International Environmental Modelling and Software Society (iEMSs)
8th International Congress on Environmental Modelling and Software
Toulouse, France, Sabine Sauvage, José-Miguel Sánchez-Pérez, Andrea Rizzoli (Eds.)
http://www.iemss.org/society/index.php/iemss-2016-proceedings

Developing a stochastic simulation model for the generation of residential water end-use demand time series

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

[a] *Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Piazza L. da Vinci, 32, I-20133 Milano, Italy (andrea.cominola@polimi.it, matteo.giuliani@polimi.it, andrea.castelletti@polimi.it)*

[b] *Department of Civil and Environmental Engineering, Utah State University, Old Main Hill, 4110, UT 84322-4110, Logan, USA (amabdallah@aggiemail.usu.edu, david.rosenberg@usu.edu)*

Abstract: Smart metering technologies allow for gathering high resolution water demand data in the residential sector, opening up new opportunities for the development of models describing water consumers' behaviors. Yet, gathering such accurate water demand data at the end-use level is limited by metering intrusiveness, costs, and privacy issues. In this paper, we contribute a stochastic simulation model for synthetically generating high-resolution time series of water use at the end-use level. Each water end-use fixture in our model is characterized by its signature (i.e., its typical single-use pattern), as well as frequency distributions of its number of uses per day, single use duration, time of use during the day, and contribution to the total household water demand. The model relies on statistical data from a real-world metering campaign across 9 cities in the US. Showcasing our model outputs, we demonstrate the potential usability of this model for characterizing the water end-use demands of different communities, as well as for analyzing the major components of peak demand and performing scenario analysis.

Keywords: end-use water demand; synthetic data; smart metering; water demand management

1 INTRODUCTION

The design of demand-side management strategies for securing reliable water supply and reducing water utilities' costs in urban contexts has been recently acquiring more and more importance as residential water demand contributes a large share of the total urban freshwater demand, with projections foreseeing a substantial increase in the next decades (UNDESA, 2014; Cosgrove and Cosgrove, 2012). The advent of smart metering technologies in the late 1990s (Mayer and DeOreo, 1999) prompted the collection of high spatial (household) and temporal (up to few seconds) resolution water consumption data, opening up new opportunities for the development of models describing water consumers' behaviors and ultimately contributing to the design of customized water demand management strategies (Rizzoli et al., 2014). First, smart-metered data advanced the development of water *end-use characterization* algorithms (e.g., Nguyen et al., 2013), finalized to breaking down total consumption data detected at the household level into the different end-use categories. In addition, these data allow improving the accuracy of residential water demand models, which support both users' segmentation by analysis of observed water consumption patterns and historical trends, and prediction of the expected water demand, under social, climate or management scenarios (for a review see

Cominola et al., 2015a, and references therein).
Yet, assimilating accurate ground-truth water demand data at the end-use level for large household samples is currently limited by intrusiveness issues and costs, which become more constraining when time and resource intensive human activities (e.g., surveys, audits, and water diaries) are put in place (Cordell et al., 2003; Stewart et al., 2010). For this reason, the implementation, calibration, and validation of most end-use water demand models have been, so far, relying on end-use data collected with ad-hoc experimental trials and research projects, involving a limited number of households, usually lower than 100 (e.g., Mayer et al., 2004; Froehlich et al., 2009; Suero et al., 2012). In addition, the data collected within these initiatives are hardly shared or made public available because of privacy and data access issues.
In this paper, we contribute an open-source, stochastic simulation model for synthetically generating high-resolution time series of water use at the end-use level. Each water end-use fixture in our model is characterized by its *signature* (i.e., its typical single-usage pattern (Cardell-Oliver, 2013)), along with the frequency distributions of its number of uses per day, single use duration, time of use during the day, and contribution to the total water demand. Such statistics, aggregated from high-efficiency water demand data gathered in 2005-2006 from over 300 single family homes located in 9 cities across the USA, are conditioned to the number of occupants of the household, the presence of water consuming fixtures, and appliance efficiency levels, which represent the only inputs required to run the model for a single household. At the best of authors knowledge, only few works provide tools for synthetically generating end-use water demand data through stochastic generation and Monte Carlo sampling (e.g, Blokker et al. (2010); Rosenberg et al. (2007); Abdallah and Rosenberg (2014); Escriva-Bou et al. (2015), or the online tool Waterville - http://waterville.hrwallingford.com/waterville/). Our model advances these tools as (i) it is built upon a dataset of high-resolution, smart-metered water demands, consistently related to a set of appliance statistics from the same household sample, (ii) it generates end-use time series of water use preserving the typical usage pattern of each fixture, thus end-use events do not consists of simple demand pulses, and (iii) it allows simulating household water demands under diverse social and technological scenarios, as well as sampling resolutions and sample size, thus allowing the generation of scenarios different form the one represented by a direct analysis of end-use calibration data.
The paper is organized as follows: we first describe the features of our model (section 2); numerical results are reported in section 3; final remarks about the model and its potential applications are presented in the last section.

2 MATERIALS AND METHODS

Our stochastic simulation algorithm operates according to the phases detailed below. In short, given a sample of houses with associated number of occupants, available water consuming fixtures, and fixture efficiency, the model simulates both the end-use time series of water use and their aggregation as total household water demand, with a time resolution (i.e., the simulation time step) up to 1 second, for the defined simulation horizon.

ASSUMPTIONS Considering a single house, in our stochastic model we assume that the time series of water use of a generic j-th water consuming fixture (e.g., toilet, faucet, shower, etc...) during a given d-th day can be characterized by the following elements: (i) number of times the j-th fixture is used during the day (each time it is used will be later referred as *consumption event*); (ii) starting time of use during the day for each consumption event; (iii) duration of each consumption event, in seconds; and (iv) volume of water required by each consumption event, in liters. In addition, we assume that each end-use consumption event can be shaped according to the specific *signature* of its fixture, i.e., the characteristic water use series over time for a single consumption event of a specific end-use, such as the ones represented in Figure 1, properly stretched according to consumption event duration and water volume. A number of stretched signatures equal to the number of daily consumption events of each fixture are then assembled to form the end-use daily time series of water use. Finally, we assume that each fixture can be operated at most only once during each simulation time step.

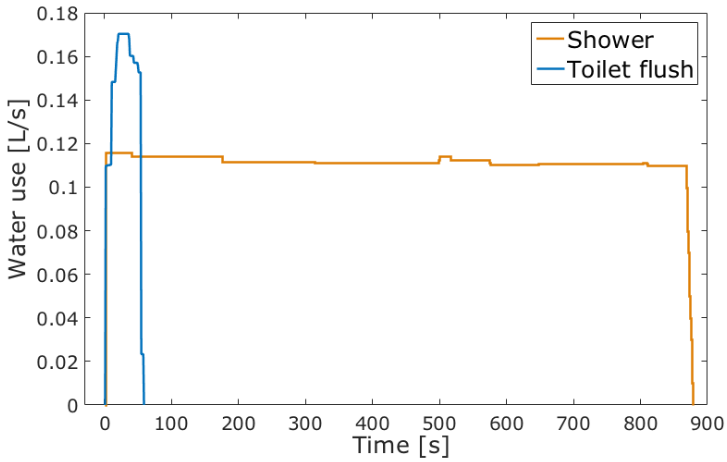

Figure 1. Sample signatures for standard-efficiency toilet and shower fixtures.

MODEL INPUTS Our stochastic model requires the specification of the following inputs:

- sample size N, i.e., number of households for which the model will simulate water demands;
- house size $H = \{h_1, h_2, ..., h_N\}, h_i > 0 \, \forall i \in [1, N]$, i.e., number of occupants for each house in the sample;
- fixture presence $K_i = \{k_1, k_2, ..., k_M\}, k_j \in \{0, 1\} \, \forall j \in [1, M]$, i.e., a binary index specifying the presence (absence) of the j-th fixture in the i-th household;
- fixture efficiency level $E_i = \{e_1, e_2, ..., e_M\}, e_j \in \{0, 1\} \, \forall j \in [1, M]$, i.e., a binary index specifying the efficiency level (*standard* or *high*) of each fixture in each household. Standard and high efficient fixtures are assumed to differ as for the water demand and duration required by their usage, as well as their signature. All other features (e.g., time of use) are assumed to be commonly shared by normal and efficient fixtures.
- length of the simulation horizon T, in days.
- time sampling resolution $r, r > 0$ for the output demand trajectories, in seconds.

END-USE WATER DEMAND TRACES GENERATION Given the input listed in the previous paragraph, let's consider the i-th house, characterized by h_i occupants, fixture presence K_i and fixture efficiencies E_i. Our stochastic model relies on a database containing M signatures (one for each fixture) and a database containing frequency distributions of the number of uses per day, event durations, water volumes, and time of use during the day for each fixture. Each distribution is conditioned to the number of house occupants (h_i) and fixture efficiencies ($e_{i,j}$). From this information, the model generates time series of water use according to the following phases:

1. **Number of consumption events.** The model samples the number of consumption events for each fixture j and each day d of the simulation horizon T as $NCE_{i,d,j} \sim$

$F(NCE_j|h_i, e_{i,j})$, where $F(NCE_j|h_i, e_{i,j})$ is the frequency distribution of the number of usages per day for appliance j, conditioned to the number of house occupants (h_i) and fixture efficiencies ($e_{i,j}$).

2. **Consumption event statistics.** For each water consumption event $l \in [1, NCE_{i,d,j}]$, the model then samples the event duration $D_{i,d,j} \sim F(D_j|h_i, e_{i,j})$, water volume $V_{i,d,j} \sim F(V_j|h_i, e_{i,j})$, and time of use $TOU_{i,d,j} \sim F(TOU_j|h_i, e_{i,j})$. $F(D_j|h_i, e_{i,j})$, $F(V_j|h_i, e_{i,j})$, and $F(TOU_j|h_i, e_{i,j})$ are, respectively, the frequency distribution of consumption event duration, water volume, and daily time of use for the j-th fixtures, conditioned to h_i and $e_{i,j}$. In the current version of the model, the frequency distributions for event duration, water volume, and time of use are assumed to be mutually independent.
3. **Consumption event trace creation.** The time series of water use of each water consumption event is generated by stretching the specific signature of the considered fixture according to the sample values $V_{i,d,j}$ and $D_{i,d,j}$. In order to do so, first randomly chosen points of the signature are iteratively removed/replicated, in order to match the desired event duration $D_{i,d,j}$. Then, the magnitude of each point of the signature is scaled so that the integral over the signature matches the desired water volume $V_{i,d,j}$. After a signature is stretched according to the above procedure, it is positioned over the end-use time series of water use $y_{i,j}$ according to $TOU_{i,d,j}$.

This procedure is iterated from step 1 to step 3 until the simulation is completed, for all the M fixtures and the simulation period T.

MODEL OUTPUT Our stochastic end-use model returns, as output, the end-use time series of water use $y_{i,j}$ for each house i and its fixtures j, as well as the total household water demand $Y_i = \sum_{j=1}^{M} y_{i,j}$.

ORIGINAL DATASET We retrieved the signatures of water consuming fixtures, as well as their associated statistics from DeOreo (2011) and Abdallah and Rosenberg (2014). In the aforementioned study, Aquacraft Inc. metered end-use data from 280 standard houses and 25 high-efficiency houses across 9 U.S. cities, for a period of over 2 weeks between 2005 and 2009 at 10 seconds resolution. Overall, Aquacraft metered 753,076 events during 4,036 days.
It is worth mentioning the following hypotheses we used in the implementation of our stochastic model:

- We only considered indoor water uses, including the following fixtures: toilet, clothes washer, shower, dishwasher, faucet. Toilet and clothes washer signatures and statistics distinguish between standard and high-efficiency, while no distinctions based on efficiency is present for the other appliances.
- We modeled houses with 1, 2, 3, 4, 5, >5 occupants. Houses with more than 5 occupants are grouped together in the last category;
- Water fixtures signatures are directly extracted from Acquacraft's report (DeOreo, 2011) using the Get Data Graph Digitizer software[1] and rescaled to 1 second resolution, according to the procedure described in Gaiardelli (2015). Moreover, we computed a median signature for each of those appliances that allowed the extraction of multiple signatures.
- Water fixture statistics (i.e., frequency distributions of number of uses per day, consumption event durations, water volume and time of use during the day) are derived, for each appliance, number of house occupants, and level of efficiency, from the data gathered in DeOreo (2011).

[1] Software available at http://getdata-graph-digitizer.com/, last visited on 30/03/2016.

3 NUMERICAL RESULTS

In this section, we showcase some outputs of the stochastic end-use simulation model. The direct output of the model consists of high-resolution end-use time series of water use for each house, such as the ones represented in Figure 2. The generated traces show the heterogeneity of the different fixtures' signatures, with the toilet and the faucet characterized by an almost instantaneous pulse, while shower and clothes washers correspond to longer consumption events (2, left panel). Moreover, it is worth noting the differences in peak flow, duration and pattern of the two faucet consumption events, which are strongly dependent on the multipurpose manual usage of this fixture. The hourly distribution of the consumption events during a sample day (2, right panel) shows a typical two-peaks pattern, with the shower and the dishwasher contributing to the morning and evening peak, respectively. The use of toilet and faucet is more equally distributed during the day, while the clothes washer is simulated overnight as this use was observed to occur during off-peak times with positive frequencies.

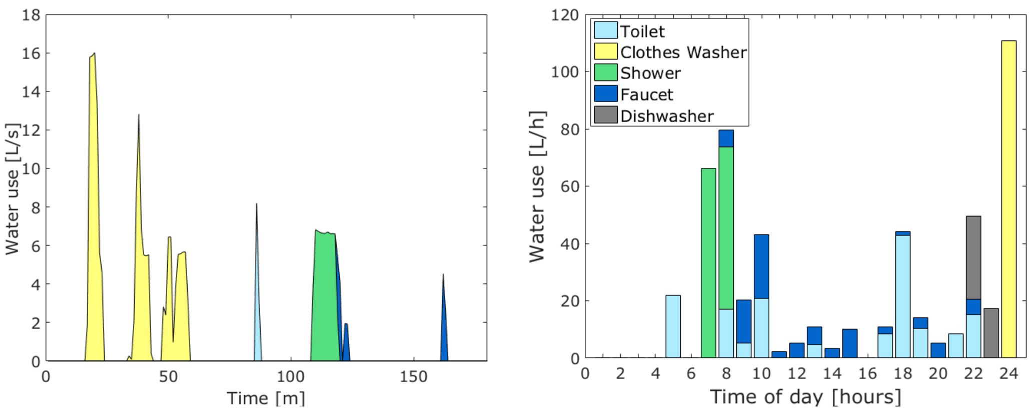

Figure 2. Samples of end-use trajectories output of the stochastic model for a single household over few hours (left panel) and one day (right panel). For the sake of visualization, the data are aggregated at one minute resolution (left panel) and one hour resolution (right panel).

The stochastic model can be exploited to generate water end-use demand time series to exploring how different aspects of water demand vary under diverse social and technological scenarios (defined in terms of sample size, households sizes, and fixture efficiency), together with their implications to water network design and water demand management. For example, characterizing water demand during peak hours is essential to inform system design, highlight opportunities for water conservation, and designing water tariffs. The plot on the left in Figure 3 illustrates the empirical cumulative distribution functions of hourly water demands, stochastically simulated for a sample of 1000 heterogeneous houses over 350 days (11% with 1 occupant, 45% with 2, 15% with 3, 19% with 4, and 10% with 5 or more occupants; 30 % of the houses with efficient devices). It clearly shows how water demand distribution changes if all hourly demands are considered or only demands during peak hours (meaning hours from 6 to 11 a.m. after preliminary analysis), thus providing probabilistic information on the variability of water demand for the defined sample, and its demographic and technological scenario. Moreover, characterizing which end-uses contribute to peak demands (right plot in Figure 3) highlights opportunities for water conservation and peak demand reductions. In the example, the largest contribution to peak demand is given by clothes washer use: this suggests that demand management strategies aimed at promoting overall savings

for this particular end-use (e.g., through more efficient devices), as well as demand shifts from peak to off-peak hours (e.g., through feedbacks and pricing schemes) can be suitable for reducing peak demand.
These numerical results represent valuable information about users' consumption patterns, which can be used for the development of demand models and for the design of water demand management strategies. In particular, the heterogeneity of consumption behaviors given by the end-use statistic frequencies allows running Monte Carlo simulations by repeatedly sampling heterogeneous water consumers' behaviors. The results of these simulations constitute representative users community, which can be used to explore the effect of different scenarios and water demand management strategies by including the response of users to changing scenarios in the model. This represents a strength of adopting a stochastic simulation approach, as it allows generating a number of scenarios that could not be represented through a direct analysis of the end-use calibration data. Moreover, the model outputs can be exploited to develop agent-based models embedding users interactive behaviors and social mechanisms.

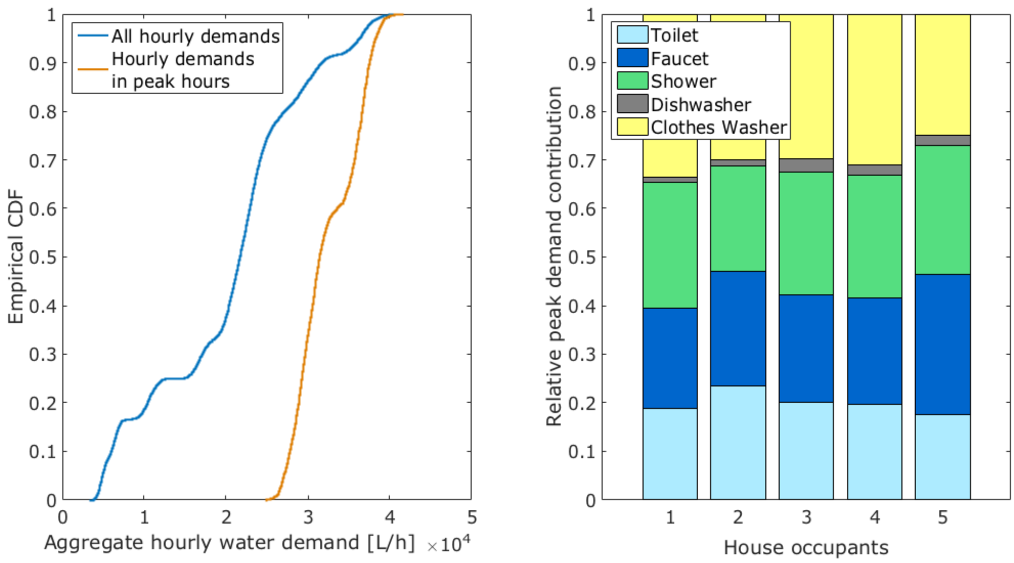

Figure 3. Empirical cumulative distribution function of all hourly demands and peak-hour demands (left plot) and relative contribution of end-uses on peak demand, for different number of house occupants (right plot). Statistics refer to data generated for a sample of 1000 houses, for a period of 350 days.

4 Discussion and conclusions

In this paper, we propose a stochastic simulation model for synthetically generating high-resolution time series of residential water use at the end-use level. Our model is based on the assumption that each water end-use fixture can be characterized by its *signature* (i.e., its typical usage pattern), as well as frequency distributions of its number of uses per day, single use duration, water demand contribution and time of use during the day. In our model, such statistics are conditioned to the number of house occupants, as well as the level of efficiency of some water consuming fixtures. The end-use statistics and signatures of our model were calibrated on real-world data collected by Aquacraft Inc. at the end-use level, with 10 seconds sampling resolution, from over 300 houses across 9 U.S. cities. Simple model applications showcase the potentialities of our model to be

considered for a number of uses related to residential water demand modeling, spanning from community testing of end-use characterization models to modeling water demand via Monte Carlo simulations and scenario analysis.
In addition, the generated data are potentially suitable as numerical benchmarks for training and comparatively testing end-use disaggregation algorithms, as they provide both total household time series of water use, as well as end-use time series, required as ground-truth for supervised disaggregation (for instance, see some recent works in the energy sector (Batra et al., 2014; Piga et al., 2015). Indeed, being synthetically generated, such data overcome problems related to intrusiveness for ground-truth end-use data gathering, metering costs, and data privacy.
The current version of the model offers many opportunities for improvement and further development. Firstly, the inclusion of outdoor end-uses would increase model representativeness in locations such as the arid western U.S. where outdoor water use is significant. Secondly, the heterogeneity of the users' behaviors that can be simulated would improve if the statistics of each end use could be differentiated on the basis time variability (e.g., weekend and weekdays, season), exogenous variables (e.g., temperature, precipitation), psychographic users characteristics (e.g., user age, economic level, water consumption attitudes) or responses to demand management interventions. This would support detailed water users' segmentation (e.g., Cominola et al., 2015b), providing potentially interesting water users' profiles for targeting for water demand management strategies. Such information would allow for better evaluation of water demands during peak or off-peak hours, as well as better understanding of the effect of water demand management strategies on targeted categories of users. In addition, an interesting development would be to calibrate the end-use statistics and signatures of the model and apply it on end-use datasets gathered in different real-world contexts, in order to allow for geo-spatial water demands analysis and comparison of the simulated demand patterns across different countries. Finally, as a further development, the model is going to be offered as a web service with GUI, in order to foster its usability and accessibility.

ACKNOWLEDGMENTS

This work was partially supported by the *SmartH2O: an ICT Platform to leverage on Social Computing for the efficient management of Water Consumption* research project funded by the EU Seventh Framework Programme under grant agreement no. 619172. The authors acknowledge Charles Rougé (University of Manchester) for his valuable review.

REFERENCES

Abdallah, A. and Rosenberg, D. (2014). Heterogeneous residential water and energy linkages and implications for conservation and management. *Journal of Water Resources Planning and Management*, 140(3):288–297.

Batra, N., Kelly, J., Parson, O., Dutta, H., Knottenbelt, W., Rogers, A., Singh, A., and Srivastava, M. (2014). NILMTK: An open source toolkit for non-intrusive load monitoring. *arXiv preprint arXiv:1404.3878*.

Blokker, E., Vreeburg, J., and van Dijk, J. (2010). Simulating residential water demand with a stochastic end-use model. *Journal of Water Resources Planning and Management*, 136(1):19–26.

Cardell-Oliver, R. (2013). Water use signature patterns for analyzing household consumption using medium resolution meter data. *Water Resources Research*, 49(12):8589–8599.

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

Cominola, A., Giuliani, M., Piga, D., Castelletti, A., Rizzoli, A., and Anda, M. (2015b). Modelling residential water consumers' behaviors by feature selection and feature weighting. In *Proceedings of the 36th IAHR World Congress*, The Hague (the Netherlands).

Cordell, D., Robinson, J., and Loh, M. (2003). Collecting residential end use data from primary sources: do's and dont's. *Proceedings of Efficient 2003*.

Cosgrove, C. E. and Cosgrove, W. J. (2012). *The Dynamics of Global Water Futures: Driving Forces 2011-2050*, volume 2 of *The United Nations World Water Development Report*. UNESCO.

DeOreo, W. B. (2011). Analysis of water use in new single family homes. *By Aquacraft. For Salt Lake City Corporation and US EPA*.

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

Froehlich, J. E., Larson, E., Campbell, T., Haggerty, C., Fogarty, J., and Patel, S. N. (2009). Hydrosense: Infrastructure-mediated single-point sensing of whole-home water activity. In *Proceedings of the 11th International Conference on Ubiquitous Computing*, Ubicomp '09, pages 235–244, New York, NY, USA. ACM.

Gaiardelli, E. (2015). Development of a flexible synthetic generator of residential water consumption traces at the end-use level. Master's thesis, Politecnico di Milano, Italy.

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

Mayer, P. W., DeOreo, W. B., Towler, E., Martien, L., and 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*.

Nguyen, K. A., Zhang, H., and Stewart, R. A. (2013). Development of an intelligent model to categorise residential water end use events. *Journal of Hydro-environment Research*, 7(3):182 – 201.

Piga, D., Cominola, A., Giuliani, M., Castelletti, A., and Rizzoli, A. (2015). Sparse optimization for automated energy end use disaggregation. *IEEE Transactions on Control Systems Technology*.

Rizzoli, A. E., Castelletti, A., Cominola, A., Fraternali, P., dos Santos, A. D., Storni, B., Wissmann-Alvese, R., Bertocchi, M., Novak, J., and Micheelg, I. (2014). The smarth2o project and the role of social computing in promoting efficient residential water use: a first analysis. In *Proceedings of the 7th International Congress on Environmental Modelling and Software*.

Rosenberg, D. E., Tarawneh, T., Abdel-Khaleq, R., and Lund, J. R. (2007). Modeling integrated water user decisions in intermittent supply systems. *Water Resources Research*, 43(7):W07425.

Stewart, R. A., Willis, R., Giurco, D., Panuwatwanich, K., and Capati, G. (2010). Web-based knowledge management system: linking smart metering to the future of urban water planning. *Australian Planner*, 47(2):66–74.

Suero, F., Mayer, P., and Rosenberg, D. (2012). Estimating and verifying united states households' potential to conserve water. *Journal of Water Resources Planning and Management*, 138(3):299–306.

UNDESA (2014). *World urbanization prospects: The 2014 Revision. Highlights (ST/ESA/SER.A/352)*. United Nations Department of Economic and Social Affairs Population Division.