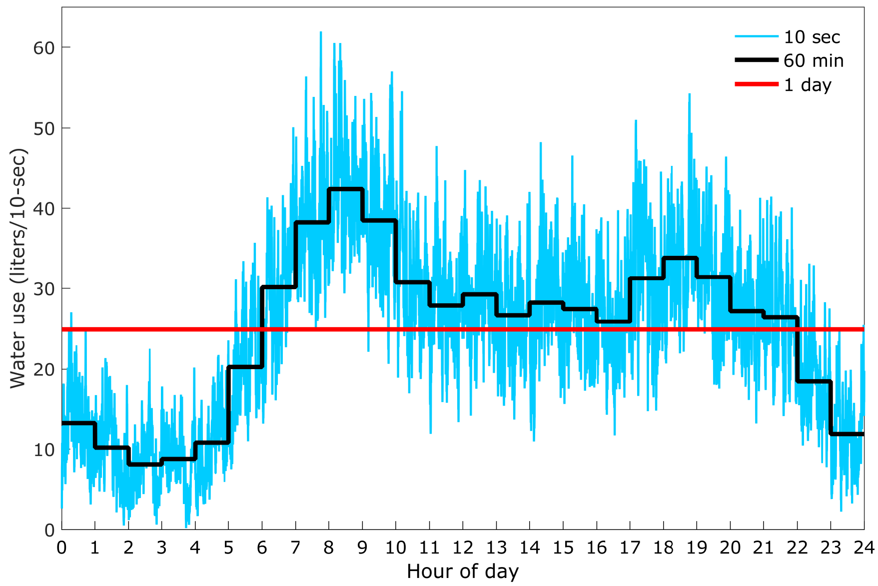

Figure 1: Sample time series of total community water use of 500 households for one day sampled at temporal resolutions of 10 seconds, 60 minutes, and 1 day.

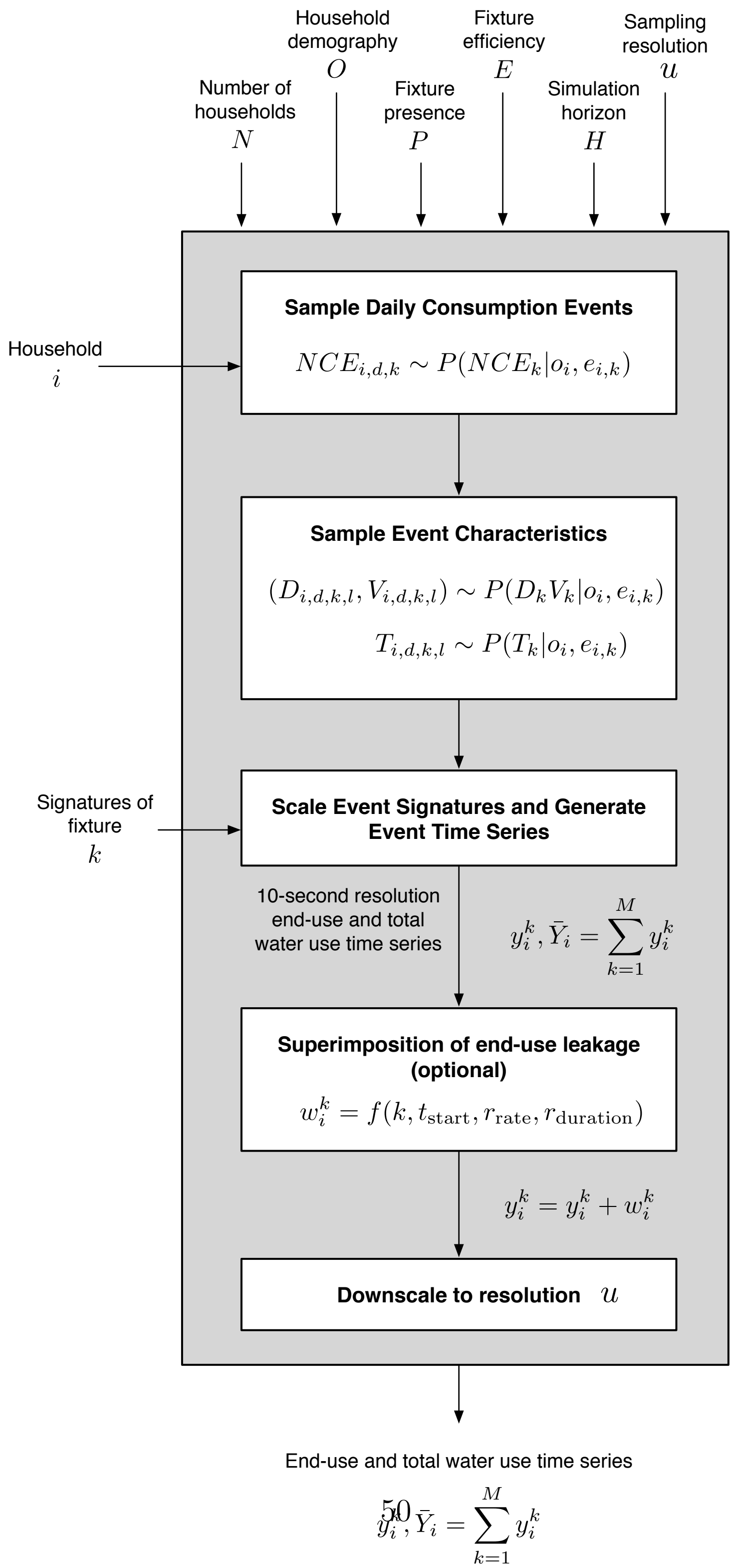

Figure 2: STREaM conceptual model flowchart.

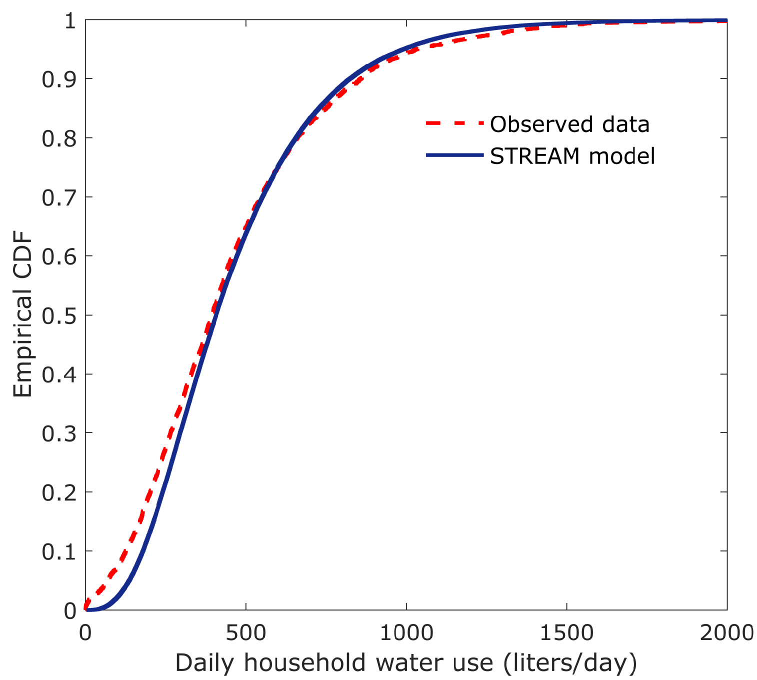

Figure 3: Comparison of Empirical Cumulative Distributions of daily household water use for 250 STREAM simulated households over one year (solid blue line; 91,250 household-days) and observed data (dashed red line; 3,413 household-days).

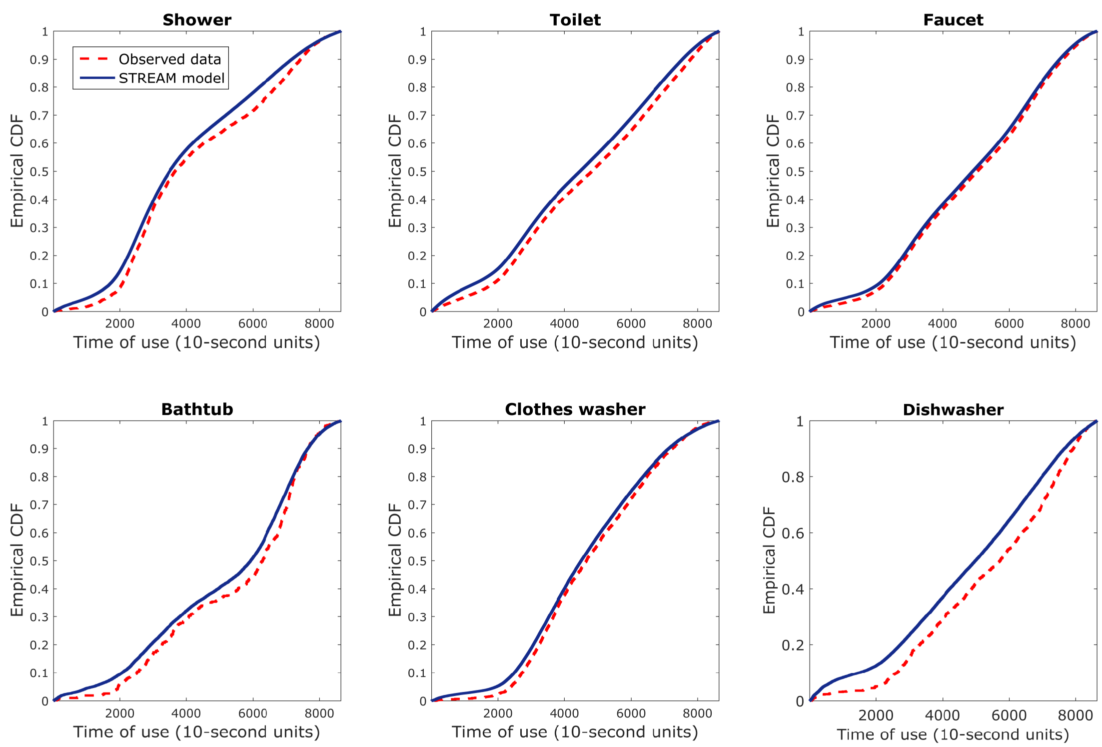

Figure 4: Comparison of Empirical Cumulative Distributions of time of use of water consumption events for six different water end uses across 250 STREAM simulated households over one year (solid blue line; 91,250 household-days) and observed data (dashed red line; 3,413 household-days).

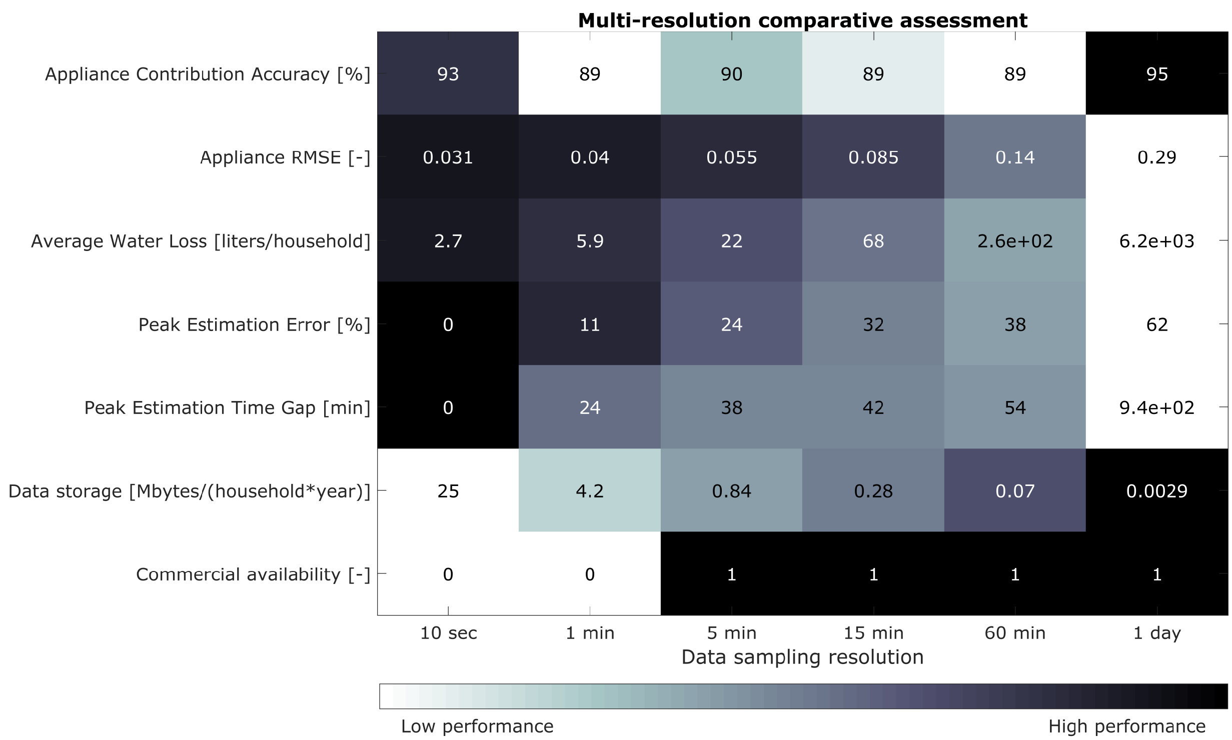

Figure 5: Multi-resolution assessment on STREaM generated water use data. Each row of the matrix refers to a performance metric (see Section 2), each column to a different data sampling resolution (see Section 4.1). Numerical labels in each matrix cell report values for each combination of performance metric and resolution. Color pattern in the figure highlights a tradeoff between the top four performance metrics and the two on the bottom (dark color refer to good performances, and vice versa).

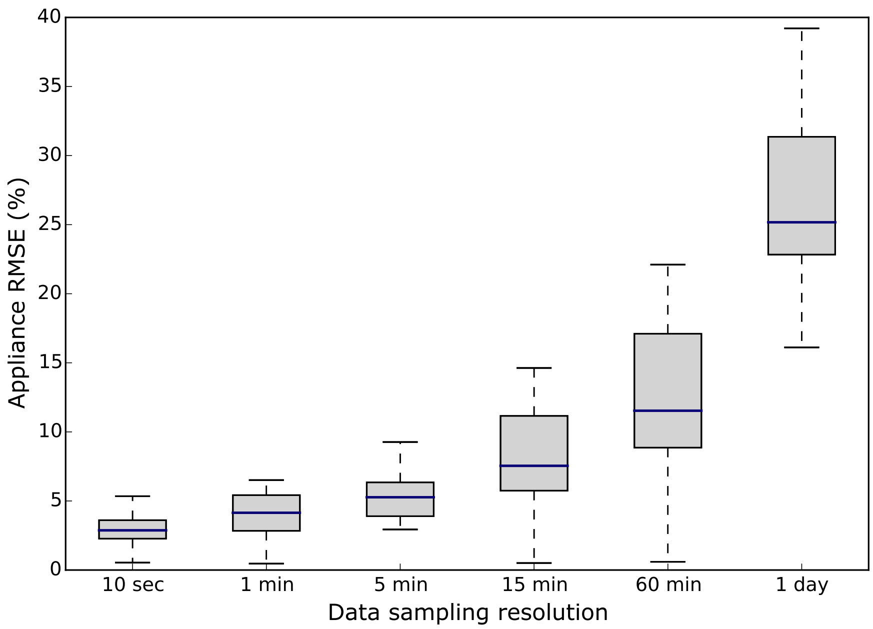

Figure 6: Boxplot showing appliance RMSE when disaggregating end-uses for decreasing data sampling resolution. We used the supervised version of HSID algorithm (Cominola et al., 2017) for end-use disaggregation of water consumption data from a set of 6 generated households with 1 to more-than-5 occupants. HSID was calibrated over 2-months data and Appliance RMSE evaluated over 1-month validation data.

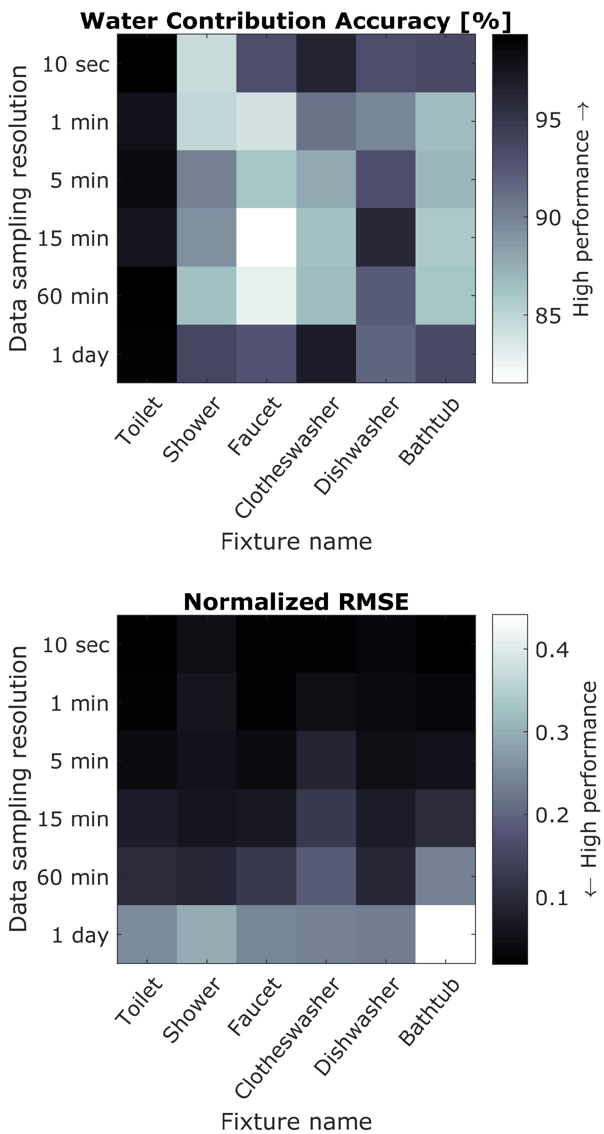

Figure 7: Water Contribution Accuracy (top) and Normalized Root-Mean Square Error (bottom) of end-use disaggregation at different data sampling resolutions. Each row of the matrices refers to a different data sampling resolution, each column to a different appliance. Color bar is proportional to the two performance metrics (dark color refer to good performances, and vice versa). For each appliance and sampling resolution performance metrics are averaged across those obtained from the end-use disaggregation of water consumption data of 6 generated households with diverse demography.

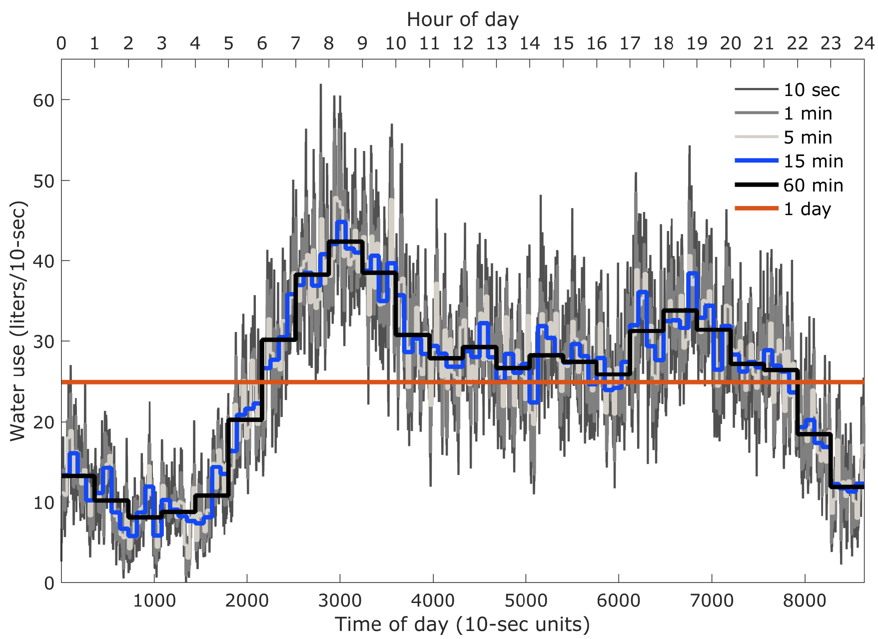

Figure 8: Time series of total community water use (500 households) for one day sampled at temporal resolutions ranging from 10 seconds to a day. Daily pattern is characterized by two peaks, at approximately 8 am and 7pm.