Residential Water Conservation: Can linear regression help reduce consumption?

Jesse Dirkman*

03 August 2012

Abstract

With the prospect of climate change and the possibility of unreliable freshwater resources on the horizon, it is necessary to reduce water consumption. Utilities and policy-makers are looking for ways to decrease demand and increase efficiency. In this paper, individual hourly meter readings and daily weather data for the city of Whitewater, Wisconsin, were used to generate multiple linear regression forecasting models. These models can be used to predict periods of high demand and to identify consumption trends. In company with the H_2Oscore.com dashboard, multiple linear regression models can serve to increase user awareness of current consumption patterns and offer solutions to motivate conservation.

1 Background of this Research Project

This paper was written as a part of the Research Experience for Undergraduates in Mathematics at Marquette University, sponsored by the National Science Foundation and the Marquette University Mathematics, Statistics and Computer Science department.

*This work was supported by the National Science Foundation under grant #CNS-1063041.

2 Problem Motivation

Although we have seen the beginning of a green revolution, it seems that the revolution has bypassed the important problem of excess water usage. Unlike gasoline or electricity, for example, there has been no large spike in prices to draw attention to our own water usage. Water is absolutely necessary for our survival, and it has no alternative. Yet at around $0.01 for five gallons, the price of water does not reflect its precious nature nor the worrisome state of our water tables, lakes, and rivers [1].

When it comes to water, we hold the illusion of an unlimited supply. Water is carried from lakes or aquifers to American homes through underground pipes. When consumers turn on the faucet, it is taken for granted that water will flow forth without delay. With a hidden and reliable system such as this, no one takes notice until the flow from the tap stops.

In many places throughout the country, aquifers are getting low and in danger of running dry. The Ogallala aquifer, which extends from South Dakota to Texas, is one of the world's largest underwater water systems, providing irrigation to one-third of the United States' corn crops and drinking water to eight states. If current rates of use continue, the aquifer is predicted to run dry as early as the year 2025 [2]. Energy production and beverage distribution companies are causing significant damage to their surrounding ecosystems. Still, prices remain low, and demand continues to grow [4].

According to the U.S. Geological Survey conducted in 2000, 48% of water usage in the United States is for thermoelectric power generation. Irrigation accounts for 34%, and public supply accounts for 11% of water used. Industrial water usage is 5% of total usage; self-supplied domestic, livestock, aquaculture, and mining combined use an additional 2% [5].

Although the largest percentage of water use is not from residential consumers, their demand still accounts for a significant portion of total water use. Encouraging residential change can occur at the grassroots level without policy changes, a major advantage in comparison to power generation or large-scale agriculture. Once individuals begin to take water conservation more seriously, there will be one less hurdle in instituting restrictions on water usage for larger users. The simplest way to reduce water use is to target individual consumers and encourage them to conserve. Reconsidering the way that water is used outside of the residence (e.g., water rides, golf courses, manicured lawns) can follow from a widespread consumer awakening. Although numerous studies have cited the oncoming perils of water insecurity, many say that the trend can be reversed [4]. By working to understand our own consumption, we can begin the process of rebuilding our nation's freshwater system.

3 H_2Oscore Dashboard

H_2Oscore is a social entrepreneurship venture that was developed out of a Marquette University Water Policy and Environmental Ethics course. H_2Oscore aims to solve the problem of water conservation through consumer awareness initiatives. Currently, H_2Oscore is collaborating with the City of Whitewater, Wisconsin, and the University of Wisconsin at Whitewater to provide city residents with a comprehensive understanding of their water usage through the H_2Oscore dashboard.

The H_2Oscore dashboard is a graphical user interface designed to give residents more information about their water use. The display includes the average number of gallons used by each residence per day and provides a graphical comparison to neighbors' consumption within an eighth of a mile and the rest of the city. Additionally, H_2Oscore users are given a ranking within their neighborhood regarding their water usage. Figure 1 is a sample dashboard. The dashboard includes a water tracker, as depicted in Figure 2, which shows a few days worth of hourly meter readings for the customer.

As household consumption has continued to rise, some utilities have begun to provide in-home displays similar to the H_2Oscore dashboard in an effort to manage demand. Results have

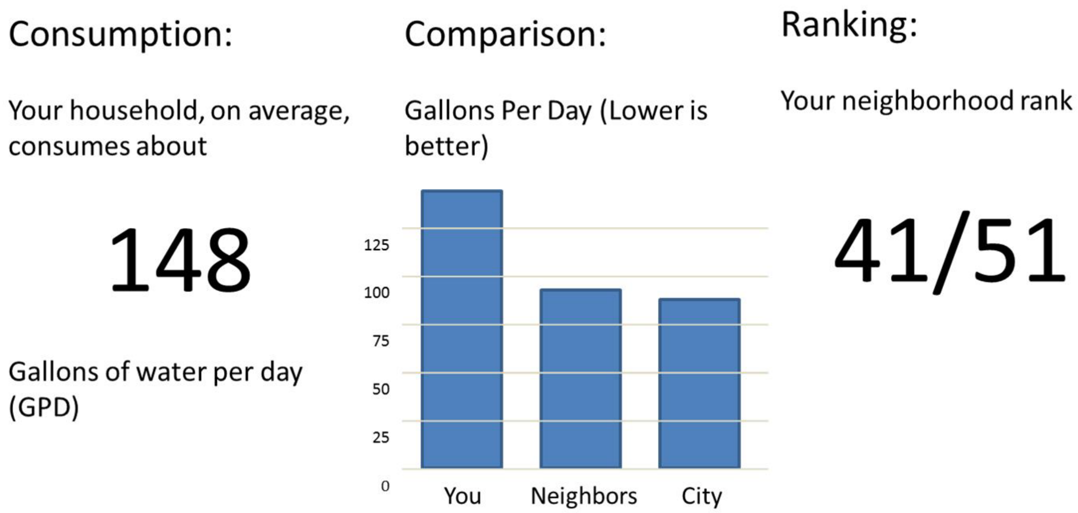

Figure 1: H_2Oscore dashboard

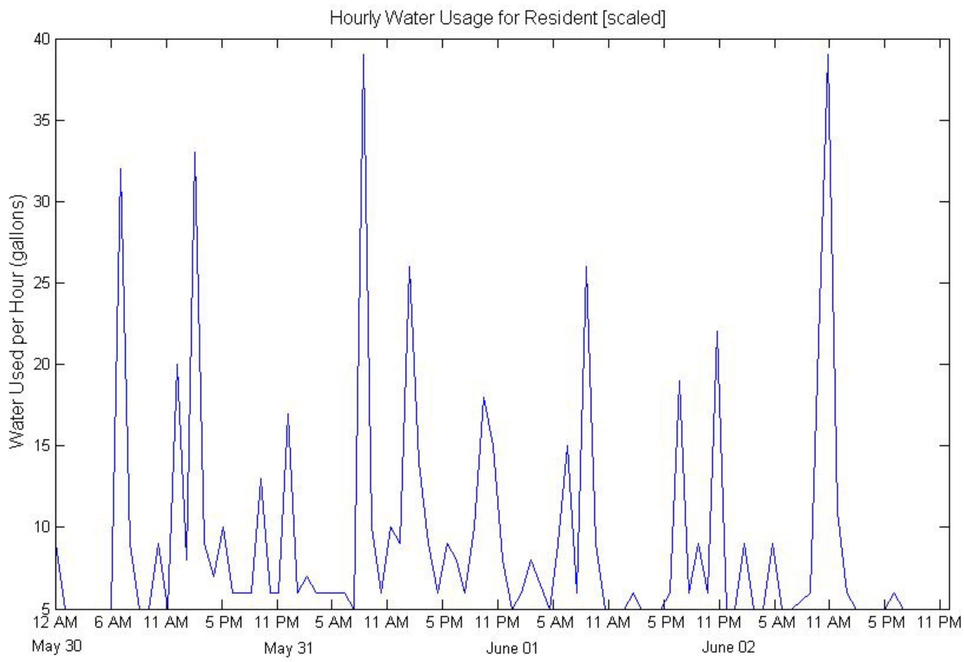

Figure 2: Sample daily water use (one customer)

varied drastically, with success rates ranging from 0-20%. Typically, user interfaces lead to an initial reduction in water demand. However, this impact tends to wane over time [10].

Additional elements are often added to the user interface in an effort to combat this decline. Some utilities have included a "traffic light" feedback system, with various levels of water use falling within the green, yellow, and red zones. Although this method is effective in encouraging consumers to reduce consumption in the red and yellow zones, there is little incentive to decrease water consumption with a green display. These displays may result in load shift, encouraging residents to distribute high-use activities throughout the day. Such a shift benefits the utility, but may not necessarily reduce overall water consumption. Additionally, user interfaces may reinforce high usage for practices that appear non-negotiable to the consumer [10].

As H_2Oscore aims to solve the problem of water conservation, previous studies indicate that a stand-alone user interface will not be enough to significantly and permanently decrease water consumption over a long period of time. This paper aims to explore how linear regression models can be a part of a long-term solution.

4 GasDay^{TM} Involvement

GasDay is a natural gas forecasting research laboratory housed at Marquette University. Our software produces multiple linear regression models, artificial neural networks, and combined models to predict natural gas flow for utilities. Through the research described in this paper, GasDay sought to apply tools and techniques of natural gas forecasting to water forecasting, exploring the possibility of expanding the business model to include water forecasts.

5 Water Forecasting Background

Although there has been a significant amount of research regarding the forecasting and modeling of water demand, old models are not necessarily reliable indicators of future usage.

Fewer residents per household and conserving appliances lead to less use per household than in the past. On the other hand, larger homes, lawns, and incomes are likely to increase water usage. These lifestyle changes have changed the level of water demand, often rendering previous models ineffective. Studies have found geographic location to be a factor in water demand. Thus, forecasting models may not be easily transferrable from one city to another [8].

As with natural gas, there are many variables that have the potential to influence water demand. Kindler and Russell [6] cite weather, precipitation, seasonal variation, population changes, industrial activity, and crop patterns as important variables in predicting water demand. Political, institutional, and legal constraints were considered as factors, as well as the price of water. According to Rockaway et. al. [8], factors include precipitation, temperature zones, and drought indexes. Income and demographic factors were also found to be relevant.

In creating forecasting models, a simple linear regression is one of the most straightforward models available. Previous scholarship used multiple linear regression models with various linear parameters to estimate water demand [3]. The use of artificial neural networks has also been explored with significant success [7], as have autoregressive and semi-parametric models [9].

These models have taken on different forms to fulfill the purposes set out by researchers in developing short and long-term forecasting models. Short-term forecasting models have been used to predict future use when water must travel for multiple days from a reservoir [9], as well as taking advantage of electricity tariffs in pumping and transporting water [3]. Long-term forecasts may be used to predict the longevity of a particular reservoir or to predict the effect of a price change on consumer demand and water supply [6].

6 Data Collection and Method of Analysis

To develop a multiple linear regression model predicting water demand, hourly meter readings for individual consumers in the city of Whitewater, Wisconsin, were obtained from

H_2Oscore. The readings were aggregated to create city-wide hourly and daily meter readings. For the purposes of this paper, a day begins at 12:00A.M. and ends at 11:59P.M.

Weather data for the city of Whitewater was obtained via the Bloomberg terminal. This data includes daily precipitation, mean temperature, high temperature, low temperature, felt air temperature, relative humidity, and wind speed[1].

Limitations: Due to restrictions in data availability, this model was trained solely with meter readings and weather information for the period of May 17, 2012 through June 19, 2012. During this period, the city of Whitewater experienced unusually high temperatures and only one day of precipitation.

In order to determine water demand, the following model was developed:

$$Y_t = \beta_0 + \beta_1(t) + \beta_2(d) + \beta_3(\varphi) + \beta_4(\varphi_{lag}) + \beta_5(T_{mean})$$

$$+\beta_6(T_{adj}) + \beta_7(T_{high}) + \beta_8(T_{low}) + \beta_9(h) + \beta_{10}(\omega) + \epsilon_t, \qquad (1)$$

where Y_t is the water demand for day t. d is a binary variable indicating days of the week Monday-Thursday, φ indicates daily precipitation, and φ_{lag} represents precipitation with a 5-day lag. T_{mean}, T_{adj}, T_{high}, and T_{low} represent mean, felt air, high, and low temperatures, respectively. h signifies relative humidity, and ω represents wind speed. β_0 is the y-intercept, and β_i values represent the coefficients of the independent variables. ϵ represents the error unaccounted for by the model.

The following model was also developed using the same variables as above but including

[1] To retrieve weather data from the Bloomberg L.P., first log in to the Bloomberg terminal. From the menu, search and select the weather tool WETR. Then search the city for which you would like to retrieve data. Click a variable along the left side of the screen and select "graph". In the next window, check the values that you would like to view, and select your desired date range. Click 2) View Table and 1) Download to export selected data to Excel.

Table 1: Preliminary results table

Independent variables	**T-statistic**	**P-value**
intercept (β_0)	-6.722	7.416e-07
date (β_1)	6.723	7.394e-07
relative humidity (β_9)	3.011	0.006
high temperature	2.028	0.054
precipitation(lag)	1.787	0.087
low temperature	1.734	0.096
mean temperature	1.604	0.122
day of week	1.087	0.288
precipitation	0.815	0.423
"feels like" temperature	0.352	0.728
wind speed	0.122	0.904

an autoregressive term:

$$Y_t = \beta_0 + \beta_1(t) + \beta_2(d) + \beta_3(\varphi) + \beta_4(\varphi_{lag}) + \beta_5(T_{mean})$$

$$+\beta_6(T_{adj}) + \beta_7(T_{high}) + \beta_8(T_{low}) + \beta_9(h) + \beta_{10}(\omega) + \beta_{11}(AR_1) + \epsilon_t, \quad (2)$$

where the AR_1 term represents an autoregressive term with lag one[2]. Because these regressions were developed using a small set of data, there were no price or population changes to incorporate into the model. Demographic and income data for the city of Whitewater were not readily available, so these factors were also not included. The multilinear regression was completed using the MATLAB function *regstats*. Various combinations of independent variables were tried to determine the best model. Coefficients are determined to be nonzero when the corresponding P-value ≤ 0.05 .

Table 2: Statistics for 2-parameter model (date and relative humidity)

R-Square	Adj. R-Square	RMSE	MAPE
0.6784	0.6576	33839.57	4.97%

7 Results and Discussion

After running a multilinear regression on Equation 1 and various combinations of independent variables, we determined that the date indicator t and relative humidity h were the only statistically significant variables in this model. Thus, the final model is:

$$\widehat{Y_t} = \beta_0 + \beta_1(t) - \beta_9(h). \tag{3}$$

For the given time period, water use was positively correlated with time (overall positive linear trend) and negatively correlated with relative humidity. No significant relationship was identified between water demand and day of the week, precipitation, temperature or wind speed.

The second model was a multilinear regression based on Equation 2. Through this model, we determined that the date indicator t and the autoregressive term AR_1 were the only statistically significant variables in this model. Thus the final model is:

$$\widehat{Y_t} = \beta_0 + \beta_1(t) + \beta_{11}(AR_1). \tag{4}$$

Again, water use was found to be positively correlated with time. Water use was also found to be positively correlated with the previous day's water use.

The positive linear trend of water usage may reflect a succession of warm days or days without rain. It is the linear trend that other dimensions of the model were unable to account for. Future models with additional data will likely provide insight into the source of this trend.

[2] An autoregressive model uses previous values as parameters in the model. Equation 2 incorporates the previous day's water demand into the forecasting model. Because meter reads were only available for May 17, 2012 through June 19th, 2012, the first forecast this model produces is for May 18, 2012.

Table 3: Preliminary results table- autoregressive model

Independent variables	T-statistic	P-value
intercept	-2.082	0.046
date	2.082	0.046
autoregressive term	4.489	9.806E-05
day of week	1.586	0.128
high temperature	1.393	0.178
relative humidity	-1.196	0.245
low temperature	0.908	0.374
mean temperature	-0.705	0.488
"feels like" temperature	-0.690	0.498
precipitation	-0.492	0.628
precipitation(lag)	-0.453	0.655
wind speed	0.171	0.866

Table 4: Statistics for 2-parameter model (date and autoregressive term)

R-Square	Adj. R-Square	RMSE	MAPE
0.6716	0.6497	34434.46	4.90%

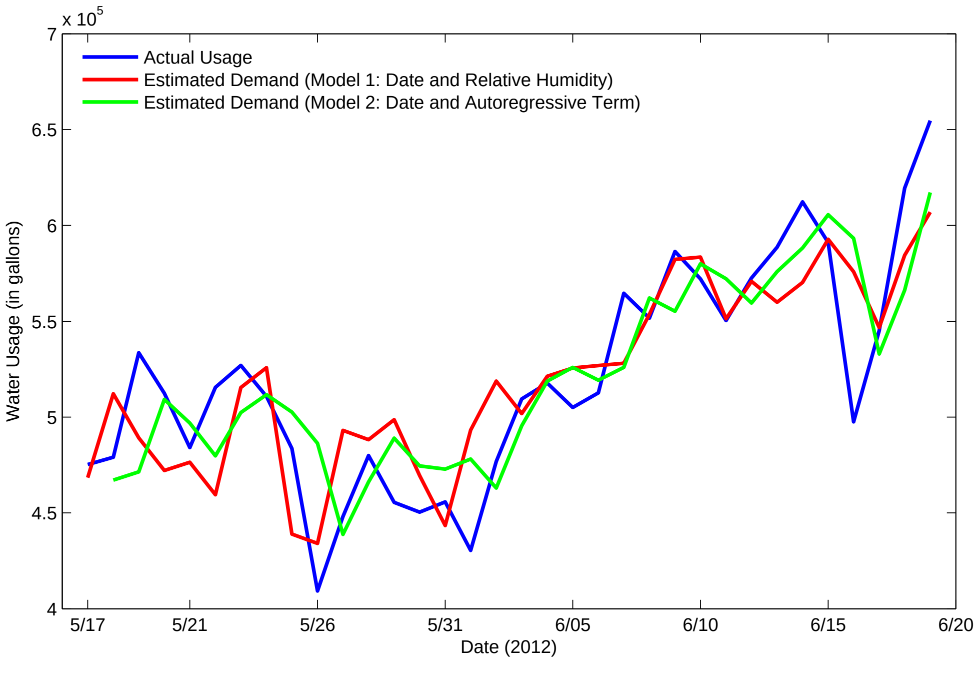

Figure 3: Predicted water demand v. actual demand

Interestingly, relative humidity was no longer a significant predictive factor once the autoregressive term was added. While there is a detectable trend in the data, these results suggest that we do not have enough readings to determine the independent variables most significantly correlated with water demand. Figure 3 depicts the predicted versus actual water demand from May 17, 2012, to June 19, 2012.

Because of the limited data used to train this model, independent variables that were not significant in this model may prove to be essential predictors in future models. Whitewater experienced little temperature variation and only one day of rainfall during the course of this study, obscuring any possible trend related to these variables.

8 The Hourly Profile

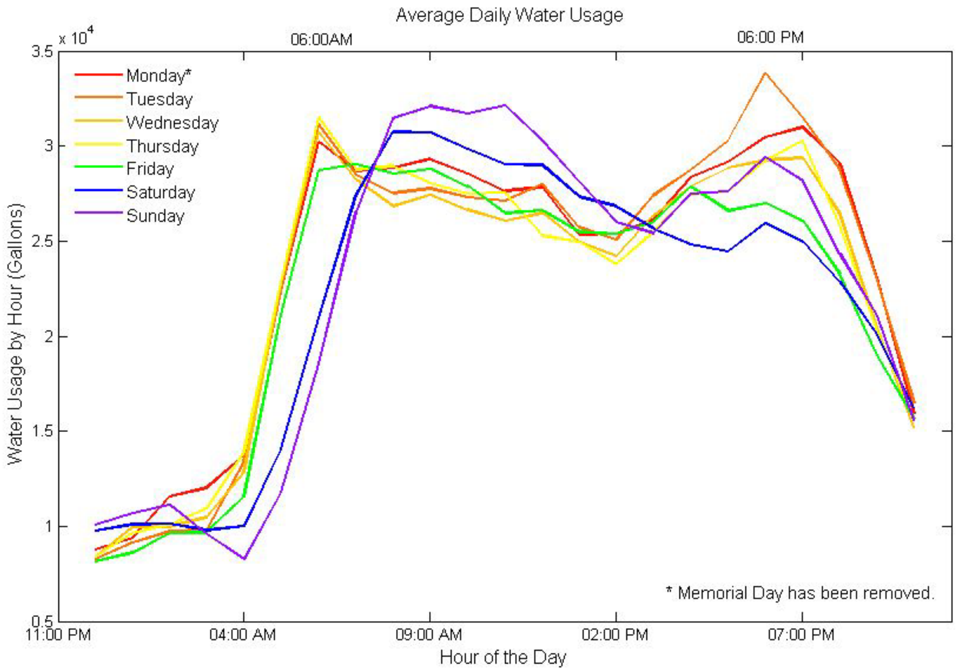

Figure 4: Hourly water usage by day of the week

In addition to multiple linear regression models, insight regarding water usage may also be gained through analyzing hourly usage data. Figure 4 shows water usage in Whitewater broken down by hour [3]. Monday through Thursday, water use increases sharply in the morning and peaks at 6:00 A.M. It declines during the day, then peaks again around 6:00 P.M. and decreases throughout the evening. Saturday and Sunday usage peaks a full two hours later than on the weekdays, and usage remains high throughout the morning. Water usage decreases in the afternoon and increases somewhat in the evening, more so on Sundays. Fridays have the same morning peak as a weekday, yet the evening trend follows more closely with that of a weekend. Water use on Memorial Day closely mirrored usage on Sundays. Given the habits of residential consumers and the structure of the work week, these results reflect expected rates of water use. Findings bear similarity to those discussed by Alvisi et. al. [3].

9 Future Work

As more data becomes available, this research can be extended to identify seasonal and long-term trends and re-evaluate the significance of day of the week and precipitation factors. Once a more refined model has been trained, a separate test data set can be used to analyze the accuracy of the model on untrained data. Future models may include artificial neural networks or combined models including multiple linear regression and autoregressive factors. Hourly models can be trained and aggregated as an alternative method to calculating daily water demand. Many of the tools and techniques that GasDay^{TM} has developed throughout the years for estimating natural gas demand will likely prove invaluable in further exploration of water forecasting.

Data collected after the launch of H_2Oscore can be analyzed to determine the impact of the H_2Oscore dashboard on residential water use. As H_2Oscore continues to add more cities to its dashboard program, predictive models for these areas may be informative and useful. While final predictive models may differ between cities, the methods and tactics used in this paper certainly

[3] Readings were taken from 12:00 A.M. to 11:59 P.M. Readings for one hour represent total water use for that hour. For example, the reading for 12:00A.M. represents the amount of water used between 12:00 A.M. and 1:00 A.M.

may be applied to other cities.

10 Conclusions

The preliminary findings of this paper indicate that there are trends in water usage, and there is a great deal more to explore in terms of creating forecasts. As water resources become less reliable and conservation becomes more and more important, it is essential to look for new and innovative ways to encourage reduced consumption. This paper explored creating multiple linear regression models to forecast water demand. By learning more about user consumption patterns and helping users become aware of their own usage patterns, utilities can suggest more targeted methods to encourage conservation efforts. Linear regression can help to identify the root causes of high water use. Rather than applying blanket solutions to complex problems, governments can then target new policies and solutions to address root causes. Producing forecasts for unusual days can help utilities to determine when to enact watering restrictions should projected demand exceed sustainable supply. Additionally, by identifying consumers with low water usage, analysis can help other users learn from their good practices. In company with the H_2Oscore.com dashboard, multiple linear regression models can help increase user awareness of current consumption patterns and offer solutions to motivate conservation.

11 Acknowledgements

The author would like to thank the National Science Foundation for providing the funding for this research as well as the Marquette University Mathematics, Statistics and Computer Science department for sponsoring the Research Experience for Undergraduates program through which this research was conducted. Additionally, the author would like to thank Dr. George Corliss and the GasDay^{TM} staff for their guidance and assistance, along with Dr. McGee Young and the H_2Oscore staff for providing water usage data for the city of Whitewater. The author would also like to thank the City of Whitewater and the University of Wisconsin at Whitewater, who are collaborating with H2Oscore in the dashboard project.

*References

[1] "Drinking water costs & federal funding," *United States Environmental Protection Agency, Office of Water*, 2004.

[2] "World's largest aquifer going dry," *U.S. Water News Online*, 2006.

[3] S. Alvisi, M. Franchini, and A. Marinelli, "A short-term, pattern-based model for water-demand forecasting," *Journal of Hydroinformatics*, vol. 9, no. 1, pp. 39–50, 2007.

[4] C. Barnett, *Blue Revolution: Unmaking America's Water Crisis.* Beacon Press, 2011.

[5] S. S. Hutson, N. L. Barber, J. F. Kenny, K. S. Linsey, D. S. Lumia, and M. A. Maupin, "Estimated use of water in the United States in 2000," *U.S. Geological Survey*, vol. USGS Circular 1268, 2004.

[6] J. Kindler and C. Russell, Eds., *Modeling Water Demands.* Academic Press, 1984.

[7] F. K. Odan and L. F. R. Reis, "Hybrid water demand forecasting model associating artificial neural network with Fourier series," *Journal of Water Resources Planning and Management*, vol. May/June, pp. 245–256, 2012.

[8] T. D. Rockaway, P. A. Coomes, J. Rivard, and B. Kornstein, "Residential water use trends in North America," *AWWA*, vol. 103, no. 2, pp. 76–89, 2011.

[9] J. A. Smith, "A model of daily municipal water use for short-term forecasting," *Water Resources Research*, vol. 24, no. 2, pp. 201–206, 1988.

[10] Y. Strengers, "Negotiating everyday life: The role of energy and water consumption feedback," *Journal of Consumer Culture*, vol. 3, pp. 319–338, 2011.