# VALUE OF HIGH FREQUENCY WATER USE DATA FOR EVALUATING PEAK WATER USE, LEAKS, AND BREAKS ON THE CUSTOMER SIDE OF THE METER

By

JOHN PAUL MCCARY

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2017

To my wife, Lorrie, and my sons, Johnathan and Jamason

ACKNOWLEDGMENTS

Actively enjoying life with family, working full time, and pursuing a doctorate are all time-consuming tasks individually, but they are especially difficult trying to manage concurrently. To do so requires the support of many people. I would like to thank Dr. James Heaney for advising me and allowing me to pursue a Ph.D. while working full time. The balance between research and practice made the work that much more gratifying as the applications of the research could be seen in engineering practice. I would also like to thank Dr. Kirk Hatfield, Dr. Ben Koopman, and Dr. John Sansalone for serving on my committee and providing guidance to improve the presentation of the research.

I would like to thank my colleagues in the Urban Water Systems research group. Ken Friedman, Scott Knight, Miguel Morales, and Randy Switt provided the on-campus support that I needed in order to be a successful off-campus student. In addition, Barbi Jackson went above and beyond to provide the support that was essential for me to adhere to the administrative requirements that I couldn't meet in person.

I would like to thank the staff at Hillsborough County Public Utilities Department for providing the opportunity to link research and utility practice. The vision of the utility to be a leader for other utilities to follow continues to provide for an exceptional working environment.

Finally, and most importantly, I would like to thank my wife, Lorrie, and my sons, Johnathan and Jamason. To my wife, Lorrie: I will forever be grateful for your support as you encouraged me to complete my Ph.D. while putting your own goals aside and handling a large portion of the parental responsibilities. I will strive to repay you, not because you expect it, but because you deserve it. To my sons, Johnathan and Jamason: I have been amazed and inspired to see you grow and challenge yourselves like few people do. You have taught me many things about life, and I hope I will inspire you the way you inspire me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

VALUE OF HIGH FREQUENCY WATER USE DATA FOR EVALUATING PEAK WATER USE, LEAKS, AND BREAKS ON THE CUSTOMER SIDE OF THE METER

By

John Paul McCary

December 2017

Chair: James P. Heaney
Major: Environmental Engineering Sciences

The United States is facing an aging infrastructure crisis, and water distribution systems are one aspect of this crisis. As options to repair and replace these systems are explored, accurate demand evaluations are necessary to improve distribution system design and operation. Technology is allowing the acquisition of large datasets and advanced analytics that can improve these demand evaluations beyond traditional analyses. With the evolution of smart systems, new applications are being developed that allow for real-time analytics to improve decision making.

While distribution system design and operation can be improved at the macro scale, homeowners and businesses face challenges at the micro scale. Water damage caused by fixture leaks and plumbing breaks accounts for extensive property damage. Smart meter systems can allow for customers to directly benefit in ways that are beyond the services traditionally offered by utilities. The deployment of smart meter systems by utilities can provide for a dual purpose system, one that can simultaneously have the ability to provide real-time feedback to utilities for demand evaluations and to notify customers of potentially damaging leak events.

This dissertation addresses key research gaps and can be used as a framework for applying high-frequency water use evaluations to both utilities and customers. Automatic meter reading meter registers are used with short-range wireless communication that allow for ease of

data collection by driving by and downloading the data from the meter registers. An evaluation of high-frequency data in the multi-family residential sector compares measured peak demands at different temporal aggregations to probabilistic peak demands assuming a normal distribution. The methods can be applied to improve design standards for the probabilistic sizing of infrastructure. In the single family residential sector, analyses are performed to quantify leaks by evaluating outlier events in terms of intensity, duration, frequency, and volume. The evaluation considers an important question: are smart meter systems worth the costs for customers? If smart meters can detect unwanted events, can the savings associated with this detection result in positive net benefits for the customers? The results show that there can be positive benefits to customers by using smart meter systems.

# CHAPTER 1
# INTRODUCTION

## Needed Investments in Water Distribution Systems

According to the U.S. Environmental Protection Agency's (USEPA's) fifth national assessment of public water system infrastructure (USEPA 2013), the nation's drinking water utilities need $384.2 billion in infrastructure investments over the 20-year period from 2011-2030. The USEPA assessment shows that over 64% of this total need is for transmission and distribution systems as shown in Figure 1-1.

Even with this seemingly high estimate, EPA recognizes that significant needs are excluded in this assessment, such as raw water dams and reservoirs, projects related primarily to population growth, and water system operation and maintenance costs. Other estimates that include these needs are significantly higher. The American Society of Civil Engineers (ASCE) 2017 report card for Drinking Water (www.infrastructurereportcard.org/) grades the nation's drinking water infrastructure as a "D", and references a $1 trillion need over the next 25 years to restore systems reaching the end of their useful lives and expanding them to serve a growing population (American Water Works Association (AWWA 2012a)). Nearly half of this total is needed in the southern United States. As shown in Table 1-1, the AWWA (2012a) report indicates that more than half of this need is for water main replacement while less than half is for population growth. The service lives of these piping systems range from 50 to over 100 years so it is essential to carefully evaluate these long-term needs for both the initial capital investment and the ongoing operational costs. Table 1-2 shows a range of expected service lives, with the range resulting from corrosive soil conditions and/or installation methods.

It is clear that major investments are needed in new infrastructure. Customer water use, peak demand in particular, is the primary decision variable for sizing this infrastructure. After

the initial capital investment has been used to install the designed infrastructure, operational costs are driven by the utilization of the infrastructure which is needed to meet demands that are estimated across the system. Operating decisions are made to meet demand measured using a top down approach at the system level because direct measurements of performance at the customer end of the network aren't available. High service pump stations supply the needed hydraulic energy required to deliver water to customers across all demand conditions. The Electric Power Research Institute (EPRI 2002) reported that pumping accounts for 80% of the energy used at most water utilities, and energy in the water and wastewater sector accounts for 3% - 4% of national energy consumption. Friedman et al. (2010) reported that the typical system expends approximately 90% of its energy use in distribution system pumping.

## Beyond the Utility: A Focus on Customer Savings

While the numbers above reference the importance of distribution system investment and the impact customer demand has on that investment, major investments and repairs also occur on the customer side of the meter. Research from the Insurance Services Office (ISO) indicates that just under 1.79 houses per 100 houses per year have claims associated with water damage, and with total property damage claims occurring at 7.15 houses per 100 houses per year, the result is water damage claims accounting for 25% of total property damage claims (ISO data reported by www.valuepenguin.com/average-cost-of-homeowners-insurance). The capabilities exist to notify customers of these events when caused by leaks or pipe breaks within the home, if proper detection algorithms can be developed and incorporated into event detection/notification systems. Outliers to expected events, starting from bottom-up evaluations of customer water use, need to be identified to incorporate these capabilities into smart meter systems.

## Bottom-Up Approach to Customer Water Use

The evaluation of potable water demand for water supply planning and resource sustainability is a key area of research in urban environmental engineering systems (House-Peters and Chang 2011). Recent research has focused on bottom-up approaches that explain how water is being used by the customer at the end use level. By developing techniques to estimate the end uses of water, analyses can be performed to quantify the impacts of replacing customer end use fixtures and modifying customer end use habits. These impacts can be aggregated from the individual customer level up to larger geographic areas, including at the utility, state, and national levels using common classification systems (Morales and Heaney 2014). The primary focus in these analyses has been on the resource and conservation, where the time scales of significance for the supply quantity are on the order of months to years. The best available data sets for actual measurements are typically monthly customer billing data which can be linked with property appraiser parcel data, demographic and housing data, and business classification data to develop indicators for water use estimates that can be applied to different strategies in demand management and forecasting. These estimates can be used to simulate water use predictions from the bottom up, both in terms of the spatial scale from the parcel level and the temporal scale from the monthly level. Bottom-up water use evaluations have been conducted on both indoor and outdoor use within the single-family residential (SFR), multi-family residential (MFR), and the commercial, industrial, and institutional (CII) sectors by our University of Florida research team at the Conserve Florida Water Clearinghouse (www.conservefloridawater.org). More specifically, Friedman et al. (2010b, 2014a) evaluated the SFR and MFR sectors, Morales et al. (2013a) evaluated indoor use in multiple sectors, and Morales et al. (2011, 2014, 2015) evaluated the CII sector. These evaluations have also been used for simulations quantifying conservation potential in the SFR indoor sector (Friedman

2011), the SFR outdoor sector (Friedman 2013a, 2014b; Knight 2015a, 2015b), across multiple indoor (Morales 2013b), and outdoor sectors (Friedman 2014c). What is lacking in these studies is the ability to precisely quantify certain end uses, which would require high-frequency data that would show a signature event-level water use pattern for specific end uses.

Interestingly, two bottom-up approaches appeared in the mid-1990s that characterized high-frequency water use events into their intensity, duration, and frequency (IDF). In general, the approaches differed in that one group focused on developing probabilistic approaches for demand simulators incorporated into water distribution system modeling and the other group focused on quantifying individual fixture-level impacts on the overall water budget. However, neither approach directly focused on unexpected customer events, like pipe breaks on the customer side of the meter. The probabilistic approaches culminated in a seminal study by Buchberger et al. (2003) with data collection on 21 homes near Cincinnati, Ohio at 1-second intervals for 252 days. Of the published high-frequency databases, this study had the best description of the distribution of flows into the home and the aggregation of use up to the neighborhood scale.

With a focus on fixture-level, end-use quantification, the initial success of DeOreo et al. (1996) led to a nationwide water use study that used the same technique for collecting water use data for 100 homes in each of 12 different cities for 4 weeks at 10-second intervals (Mayer et al. 1999). This seminal study has provided a solid foundation for estimating end uses at the individual fixture level. An update, titled *Residential End Uses of Water, Version 2* (DeOreo et al. 2016), included both the original and additional data sets with more varied study site locations, hot water usage data, more detailed landscape analysis, and additional water rate analysis.

## Smart Meters and Real-Time Analytics

Early justifications of automatic meter reading (AMR) systems were based on savings in meter-reading costs due to the reduced time to manually read each meter. As the technology has progressed, utilities have transitioned from AMR to smart meter systems where additional meter-reading savings have been realized (Thiemann et al., 2011; Daigle and Jackson, 2013). Additional savings to customers have been realized through smart meter applications that can alert customers to potential leakage and overall water use quantities (Cardell-Oliver, 2013; Davies et al., 2014; Daigle and Jackson, 2013). In these applications, the smart meters are working for both the customer and the utility.

Real-time analytics is the use of, or the capacity to use, all available enterprise data and resources when they are needed. It consists of dynamic analysis and reporting, based on data entered into a system less than one minute before the actual time of use. Real-time analytics is also known as real-time data analytics, real-time data integration, and real-time intelligence (http://searchcrm.techtarget.com/definition/real-time-analytics). The 2014 3rd edition of the AWWA M22 Manual of Water Supply Practices, *Sizing Water Service Lines and Meters,* includes new methods that incorporate the anticipated increasing ability to manage water demand using advanced sensing and database management systems to develop more efficient smart systems (AWWA 2014). A key policy question is whether utilities should invest in costly smart meter systems.

My experimental design uses available AMR technology to generate one minute databases, wherein datasets are aggregated and evaluated ranging from one minute up to one hour for use in simulating a smart meter system. This allows a prototype system to be evaluated without the major cost of purchasing one. The purpose of this research focuses on the cost and water savings to customers based on real-time event detection at the home, and the cost savings

to utilities based on improved understanding of aggregate demands for system design. With smart meter systems, the needs of both can be met while providing the right level of aggregation for the targeted end uses, either the customer or the utility. The following chapters address key research gaps needed to evaluate the benefits of smart meter systems.

Chapter 2 evaluates high-frequency AMR data in the MFR sector. It compares measured peak demand at different temporal aggregations to probabilistic peak demands assuming a normal distribution around the average use. In addition, it evaluates the overall distribution of the data for comparison with meter accuracy curves. The methods in this paper can be applied to improve design standards for master meters in the multi-family residential sector.

Chapter 3 presents the results of a prototype high-frequency water use evaluation using one-minute data collected for three single family homes in Hillsborough County, Florida over a period of one year. AMR meter registers are used with short-range wireless communication that allow for ease of data collection by driving by and downloading the data from the meter registers. This analysis quantifies leaks by looking at outlier events in terms of intensity, duration, frequency, and volume. These homes have separate indoor and outdoor meters, so the analysis can compare outliers across both indoor events and irrigation. The events are summarized over continuous durations in order to analyze the most significant leak/break events.

Chapter 4 applies the approach developed in Chapter 3 to a district meter area (DMA) evaluation to quantify leaks and potential pipe breaks on the customer side of the meter. Previous research has described benefits to utilities without rigorous cost analysis. This chapter looks at the savings potential to customers, assuming the utility passes the net cost of the meter installation on to the customer. This chapter presents an evaluation of high-frequency data to see if smart meters can detect unwanted events, and if so, can the savings associated with this

detection result in positive net benefits for the customer?  Previous research didn't focus on detecting unwanted customer-level events, at least not within short time scales (e.g. minutes).

Chapter 5 concludes with a summary of the evaluations presented in this dissertation, the resulting conclusions that can be made from these evaluations, and the future work that can build upon these evaluations to advance the state-of-the-art in the water utility industry.

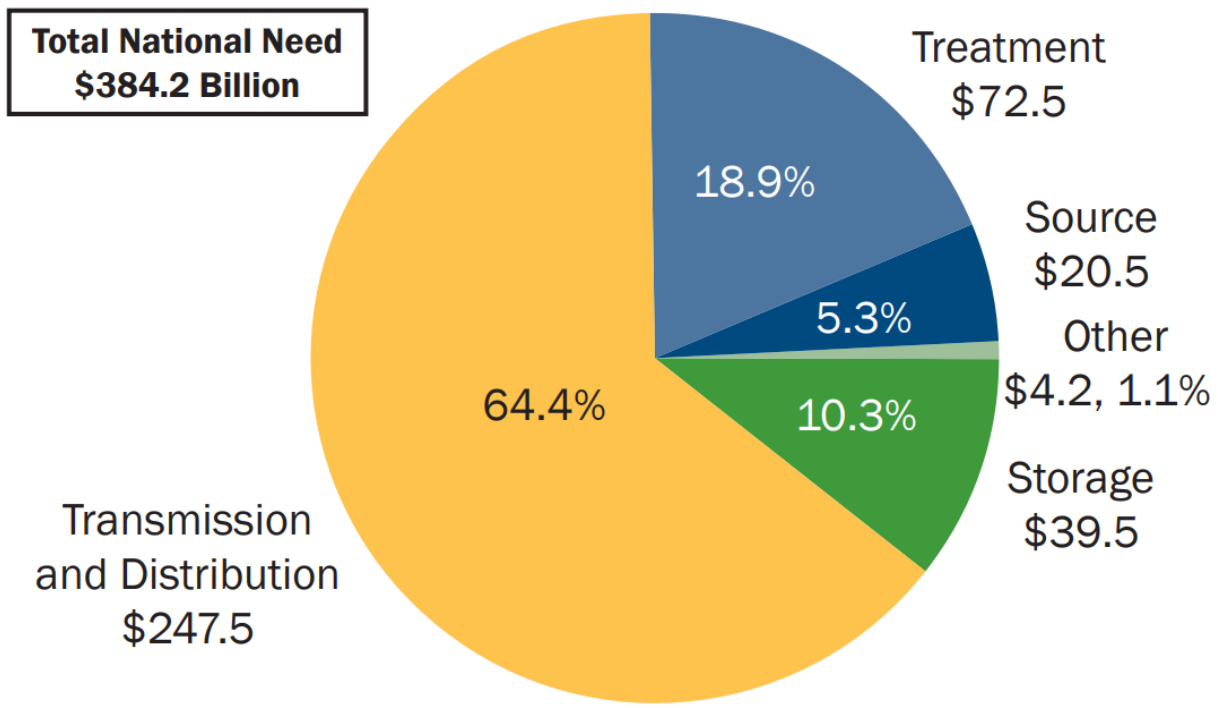


Figure 1-1. Total 20-year (2011-2030) need by project type in billions of January 2011 dollars (USEPA 2013)

Table 1-1. Regional investment needs in water mains from 2011 through 2035 (AWWA 2012a)

| Region | Replacement Need ($M) | Growth Need ($M) | Total ($M) |
|---|---|---|---|
| Northeast | $92,218 | $16,525 | $108,744 |
| Midwest | $146,997 | $25,222 | $172,219 |
| South | $204,357 | $302,782 | $507,139 |
| West | $82,866 | $153,756 | $236,622 |
| Total | $526,438 | $498,285 | $1,024,724 |

Table 1-2. Estimated service lives in years of distribution mains for various regions of the United States (AWWA 2012a)

| Region and Size | Cast Iron | Cast Iron with Cement Lining | Ductile Iron | Asbestos Cement | PVC | Steel | Concrete and PCCP |
|---|---|---|---|---|---|---|---|
| Northeast Large | 130 | 100 to 120 | 50 to 110 | 80 | 100 | 100 | 100 |
| Midwest Large | 125 | 85 to 120 | 50 to 110 | 85 to 100 | 55 | 80 | 105 |
| South Large | 110 | 100 | 55 to 105 | 80 to 100 | 55 | 70 | 105 |
| West Large | 115 | 75 to 100 | 60 to 110 | 75 to 105 | 70 | 95 | 75 |
| Northeast Medium & Small | 115 | 100 to 120 | 55 to 110 | 85 to 100 | 100 | 100 | 100 |
| Midwest Medium & Small | 125 | 85 to 120 | 50 to 110 | 70 | 55 | 80 | 105 |
| South Medium & Small | 105 | 100 | 55 to 105 | 80 to 100 | 55 | 70 | 105 |
| West Medium & Small | 105 | 75 to 100 | 60 to 110 | 75 to 105 | 70 | 95 | 75 |
| Northeast Very Small | 115 | 100 to 120 | 60 to 120 | 85 to 100 | 100 | 100 | 100 |
| Midwest Very Small | 135 | 85 to 120 | 60 to 110 | 75 to 80 | 55 | 80 | 105 |
| South Very Small | 130 | 100 to 110 | 55 to 105 | 80 to 100 | 55 | 70 | 105 |
| West Very Small | 130 | 75 to 100 | 60 to 110 | 65 to 105 | 70 | 95 | 75 |

# CHAPTER 2
# STATISTICAL ANALYSIS OF AUTOMATIC METER READING IN THE MULTI-FAMILY RESIDENTIAL SECTOR

## Scope and Overview

In August 2013, an automatic meter reading (AMR) data collection and analysis case study began for Hillsborough County Public Utilities Department (HCPUD) in the Tampa, Florida area. The entire study group consisted of one large single-family residential (SFR) neighborhood, two multi-family residential (MFR) complexes, one commercial "big-box" retail store, and one hospital. This analysis focuses on the two MFR complexes, and data collection for these two study areas began in September 2013 with data downloaded through February 2015.

The reason the analysis on the MFR complexes was selected for this study is because of the return on data investment: One meter indicated the combined water use habits of a large number of individuals as opposed to looking at single-family residences. In addition, limited research has been done on high-frequency water use in the MFR sector, as opposed to several well-documented studies that have been completed on the SFR sector as described in Chapters 3 and 4 (DeOreo et al., 1996; Buchberger and Wells, 1996; Mayer et al., 1999; Blokker et al., 2010; Buchberger et al., 2003).

Outside of these studies in the SFR sector, Blokker et al. (2011) presented measurements for an office building, a hotel, and a nursing home. The office building was monitored for 14 days with data measured at 1-minute intervals. The hotel and nursing home were monitored for 30 days with data measured at 5-minute intervals. The data were presented for weekdays with a cumulative distribution of all flow measurements for each facility. Also, the average of all data for each time interval of the weekday for each facility was presented to show an average weekday water use pattern for each facility. Similar to the approaches in the residential sector,

non-residential data collection efforts have focused on indoor water use but with fewer data collection efforts. Dziegielewski et al. (2000) collected data for 25 commercial and institutional (CI) establishments, five each for the categories of schools, hotels/motels, office buildings, restaurants, and food stores. Sub-meters were installed at three sites to better measure individual end uses downstream of the master meter. The data were recorded at 10-second intervals for approximately five days. Peak flows were summarized, and a limited time series of a few hours was presented for two of the CI establishments and two of the sub-metered areas downstream of the master meters.

From an indoor use perspective, other studies have assumed that MFR water use is similar to SFR water use on a per unit basis. Instead of making such an assumption, the current research provides measured data to support analysis on a larger spatial scale at a high temporal frequency. One key application for the MFR data presented in this chapter, as well as the SFR data presented in Chapters 3 and 4, is meter sizing based on customer end use flow rates (Buchberger et al. 2012; Blokker et al. 2012; AWWA 2014). Beyond meter sizing, the high-frequency and peak demand evaluations can be aggregated up to larger spatial and temporal scales that affect larger distribution system design and operation issues.

The ability to start with high-frequency data and aggregate up in both temporal and spatial scale is important because these data aren't typically available. The design of water systems uses high and low values of water demand to size these systems, e.g., peak hourly water demand. In order to provide this information, a duration needs to be specified, e.g., peak hourly demand for a specified year. The analysis is limited by the time steps for the data. Typical customer demand data are based on monthly measurements. For those typical cases, the analyst can report statistics for monthly or larger measuring periods. However, it is necessary to

extrapolate to estimate the statistics for shorter periods of interest, e.g., hourly values. These shorter time step data are seldom available for individual urban water customers. Thus, a major high-frequency data collection effort was conducted for about 19 months to be able evaluate actual data for comparison with extrapolations.

**Study Area 1**

Shown in Figure 2-1, this MFR complex has 440 MFR units on a parcel classified by the Department of Revenue (DOR) Code 0310 (Multi-Family Residential > 9 Units, Class A). There are 22 residential buildings, resulting in an average of 20 units per building. According to the American Community Survey (ACS) data, the rolling 5-year average of persons-per-household (pph) for the Census Tract that encompasses this study area is 2.03. Assuming that the 2.03 pph is an appropriate average for the 440 units, the resulting population is 893 residents.

The MFR complex has one 8-inch master-meter with an AMR data logger with recording capability in 10-gallon increments. The data storage was limited to 16,000 data points, which required downloading every 11 days in order to avoid gaps in the data. Over the period of record from September 2013 to March 2015, 744,785 one minute data points have been collected. The average flow during the period of record is 36.1 gallons per minute (gpm) with a standard deviation of 19.7 gpm, resulting in a coefficient of variation of 0.51. The calculated gallons per capita per day (gpcd) is 58.

**Study Area 2**

Shown in Figure 2-2, this MFR complex has 257 multi-family residential units on a parcel classified by the DOR Code 0621 (Retirement Independent Living Facility, Class B). There are 10 residential buildings, resulting in an average of 25.7 units per building. According to the ACS data, the rolling 5-year average of pph for the Census Tract that encompasses this study area is 1.74. Assuming that the 1.74 pph is an appropriate average for the 257 units, the

resulting population is 447 residents. The MFR complex has one 8-inch master-meter with an AMR data logger with the same recording capability as Study Area 1. Over the period of record from September 2013 to March 2015, 700,628 one minute data points have been collected. The average flow during the period of record is 19.6 gpm with a standard deviation 10.5 gpm, resulting in a coefficient of variation of 0.53. The calculated gpcd is 63.

The values of 58 and 63 gpcd reported for Study Areas 1 and 2, respectively, are typical values for indoor water use in the MFR sector (Friedman et al., 2010). These values are also consistent with the range of 50 to 65 gpcd reported for previous studies in the SFR sector (Mayer et al., 1999; Buchberger et al., 2003). This is important to note for future studies comparing the MFR sector data to aggregated SFR sector data.

## Water Use for Study Areas

Water use data were initially available from monthly meter reads used for billing purposes. These are presented for historical perspective on water use prior to the AMR study period. However, the installation of new meters with AMR data loggers allowed for 1-minute water use data to be evaluated at higher frequencies and up to the aggregated, more commonly collected monthly billing data.

### Monthly and Daily Averages

Figure 2-3 shows the monthly average water use for both study areas obtained from billing data starting in October 2010 and reported through December 2014. Prior to the AMR data collection starting in September 2013, the meters were changed because of questionable readings. These questionable readings can be seen in Figure 2-3 with wide variations in reported water use prior to the meters being replaced. For both meters, the data has been more consistent once replaced with a meter and an AMR data logger.

Figure 2-4 shows the daily average water use for both Study Area 1 and Study Area 2. Each point on the graph is calculated by averaging the flow for each minute of the day, i.e. the average of 1,440 data points. Study Area 1 doesn't show any noticeable seasonal variation in flow, meaning there is little or no irrigation relative to the quantity of indoor water use. Study Area 2 indicates that there is some seasonality, with the rolling 30-day average increasing from mid-spring through end of summer.

**Demand Patterns**

Figures 2-5 and 2-6 show the average values for both study areas reflected in a weekly time series. Each point on the graph represents all data available for that time of day and day of week averaged together. For one year of data, each 1-minute value on the graph represents the mean of each of the individual 52 weekly 1-minute data points for that minute and day of the week. When aggregated up to the 1-hour time step, each 1-hour value on the graph represents the mean of 3,120 data points (52 weeks multiplied by 60 minutes) for that hour and day of the week. This level of aggregation shows the time-averaged smoothing when transitioning from 1-minute to 1-hour time steps. However, as noted previously, the averaging across the entire dataset doesn't account for any seasonality throughout the year that would be required to compare changes in seasonal patterns.

Of note is that Study Area 1 is indicative of a younger demographic, with early morning and evening peaks as the residents prepare for and return from work or school. This is also evident by the similar pattern for Monday through Friday; however, there are noticeably different patterns for Saturday and Sunday. Study Area 2 is indicative of an older, retired demographic, with peaks occurring later in the morning and use slowly declining over the rest of the day. What is also evident is that the pattern for each day of the week, whether weekday or weekend, shows a similar pattern.

The key element to take from the pattern comparison is that the two study areas have significantly different, repetitive demand patterns. However, the flow distribution analysis discussed in the following sections can be applied regardless of knowing the actual time-varying demand patterns.

## Comparisons of Measured Data with Normal Distribution

An evaluation of the one minute datasets for both study areas using Minitab 17 Statistical Software (2010) indicated that of the more common probability distributions, the normal (aka Gaussian) distribution had the best fit. This was based on distributions using the actual flow values, resulting in the means and standard deviations as previously reported. Conceptually, this makes sense because of the Central Limit Theorem, which basically states that when you combine many random variables each having independent distributions, the combined distribution approaches a normal distribution. The data were aggregated and distributions generated at different time steps with the mean values preserved and the resulting changes in standard deviations and coefficients of variation as shown in Table 2-1. Rather than focusing on various distribution fitting tests and confidence intervals to evaluate the fit of the entire dataset, a simple comparison was performed to only evaluate the peak predictions from the normal distribution compared to the actual dataset values. Prior to performing this analysis, the flow distributions are presented to visualize how well the normal distribution approximation matches the measured data.

From this point forward, any application of the normal distribution is used to distribute predicted values around the actual mean flow with an assumed standard deviation equal to one-half the mean flow, i.e., a coefficient of variation equal to 0.5. This coefficient of variation value is based on the actual calculated values from the one-minute dataset, as previously reported. Equation 2-1 shows the notation for a random variable "$X$" that is normally distributed, with "$\mu$"

representing the mean and "$\sigma$" representing the standard deviation. Equation 2-2 shows the modified notation used for the distributions discussed in the following sections, with the standard deviation assumed be one-half the mean.

$$X \sim N(\mu, \sigma^2) \tag{2-1}$$

$$X \sim N(\mu, [0.5\mu]^2) \tag{2-2}$$

The analysis discussed in the following sections compares the actual flow rates to the assumed flow rates estimated from distributing high-frequency flow values around the mean flow value. This was done assuming that only the mean flow values were available as would be the case from collecting a single meter read during traditional meter reading applications. The distributions were then generated with an assumed standard deviation since the actual standard deviation couldn't be calculated from the single meter read data point. For presentation purposes, only Study Area 1 is shown graphically, although the flow distributions are similar for Study Area 2 with a distribution around a lower mean flow value.

**Flow Distributions**

Figure 2-7 shows the distribution of one minute flow values for the entire period of record (total of 744,785 data points) for Study Area 1, which has a mean flow value of 36.1 gpm. For display purposes, the x-axis is limited to a flow rate of 100 gpm. The actual peak flow rate of 1,200 gpm occurred during only one minute during the total period of record, and only 28 data points exceeded a flow rate of 130 gpm. These high flow rates occurred during short durations on two separate days, so this is likely a result of onsite fire hydrant testing. Outside of these two periods, the peak flow was 130 gpm, but this flow rate occurred so infrequently that it was invisible for graphing purposes. During the period of record, flow was recorded for 98.3% of the minutes with the remaining 1.7% of the minutes resulting in zero flow.

Because these particular meter registers record the data in discrete 10-gallon increments, the data columns in Figure 2-7 are displaying the actual data reported by the data logger and is not the result of binning the database. The reported value for each 1-minute interval carries the remainder of the value forward from the previous time step if it didn't result in a discrete 10-gallon increment. The following example illustrates this concept: Assume that for three consecutive minutes, the actual flow values are 1 gallon, 21 gallons, and 8 gallons, respectively. The data logger would report the flow values as 0 gallons, 20 gallons, and 10 gallons, respectively. In this manner, the total flow over the three minutes is conserved although the reported values vary slightly during the actual time of use. Because of the way the remainders are carried forward, the maximum error for any one value is +/-10 gallons; however, the maximum cumulative error over any period of record is -10 gallons.

Figure 2-8 shows both the probability and cumulative distributions using the measured data and the normal distribution approximation. Since the actual data is based on discrete points, and the normal distribution is continuous, the points used for plotting the normally distributed probability distribution used +/-5 gallons around the discrete 10-gallon increment. As an example, the data point used for graphing the probability at 10 gallons used the difference between the cumulative probability at 15 gallons and 5 gallons. This affects the display of the results only; it doesn't have any impact on the normal distribution calculations.

**High-Frequency Peak Predictions**

As previously noted, the primary goal of the normal distribution approximation was to be able to test the ability of using traditionally collected billing data to predict high-frequency peak flows. If successful, this would allow for a method of predicting high-frequency flow values from only single average values measured over longer durations. In order to perform this test, a simple question was asked: What flow would result for a given time-period statistic, e.g.,

the peak hour flow for a given week, assuming the probability of occurrence is consistent with the actual percentage of time that the period of interest occurs? Equation 2-3 through Equation 2-6 show the time period percentage, represented by "$\rho$", calculated for the four time period statistics that will be used for the comparisons.

1 Minute: $$\rho = 1\ minute \times \frac{1\ hour}{60\ minutes} \times \frac{1\ day}{24\ hours} \times \frac{1\ week}{7\ days} = 0.0001 = 0.01\% \quad (2\text{-}3)$$

5 Minutes: $$\rho = 5\ minutes \times \frac{1\ hour}{60\ minutes} \times \frac{1\ day}{24\ hours} \times \frac{1\ week}{7\ days} = 0.0005 = 0.05\% \quad (2\text{-}4)$$

15 Minutes: $$\rho = 15\ minutes \times \frac{1\ hour}{60\ minutes} \times \frac{1\ day}{24\ hours} \times \frac{1\ week}{7\ days} = 0.0015 = 0.15\% \quad (2\text{-}5)$$

1 Hour: $$\rho = 1\ hour \times \frac{1\ day}{24\ hours} \times \frac{1\ week}{7\ days} = 0.006 = 0.6\% \quad (2\text{-}6)$$

The question was tested for both study areas for 76 weeks, with each week tested independently. For each week, a normally distributed cumulative distribution was generated using the actual mean flow and an assumed standard deviation equal to one-half the mean flow as described previously. After the distribution was generated, the minimum and peak flows were calculated and compared to the measured values at each level of aggregation. As an example, the minimum and peak 1-minute flow values during the week were assumed to occur over exactly one minute, which would equate to a frequency of 0.01% of time during the week as calculated in Equation 2-3. Using the cumulative distributions that were generated, the minimum 1-minute value for each week was selected from the cumulative distribution whose flow value corresponded to 0.01%, and the peak 1-minute flow value was selected from the corresponding value at 99.99%. For a random variable "$Z$" that is normally distributed as indicated in Equation 2-2, the minimum and peak flow values are determined based on the following equations wherein "$\rho$" was calculated in Equation 2-3 through Equation 2-6 and "$x$" is the value being solved for.

Minimum: $Probability\ (Z \leq x) = \rho$ (2-7)

Peak: $Probability\ (Z \leq x) = 100\% - \rho$ (2-8)

Referring to Figure 2-8, the expected peak flow values are not visually evident because of the "flattened" curve above the 99% cumulative probability. However, what is visible from the overall graph is that the normal distribution would predict minimum flows of zero for all four levels of aggregation when truncating the distribution at a minimum of zero flow. In a true normal distribution, the probability of any single value occurring is zero. However, by truncating the distribution at zero flow, the probability for the occurrence of zero flow is calculated by summing the cumulative probability of all values less than or equal to zero. While Figure 2-8 is representative of the entire dataset, this is consistent with the individual weekly distributions as well. Therefore, Table 2-2 doesn't summarize the minimum values, but it is important to note that the actual data recorded a zero value every week for the 1- and 5-minute levels of aggregation for both study areas. At the 15-minute and 1-hour levels of aggregation, the actual data showed that there were weeks with minimum flow values of zero but on average there was flow. Table 2-2 shows the weekly summary of all 76 weeks with peak flows at 1-minute, 5-minute, 15-minute, and 1-hour levels of aggregation. The "% Difference" values in the table reflect the summary of all 76 weeks, not the percent difference between the measured and predicted values already summarized in the table. As an example, the maximum value of 21% reported under the "Peak 1-Minute" column for Study Area 1 indicates that the maximum difference for any of the 76 weeks results in a measured peak flow that is 21% greater than the predicted peak flow.

**Comparison of Measured Data with Meter Accuracy**

Another application of the flow distribution data is for estimating meter accuracy. One area of concern for meter accuracy has been the use of compound meters considering the

transition between the low- and high-flow meter registers. In order to test this concern, the collected data were used and compared against meter accuracy curves. The collected data were assumed to be 100% correct, and these data were applied to the meter accuracy curves published for the twenty-three meters currently approved for use by Hillsborough County Public Utilities Department at the sizes of 4-, 6-, and 8-inch. For each flow value recorded for the two study areas, the meter accuracy error for each of the twenty-three meters was individually applied and the cumulative error for each meter type was calculated.

Figure 2-9 shows the measured probability distribution and the meter accuracy error curves for three meters of interest for Study Area 1. The three meters of interest are: the actual 8-inch meter used at the study area (the black line), the meter that resulted in the highest cumulative negative error (the red line), and the meter that resulted in the highest cumulative positive error (the green line). In this case, both the meters with highest negative and positive cumulative errors are compound meters. As can be seen in Figure 2-9, both meters underestimate the lower flow rates up through the transition to the high-flow meter, and after the transition, they slightly overestimate the higher flows. The actual 8-inch meter used resulted in a -0.2% error, and the meters with the highest negative and positive cumulative errors resulted in -2.3% and +0.4%, respectively. While not graphed, Study Area 2 had similar results with the actual 8-inch meter resulting in 0% error, and the meters with the highest negative and positive cumulative errors resulting in -1.8% and +0.6%, respectively.

## Synopsis

The high-frequency water use data collected from the AMR data loggers provide excellent insight into the demand patterns and overall flow distributions for two MFR complexes representing a combined population estimated at 1,340 residents. An analysis of the 1.5 million data points between the two study areas indicates that the normal distribution with a standard

deviation of one-half the mean flow produces an excellent approximation to the actual data. This conclusion is subjective, as it is up to the individual depending on application to determine how close of an approximation is needed. It is unlikely that additional data collection efforts would result in a quantitative improvement in the analysis for either the total distribution or the peak flow estimates. However, future research will involve evaluating how much data collection is necessary to accurately forecast demand patterns and account for seasonal variations.

The AMR data also provided an excellent dataset for evaluating meter accuracy. While there weren't significant cumulative meter accuracy errors, in an application where the water use would occur more at one extreme or much more frequently at the transition period, the errors would be more significant. For a total of 46 comparisons, consisting of each of the two study areas being tested against the 23 approved meters, the accuracy ranged from 97.7% to 100.6%.


Figure 2-1. Aerial view of Study Area 1

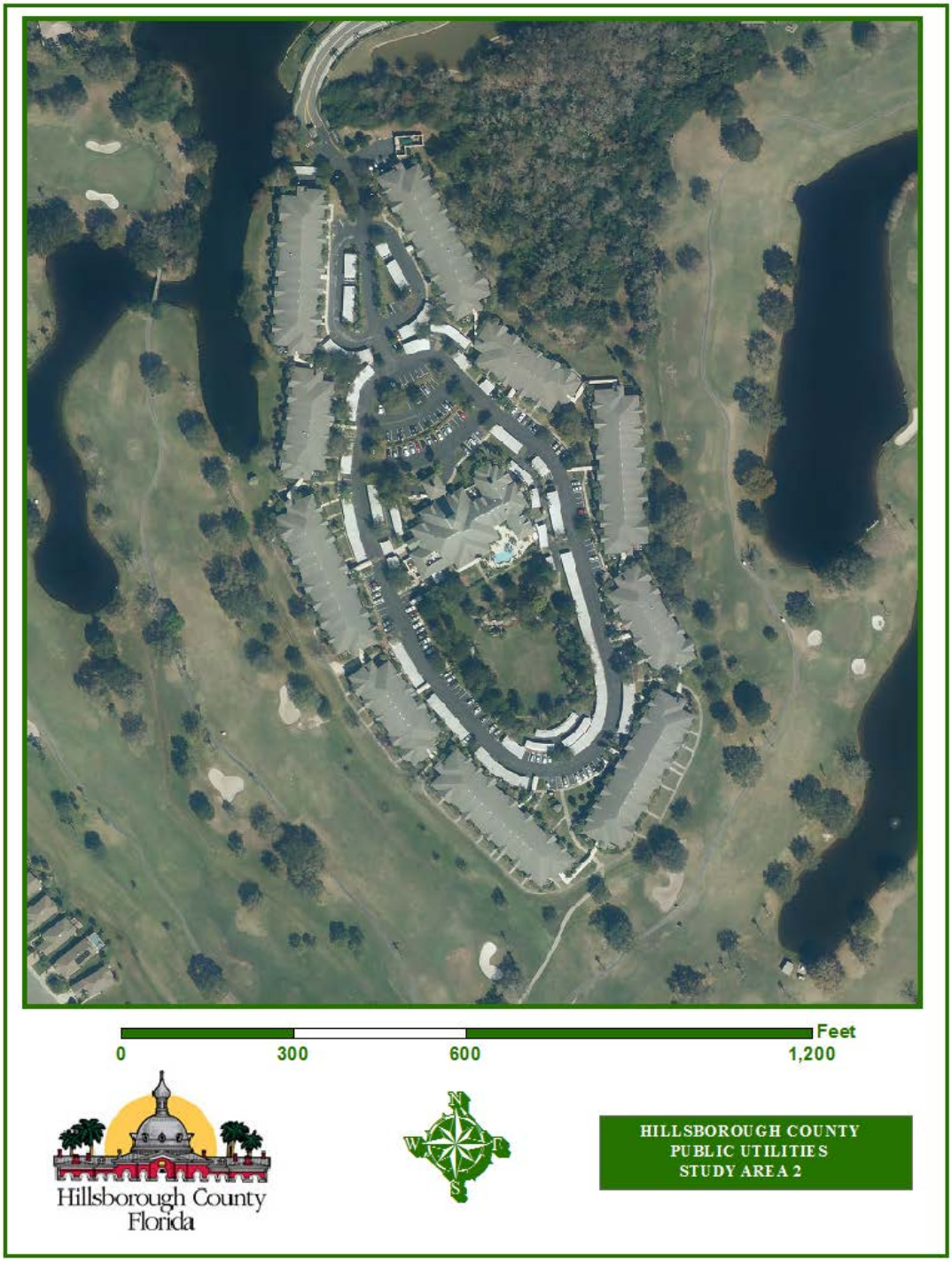


Figure 2-2. Aerial view of Study Area 2

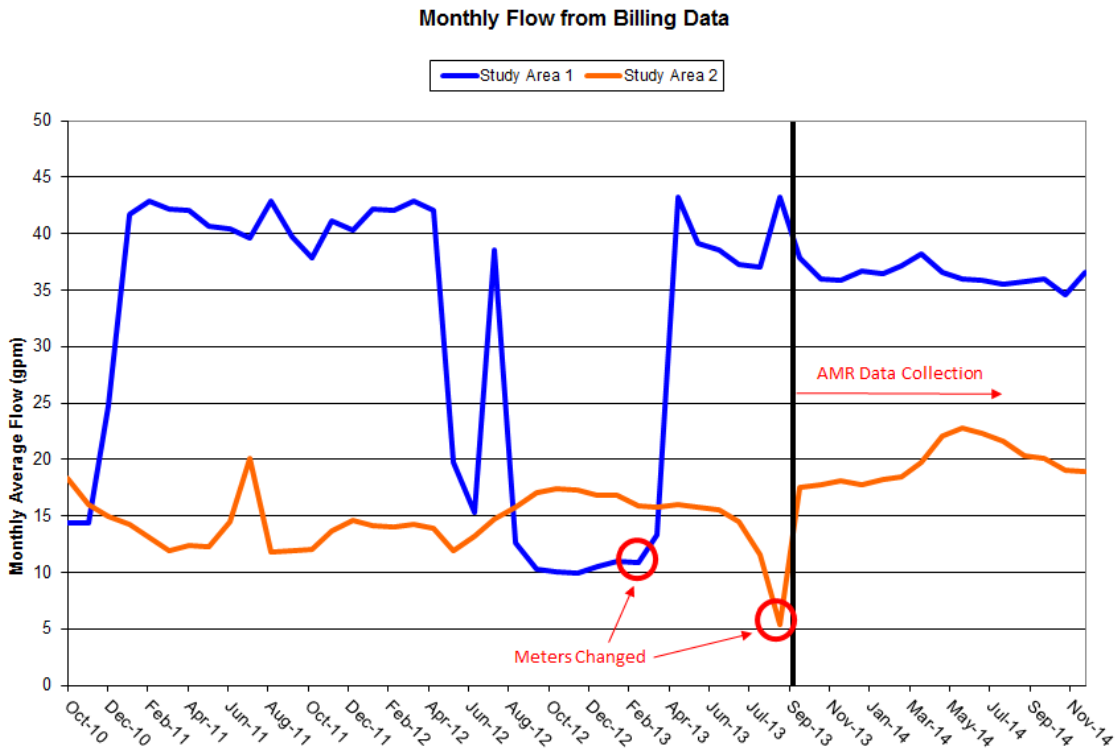


Figure 2-3. Average monthly flow from billing data for both study areas

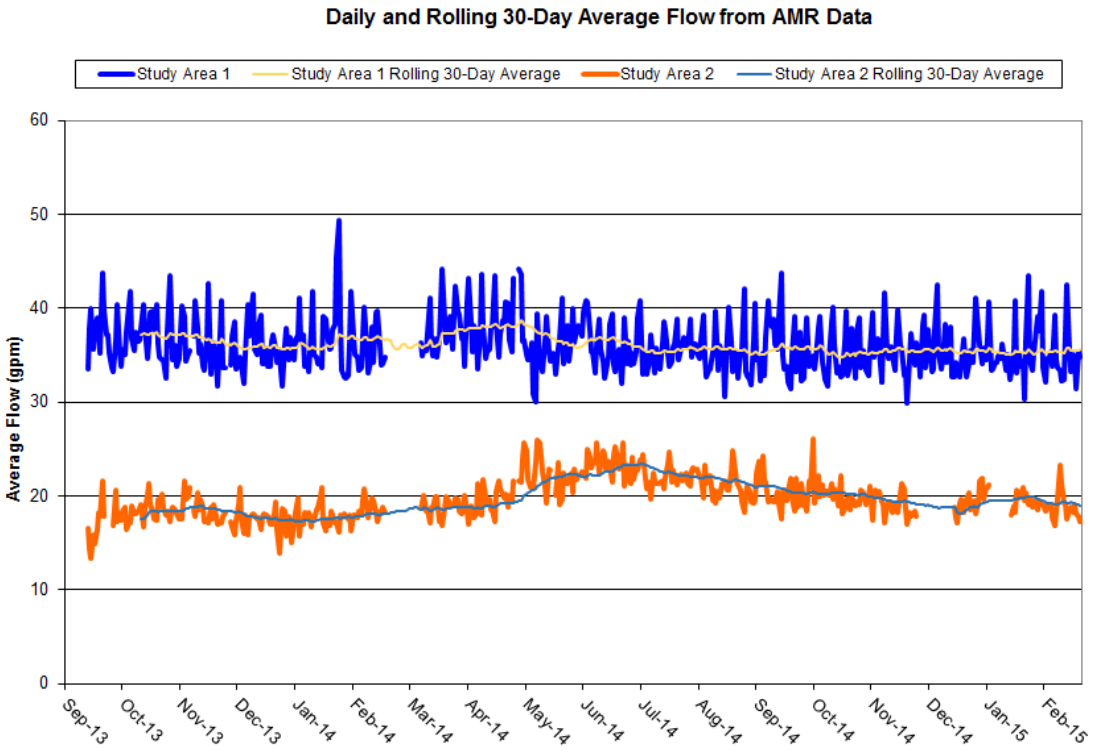


Figure 2-4. Average daily and monthly flow from AMR data for both study areas

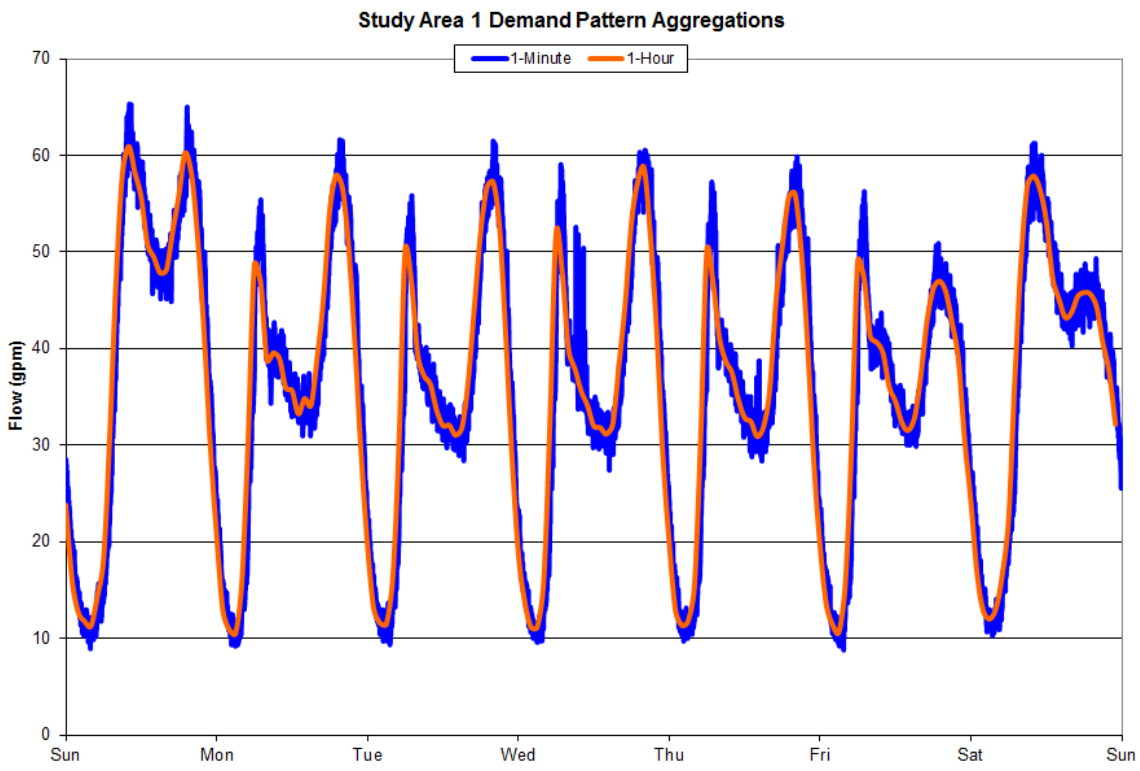


Figure 2-5. Aggregate demand patterns for Study Area 1

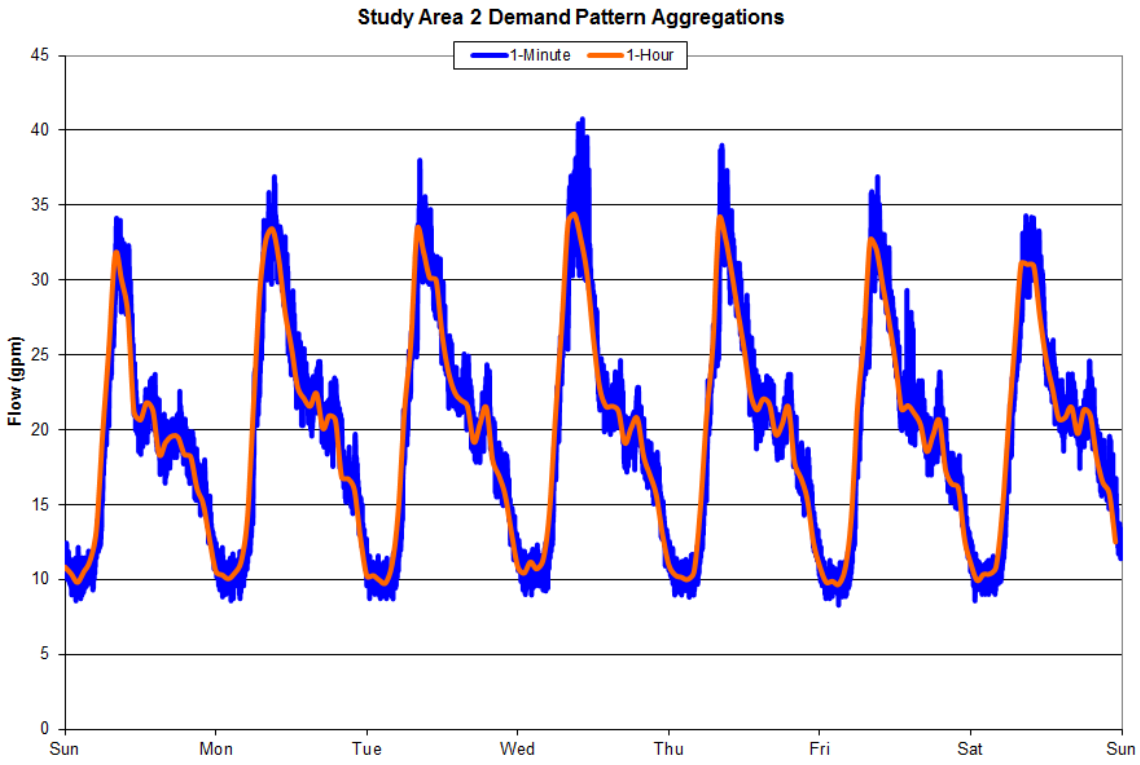


Figure 2-6. Aggregate demand patterns for Study Area 2

Table 2-1. Standard deviation and coefficient of variation as a function of time step

| | Study Area 1<br>Mean Flow = 36.1 gpm | | Study Area 2<br>Mean Flow = 19.7 gpm | |
|---|---|---|---|---|
| Time Step | Standard Deviation | Coefficient of Variation | Standard Deviation | Coefficient of Variation |
| 1 Minute | 18.5 | 0.51 | 10.5 | 0.53 |
| 5 Minutes | 18.2 | 0.50 | 10.3 | 0.52 |
| 15 Minutes | 17.3 | 0.48 | 9.5 | 0.48 |
| 1 Hour | 15.6 | 0.43 | 8.0 | 0.41 |

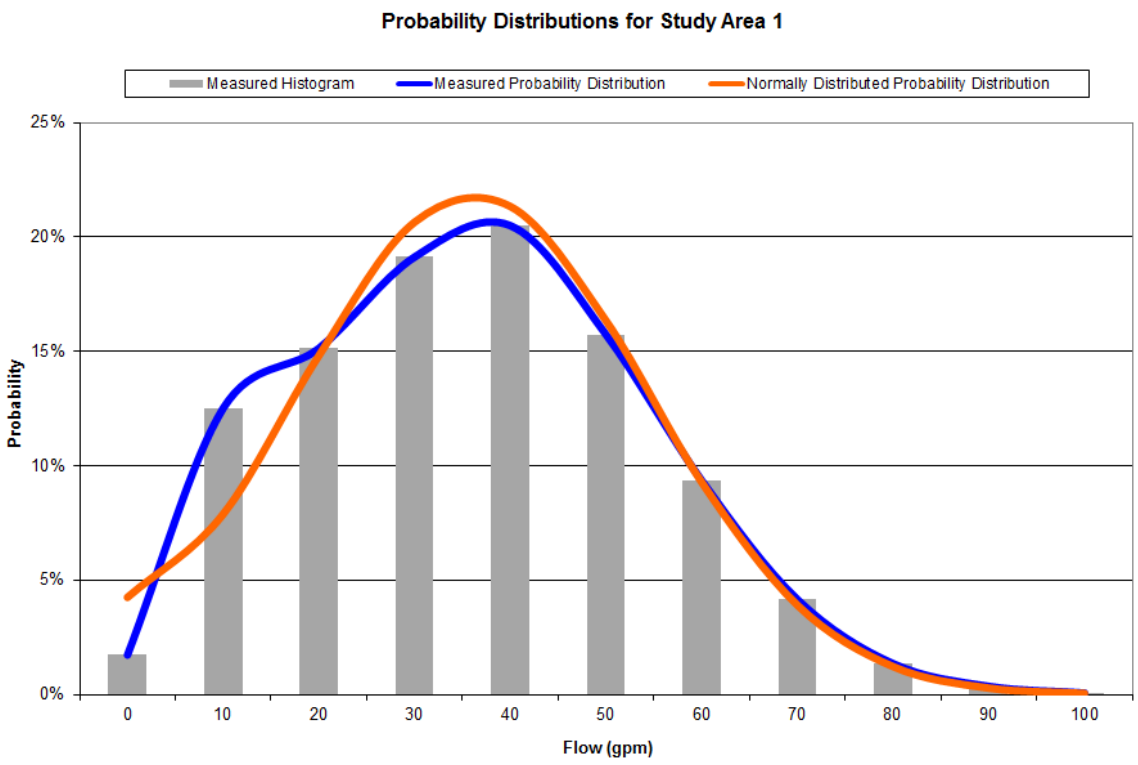


Figure 2-7. Probability distributions of 744,785 1-minute flows for Study Area 1

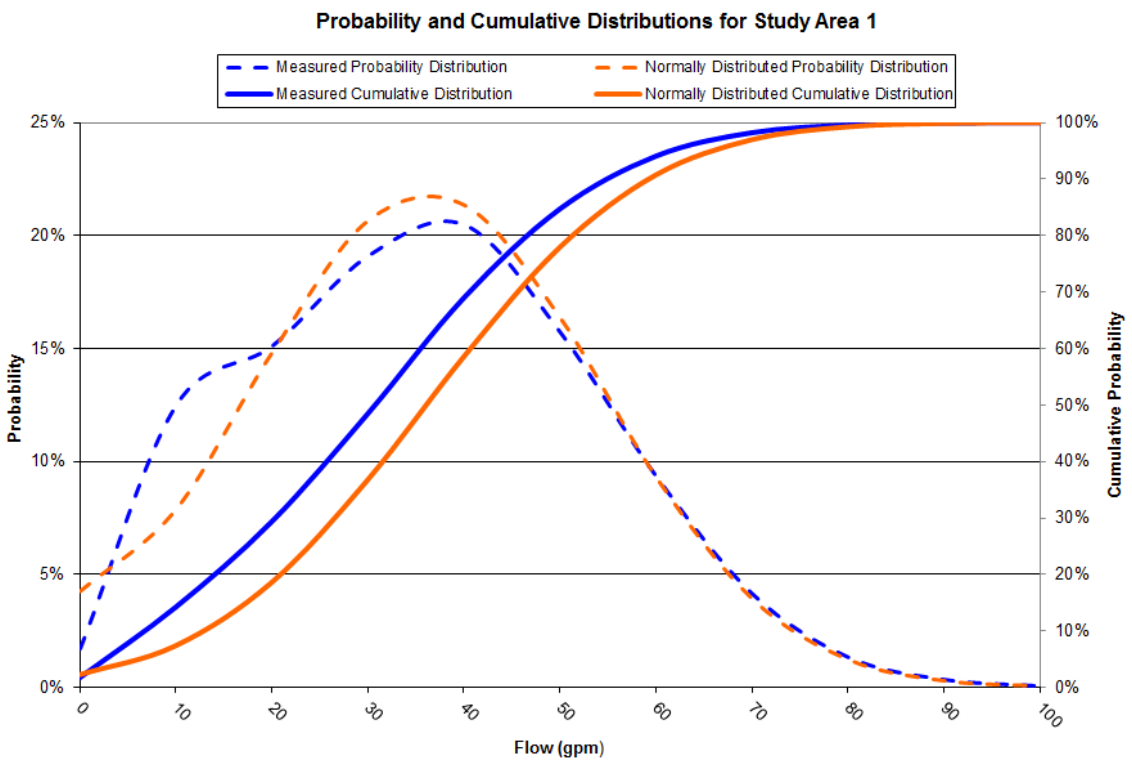


Figure 2-8. Probability and cumulative distributions of 744,785 1-minute flows for Study Area 1

Table 2-2. Summary of weekly measured and predicted values for 76 weeks

| Location | | Statistic | Weekly Average | Peak 1 Hour | Peak 15 Minute | Peak 5 Minute | Peak 1 Minute |
|---|---|---|---|---|---|---|---|
| Study Area 1 | Measured | Minimum Flow | 33.9 | 60.0 | 75.0 | 80.0 | 90.0 |
| | | Average Flow | 36.1 | 71.1 | 88.9 | 96.1 | 107.2 |
| | | Peak Flow | 39.1 | 88.6 | 120.0 | 120.0 | 130.0 |
| | Predicted | Minimum Flow | n/a | 76.6 | 84.4 | 89.8 | 97.1 |
| | | Average Flow | n/a | 81.4 | 89.6 | 95.4 | 103.1 |
| | | Peak Flow | n/a | 87.6 | 96.4 | 102.7 | 111.0 |
| | Percent Difference | Minimum | n/a | -35% | -20% | -20% | -13% |
| | | Average | n/a | -15% | -1% | 0% | 3% |
| | | Maximum | n/a | 8% | 25% | 20% | 21% |
| Study Area 2 | Measured | Minimum Flow | 16.6 | 34.3 | 40.0 | 50.0 | 60.0 |
| | | Average Flow | 19.6 | 40.7 | 53.4 | 59.3 | 67.1 |
| | | Peak Flow | 23.9 | 55.7 | 70.0 | 80.0 | 80.0 |
| | Predicted | Minimum Flow | n/a | 38.1 | 41.9 | 44.6 | 48.3 |
| | | Average Flow | n/a | 44.4 | 48.9 | 52.1 | 56.3 |
| | | Peak Flow | n/a | 54.0 | 59.4 | 63.3 | 68.4 |
| | Percent Difference | Minimum | n/a | -34% | -20% | -21% | -11% |
| | | Average | n/a | -10% | 8% | 12% | 16% |
| | | Maximum | n/a | 20% | 28% | 26% | 33% |

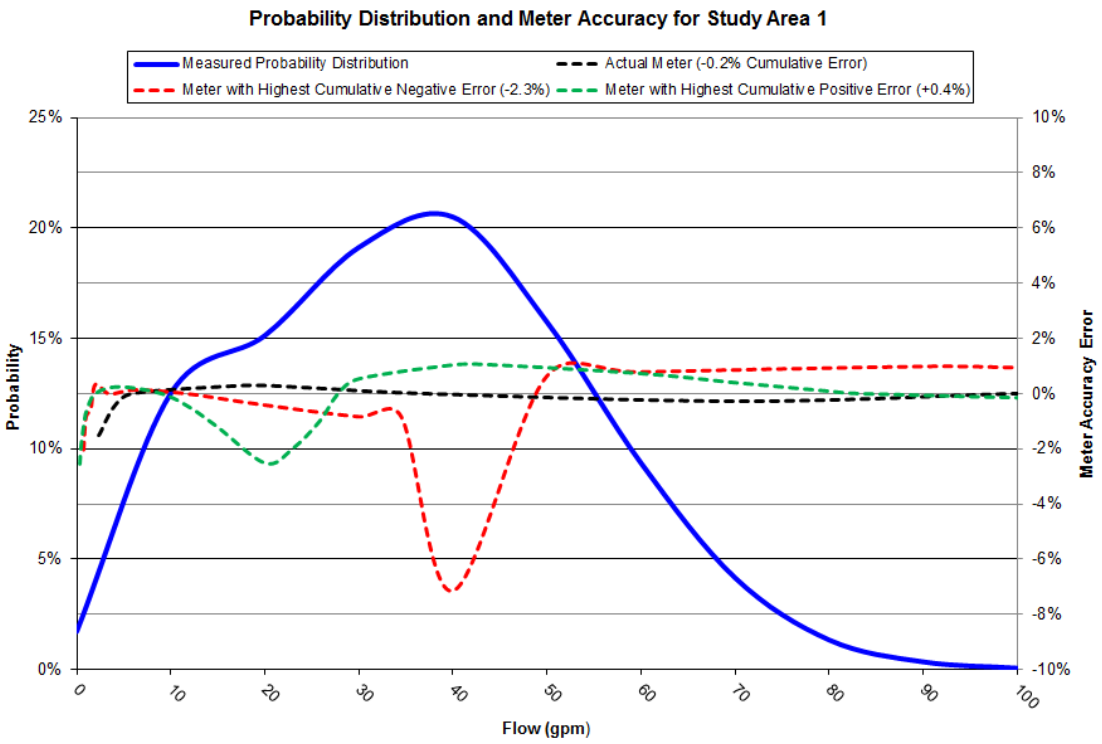


Figure 2-9. Measured probability distribution vs. meter accuracy

# CHAPTER 3
# USE OF HIGH-FREQUENCY DATA TO DETECT HOUSEHOLD LEAKS AS OUTLIERS TO EXPECTED EVENT INTENSITY, DURATION, FREQUENCY, AND VOLUME

## Scope and Overview

This paper presents the results of a prototype high-frequency water use evaluation using one-minute data collected for three single family homes in Hillsborough County, Florida over a period of one year. Automatic meter reading (AMR) meter registers are used with short-range wireless communication that allow for ease of data collection by driving by and downloading the data from the meter registers. This analysis quantifies leaks by looking at outlier events in terms of intensity, duration, frequency, and volume. These homes have separate indoor and outdoor meters so the analysis can compare outliers across aggregate indoor events and irrigation. The term "aggregate" event is used as specific end uses, e.g. toilet flushes, are not quantified. The aggregates of end use events are summarized over continuous durations in order to analyze the most significant leak/break events. The results are promising and the techniques have been applied to the larger study area presented in Chapter 4.

Leaks can have a significant impact on the overall water budget of residential end use, but limited research has focused on the precise quantification of this leakage. In addition, extreme leaks caused by pipe breaks may be minor contributors to the overall water budget but cause costly damage. For precise quantification of leaks, high-frequency water use data must be analyzed at the individual household level. However, limited high-frequency water use evaluations have been published in the literature that summarize water use at this refined scale. The published evaluations have focused on defining anticipated end use events and separating them into their own intensity, duration, and frequency (IDF). These evaluations are considered bottom-up analysis in that they summarize water use at the end use level and can be aggregated up to the household level or larger spatial scales. The aggregation up to the household level

allows for the probabilistic leakage evaluation discussed in this paper. Available research is described below to understand the difference between previous approaches and this current research. The goal of previous research efforts was to define the anticipated events, whereas the goal of the current research is to determine outlying events as indicators of unanticipated events, i.e. leaks. These events that occur with the lowest frequency can have the biggest consequences.

## Previous High-Frequency Evaluations for Individual Homes

Two bottom-up approaches appeared in the mid-1990s using high-frequency data for single-family residences. One approach focused on probabilistic demands for distribution system modeling, and the other approach focused on end use identification for conservation and water use efficiency purposes. Buchberger and Wells (1996) proposed a method of characterizing one second data sets into their IDF's for the purpose of developing a probabilistic demand simulator for water distribution system simulation modeling with an emphasis on estimating water quality as a function of residence time. They performed water use data collection and analysis for one year on four homes at one-second intervals for a neighborhood near Cincinnati, Ohio. They logged data and classified single equivalent rectangular pulses (SERPs) by type (deterministic or random), location (indoor or outdoor), and day (weekday or weekend). They used the pulses to test the previously proposed hypothesis that residential demand can be simulated using a nonhomogeneous Poisson rectangular pulse (PRP) process (Buchberger & Wu 1995). In addition, they presented the data for two residences showing the distribution of the data, both from a total cumulative distribution perspective and average weekday/weekend hourly patterns. None of these houses used irrigation systems, so the aggregated data reflected indoor water use events. They did not attempt to define the actual indoor end uses such as toilet flushes and showers. Buchberger et al. (2003) built on the initial PRP process and data collection effort and followed with data collection on 21 homes at 1-

second intervals for 252 days near Cincinnati, Ohio. Figure 3-1 shows an example of SERPs for three homes as illustrated by Buchberger et al. (2003). The SERPs are shown by the color-coded pulses for the three homes that indicate a time series of separate, fixture-level, events. When defining the IDF, the statistics summarized these individual events and not the aggregate event at the household level. In Figure 3-1, the black bars underneath the time series show the continuous duration of the aggregate event at the household level. This will be discussed in more detail in the next section as it is the basis for the current research.

Other studies have built on this research but have primarily looked at the aggregation up to many homes for the purpose of distribution system modeling, although one subsequent study (Vertommen et al. 2014) presented data collected for indoor water use for 82 single-family residences from the town of Latina, Italy. Each home was monitored for 4 total days, consisting of 4 consecutive Mondays, with a temporal resolution of 1 second. The purpose of this study was to compare measured data to theoretical scaling laws.

At the same time as the early work by Buchberger et al. in the 1990s, DeOreo et al. (1996) developed a bottom-up, end use, approach with measured water use data for 16 homes in Boulder, Colorado at 10-second intervals for 21 days. The focus of this research was water conservation wherein an end use inventory is a critical part of the study since the end uses are the decision variables for finding the optimal blend of investments in water conservation (Friedman et al. 2014, Morales et al., 2013). Software called “Trace Wizard” was developed that could estimate the type of end use as illustrated in Figure 3-2. The software was used to quantify individual fixture-level events that were determined to have a distinct signature in terms of IDF. In this manner, they could temporally aggregate end uses for each customer to quantify the

relative importance that each end use has on total water use, including water use for toilets, clothes washers, showers, faucets, and irrigation.

The initial success of this process oriented, end use, approach led to a nationwide water use study that used the same technique for collecting water use data for 100 homes in each of 12 different cities for 4 weeks at 10-second intervals (Mayer et al. 1999). They reported end use statistics for fixture-level events and presented hourly use patterns based on the average data for all homes, showing both indoor and outdoor use as well as the hourly pattern for each component that was added to calculate the total indoor use. Key results include the observation that single family indoor residential water use is quite consistent from city to city and that individual water use patterns, e.g., toilet flushes per person per day, are very similar. Numerous follow-up studies have further confirmed these findings, e.g., DeOreo and Mayer (2012). This 1999 seminal study has provided a solid foundation for estimating end uses at the individual fixture level. An update, titled *Residential End Uses of Water, Version 2* (DeOreo et al. 2016), included both the original and additional data sets. The additional 10 second data included 762 homes randomly selected from 9 study areas. The data were collected for about 2 weeks. The updated study had more varied study site locations, hot water usage data, more detailed landscape analysis, and additional water rate analysis. Similar to the work by Buchberger et al. (2003), the events were summarized at the fixture level and not by aggregate events at the household level.

Blokker et al. (2010) looked at behavioral statistics and developed a simulation approach that bridges the gap between end use processes and probabilistic demands for modeling. The approach used process statistics based on survey data of water use habits as this was available for a larger population than were direct water use measurements. For comparison with the process-driven approach, Blokker et al. (2010) presented water use data for 43 homes dispersed over the

city of Amsterdam in the Netherlands at 5-minute intervals for 7 days. The data were aggregated from data collected at 1-minute intervals in order to dampen errors caused from the volumetric resolution of the raw measurements, which were available in 1-liter increments. The cumulative distribution of the entire data set as well as the maximum flow measurements were presented as a composite of data for all homes. Also, the average of the summation of all 43 homes for each 5-minute interval of the weekday was presented to show the average weekday water use pattern for the summation of the 43 homes. The measurements were compared to results from the simulation model that was developed using the process statistics, called SIMDEUM. One key feature that was excluded in SIMDEUM was leaks, and one of the homes in the analysis was excluded because it had a continuous leak of 0.2 L/min.

The idea of leak quantification highlights one of the key differences in probabilistic demand simulators vs. end use identification. For the probabilistic demand simulators, the concept defined water use "pulses" as events, and therefore lumped all measured data on water use events together when developing their IDFs. As such, unless the probability of continuous leaks is included separate from "pulses", there isn't a way to include leaks in the probabilistic demand simulators. For end use identification, all uses of water, including leaks, would need to be quantified. Events are identified by specific end use types, and therefore each type has its own series of IDFs. If the IDFs for the end use types are properly quantified, it is much easier to determine if a water use event is anticipated based on how long and frequent a certain intensity occurs.

## Definition of an Event

An important concept is to define an "event". For the previously referenced studies, they defined an "event" as something that is an anticipated, normal, use. Therefore, they focused on quantifying the intensity and duration of anticipated uses for either simulating residential demand

starting from the fixture level or for determining fixture-level water budgets, e.g. percent of water used by showers. The studies collected data at a temporal frequency ranging from 1 to 10 seconds. This is necessary to measure individual end use events because many events occur on the order of seconds as can be seen in Table 3-1.

Because the fixture-level events as indicated in Table 3-1 were the focus of previous research, the statistics for the aggregate events where more than one of these fixture-level events were occurring were not presented. As can be seen in Figures 3-1 and 3-2, these fixture-level events can occur at the same time or in close proximity to other events. The previous research efforts focused on splitting these into individual events in order to determine the IDF of these anticipated events. While previous studies identified leakage, the event statistics for leakage were not presented because of the variability in the types of leaks and how their IDFs can differ drastically. Because the event statistics were not presented, methods to identify leakage were not clearly defined other than the description that they could be identified because they didn't fit into other categories.

The current research looks at aggregate event statistics to determine outlying events as indicators of leakage. The analysis looks at outliers to identify two types of leaks: 1) continuous leak with a high duration and low frequency of occurrence, and 2) intermittent leak with a short duration and high frequency of occurrence. The detection of continuous leaks is especially important because it could be an indicator of damage-causing pipe break events within the home. In the current research, aggregate events are defined by consecutive data points where water use is greater than zero, and the event statistics will report the number of events along with the duration, volume, and average intensity of each event. This is considered average intensity because the aggregate events summarize all periods of time with continuous use, effectively

creating a weighted average of all individual fixture-level event intensities that occur within the aggregate period without being able to quantify these individual events. Another important concept is that in this definition, the minimum inter-event time is the resolution of the data, i.e. one minute. The first data point with water use greater than zero starts the event, and the subsequent data point where water use is not greater than zero will end the event. The example in Figure 3-3 shows how a time series of water use data is split into different aggregated events when separated by a data point with no water use. The aggregate event matches the duration and volume of the individual data points that make up the event. The average intensity of the event is calculated by dividing the volume by the duration.

In the current research, the data points are at 1-minute frequencies. This will be discussed further in the next sections. The discussion up to this point has focused on defining aggregate events and potential leakage, not individual fixture-level events. Part of the reasoning is because the temporal resolution is greater than that of previous research and can't identify individual fixture-level end uses. Consider the events in Figures 3-1 and 3-2 with data at 1- and 10-second intervals, and Figure 3-3 with data at 1-minute intervals. As the time step increases, the ability to distinguish any one of these individual fixture-level events becomes increasingly difficult. More importantly and to the point of the current research, the higher frequencies used in previous research efforts aren't necessary for determining outliers to the aggregate event data that are used to identify potential leakage.

### Process to Identify and Review Potential Unanticipated Events

The following steps outline the overall approach to identify and review potential unanticipated events. The potential unanticipated events will be summarized in order to quantify their water use and classify them by volumetric ranges that indicate low intensity leaks or high intensity pipe breaks. The step-by-step process is important because previous research efforts

didn’t explicitly focus on ways to identify leaks or pipe breaks. The process described below allows for the identification and quantification of these events.

Step 1: Collect data and develop a database that has the time series across the period of record for each individual meter.

Step 2: Aggregate all consecutive data points with water use into individual events for each meter using the database from Step 1.

Step 3: Summarize events for each meter by month showing ranges of event statistics.

Step 4: Plot the cumulative distribution of all data points with water use as a potential indicator of any obvious distribution outliers.

Step 5: Classify and plot each individual event from Step 2 by the duration, volume, and average intensity.

Step 6: Define anticipated event ranges using the intensity, duration, and volume from Table 3-1.

Step 7: Summarize all events from Step 5 that are outside of the anticipated event ranges in Step 6 by including the total count and volume within specified ranges.

Step 8: Split the total volume of the events in Step 7 into two sub-categories based on the intensities of the individual data points: the individual data points with intensities inside the anticipated range, and those outside of the anticipated range.

Step 9: Evaluate the results and refine the ranges used in Step 6, if necessary.

Repeat steps 6 through 9 until the user is satisfied that the ranges used in Step 6 are appropriate.

## Implementation of Process to Identify Unanticipated Events

### Step 1 – Data Collection and Database Development

A pilot study using automatic meter reading (AMR) data was conducted using data for 3 single family homes in Hillsborough County, Florida. The AMR data loggers used in this research replaced the analog registers on the meters. No internal mechanical components of the meter itself were replaced, and the resolution of the gallons reported by the AMR data loggers was as accurate as the registering capability of the mechanical components of the meter. The internal mechanical components of the meters used nutating discs capable of reading in increments of 0.017 gallons. The local data storage on the AMR was limited to 32,000 data points which meant that the data had to be downloaded every 22 days in order to avoid data loss. The data files were collected by driving to each meter and downloading the data from the loggers through short-range wireless communication. The vehicle was equipped with a radio that communicated with a local radio transmitter on each of the data loggers. Each data file took approximately five minutes to download. A database was built that allowed each data file to be uploaded to the appropriate dataset for each meter. The resulting database allowed easy access to water use data by time of day, day of week, and any combination of these two.

The 3 homes were targeted because they had separate indoor and outdoor meters, allowing for a clear distinction between indoor and outdoor events. The data were recorded at 1-minute frequencies, and the collection effort covered a period from April 2014 to August 2015. A subset of 365 days was analyzed in order to reduce the potential for skewing results based on seasonality and to summarize data based on annual statistics. Over 4 million data points were collected, with a subset of over 3 million data points used for the analysis. An aerial map of the pilot area is shown in Figure 3-4. The housing and annual water use statistics for each home are shown in Table 3-2. According to American Community Survey (ACS) data for 2015, the

rolling 5-year average persons per household (pph) for the Census Tract that includes the study area is 2.89. Assuming that the 2.89 pph is an appropriate average for the 3 homes, the resulting gallons per capita per day (gpcd) for indoor water use are 56, 67, and 92 gpcd, respectively. For comparison, the two *Residential End Uses of Water* studies showed a decline in average per capita water use from 69.3 gpcd (Mayer et al. 1999) to 58.6 gpcd (DeOreo et al. 2016). Buchberger et al. (2003) reported an average of 55 gpcd; however, this was after excluding leaks from the dataset.

The irrigable areas for the three houses are 10,715; 9,145; and 10,985 square feet, respectively. Knight et al. (2015) present the pdf and cdfs for 6,305 single family residences in central Florida. Based on this data, the median irrigable area is about 7,000 square feet and the three houses with irrigable areas of about 10,000 square feet would be in the 60 percentile range. The application rates for irrigation for the three homes were 21, 96, and 21 inches per year. The benchmark application rate for this study area is about 25 inches per year (Knight et al. 2015). Thus, house 2 is applying about four times the needed application rate. The other two homes are applying about the benchmark application rate. All of these three homes have pools. They were built in 2006 and have about 4,000 square feet of heated area, much larger than a typical newer home with about 2,500 square feet. The 2016 market values of the three houses are $482,000, $411,000, and $376,000 respectively. Thus, overall these three houses are well above average in value, size, and features.

**Step 2 – Data Aggregation into Aggregate Events**

The data for the 3 homes were aggregated into individual events based on data points with continuous water use. Figure 3-3 shows a graphical example of this process and how the duration, volume, and average intensity are calculated. The frequency of events is limited by the resolution of the data. In general, the maximum number of aggregate events that can occur in a

day is one-half of the data points recorded for the day. Specific to the current research with 1-minute data points, the maximum number of events that can occur within one day is 720. Table 3-3 lists the possible number of events that can occur for a few examples.

In order to put the frequency of events in context, consider the following example for differing leak types. If water is used continuously for 12 hours and then is shut off for 12 hours, the percent of time water is used for the day would be 50%. This would result in 1 event for the day and could indicate a continuous leak. If water use occurs every other minute, with no water use recorded in between, this would result in 720 events for the day and could indicate an intermittent leak. If the intermittent leak occurs with a high frequency, then as the time step increases, the more likely the intermittent leak will appear as a continuous leak. This is not necessarily a problem as long as the leak can be detected. These are extreme cases, but they do provide information on understanding the number of events that can occur.

**Step 3 – Aggregate Event Summaries for Each Meter**

Table 3-4 provides a summary of the data and event statistics for each month of the analysis. There were some gaps in the data for March 2015, so additional days were used in April 2015 in order to complete the 365 days of record used for the analysis. The table shows the first documented results of a high-frequency evaluation summarized for each meter by month for an entire year, allowing for a longer period to evaluate annual leakage and to account for seasonality. The individual event outliers will be discussed in a later section; however, the following are some observations from looking at how the data summaries vary in Table 3-4.

The House 1 indoor event summaries show several indicators of continuous leaks for the months of March and April 2015. The percent of data with water use, water use per day, event volume, and event duration all increase significantly. Likewise, the event starts per day and event intensity all decrease significantly, indicating that the predominant water uses during the

events are prolonged leaks that reduce the detection of new event starts. By comparison, the House 3 indoor event summaries show several indicators of intermittent leaks for the months of April through June 2014. The percent of data with water use and event starts per day are significantly higher than the rest of the period; however, event volume and event duration are both lower. The House 1 outdoor event summaries shown prolonged continuous leaks over most of the dataset, with a continuous leak occurring from August 2014 to December 2014, and another occurring from January 2015 through the end of the period of record ending on April 27, 2015. When comparing the event starts per day to the anticipated values in Table 3-1, the House 3 outdoor event summaries are the cleanest ranging from 0.4 to 2.6 event starts per day. The House 2 indoor event summaries are the cleanest ranging from 42 to 65 event starts per day. However, the House 2 outdoor event starts per day average 19 whereas typical watering intervals are two to three times per week. Similarly, the House 2 outdoor event durations average 5.4 minutes, far less than anticipated irrigation durations of 30 to 120 minutes. One possible explanation is that there is a time delay between the starting/stopping of multiple irrigation zones, thereby splitting one continuous event into multiple events

**Step 4 – Cumulative Distributions of All Data Points with Water Use**

Probability and cumulative distributions are used to evaluate the probability of values within specified ranges. If the distributions represent all the individual data points, there is no indication of how one data point occurs relative to another. As an example, the probability of any one data point exceeding the 99% cumulative distribution could be of interest for investigating peak flow rates, but there would be no indication of how these peak values occur relative to one another. Cumulative distributions can be especially useful for analyzing the most frequent flow rates, which can be observed as near vertical portions of the curve (i.e. a small change in flow on the x-axis with a large range of cumulative occurrence on the y-axis). Water

use is anticipated to occur over a small period of time during the day, resulting in a high percentage of zero data points. A cumulative distribution of these data would show a vertical line at zero. Rather than plotting all of these zero data points, Figures 3-5 and 3-6 plot only the data points with water use, as summarized in Table 3-4. The results show that indoor water use for the three homes occurred 23%, 11%, and 22% of the time, respectively. These are higher than the values reported by Buchberger et al. (2003) based on one second data where water use occurred 4.5% of the time. The current research shows a higher percent of time with water use because of the effect of time averaging when using a larger time step (e.g. 1-minute data points compared to 1-second data points).

Figures 3-5 and 3-6 show the probability and cumulative distributions of the data points that are greater than zero. Figure 3-5 shows that for indoor water use, flow rates are normally in the 0 to 5 gpm range with only the top 10% of individual data points exceeding 1.5, 2.1, and 2.5 gpm for the three homes. By comparison, Figure 3-6 shows that a majority of outdoor water use exceeds 10 gpm. The high percent of time that House 1 indicates a low flow rate in both figures is indicative of prolonged, continuous, leaks. This is confirmed by the values in Table 3-4, where House 1 outdoor water use occurs 93% of the time, compared to 7% and 1% for the other two homes. Aside from House 1, Table 3-4 and Figure 3-6 show anticipated on/off distributions for irrigation systems, with a majority of the time at zero flow and the remainder of the time at flow rates greater than 10 gpm.

**Steps 5 and 6 – Plot Aggregate Events and Show Anticipated/Unanticipated Event Ranges**

As noted previously, the cumulative distributions don't indicate how the individual data points occur relative to one another. They do provide insight into the anticipated flow rates of indoor and outdoor water use, as well as how leaks will skew the data. A different summation is presented to analyze aggregate events so that outliers can be used to quantify the number of

potential leak events and determine how quickly these events can be detected. As defined previously, aggregate events are summations of the consecutive data points where water use is greater than zero, and the aggregate events in Figures 3-7 and 3-8 show the duration, volume, and average intensity of every event in the House 1 dataset. In addition, the anticipated event ranges utilizing the criteria in Table 3-1 are applied to the table, with the shaded regions indicating where unanticipated events have occurred within the defined volumetric ranges.

The average event intensities plotted in Figures 3-7 and 3-8 are weighted averages of all individual fixture-level intensities that occur within the aggregate period of each event. It is calculated by summing the total volume over the aggregate event and dividing it by the duration. This means that the longer the duration of a continuous leak, the more weight the leakage rate will have on the average intensity of the aggregate event. High intensity continuous leaks will show as outliers by some combination of high intensity, long duration, and large volume. Low intensity continuous leaks will show as outliers by some combination of low intensity, long duration, and large volume.

**Step 7 – Summarize the Unanticipated Events by the Total Number and Volume within Specified Volumetric Ranges**

Table 3-5 shows the total number of unanticipated events and the cumulative volume of those events that occur within the defined volumetric ranges. The defined volumetric ranges correspond to the bounds of the isovolume lines shown in Figures 3-7 and 3-8.

**Step 8 –Split the Total Volume of Unanticipated Events into Volumes within Anticipated and Unanticipated Intensity Ranges**

In order to further evaluate the unanticipated event volumes, the individual data points must be evaluated to see if shorter-duration anticipated event volumes are being masked by the longer-duration unanticipated aggregate event volumes. Table 3-5 shows the percentage split of the individual 1-minute intensities that comprise the aggregate events within each category. The

percentage split indicates the cumulative volumetric percentage of all individual data points that occur within the anticipated and unanticipated intensity ranges. This reporting isn't used to directly indicate whether sub-events to the larger aggregate event are anticipated or unanticipated, only to indicate how much of the data occurs within anticipated and unanticipated intensity ranges. This is valuable for the purpose of identifying potential low intensity or high intensity leaks.

**Step 9 – Evaluate the Results and Refine the Ranges Used in Step 6, if Necessary**

For the unanticipated events that have been summarized in Steps 7 and 8, it is likely that aggregate events with many individual data points inside the anticipated intensity range are either high intensity leaks, like a pipe break, or an anticipated use with a longer duration, like adding water to a pool. This is critical to understand so that future research efforts can balance the reward of providing rapid and early detection of an unanticipated event with the risk of providing too many notifications or false alarms.

As an example, Figure 3-7 shows three aggregate events for House 1 that each occur with a volume greater than 1,000 gallons (refer to the rose-colored region in Figure 3-7). All three of these events occur with an aggregate intensity just under 0.2 gallons per minute. Referring to Table 3-5, 77.0% of the total volume that occurred for these three events was calculated from individual data points within anticipated intensity ranges (refer to the rose-colored region in Table 3-5 for House 1). Note that for all three homes, there is a trend that correlates an increasing total unanticipated event volume with an increasing percentage of volume occurring within anticipated intensity ranges. This indicates that as the unanticipated event volume increases, the events are likely caused by either pipe breaks or high intensity anticipated uses that have a longer duration than what is normally anticipated to occur.

Because the ranges used in this paper for determining anticipated events have been developed and verified through multiple research studies, the ranges used for Step 6 are not being modified and Step 10 (repeat Steps 6 through 9) is not needed. As noted previously, it is possible that some of the unanticipated events detected using these ranges are actually anticipated uses that have exceeded the defined ranges for duration. The risk of falsely classifying a few anticipated uses as unanticipated events is not addressed in the current study but should be addressed in future research. Of note is that for the three homes, there are 79,789 unanticipated events with an individual event volume less than 100 gallons. The combined volume of these events is 5,345 gallons. By comparison, there are 87 unanticipated events with an individual event volume greater than or equal to 100 gallons. The combined volume of these events is 137,492 gallons. This means that 0.1% of the unanticipated events yield 96.3% of the unanticipated volume.

## Synopsis

The research described in this paper presents an approach for finding unanticipated events for a home and quantifies the annual statistics for three homes with 1-minute water use data. Previous research efforts didn't have data formulated in the process described in this paper or for the duration needed to quantify annual statistics. Therefore, a major data collection effort was needed and launched as described in this paper. This evaluation was the first to explicitly search for leaks and pipe breaks using high-frequency customer water use data. Volumetric ranges are used as summarization categories because the volume of water is both what causes damage in the event of a pipe break and what needs to be conserved, i.e. the intensity and duration are not what drive conservation efforts or damage, it is total volume. One key finding in the current study is that 0.1% of the unanticipated events, specifically those with an individual event volume greater than or equal to 100 gallons, yield 96.3% of the unanticipated volume. The

approach used in this study is being applied to a larger study area with a goal of identifying criteria by which immediate notification to customers could help reduce costly damages to the home in addition to providing data on leakage quantities for a larger study area.

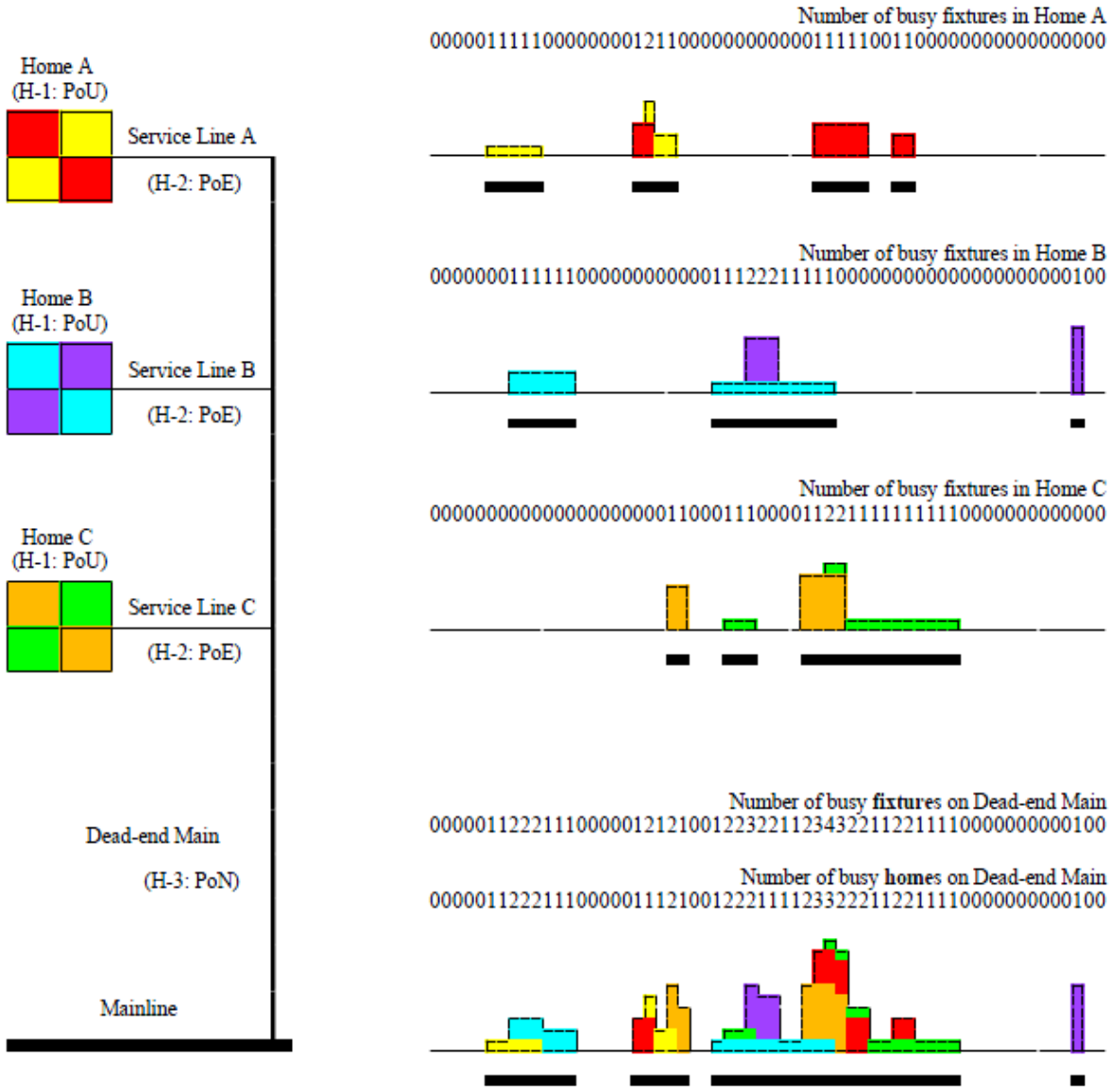


Figure 3-1. Representation of urban water supply end use events by Buchberger et al. (2003)

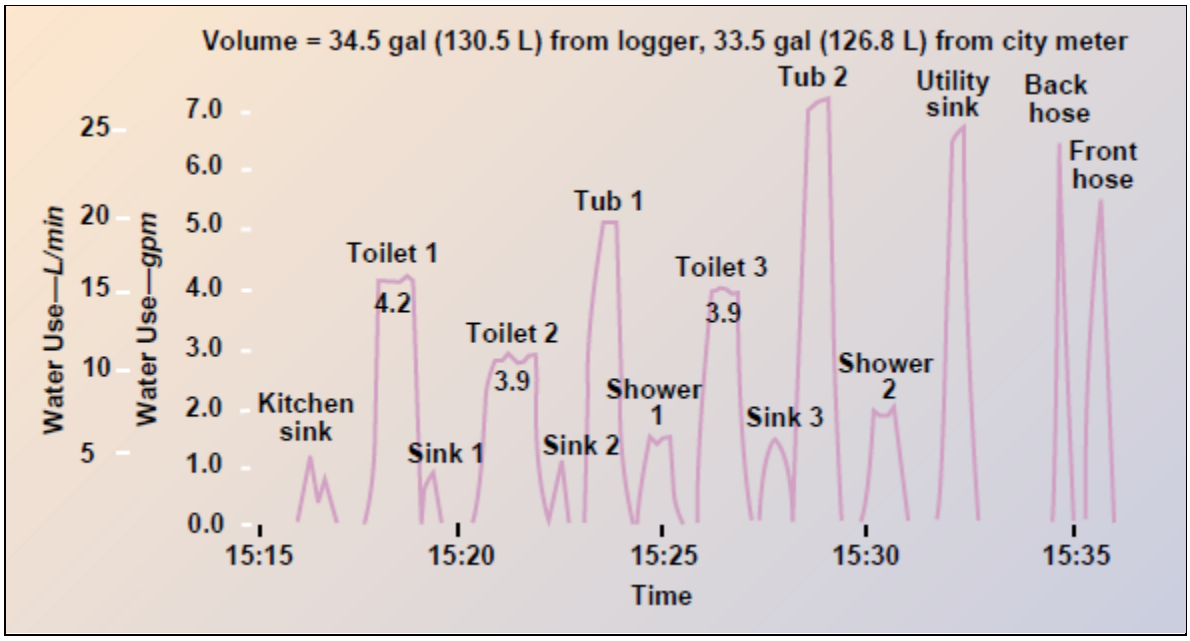


Figure 3-2. Flow trace showing signature end use intensity and duration by DeOreo et al. (1996)

Table 3-1. Fixture-level water use benchmark values for single family residences

| Anticipated Indoor Event End Use[+] | Intensity (gpm) | Duration (min) | Volume (gal) | Events per Day |
|---|---|---|---|---|
| Toilet | 2 to 6 | 0.5 to 1 | 1 to 6 | 15 |
| Shower | 2 to 5 | 5 to 20 | 10 to 100 | 3 |
| Bath | 2 to 6 | 5 to 20 | 25 to 100 | 0.3 |
| Faucet | 0.1 to 3 | 0.5 to 5 | 0.05 to 15 | 30 |
| Clothes Washer[*] | 2 to 4 | 20 to 60 | 20 to 40 | 0.8 |
| Dishwasher[*] | 1 to 3 | 30 to 120 | 5 to 30 | 0.8 |
| **Anticipated Outdoor Event End Use[++]** | **Intensity (gpm)** | **Duration (min)** | **Volume (gal)** | **Events per Day** |
| Automatic Irrigation[**] | 5 to 20 | 30 to 240 | 150 to 4800 | 0.3 |
| Manual Irrigation[**] | 2 to 10 | 5 to 100 | 10 to 1000 | 0.3 |
| **Unanticipated Event End Use** | **Intensity (gpm)** | **Duration (min)** | **Volume (gal)** | **Events per Day** |
| Low Intensity, Itermittent Leaks | <0.1 | 0.5 to 30 | <3 | ? |
| Low Intensity, Continous Leaks | <0.1 | 1440 | <144 | ? |
| High Intensity Pipe Breaks | 1 to 20 | >5 | >100 | ? |

[+]Ranges adapted from previous studies: Buchberger et al. (2003); Blokker et al. (2010); DeOreo et al. (2016).
[*]Flows are intermittent. Reported flow rates are averages over the water use periods.
[++]Ranges adapted from sprinkler system design and maximum flow limitations through residential meters.
[**]Assumes twice-per-week irrigation restrictions.

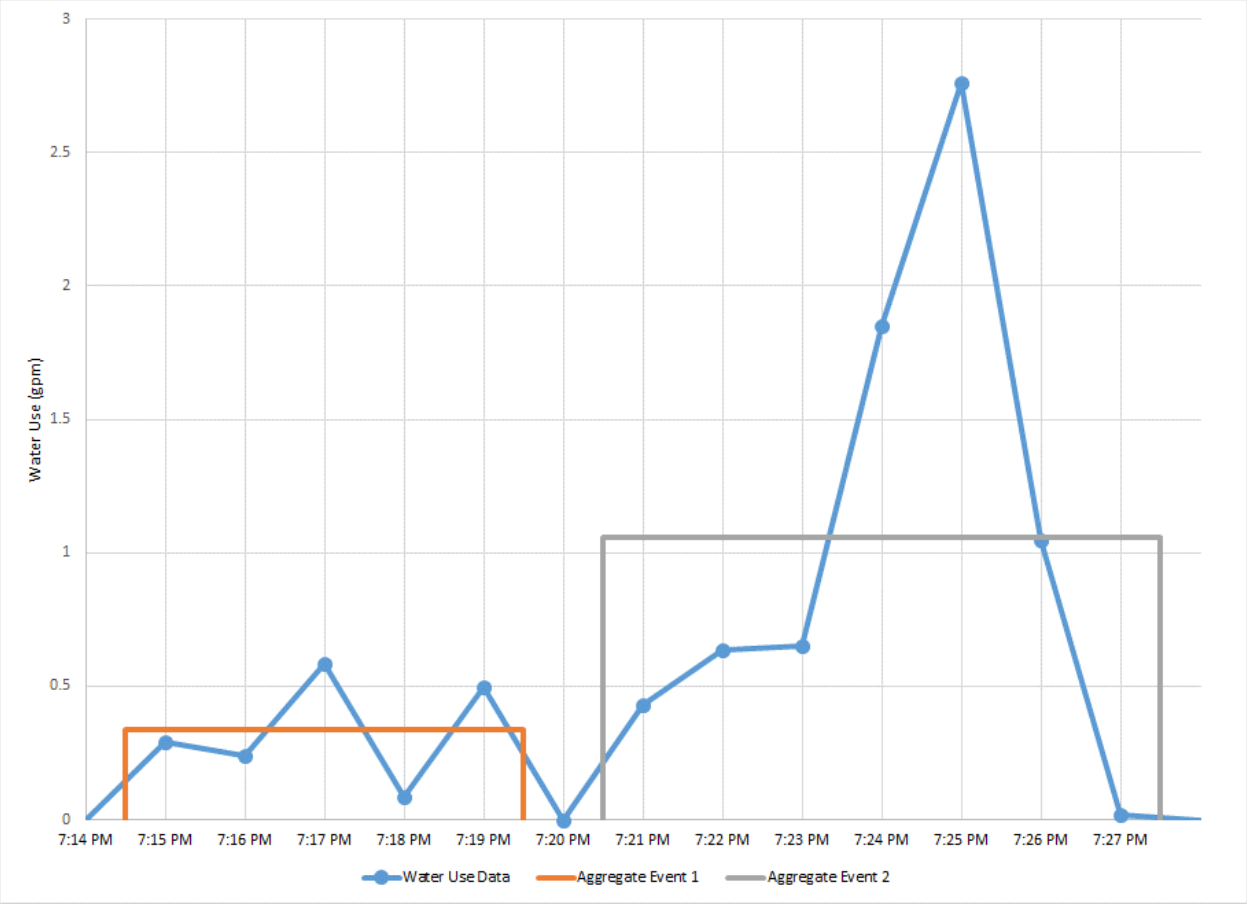


Figure 3-3. Water use data separated into aggregate events showing duration and average intensity

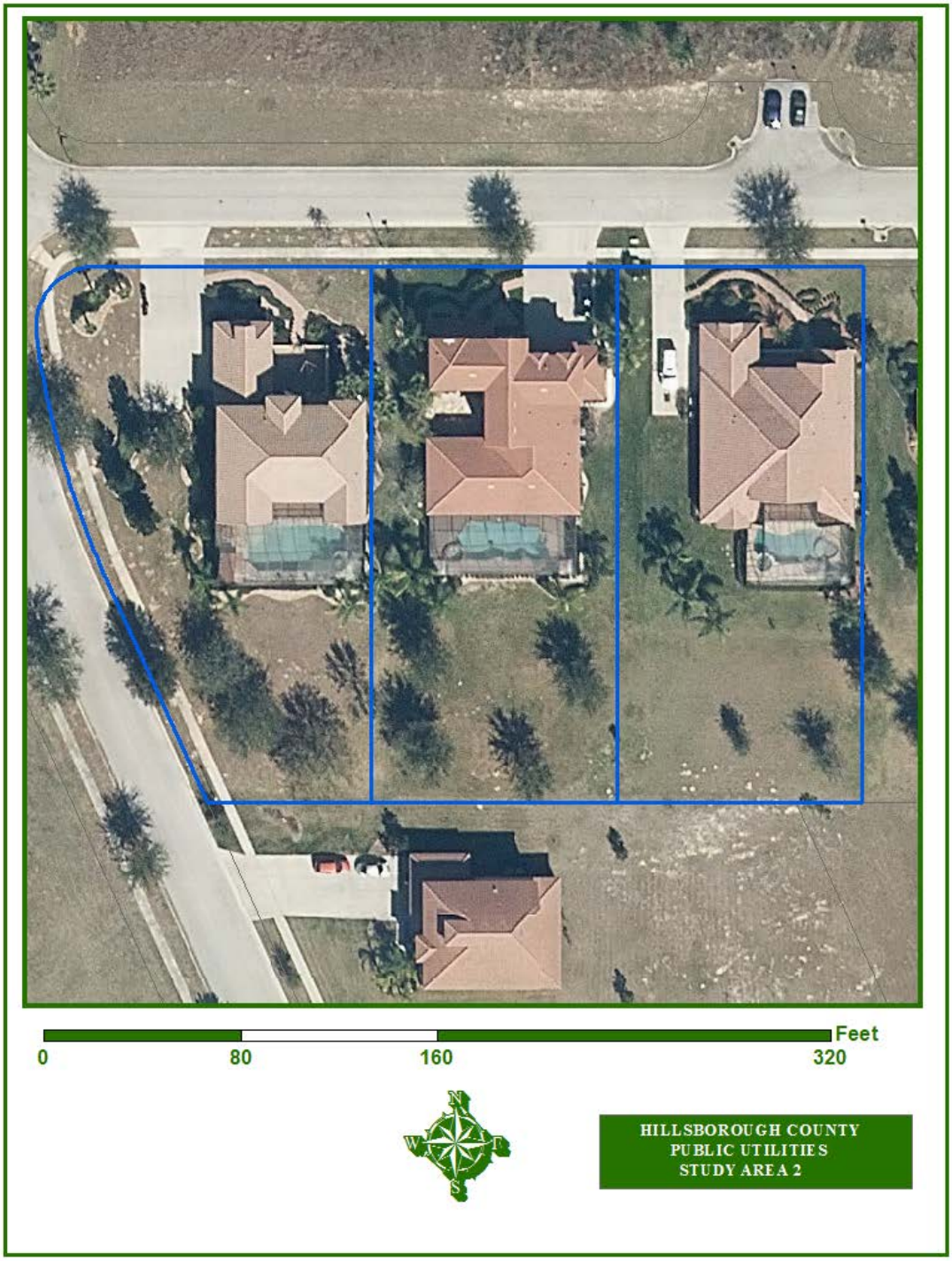


Figure 3-4. Aerial view of 3-home study area in Hillsborough County, Florida

Table 3-2. Housing and annual water use statistics for 3-home study area

| Housing Information | House 1 | House 2 | House 3 |
|---|---|---|---|
| Year Built | 2006 | 2006 | 2006 |
| Heated Area (sq. ft.) | 4,413 | 4,219 | 3,605 |
| Lot Area (sq. ft.) | 23,954 | 21,800 | 21,800 |
| Irrigable Area (sq. ft.) | 10,715 | 9,143 | 10,985 |
| Market Value | $482,282 | $410,571 | $376,478 |
| Annual Average Indoor Use (gpd) | 162 | 194 | 266 |
| Indoor Per Capita Use (gpcd) | 56 | 67 | 92 |
| Annual Average Outdoor Use (gpd) | 376 | 1,441 | 391 |
| Inches per Year of Irrigation | 21 | 92 | 21 |
| Outdoor Per Capita Use (gpcd) | 130 | 499 | 135 |

Table 3-3. Number of possible events per day

| Data Points with Water Use | Data Points with No Water Use | Number of Possible Events | Maximum Possible Events |
|---|---|---|---|
| 1,440 | 0 | 1 | 1 |
| 1,439 | 1 | 1, 2 | 2 |
| 1,438 | 2 | 1, 2, 3 | 3 |
| … | … | … | … |
| 720 | 720 | 1, 2, …, 719, 720 | 720 |
| … | … | … | … |
| 2 | 1,438 | 1, 2 | 2 |
| 1 | 1,439 | 1 | 1 |
| 0 | 1,440 | 0 | 0 |

Table 3-4. Summary of data and average event statistics for each meter by month.

| Summary of Data for Each Home | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Days of Record | 29 | 31 | 30 | 31 | 31 | 30 | 31 | 30 | 31 | 31 | 28 | 9 | 23 | 365 |
| Data Points | 41,760 | 44,640 | 43,200 | 44,640 | 44,640 | 43,200 | 44,640 | 43,200 | 44,640 | 44,640 | 40,320 | 12,960 | 33,120 | 525,600 |
| House 1 Indoor | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Average |
| Percent of Data with Water Use | 18% | 15% | 18% | 14% | 4% | 16% | 16% | 17% | 19% | 19% | 26% | 97% | 96% | 23% |
| Water Use per Day (gallons) | 175 | 179 | 167 | 119 | 55 | 158 | 135 | 173 | 157 | 163 | 173 | 279 | 286 | 162 |
| Events Starts per Day | 115 | 84 | 86 | 92 | 28 | 96 | 100 | 101 | 135 | 132 | 163 | 45 | 7 | 95 |
| Event Volume (gallons) | 1.5 | 2.1 | 1.9 | 1.3 | 2.0 | 1.7 | 1.4 | 1.7 | 1.2 | 1.2 | 1.1 | 6.8 | 31.2 | 1.7 |
| Event Intensity (gpm) | 0.7 | 0.8 | 0.7 | 0.6 | 1.0 | 0.7 | 0.6 | 0.7 | 0.6 | 0.6 | 0.5 | 0.2 | 0.2 | 0.6 |
| Event Duration (minutes) | 2.2 | 2.7 | 3.0 | 2.1 | 1.9 | 2.4 | 2.2 | 2.4 | 2.0 | 2.0 | 2.3 | 35.4 | 154.7 | 3.4 |
| House 2 Indoor | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Average |
| Percent of Data with Water Use | 14% | 14% | 9% | 10% | 11% | 9% | 10% | 11% | 12% | 10% | 9% | 10% | 9% | 11% |
| Water Use per Day (gallons) | 207 | 251 | 192 | 195 | 196 | 146 | 195 | 210 | 211 | 198 | 156 | 173 | 171 | 194 |
| Events Starts per Day | 65 | 57 | 42 | 49 | 54 | 49 | 52 | 54 | 56 | 48 | 44 | 44 | 42 | 51 |
| Event Volume (gallons) | 3.2 | 4.4 | 4.6 | 4.0 | 3.6 | 3.0 | 3.8 | 3.9 | 3.8 | 4.1 | 3.6 | 3.9 | 4.0 | 3.8 |
| Event Intensity (gpm) | 1.0 | 1.3 | 1.5 | 1.4 | 1.2 | 1.1 | 1.3 | 1.3 | 1.2 | 1.3 | 1.2 | 1.2 | 1.3 | 1.3 |
| Event Duration (minutes) | 3.1 | 3.5 | 3.0 | 2.9 | 3.0 | 2.6 | 2.8 | 3.0 | 3.1 | 3.1 | 2.9 | 3.3 | 3.1 | 3.0 |

Table 3-4. Continued

| Summary of Data for Each Home | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Days of Record | 29 | 31 | 30 | 31 | 31 | 30 | 31 | 30 | 31 | 31 | 28 | 9 | 23 | 365 |
| Data Points | 41,760 | 44,640 | 43,200 | 44,640 | 44,640 | 43,200 | 44,640 | 43,200 | 44,640 | 44,640 | 40,320 | 12,960 | 33,120 | 525,600 |
| House 3 Indoor | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Average |
| Percent of Data with Water Use | 31% | 49% | 31% | 25% | 25% | 14% | 15% | 20% | 10% | 13% | 19% | 24% | 14% | 22% |
| Water Use per Day (gallons) | 253 | 283 | 391 | 281 | 312 | 225 | 224 | 325 | 170 | 228 | 255 | 241 | 245 | 266 |
| Events Starts per Day | 259 | 399 | 218 | 206 | 195 | 68 | 77 | 80 | 49 | 63 | 64 | 63 | 71 | 146 |
| Event Volume (gallons) | 1.0 | 0.7 | 1.8 | 1.4 | 1.6 | 3.3 | 2.9 | 4.0 | 3.5 | 3.6 | 4.1 | 3.3 | 3.5 | 1.8 |
| Event Intensity (gpm) | 0.6 | 0.4 | 0.9 | 0.8 | 0.9 | 1.2 | 1.1 | 1.1 | 1.2 | 1.2 | 0.8 | 1.1 | 1.2 | 0.8 |
| Event Duration (minutes) | 1.7 | 1.8 | 2.1 | 1.7 | 1.8 | 2.9 | 2.7 | 3.7 | 2.9 | 3.0 | 5.0 | 2.9 | 2.9 | 2.2 |
| House 1 Outdoor | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Average |
| Percent of Data with Water Use | 99% | 92% | 76% | 70% | 100% | 100% | 100% | 100% | 81% | 100% | 100% | 100% | 100% | 93% |
| Water Use per Day (gallons) | 458 | 586 | 562 | 144 | 65 | 220 | 929 | 677 | 130 | 209 | 173 | 152 | 433 | 376 |
| Events Starts per Day | 9 | 80 | 282 | 291 | 6 | | | | 0.35 | 0.03 | | | | 56 |
| Event Volume (gallons) | 53.1 | 7.3 | 2.0 | 0.5 | 307.8 | | | | 1.9 | 25,899 | | | | 6.7 |
| Event Intensity (gpm) | 0.3 | 0.4 | 0.5 | 0.1 | 0.3 | | | | 1.3 | 0.2 | | | | 0.3 |
| Event Duration (minutes) | 166.1 | 16.4 | 3.9 | 3.5 | 954.8 | | | | 1.5 | 160,464 | | | | 23.9 |

Table 3-4. Continued

| Summary of Data for Each Home | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Days of Record | 29 | 31 | 30 | 31 | 31 | 30 | 31 | 30 | 31 | 31 | 28 | 9 | 23 | 365 |
| Data Points | 41,760 | 44,640 | 43,200 | 44,640 | 44,640 | 43,200 | 44,640 | 43,200 | 44,640 | 44,640 | 40,320 | 12,960 | 33,120 | 525,600 |
| House 2 Outdoor | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Average |
| Percent of Data with Water Use | 12% | 10% | 8% | 6% | 7% | 6% | 10% | 6% | 5% | 4% | 5% | 7% | 5% | 7% |
| Water Use per Day (gallons) | 2,763 | 2,189 | 1,750 | 1,207 | 1,550 | 1,154 | 2,062 | 1,197 | 915 | 754 | 770 | 1,272 | 900 | 1,441 |
| Events Starts per Day | 20 | 18 | 15 | 14 | 12 | 15 | 23 | 23 | 19 | 19 | 22 | 26 | 25 | 19 |
| Event Volume (gallons) | 146.0 | 115.5 | 120.4 | 87.8 | 142.9 | 69.4 | 89.0 | 51.2 | 48.4 | 39.7 | 35.0 | 48.1 | 35.6 | 77.0 |
| Event Intensity (gpm) | 15.6 | 15.1 | 15.6 | 15.2 | 16.0 | 14.1 | 14.8 | 13.4 | 13.1 | 12.0 | 11.6 | 12.7 | 11.4 | 13.8 |
| Event Duration (minutes) | 9.4 | 7.7 | 7.7 | 5.8 | 9.0 | 4.9 | 6.0 | 3.8 | 3.7 | 3.3 | 3.0 | 3.8 | 3.1 | 5.4 |
| House 3 Outdoor | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Average |
| Percent of Data with Water Use | 2% | 1% | 1% | 1% | 1% | 1% | 2% | 2% | 1% | 2% | 2% | 1% | 2% | 1% |
| Water Use per Day (gallons) | 494 | 374 | 264 | 220 | 376 | 306 | 436 | 438 | 367 | 480 | 533 | 228 | 487 | 391 |
| Events Starts per Day | 0.7 | 0.4 | 0.6 | 0.9 | 1.1 | 1.3 | 2.6 | 1.6 | 1.0 | 0.9 | 0.9 | 0.6 | 0.7 | 1.1 |
| Event Volume (gallons) | 716.2 | 892.8 | 416.3 | 234.8 | 353.2 | 229.8 | 166.9 | 268.2 | 379.2 | 550.9 | 574.2 | 409.8 | 700.6 | 367.5 |
| Event Intensity (gpm) | 19.3 | 19.4 | 19.2 | 18.1 | 18.9 | 18.4 | 17.9 | 18.3 | 19.0 | 18.9 | 18.9 | 19.5 | 19.1 | 18.6 |
| Event Duration (minutes) | 37.2 | 46.0 | 21.6 | 13.0 | 18.7 | 12.5 | 9.3 | 14.6 | 20.0 | 29.2 | 30.3 | 21.0 | 36.6 | 19.6 |

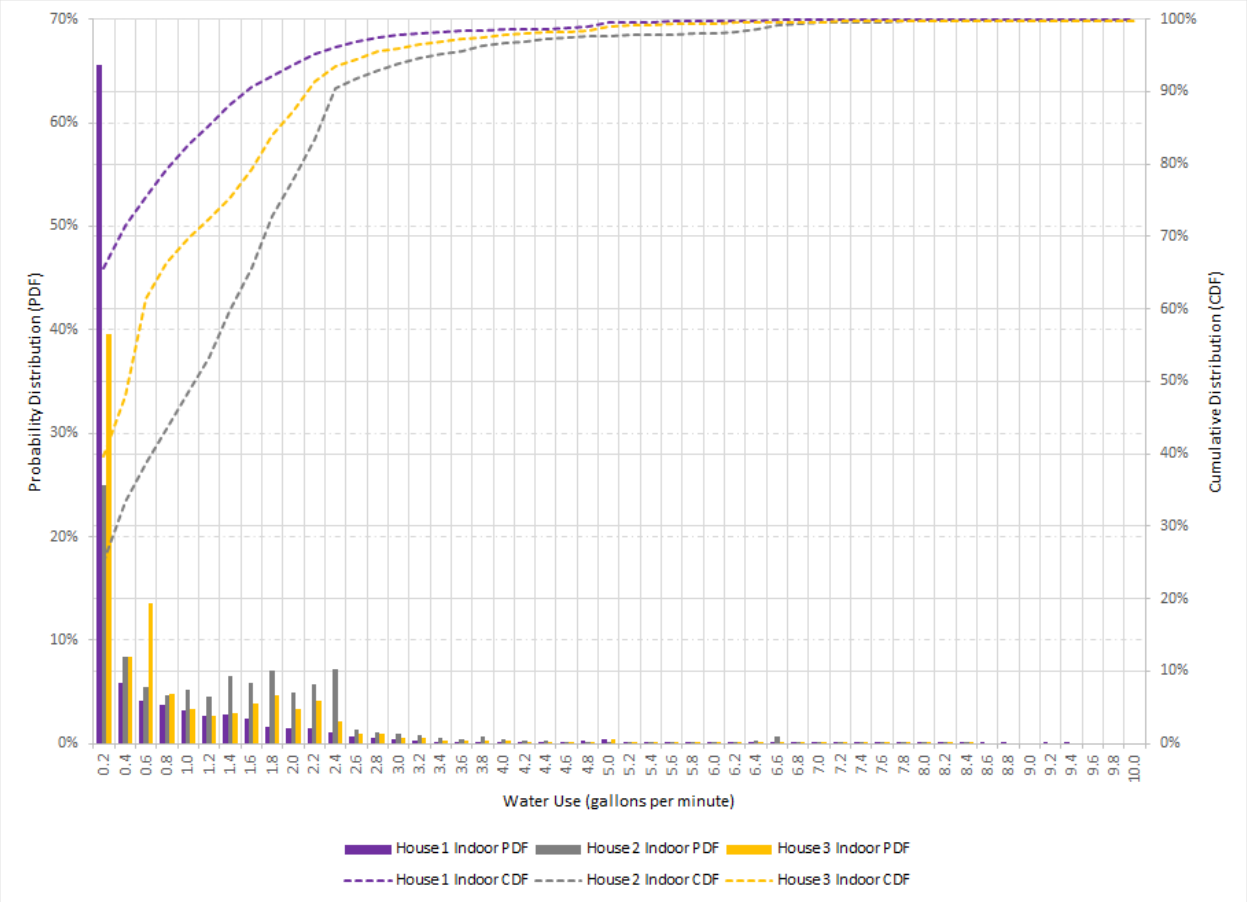


Figure 3-5. Distributions of indoor data points where water use is greater than zero

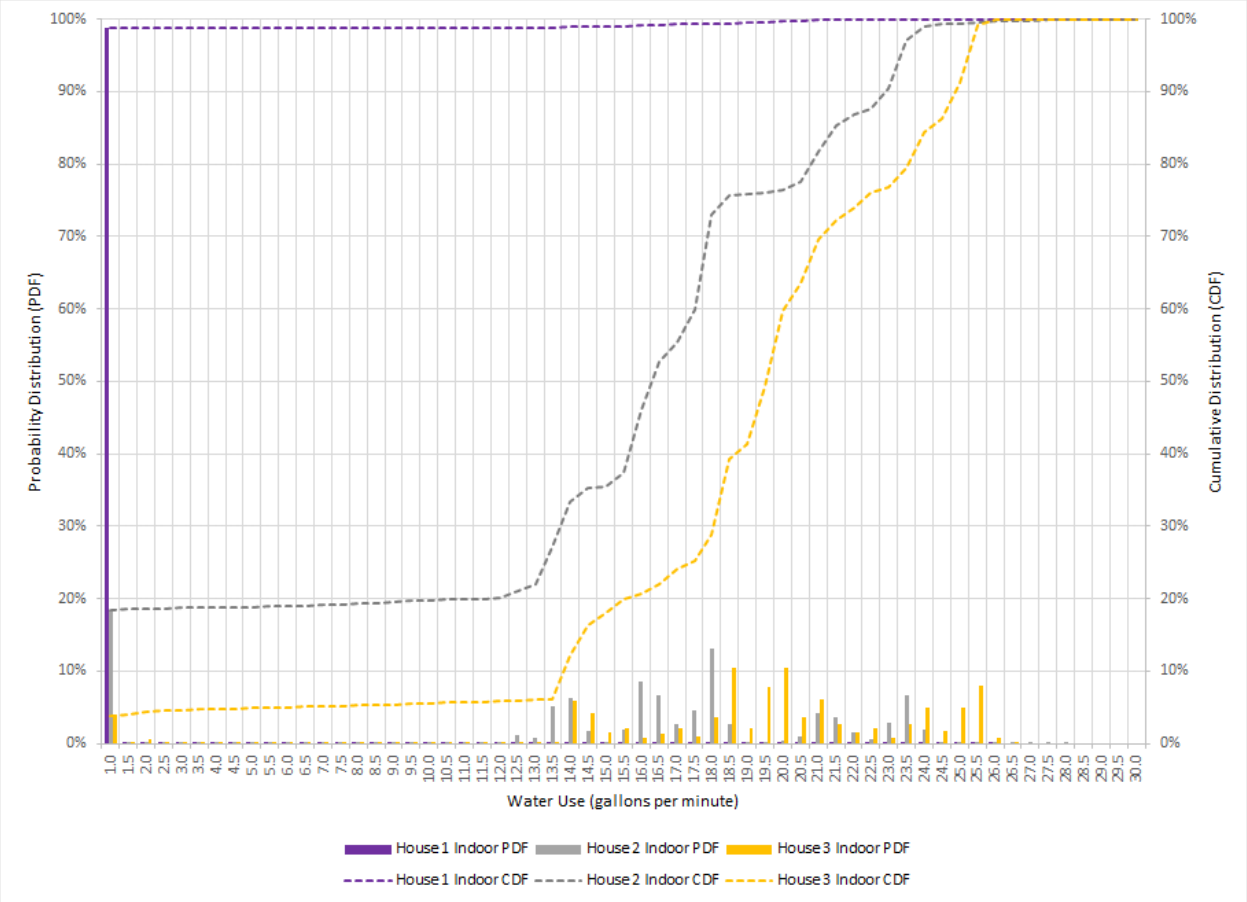


Figure 3-6. Distributions for outdoor data points where water use is greater than zero

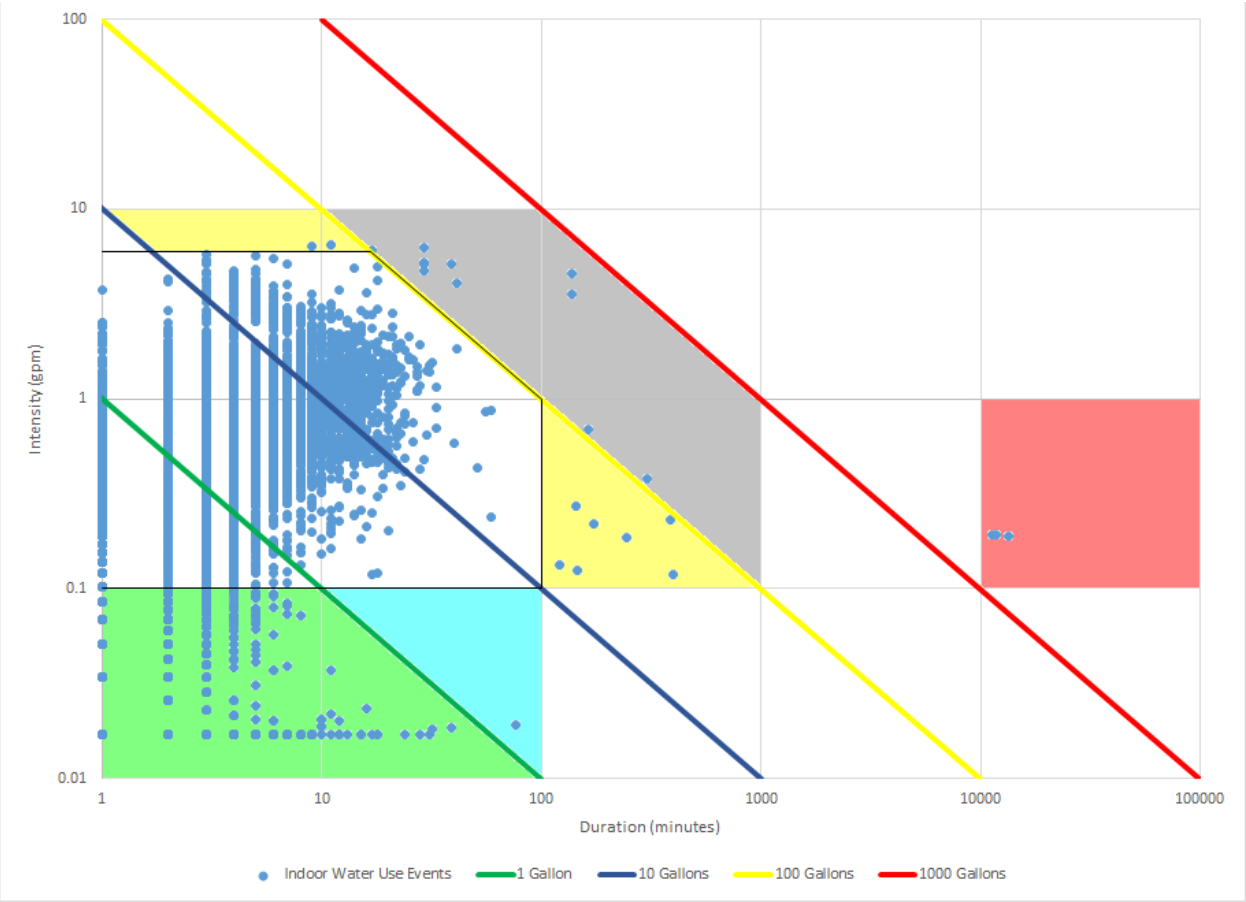


Figure 3-7. Aggregate events with anticipated event ranges for House 1 indoor water use

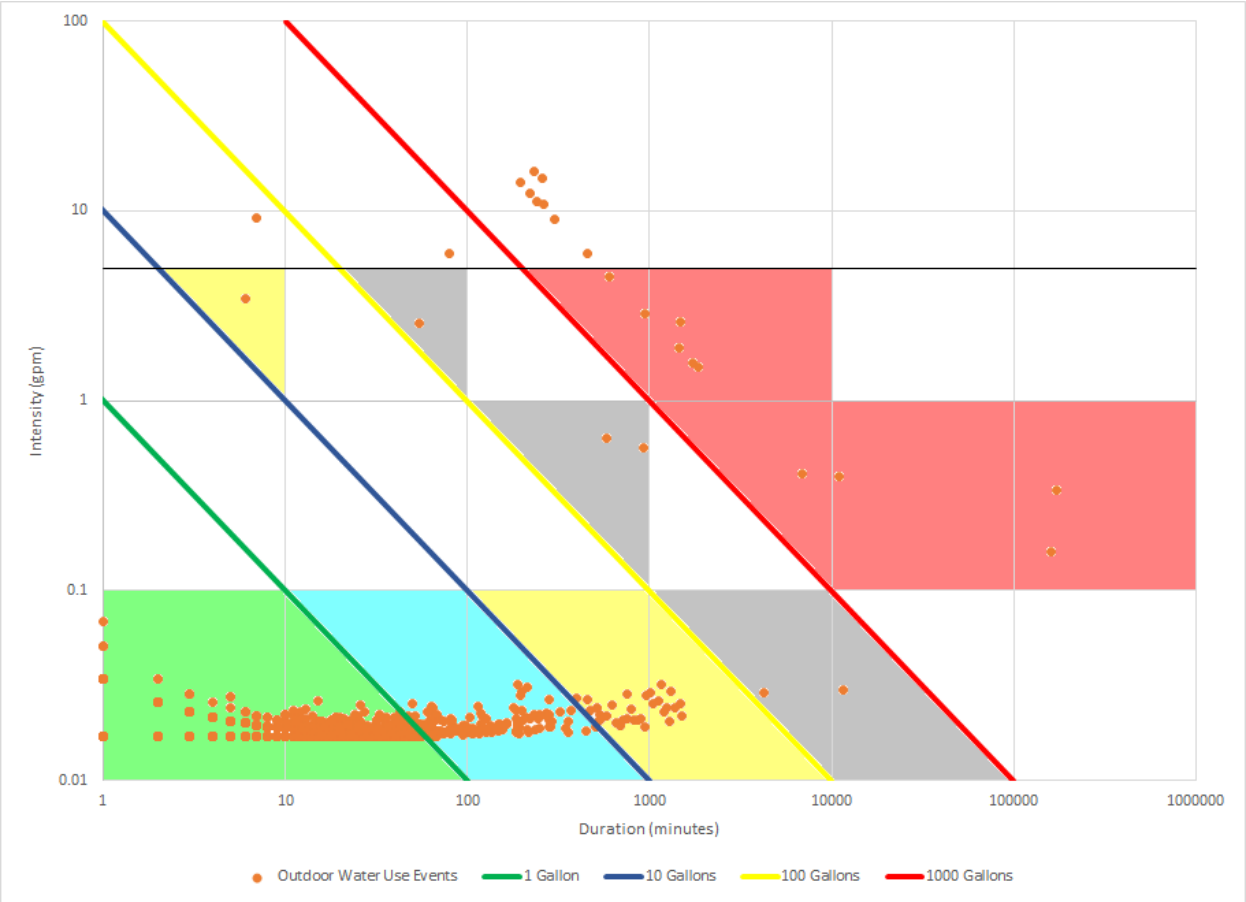


Figure 3-8. Aggregate events with anticipated event ranges for House 1 outdoor water use

Table 3-5. Unanticipated aggregate events summarized within defined volumetric ranges

| Volumetric Ranges | | House 1 | | House 2 | | House 3 | |
|---|---|---|---|---|---|---|---|
| Units in Gallons | | Indoor | Outdoor | Indoor | Outdoor | Indoor | Outdoor |
| V<1 | Total Events | 22,965 | 20,218 | 5,956 | 6,575 | 23,578 | 278 |
| | Total Volume (gal) | 607 | 1,112 | 242 | 116 | 590 | 5 |
| | Percent of Volume Within: | --------- | --------- | -------- | --------- | -------- | --------- |
| | Anticipated Intensity Ranges | 9.5% | 0.0% | 10.9% | 0.0% | 9.5% | 0.0% |
| | Unanticipated Intensity Ranges | 90.5% | 100.0% | 89.1% | 100.0% | 90.5% | 100.0% |
| 1<=V<10 | Total Events | 1 | 139 | 2 | 1 | 2 | 0 |
| | Total Volume (gal) | 1 | 408 | 4 | 10 | 5 | 0 |
| | Percent of Volume Within: | --------- | --------- | -------- | --------- | -------- | --------- |
| | Anticipated Intensity Ranges | 0.0% | 0.0% | 64.8% | 80.6% | 55.2% | N/A |
| | Unanticipated Intensity Ranges | 100.0% | 100.0% | 35.2% | 19.4% | 44.8% | N/A |
| 10<=V<100 | Total Events | 9 | 33 | 16 | 0 | 16 | 0 |
| | Total Volume (gal) | 422 | 690 | 685 | 0 | 448 | 0 |
| | Percent of Volume Within: | --------- | --------- | -------- | --------- | -------- | --------- |
| | Anticipated Intensity Ranges | 63.1% | 3.0% | 28.0% | N/A | 45.1% | N/A |
| | Unanticipated Intensity Ranges | 36.9% | 97.0% | 72.0% | N/A | 54.9% | N/A |
| 100<=V<1000 | Total Events | 11 | 5 | 36 | 0 | 21 | 0 |
| | Total Volume (gal) | 2,429 | 1,503 | 9,339 | 0 | 5,347 | 0 |
| | Percent of Volume Within: | --------- | --------- | -------- | --------- | -------- | --------- |
| | Anticipated Intensity Ranges | 58.0% | 66.0% | 34.9% | N/A | 65.2% | N/A |
| | Unanticipated Intensity Ranges | 42.0% | 34.0% | 65.1% | N/A | 34.8% | N/A |
| V>=1000 | Total Events | 3 | 10 | 0 | 0 | 1 | 0 |
| | Total Volume (gal) | 6,963 | 108,861 | 0 | 0 | 3,051 | 0 |
| | Percent of Volume Within: | --------- | --------- | -------- | --------- | -------- | --------- |
| | Anticipated Intensity Ranges | 77.0% | 68.7% | N/A | N/A | 99.7% | N/A |
| | Unanticipated Intensity Ranges | 23.0% | 31.3% | N/A | N/A | 0.3% | N/A |

# CHAPTER 4
# USE OF AUTOMATIC METER READING DATA FOR RAPID EVENT DETECTION AND LONG-TERM LEAKAGE QUANTIFICATION IN A DISTRICT METERING AREA

## Scope and Overview

This study presents the results of a prototype high-frequency water use evaluation using one-minute and five-minute data collected from meters at 194 single family homes in Hillsborough County, Florida over a period of 2 years. Of the 194 homes, 191 are located in a single District Metering Area (DMA) that is hydraulically separate from the rest of the network. Automatic meter reading (AMR) meter registers are used with short-range wireless communication that allow for ease of data collection by driving by and downloading the data from the meter registers. The purpose of this study is to evaluate the data collected from the AMR registers to see if installing advanced metering infrastructure (AMI) "smart meters" would provide cost and water savings to customers if the smart meters can detect and notify customers of unwanted leakage events.

Recent advancements in the utility industry have made high-frequency water use data more readily available as the use of "AMR", "AMI", and "Smart Meters" become more prevalent. Since these terms have been used referring to a broad range of applications, a definition is presented for each that defines their capabilities for use in the current research. AMR allows local storage of data whereby a human activity is necessary to download the data. This typically involves driving by and downloading data with short-range wireless equipment. AMI allows this data transfer to occur without human intervention using telemetry systems where the local data can be transmitted to a centralized data storage system. Smart meter systems go beyond the transmittal of data and involve some level of analytics, either at the local meter itself or at the centralized operational system.

Historically, most utilities read the customer's meter at monthly or longer intervals. AMR and AMI are making it possible to have high-frequency (1 second to 1 day) meter reads for every customer in the water system. AMI is allowing communication between the meters and operational systems that can store and use these high-frequency reads for decision support services. This transition from monthly to high-frequency water use data allows operating decisions to be made with near real-time demand analysis. However, serious consideration needs to be given to the value added by such data and systems. Analyses need to be performed to determine the potential savings of installing such systems prior to utilities making major investments to upgrade telemetry networks, decision support infrastructure, and customer meters.

When evaluating smart meter systems, several key questions need to be analyzed from the perspective of the utility as described next. Does the utility need to know individual water use habits to improve system design/operation, or does the utility only need to know the aggregate effect of many customers on large areas of the system? If the utility knew what every customer was using at every instant, would the utility do anything different? If the utility doesn't make operational decisions based on individual customer's real-time use, then is there any savings potential for the utility to get updates of individual customer use in real-time? From the utility perspective, real-time access to data may be more important at larger spatial scales where the data summarizes impacts to many customers.

The above questions are traditional utility-centric considerations. However, when considering the potential savings to the customer, the installation of smart meter systems becomes more attractive. From the customer perspective, the value of real-time or near real-time updates is more important than the overall evaluation of the entire data set, including real-time notification of potential leakage events. Considering this dual-savings approach, Figure 4-1

shows a summary of the potential areas for cost and water savings to both the utility and the customers. A subset of these topics is discussed in this paper, and based on these topics, a question is posed: if the net cost of the AMI installation to the utility is passed on to the customer, can the cost savings provided by the AMI installation result in at least cost neutrality for the customer?

## Savings Topics Discussed for Cost Analysis

### Continuous Leak Detection (Low Intensity, Long Duration) and Conservation

Because normal residential water use is intermittent, it is easy to identify continuous leaks as they will show up as a continuous flow. Cardell-Oliver (2013) indicated that alarms were set to notify the utility for continuous customer use at a utility in Kalgoorlie-Boulder, Australia. These alarms were based on data collected at 1-hour intervals, and the alarms trigger interaction with the residents from the utility as appropriate for the amount of the flow. For high flow rates, the residents can be contacted immediately by telephone. Medium flows may trigger a letter and the least significant flows may simply receive advice in the regular water bill. The next step in the evolution of utility/customer interaction is for customer notification to come directly from the smart meter, with multiple utilities unveiling systems including smart meter analytics (Anderson 2015).

Beyond leak detection only, additional research has focused on the “self-awareness” factor, i.e. that water use awareness brings customer-initiated conservation. This “self-awareness” is noted by Davies et al. (2014) who investigated the impact of smart meters on reducing residential water use in the long term. A key finding was that households with an in-home display that could be used to track water usage reduced their usage by an average of over 6.8% when compared with the control group that did not have an in-home display. The “self-awareness” factor was also used to support the long-term conservation goal of Albuquerque

Bernalillo County Water Utility Authority in New Mexico as indicated by Daigle and Jackson (2013), who described the implementation of AMI, meter data management, and customer engagement software that put the power in the hands of the consumers. It was used to identify leaks and also allowed customers to view their consumption patterns on a near real-time basis; customize and receive usage reports via e-mail, text, or phone; create personal conservation goals and water budgets; and download targeted educational material regarding conservation.

**Customer Pipe Break Detection (High Intensity, Short Duration) and Insurance Damages**

While pipe breaks have been evaluated at the distribution system level, there is potential to provide significant savings to customers if pipe breaks in residential plumbing can be detected and the customer notified prior to significant damage. Expected event ranges must be defined so that rapid notification can occur through "report by exception", where the flow data is monitored at the local device level and reporting only takes place if there is an exception to expected data. For this to be successful at the individual customer level, the event must be detected and notification provided as quickly as possible.

Approximately 25% of insurance claims are the result of water damage (see Table 4-1), with claims from faulty plumbing averaging over $17,000 per claim (see Table 4-2). If pipe break events can be detected as discussed in the previous topic, and smart meters can provide notification to customers and automatic shutoff valves, then the damage from these pipe break events can be minimized. The use of automatic shutoff valves in homes has become more prevalent in recent years; however, they are typically linked to sensors in the home that have to detect the presence of water (e.g., a sensor in a laundry room that detects water on the floor). The use of automatic shutoff valves can be coupled with smart meters if the data can be analyzed at the local level and the exception to the expected demand can be detected. This requires an understanding of expected demand obtained through the analysis of high-frequency databases.

### Utility Staffing for Meter Reading, Inspections, and Code Enforcement

Initial focus on AMR/AMI systems was on reducing the staffing needed for meter reading. For AMR systems, this would involve driving by and downloading the data using short range radio communication as opposed to manually reading each meter. For AMI systems, this would involve the data being automatically uploaded to a central database system used for billing. The Kansas City, Missouri, Water Services Department was able to eliminate 33 meter reading positions and use daily AMI reading to reduce meter re-reads and leakage inspections by 90% as well as reduce meter shut offs by instead monitoring and billing vacant home use (Thiemann et al., 2011). In addition, the customers could view their own water use via website with future plans to allow customers to receive automatic notifications of high consumption via e-mail or phone call. Daigle and Jackson (2013) noted the benefit of the utility being able to detect irrigation events for code enforcement purposes, and this could eliminate the need for an employee to drive to multiple locations to inspect irrigation behavior when it can be detected by a smart meter.

## Cost Framework for Study Areas

Prior to evaluating the cost and water savings potential, a basic framework needed to be established to compare savings to costs. The cost framework is based on actual costs for Hillsborough County Public Utilities Department, located in Hillsborough County, Florida near Tampa Bay. The cost for each AMI data logger with smart-meter capability is $250, which covers the data storage and reporting to both customers and utilities for 10 years with data accessible at 5-minute intervals. While the current research is not focused on “interface screens” or “dashboards” available through webpage and smartphone applications, the research will focus on how the data from the smart meters can be used to detect unwanted events and be used to notify customers through these applications.

For the cost comparison, the $250 is assumed to take the place of any meter-reading cost for 10 years. The AMI data loggers replace the analog registers on the meters. However, no internal mechanical components of the meter are replaced or impacted in any way. As such, the addition of the AMI data loggers doesn't impact the normal replacement schedule for the meters, so no additional costs or savings are included with the addition of the AMI data loggers. Table 4-3 shows how this cost breaks down from the 10-year total to annual, monthly, and daily costs. For comparison, actual costs for the utility per meter read range from $0.56 to $0.99. The low end of these costs is for contract meter-reading with no other services provided. The high end of these costs includes overhead and other services by utility workers, like reporting and fixing anomalies in the field. The normal meter-read frequency based on standard meters is once per month. Table 4-3 shows the costs and differences when comparing the range of standard meter read costs to the AMI costs. The resulting range of cost differences shown in Table 4-3 is what is being proposed to be passed on to the customer to result in cost neutrality for the utility. While the utility could realize other potential savings which would reduce these differences, those are not being discussed in the current research so no additional savings are being included. Assuming that the "smart meters" could be used to detect pipe break events and notify the customer in order to prevent or reduce damage, thereby reducing the risk for significant property damage, there is potential for insurance companies to incentivize the use of these "smart meters". Insurance policies are typically written on an annual basis, so the required annual premium reduction would need to range from $13 to $18 in order to result in cost neutrality for the customer without any other savings considerations.

Aside from pipe break or leak detection, the customer can realize other potential savings like conservation through self-awareness as discussed above and more specifically through

fixture leak detection. Hillsborough County uses a conservation block structure for water rates which is shown in Table 4-4. Assuming that no savings are realized through the insurance premium reduction, Table 4-3 shows the resulting water savings that would be required in order to result in cost neutrality for the customer. In order to show a high and low end for the range, it was assumed that the highest cost difference for meter read options was applied to a customer with water use in the lowest range, thereby paying the lowest block rate. Comparatively, the lowest cost difference was used assuming the savings would occur in the highest block rate. The resulting water savings required in order for the customer to result in cost neutrality ranges from 5 to 14 gallons per household per day (gphd). A key question is if leakage quantities are in this range so that leakage reduction can result in cost neutrality for the customer. A recent nationwide study (DeOreo et al. 2016) built upon an earlier nationwide study (Mayer et al. 1999) showed that average leakage was 17 gphd, so there is data to support the potential for these savings. The following case study builds upon work completed in Chapter 3 and shows where fixture leaks can be easily detected and quantified, and where pipe breaks could be easily and quickly detected.

## Case Study and Comparison with Previous Studies

A pilot study for AMR data collection and analysis began in June 2013 for Hillsborough County Public Utilities Department. The pilot included 194 single-family homes, of which 191 were located in one hydraulically connected neighborhood; two master-metered multi-family communities; one “big box” retail store; and one small hospital facility. This study focuses on the evaluation of the 191 single-family homes along with a comparison to previous studies and the 3 homes evaluated in Chapter 3. The data were collected at either 1-minute or 5-minute recording intervals, and while the period of record was different for each home, each home had at least one year of data in the range of June 2013 to August 2015. Similar to the AMI data

loggers noted previously, the AMR data loggers used in this study only replaced the analog registers on the meters. No internal mechanical components of the meter itself were replaced, and the resolution of the gallons reported by the AMR data loggers was as accurate as the registering capability of the mechanical components of the meter. The internal mechanical components of the meters used nutating discs capable of reading in increments of 0.017 gallons. The local data storage on the AMR was limited to 32,000 data points. For the data collected in this study, a data file had to be collected by driving to each meter and downloading the data from the loggers through short-range wireless communication. The vehicle was equipped with a radio that communicated with a local radio transmitter on each of the data loggers. Each data file took approximately five minutes to download. A database was built that allowed each data file to be uploaded to the appropriate dataset for each meter. The resulting database allowed easy access to water use data by time of day, day of week, and any combination of these two.

**DMA Study Area**

A single family residential (SFR) neighborhood of predominantly indoor-use-only customers was selected as a study area in order to collect a large dataset within an isolated district metered area (DMA). This study area of 191 SFRs was selected to perform hydraulic analyses that are the subject of a future study. There were 166 SFRs programmed with a 5-minute recording interval, and at this interval, the data must be downloaded every 111 days in order to avoid gaps in the data. The other 25 homes were programmed with a 1-minute recording interval with the data needing to be downloaded every 22 days. An aerial map of the pilot area is shown in Figure 4-2. The blue parcels indicate the 166 SFRs with 5-minute recorded intervals, and the orange parcels indicate the 25 SFRs with 1-minute recorded intervals. Table 4-5 shows a summary of housing and water-use statistics. American Community Survey (ACS) data for 2015 were used to estimate the persons per household (pph) for the

neighborhood. According to ACS data, the rolling 5-year average of pph for the Census Tract that includes the study area is 3.43. Assuming that the 3.43 pph is an appropriate average for the 191 SFRs, the resulting gallons per capita per day (gpcd) is shown in Table 4-5. These SFRs were primarily built in the late 1970s before water use efficiencies were improved. The longest lived indoor appliances are toilets with an average service life of 35-40 years. Using an average year built of 1980, then the average house would be 35 years old in 2015 and would be expected to have replaced the original fixtures.

### Comparison with Previous High-Frequency Studies

The DMA used for the study area provides for a larger test area than what was presented in Chapter 3, as well as a lower per capita water use that spans a range across previous research studies. Table 4-6 shows how this study compares to previous studies as well as what was presented in Chapter 3.

## Evaluation of Water Use Data and Event Outliers at Different Time Steps

The framework for identifying unexpected events for the purpose of rapid pipe break detection and overall leak quantification was developed in Chapter 3. The previous researchers noted in Table 4-6 focused on other research areas, while the current research described in Chapters 3 and 4 explicitly looks at the identification of leaks and pipe breaks. A reduced dataset was used to limit the evaluation to one year in order to evaluate continuous data and report based on annual statistics. Over 20 million data points were collected for the 166 homes with 5-minute data, and over 13 million data points were used for the final dataset with 128 homes that had continuous water use data for a one year period. Likewise, over 17 million data points were collected for the 25 homes with 1-minute data, and while these data are summarized in Table 4-5, the high-frequency data for these 25 homes weren’t used for the leak evaluation discussed in this chapter. The 128-home dataset was compared with the 3 homes in Chapter 3 to

develop statistics on a per home basis at different levels of temporal aggregation of the data. The different levels of aggregation allow for a comparison between the detection capabilities of increasing time steps from 1 minute to 1 hour. Tables 4-7 and 4-8 show the monthly statistics for the two areas on a per home basis, and Tables 4-9 and 4-10 show the statistics for aggregate event outliers on a per home basis. These tables were created following the process outlined in Chapter 3.

**Potential Cost Savings from Mitigating Event Outliers**

From looking at only the conservation perspective, Table 4-3 indicates that an annual water savings of 1,697 – 5,050 gallons per home is required to result in cost neutrality for the customer. This could be achieved by preventing only the larger events greater than 1,000 gallons. However, aside from the conservation perspective, the cost of damage prevention could be the most attractive benefit. If these large events are internal pipe or fixture breaks within the home, being able to mitigate these events as a result of early detection could more than offset the cost. As an example, Table 4-3 indicates that an annual cost savings of $13 - $18 per home is required to result in cost neutrality for the customer. Table 4-1 indicates that there are approximately 1.79 water damage claims per 100 homes, resulting in approximately 2.29 claims per year in the 128-home subset in the DMA used for this study. Table 4-2 indicates that the lowest cost of claims caused by leaks averages $3,642 for damage from internal water heater leaks. If only one of these average events could be detected and prevented in the 128-home study area, the average cost savings per home would be $28. From reviewing the 5-minute data in Table 4-9, there are 2 events per home greater than 1,000 gallons with an average event volume of 13,900 gallons. If only 1 of these events for 1 home was an internal fixture or pipe break event, and it was prevented from the use of a smart meter, the average cost savings per home would cover the cost of the smart meter installation.

## Synopsis

The current study builds upon the earlier evaluation from Chapter 3 wherein aggregate event outliers are quantified based on volumetric ranges. The results show that as the time step increases, there is an overall decrease in the number of events which is intuitive as the larger time steps capture many smaller events within a single larger event. Likewise, the larger time steps result in an increase in the number of unanticipated events although the extreme events (greater than 1,000 gallons) are only slightly more prevalent. While the smaller time steps capture many more of the smaller events, these are not significant in terms of overall volumetric contribution. The current study makes a case for a framework wherein smart meter systems can directly benefit customers by detecting these larger events. This should be evaluated in future smart system evaluations instead of using the traditional benefit analysis for utility savings only.

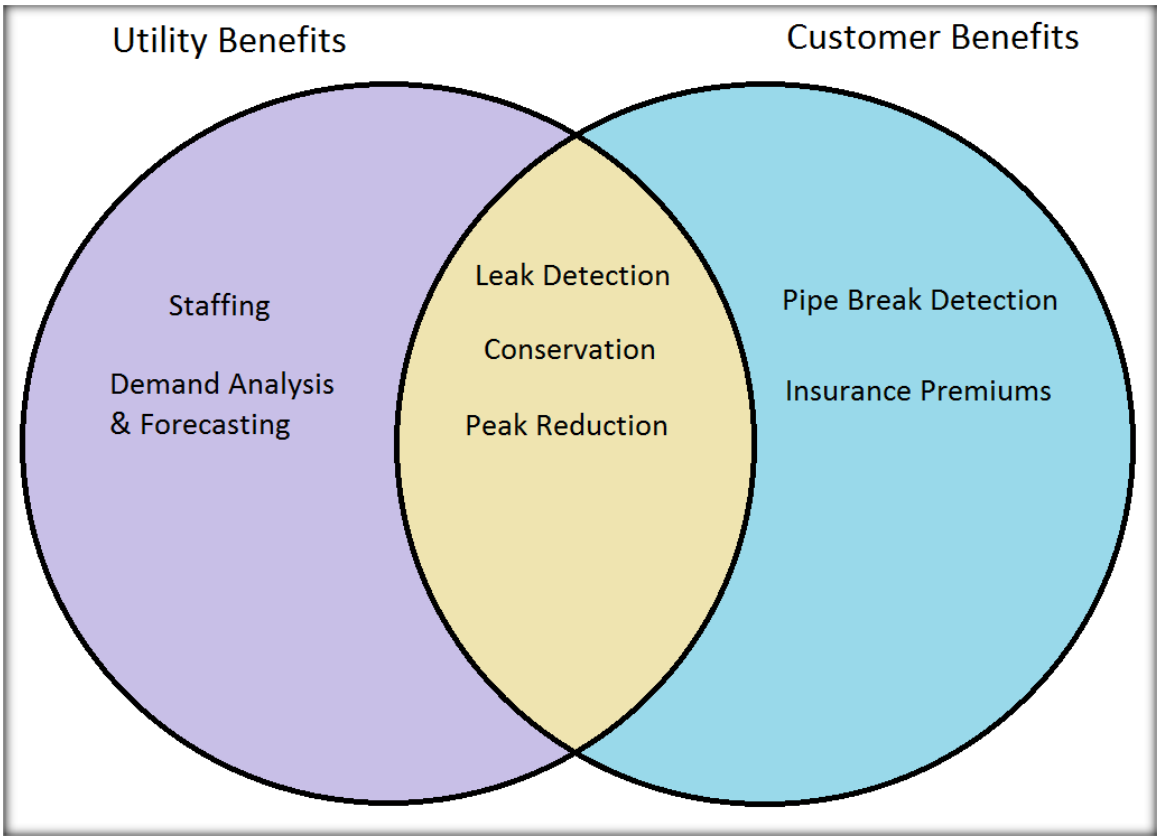


Figure 4-1. Potential savings of residential smart metering for utilities and customers

Table 4-1. Insurance claims by type of damage events

| Type of Event | Annual Claims per 100 Houses | Claim Frequency per House in Years | Percent of Total |
|---|---|---|---|
| Wind and Hail | 3.37 | 29.7 | 47.1% |
| Water Damage and Freezing | 1.79 | 55.9 | 25.0% |
| Other Property Damage | 1.04 | 96.2 | 14.5% |
| Theft | 0.52 | 192.3 | 7.3% |
| Fire, Lightning, and Debris | 0.43 | 232.6 | 6.0% |
| Total | 7.15 | 14 | 100.0% |

Source: Insurance Services Office as reported by www.valuepenguin.com/average-cost-of-homeowners-insurance

Table 4-2. Repair costs for different types of water damage

| Cause of Leak | Average 2013 Repair Cost |
|---|---|
| Water Heaters – Internal Leaks | $3,642 |
| Water Heaters – Valve Failures | $4,218 |
| Washing Machine Failures – Occupied Homes | $4,959 |
| Water Heaters – Supply Line Failure | $5,825 |
| Flooded House – 1 to 4 Inches of Water[+] | $7,800 |
| Frozen Pipe Related Failures | $8,189 |
| Bathroom Fixtures | $10,799 |
| Washing Machine Failures – Unoccupied Homes | $12,308 |
| Appliance Leaks – Overall | $13,467 |
| Faulty Plumbing | $17,250 |

[+]Water may be from leak or flooding

Source: www.waterdamagedefense.com/pages/water-damage-by-the-numbers

Table 4-3. Comparison of AMI to standard meter reading costs per single family residential customer for Hillsborough County Public Utilities Department

| Costs | 10-Year Total | Per Year | Per Month | Per Day |
|---|---|---|---|---|
| AMI Installation Cost | $250.00 | $25.00 | $2.08 | $0.07 |
| Meter Read Cost, Option 1 | $67.20 | $6.72 | $0.56 | $0.02 |
| Meter Read Cost, Option 2 | $118.80 | $11.88 | $0.99 | $0.03 |
| Cost Difference Option 1 | $182.80 | $18.28 | $1.52 | $0.05 |
| Cost Difference Option 2 | $131.20 | $13.12 | $1.09 | $0.04 |
| **Water Savings** | **10-Year Total** | **Per Year** | **Per Month** | **Per Day** |
| Option 1, Block 1 | 50,497 | 5,050 | 421 | 14 |
| Option 2, Block 4 | 16,973 | 1,697 | 141 | 5 |

Table 4-4. Monthly conservation block rate for Hillsborough County Public Utilities for 2016

| Block | Gallons per Month | Rate per 1,000 Gallons |
|---|---|---|
| 1 | 0 to 5,000 | $3.62 |
| 2 | 5,001 to 15,000 | $4.85 |
| 3 | 15,001 to 30,000 | $6.14 |
| 4 | 30,001 and higher | $7.73 |

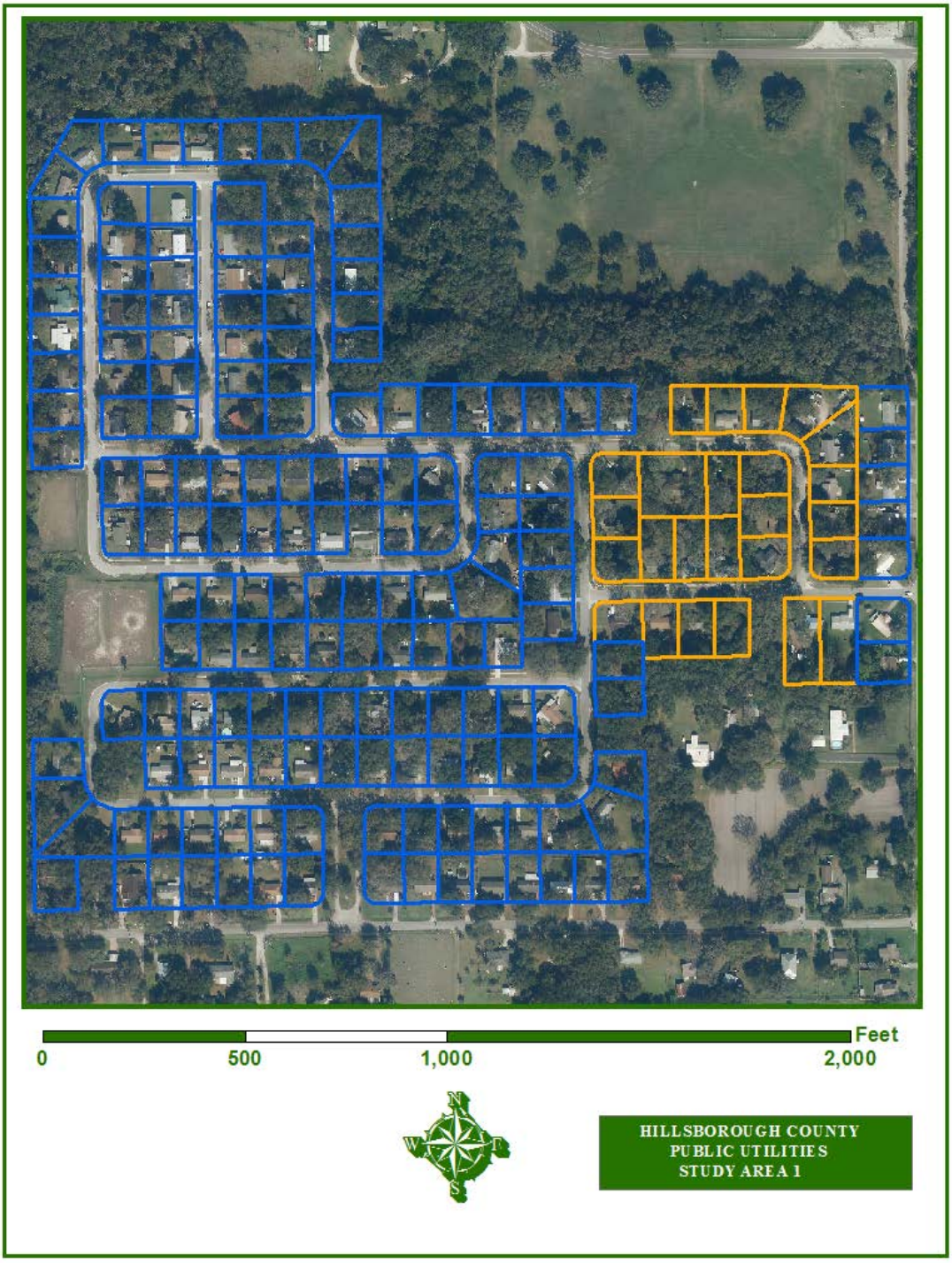


Figure 4-2. Aerial view of 191 Single Family Residential Parcels within Study Area 2

Table 4-5. Housing statistics for the 191 homes within the DMA for Study Area 2

| Housing Information | 25 Homes 1-Minute Recording Intervals From October 2013 to June 2015 | | | 166 Homes 5-Minute Recording Intervals From June 2013 to August 2014 | | |
|---|---|---|---|---|---|---|
| | Minimum | Average | Maximum | Minimum | Average | Maximum |
| Year Built | 1976 | 1979 | 1995 | 1974 | 1978 | 1982 |
| Heated Area (sq. ft.) | 1,092 | 1,367 | 1,886 | 968 | 1,254 | 2,458 |
| Lot Area (sq. ft.) | 10,512 | 13,466 | 26,416 | 9,156 | 11,133 | 20,687 |
| Market Value | $51,841 | $70,795 | $107,673 | $48,894 | $62,368 | $112,506 |
| Annual Average Use (gpd) | 16 | 187 | 542 | 2 | 159 | 889 |
| Per Capita Use (gpcd) | --- | 54 | --- | --- | 46 | --- |

Table 4-6. High-frequency water use studies on single-family residences

| Study | Location | Interval | Homes | Days | Indoor Water Use (gpcd) | Purpose |
|---|---|---|---|---|---|---|
| Buchberger and Wells, 1996 | Cincinnati, Ohio | 1 sec | 4 | 273 to 365 | 58.5+ | Demand Simulation for Modeling |
| DeOreo et al., 1996 | Boulder, Colorado | 10 sec | 16 | 21 | 58.8 | Fixture Level Water Balance |
| Mayer et al., 1999 | 12 Cities in US and Canada | 10 sec | 1,188 | 28 | 69.3 | Fixture Level Water Balance |
| Buchberger et al., 2003 | Cincinnati, Ohio | 1 sec | 21 | 252 | 55+ | Demand Simulation for Modeling |
| Blokker et al., 2010 | Amsterdam, Netherlands | 1 min | 43 | 7 | not reported | Demand Simulation for Modeling |
| DeOreo et al., 2016 | 21 Cities in US and Canada | 10 sec | 1,950 | 14 to 28 | 58.6 | Fixture Level Water Balance |
| Chapter 3 | Hillsborough County, Florida | 1 min | 3 | 400 | 71.7 | |
| --- subset of Chapter 3 data | | 1 min | 3 | 365 | | Leakage and Plumbing Breaks |
| Chapter 4 | Hillsborough County, Florida | 1 - 5 min | 194 | 401 | 47 | |
| --- subset of Chapter 4 data | | 5 min | 128 | 365 | | Leakage and Plumbing Breaks |

+Reported values exclude leaks.

Table 4-7. Summary of per home data for DMA study area

| Summary of Data for 128 Homes (per Home) | Aug-13 | Sep-13 | Oct-13 | Nov-13 | Dec-13 | Jan-14 | Feb-14 | Mar-14 | Apr-14 | May-14 | Jun-14 | Jul-14 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Days of Record | 31 | 30 | 31 | 30 | 31 | 31 | 28 | 31 | 30 | 31 | 30 | 31 | 365 |
| Water Use per Day (gallons) | 174 | 175 | 171 | 183 | 181 | 177 | 177 | 174 | 186 | 179 | 189 | 165 | 177 |
| 5-Minute Data | Aug-13 | Sep-13 | Oct-13 | Nov-13 | Dec-13 | Jan-14 | Feb-14 | Mar-14 | Apr-14 | May-14 | Jun-14 | Jul-14 | Average |
| Data Points | 8,928 | 8,640 | 8,928 | 8,640 | 8,928 | 8,928 | 8,064 | 8,928 | 8,640 | 8,928 | 8,640 | 8,928 | |
| Percent of Data with Water Use | 32% | 32% | 31% | 34% | 35% | 34% | 31% | 30% | 29% | 30% | 30% | 29% | 31% |
| Events Starts per Day | 22 | 23 | 22 | 24 | 23 | 25 | 24 | 24 | 24 | 23 | 23 | 23 | 23 |
| Event Volume (gallons) | 8.1 | 7.3 | 7.0 | 8.2 | 7.4 | 8.7 | 6.5 | 7.5 | 6.8 | 8.1 | 7.2 | 6.3 | 7.4 |
| Event Intensity (gpm) | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 | 0.4 |
| Event Duration (minutes) | 19.6 | 19.6 | 17.0 | 21.4 | 19.9 | 23.1 | 16.4 | 16.8 | 17.7 | 19.1 | 17.2 | 16.0 | 18.7 |
| 15-Minute Data | Aug-13 | Sep-13 | Oct-13 | Nov-13 | Dec-13 | Jan-14 | Feb-14 | Mar-14 | Apr-14 | May-14 | Jun-14 | Jul-14 | Average |
| Data Points | 2,976 | 2,880 | 2,976 | 2,880 | 2,976 | 2,976 | 2,688 | 2,976 | 2,880 | 2,976 | 2,880 | 2,976 | |
| Percent of Data with Water Use | 45% | 45% | 44% | 47% | 49% | 48% | 45% | 44% | 43% | 43% | 43% | 42% | 45% |
| Events Starts per Day | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 11 | 10 | 10 | 10 | 10 |
| Event Volume (gallons) | 17.0 | 18.3 | 16.6 | 18.2 | 17.4 | 21.2 | 14.6 | 19.7 | 15.6 | 17.0 | 16.1 | 13.7 | 17.1 |
| Event Intensity (gpm) | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 |
| Event Duration (minutes) | 56.5 | 65.7 | 64.1 | 70.0 | 63.9 | 75.5 | 54.5 | 67.2 | 57.7 | 56.8 | 57.2 | 51.6 | 61.7 |
| 60-Minute Data | Aug-13 | Sep-13 | Oct-13 | Nov-13 | Dec-13 | Jan-14 | Feb-14 | Mar-14 | Apr-14 | May-14 | Jun-14 | Jul-14 | Average |
| Data Points | 744 | 720 | 744 | 720 | 744 | 744 | 672 | 744 | 720 | 744 | 720 | 744 | |
| Percent of Data with Water Use | 67% | 66% | 65% | 68% | 69% | 68% | 66% | 65% | 66% | 65% | 64% | 64% | 66% |
| Events Starts per Day | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 2 | 3 |
| Event Volume (gallons) | 73.7 | 77.1 | 55.3 | 75.7 | 63.4 | 59.7 | 53.6 | 72.4 | 65.2 | 59.5 | 58.1 | 53.0 | 64.0 |
| Event Intensity (gpm) | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| Event Duration (minutes) | 345.0 | 407.2 | 311.7 | 419.8 | 376.1 | 347.2 | 310.3 | 357.8 | 352.5 | 322.6 | 314.1 | 316.9 | 348.3 |

Table 4-8. Summary of per home data for 3 homes evaluated in Chapter 3

| Summary of Data for 3 Homes (per Home) | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Days of Record | 29 | 31 | 30 | 31 | 31 | 30 | 31 | 30 | 31 | 31 | 28 | 9 | 23 | 365 |
| Water Use per Day (gallons) | 212 | 238 | 250 | 198 | 187 | 176 | 185 | 236 | 180 | 197 | 195 | 231 | 234 | 207 |
| 1-Minute Data | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Average |
| Data Points | 41,760 | 44,640 | 43,200 | 44,640 | 44,640 | 43,200 | 44,640 | 43,200 | 44,640 | 44,640 | 40,320 | 12,960 | 33,120 | |
| Percent of Data with Water Use | 21% | 26% | 19% | 16% | 13% | 13% | 13% | 16% | 14% | 14% | 18% | 44% | 40% | 19% |
| Events Starts per Day | 147 | 180 | 115 | 116 | 92 | 71 | 76 | 78 | 80 | 81 | 90 | 51 | 40 | 97 |
| Event Volume (gallons) | 1.4 | 1.3 | 2.2 | 1.7 | 2.0 | 2.5 | 2.4 | 3.0 | 2.2 | 2.4 | 2.2 | 4.5 | 5.8 | 2.1 |
| Event Intensity (gpm) | 0.7 | 0.6 | 0.9 | 0.9 | 1.0 | 1.0 | 1.0 | 1.0 | 0.9 | 1.0 | 0.7 | 0.4 | 0.4 | 0.8 |
| Event Duration (minutes) | 2.0 | 2.1 | 2.4 | 2.0 | 2.1 | 2.6 | 2.5 | 3.0 | 2.5 | 2.5 | 3.0 | 12.7 | 13.7 | 2.8 |
| 5-Minute Data | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Average |
| Data Points | 8,352 | 8,928 | 8,640 | 8,928 | 8,928 | 8,640 | 8,928 | 8,640 | 8,928 | 8,928 | 8,064 | 2,592 | 6,624 | |
| Percent of Data with Water Use | 48% | 51% | 40% | 41% | 30% | 28% | 29% | 33% | 31% | 32% | 36% | 52% | 48% | 37% |
| Events Starts per Day | 39 | 26 | 35 | 33 | 21 | 30 | 32 | 33 | 36 | 38 | 33 | 16 | 17 | 31 |
| Event Volume (gallons) | 5.4 | 9.2 | 7.1 | 7.7 | 6.4 | 5.8 | 5.9 | 7.1 | 5.0 | 5.2 | 5.7 | 22.0 | 11.5 | 6.7 |
| Event Intensity (gpm) | 0.3 | 0.3 | 0.4 | 0.3 | 0.5 | 0.4 | 0.4 | 0.5 | 0.4 | 0.4 | 0.4 | 0.3 | 0.4 | 0.4 |
| Event Duration (minutes) | 17.7 | 28.4 | 16.5 | 23.5 | 11.9 | 13.2 | 13.2 | 14.3 | 12.5 | 12.2 | 14.9 | 86.0 | 28.3 | 17.4 |

Table 4-8. Continued

| Summary of Data for 3 Homes (per Home) | Apr-14 | May-14 | Jun-14 | Jul-14 | Aug-14 | Sep-14 | Oct-14 | Nov-14 | Dec-14 | Jan-15 | Feb-15 | Mar-15 | Apr-15 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Days of Record | 29 | 31 | 30 | 31 | 31 | 30 | 31 | 30 | 31 | 31 | 28 | 9 | 23 | 365 |
| Water Use per Day (gallons) | 212 | 238 | 250 | 198 | 187 | 176 | 185 | 236 | 180 | 197 | 195 | 231 | 234 | 207 |
| **15-Minute Data** | **Apr-14** | **May-14** | **Jun-14** | **Jul-14** | **Aug-14** | **Sep-14** | **Oct-14** | **Nov-14** | **Dec-14** | **Jan-15** | **Feb-15** | **Mar-15** | **Apr-15** | **Average** |
| Data Points | 2,784 | 2,976 | 2,880 | 2,976 | 2,976 | 2,880 | 2,976 | 2,880 | 2,976 | 2,976 | 2,688 | 864 | 2,208 | |
| Percent of Data with Water Use | 67% | 64% | 59% | 57% | 43% | 45% | 46% | 51% | 50% | 52% | 54% | 62% | 58% | 54% |
| Events Starts per Day | 9 | 8 | 10 | 8 | 10 | 13 | 12 | 13 | 11 | 12 | 11 | 8 | 8 | 10 |
| Event Volume (gallons) | 23.1 | 30.9 | 31.6 | 21.1 | 13.0 | 14.1 | 15.5 | 18.0 | 16.3 | 16.7 | 16.6 | 59.4 | 20.9 | 19.9 |
| Event Intensity (gpm) | 0.2 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.3 | 0.2 | 0.3 | 0.3 | 0.2 | 0.4 | 0.3 |
| Event Duration (minutes) | 105.4 | 122.6 | 111.4 | 82.1 | 41.0 | 51.2 | 56.5 | 56.0 | 65.8 | 64.0 | 64.5 | 271.6 | 58.0 | 74.7 |
| **60-Minute Data** | **Apr-14** | **May-14** | **Jun-14** | **Jul-14** | **Aug-14** | **Sep-14** | **Oct-14** | **Nov-14** | **Dec-14** | **Jan-15** | **Feb-15** | **Mar-15** | **Apr-15** | **Average** |
| Data Points | 696 | 744 | 720 | 744 | 744 | 720 | 744 | 720 | 744 | 744 | 672 | 216 | 552 | |
| Percent of Data with Water Use | 82% | 80% | 78% | 72% | 65% | 69% | 71% | 78% | 73% | 77% | 76% | 78% | 75% | 75% |
| Events Starts per Day | 2 | 2 | 2 | 2 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 3 | 2 |
| Event Volume (gallons) | 336.3 | 79.5 | 67.7 | 112.1 | 50.3 | 54.1 | 63.3 | 87.7 | 80.5 | 70.1 | 69.6 | 195.2 | 63.7 | 88.0 |
| Event Intensity (gpm) | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 | 0.2 |
| Event Duration (minutes) | 1613.5 | 367.6 | 359.6 | 579.3 | 281.0 | 306.6 | 349.4 | 414.3 | 474.6 | 391.4 | 379.3 | 1047.3 | 273.7 | 456.1 |

Table 4-9. Summary of event outlier detection per home for DMA study area

| Volumetric Ranges Units in Gallons | | 5 Minute | 15 Minute | 60 Minute |
|---|---|---|---|---|
| V<1 | Events per Home | 3,498 | 1,107 | 127 |
| | Total Volume per Home (gal) | 483 | 245 | 29 |
| | Percent of Volume Within: | ---------- | ---------- | ---------- |
| | Anticipated Intensity Ranges | 19.2% | 0.0% | 0.0% |
| | Unanticipated Intensity Ranges | 80.8% | 100.0% | 100.0% |
| 1<=V<10 | Events per Home | 151 | 576 | 312 |
| | Total Volume per Home (gal) | 319 | 1,496 | 1,232 |
| | Percent of Volume Within: | ---------- | ---------- | ---------- |
| | Anticipated Intensity Ranges | 74.8% | 58.3% | 8.1% |
| | Unanticipated Intensity Ranges | 25.2% | 41.7% | 91.9% |
| 10<=V<100 | Events per Home | 37 | 233 | 333 |
| | Total Volume per Home (gal) | 1,780 | 10,590 | 13,218 |
| | Percent of Volume Within: | ---------- | ---------- | ---------- |
| | Anticipated Intensity Ranges | 93.5% | 95.4% | 81.6% |
| | Unanticipated Intensity Ranges | 6.5% | 4.6% | 18.4% |
| 100<=V<1000 | Events per Home | 40 | 76 | 117 |
| | Total Volume per Home (gal) | 8,972 | 15,424 | 26,342 |
| | Percent of Volume Within: | ---------- | ---------- | ---------- |
| | Anticipated Intensity Ranges | 85.6% | 92.9% | 92.1% |
| | Unanticipated Intensity Ranges | 14.4% | 7.1% | 7.9% |
| V>=1000 | Events per Home | 2 | 3 | 4 |
| | Total Volume per Home (gal) | 13,900 | 17,369 | 23,203 |
| | Percent of Volume Within: | ---------- | ---------- | ---------- |
| | Anticipated Intensity Ranges | 87.5% | 89.1% | 89.6% |
| | Unanticipated Intensity Ranges | 12.5% | 10.9% | 10.4% |

Table 4-10. Summary of event outlier detection per home for 3 homes evaluated in Chapter 3

| Volumetric Ranges Units in Gallons | | 1 Minute | 5 Minute | 15 Minute | 60 Minute |
|---|---|---|---|---|---|
| V<1 | Events per Home | 17,500 | 5,302 | 1,056 | 74 |
| | Total Volume per Home (gal) | 480 | 565 | 118 | 5 |
| | Percent of Volume Within: | --------- | ---------- | ---------- | ---------- |
| | Anticipated Intensity Ranges | 9.7% | 12.9% | 0.0% | 0.0% |
| | Unanticipated Intensity Ranges | 90.3% | 87.1% | 100.0% | 100.0% |
| 1<=V<10 | Events per Home | 2 | 202 | 392 | 245 |
| | Total Volume per Home (gal) | 3 | 424 | 1,234 | 891 |
| | Percent of Volume Within: | --------- | ---------- | ---------- | ---------- |
| | Anticipated Intensity Ranges | 50.5% | 60.6% | 80.8% | 8.9% |
| | Unanticipated Intensity Ranges | 49.5% | 39.4% | 19.2% | 91.1% |
| 10<=V<100 | Events per Home | 14 | 48 | 363 | 304 |
| | Total Volume per Home (gal) | 519 | 2,537 | 16,865 | 13,475 |
| | Percent of Volume Within: | --------- | ---------- | ---------- | ---------- |
| | Anticipated Intensity Ranges | 42.4% | 96.2% | 96.8% | 85.6% |
| | Unanticipated Intensity Ranges | 57.6% | 3.8% | 3.2% | 14.4% |
| 100<=V<1000 | Events per Home | 23 | 59 | 113 | 194 |
| | Total Volume per Home (gal) | 5,705 | 13,052 | 22,569 | 41,299 |
| | Percent of Volume Within: | --------- | ---------- | ---------- | ---------- |
| | Anticipated Intensity Ranges | 47.6% | 75.6% | 89.3% | 90.3% |
| | Unanticipated Intensity Ranges | 52.4% | 24.4% | 10.7% | 9.7% |
| V>=1000 | Events per Home | 2 | 2 | 2 | 3 |
| | Total Volume per Home (gal) | 3,841 | 6,895 | 12,807 | 19,300 |
| | Percent of Volume Within: | --------- | ---------- | ---------- | ---------- |
| | Anticipated Intensity Ranges | 83.0% | 91.7% | 89.1% | 92.0% |
| | Unanticipated Intensity Ranges | 17.0% | 8.3% | 10.9% | 8.0% |

# CHAPTER 5
# SUMMARY, CONCLUSIONS, AND FUTURE WORK

This dissertation presents, analyzes, and summarizes high-frequency water use data using 18 million data points collected from residential end users in Hillsborough County, Florida, in the Tampa Bay area. This is a subset of the 48 million data points collected with the overall AMR pilot. In the emerging world of "big data", this dissertation describes methods for formulating large datasets into useful databases that can be used for demand evaluations and event detection. The high-frequency evaluations discussed in Chapters 2 through 4 provide a framework for evaluating customer demand at varying temporal aggregations and designing event detection systems for unexpected customer events. As the data are aggregated up to larger spatial and temporal scales, the data can be used for system design and operation. This dissertation demonstrates a dual benefit approach to smart meter systems wherein both the utility and the customer can directly benefit.

In Chapter 2, high-frequency data for two master-metered multi-family residential complexes are evaluated at varying temporal aggregations. The evaluation shows through the analysis of large datasets collected for two complexes, that traditional meter sizing applications can be improved by assuming the high-frequency data are normally distributed around the mean with a standard deviation of one-half the mean. This assumption allows for accurate approximations of peak water use at varying temporal aggregations as well as accurate representations of the overall distribution of water use.

In Chapters 3 and 4, high-frequency water use for individual homes is evaluated. The evaluations include analysis of the peaks and the distribution of the overall data. A major concept developed is that of the aggregate event, wherein all consecutive data points with water use are part of the same aggregate event. Aggregate events are evaluated based on their

intensity, duration, frequency, and volume. The time step for creating aggregate events is increased from one minute to one hour in order to evaluate the effects of time averaging on overall event statistics and unanticipated event detection. The evaluations presented in these chapters are the first that directly search for identifying leaks and pipe breaks within customer homes based on outliers to anticipated events. These individual aggregate events that are outliers in terms of intensity, duration, and volume are identified as unanticipated events. Based on the definition of unanticipated events described in Chapters 3 and 4, larger events are relatively infrequent and easy to identify. Further analysis is needed to evaluate the tradeoff of threshold values for identifying unanticipated events, specifically to compare the risk of too many alarms versus the reward of providing an alarm that prevents expensive damage or reduces wasted water. This dissertation provides a framework which future evaluations can follow and provides the first event statistics for these unanticipated events.

Future work can be broken into research in three key areas: 1) using high-frequency water use data and probability distributions to improve demand evaluations for infrastructure sizing, especially for master meters; 2) using smart systems to quickly notify customers of unanticipated events based on algorithms that detect abnormal water use behavior; and 3) linking high-frequency data and real-time distribution system modeling to improve distribution system operation. The work from this dissertation is currently being applied in all three areas.

The district metered area discussed in Chapter 4 provides an excellent test network for future distribution system modeling as the collected high-frequency water use data encompasses every home within the district metered area. This allows for high-frequency water use to be allocated with known quantities at the individual customer level, as opposed to traditional demand allocation that requires estimations. Contemporary urban water systems with smart

meters can generate massive amounts of data. A major challenge is how to manage and analyze this complex information in a timely manner for real-time control. Much of water supply analytics are embedded in state-of-the-art water distribution systems simulation models. Linking these models with real-time data for real-time simulations will provide operational control that is not currently available in the industry. An emerging research area in the field of real-time data analytics is measuring energy efficiency, and linking real-time water use to real-time modeling applications will allow for direct evaluations of distribution system performance and the energy needed to provide such performance.

LIST OF REFERENCES

AWWA, 1999. *Engineering Computer Applications Committee: Calibration Guidelines for Water Distribution System Modeling*. AWWA, Denver.

AWWA, 2008 (2nd ed.). Manual of Water Supply Practices, M36. *Water Audits and Loss Control Programs*. AWWA, Denver.

AWWA, 2012a. *Buried No Longer: Confronting America's Water Infrastructure Challenge*. AWWA, Denver.

AWWA, 2012b (3rd ed.). Manual of Water Supply Practices, M32. *Computer Modeling of Water Distribution Systems*. AWWA, Denver.

AWWA, 2014 (3rd ed.). Manual of Water Supply Practices, M22. *Sizing Water Service Lines and Meters*. AWWA, Denver.

Blokker, E.J.; Vreeburg, J.H.; & Van Dijk, J.C., 2010. Simulating Residential Water Demand with a Stochastic End-Use Model. *Journal of Water Resources Planning and Management*, 136:1:19.

Blokker, E. J.; Pieterse-Quirijns, E. J.; Vreeburg, J. H.; & Van Dijk, J. C., 2011. Simulating Non-Residential Water Demand with a Stochastic End-Use Model. *Journal of Water Resources Planning and Management*, 137:6:511.

Blokker, M.; Vloerbergh, I; & Buchberger, S., 2012. Estimating Peak Water Demands in Hydraulic Systems II – Future Trends. Proc. 2012 Water Distribution Systems Analysis Conference, Adelaide, Australia.

Buchberger, S.; Blokker, M.; & Cole, D., 2012. Estimating Peak Water Demands in Hydraulic Systems I – Current Practice. Proc. 2012 Water Distribution Systems Analysis Conference, Adelaide, Australia.

Buchberger, S.G.; Carter, J.T.; Lee, Y.; & Schade, T.G., 2003. Random Demands, Travel Times, and Water Quality in Deadends. Water Research Foundation, Denver.

Buchberger, S.G.; & Wells, G.J., 1996. Intensity, Duration, and Frequency of Residential Water Demands. *Journal of Water Resources Planning and Management*, 122:1:11.

Buchberger, S.G.; & Wu, L., 1995. Model for Instantaneous Residential Water Demands. *Journal of Hydraulic Engineering*, 121:3:232.

Buchberger, S.G.; & Li, Z., 2007. PRPsym: A modeling system for simulation of stochastic water demands. Proc. 2007 World Environmental and Water Resources Congress. Reston, Virginia.

Cardell-Oliver, R., 2013. Water Use Signature Patterns for Analyzing Household Consumption Using Medium Resolution Meter Data. *Water Resources Research*, 49:12:8589.

Daigle, N.; & Jackson, A., 2013. New Mexico Utility Rolls Out Smart Grid Infrastructure. *Journal of the American Water Works Association*, 105:2:51.

Davies, K.; Doolan, C.; Van Den Honert, R.; & Shi, R, 2014. Water-Saving Impacts of Smart Meter Technology: An Empirical 5 Year, Whole-of-Community Study in Sydney, Australia. *Water Resources Research*, 50:9:7348.

DeOreo, W.B.; Heaney, J.P.; & Mayer, P.W., 1996. Flow Trace Analysis to Assess Water Use. *Journal of the American Water Works Association*, 88:1:79.

DeOreo, W.B.; & Mayer, P.W., 2012. Insights Into Declining Single Family Residential Water Demands. *Journal of the American Water Works Association*, 104:6.

DeOreo, W.B.; Mayer, P.W.; Dziegielewski, B.; & Kiefer, J., 2016. *Residential End Uses of Water, Version 2*. Water Research Foundation, Denver.

Dziegielewski, B.; Kiefer, J.C.; Opitz, E.M.; Poerter, G.A.; Lantz, G.L.; DeOreo, W.B.; Mayer, P.W.; & Nelson, J.O., 2000. *Commercial End Uses of Water*. Water Research Foundation, Denver.

EPRI (Electric Power Research Institute), 2002. *Water and Sustainability (Volume 4): U.S. Electricity Consumption for Water Supply and Treatment – The Next Half Century*. Technical Report 1006787, Palo Alto, California.

Friedman, M.; Kirmeyer, G.; Limeux, J.; LeChevallier, M.; Seidl, S.; & Routt, J., 2010a. *Criteria for Optimized Distribution Systems*. Water Research Foundation, Denver.

Friedman, K.; Heaney, J.P.; Morales, M.; & Switt, R., 2010b. Water Use and Demand Management Options for the Multi-family Residential Sector. Proc. 2010 Florida Section of AWWA Fall Conference, Orlando, Florida.

Friedman, K.; Heaney, J.P.; Morales, M.; & Palenchar, J., 2011. Water Demand Management Optimization Methodology. *Journal of the American Water Works Association*, 103:9:74.

Friedman, K.; Heaney, J.P.; Morales, M.; and Palenchar, J., 2013a. Predicting and Managing Residential Potable Irrigation Using Parcel-level Databases. *Journal of the American Water Works Association*, 105:7:372.

Friedman, K.; Heaney, J.; & Morales, M., 2013b. *Evaluation of water demand management and water loss control in Sanford, Florida*. Final Report to Sanford, Florida and the St. Johns River Water Management District.

Friedman, K.; Heaney, J.P.; & Morales, M., 2014a. Using Process Models to Estimate Residential Water Use and Population Served. *Journal of the American Water Works Association*, 106:6:264.

Friedman, K.; Heaney, J.P.; Morales, M.; and Palenchar, J., 2014b. Estimation of Single-Family Residential Irrigation Demand Management Effectiveness. *Journal of the American Water Works Association*, 106:5:253.

Friedman, K.; Heaney, J.P.; Morales, M.; & Palenchar, J., 2014c. Analytical Optimization of Demand Management Strategies Across All Urban Water Use Sectors. *Water Resources Research*, 50:7. http://dx.doi.org/10.1002/2013WR014261.

House-Peters, L.; & Chang, L., 2011. Urban Water Demand Modeling: Review of Concepts, Methods, and Organizing Principles. *Water Resources Research*, 47:5. http://dx.doi.org/10.1029/2010WR009624.

Knight, S.L.; Morales, M.A.; and Heaney, J.P., 2015a. Effect of Commodity Charges on the Demand for Reclaimed Water. *Journal of the American Water Works Association*, 107:11.

Knight, S.L.; Heaney, J.P.; & Morales, M.A., 2015b. Flat-Rate Reclaimed Use and Savings in Single-family Homes. *Journal of the American Water Works Association*, 107:5. http://dx.doi.org/10.5942/jawwa.2015.107.0054.

Mayer, P.W.; DeOreo, W.B.; Opitz, E.M.; Kiefer, J.C.; Davis, W.Y.; Dziegielewski, B.; & Nelson, J.O., 1999. *Residential End Uses of Water*. Water Research Foundation, Denver.

McCary, J.P., 2015. Statistical Analysis of Automatic Meter Reading in the Multifamily Sector. *Florida Water Resources Journal*, 67:9:32.

Minitab 17 Statistical Software, 2010. [Computer Software]. Minitab, Inc., State College, Pennsylvania.

Morales, M.A.; Heaney, J.P.; Friedman, K.R.; & Martin, J.M., 2011. Estimating Commercial, Industrial, and Institutional Water Use on the Basis of Heated Building Area. *Journal of the American Water Works Association*, 103:6:84.

Morales, M.A.; & Heaney, J.P., 2014. Classification, Benchmarking, and Hydro-Economic Modeling of Nonresidential Water Users. *Journal of the American Water Works Association*, 106:12. http://dx.doi.org/10.5942/jawwa.2014.106.0150.

Morales, M.; Martin, J.; Heaney, J.; & Friedman, K., 2013a. Parcel-Level Modeling of End-Use Water Demands in Public Supply. *Journal of the American Water Works Association*, 105:8. http://dx.doi.org/10.5942/jawwa.2013.105.0107.

Morales, M.; & Heaney, J., 2015. Benchmarking Nonresidential Water Use Efficiency Using Parcel-Level Data. *Journal of Water Resources Planning and Management*, 142.3.

Morales, M.A,; Heaney, J.P.; Friedman, K.R.; & Martin, J.M., 2013b. Parcel-Level Model of Water and Energy End Use: Effects of Indoor Water Conservation. *Journal of the American Water Works Association*, 105:9. http://dx.doi.org/10.5942/jawwa.2013.105.0103.

Thiemann, R.; Haas, J.; & Schlenger, D., 2011. Reaping the Benefits of AMI: A Kansas City Case Study. *Journal of the American Water Works Association*, 103:4:38.

USEPA (US Environmental Protection Agency), 2013. *Drinking Water Infrastructure Needs Survey and Assessment, Fifth Report to Congress*. EPA 816-R-13-006, Washington.

Vertommen, I.; Magini, R.; & Cunha, M., 2014. Scaling Water Consumption Statistics. *Journal of Water Resources Planning and Management*, 141:5. http://dx.doi.org/10.1061/(ASCE)WR.1943-5452.0000467.

BIOGRAPHICAL SKETCH

John McCary was born in 1980 in Tampa, Florida, to parents John and Melonny. He lived in Tampa until the age of six when he moved to Clearwater, Florida. As a kid, he participated in many sports and continues to do so as an adult while playing with his sons and occasionally coaching their teams. He lived in Clearwater until graduating from Clearwater High School in 1998.

John moved back to Tampa when he decided to attend the University of South Florida to study civil engineering. During his time at the University of South Florida, he met his future wife, Lorrie. Their first son, Johnathan, was born in July 2000. In order to gain engineering experience and provide financial support for his family while finishing his undergraduate program, he worked as an Environmental Co-op for Cargill Fertilizer. His work with hydraulic systems while at Cargill Fertilizer led to a desire to pursue a career working with hydraulic systems and ultimately led to changing his engineering concentration from structural to water resources. He completed his bachelor's degree in 2002 and was awarded *Outstanding Student of the Year* by the Engineering Alumni Society. While working on his bachelor's degree, he was accepted into the *Research Experiences for Undergraduates* program that allowed him to start work on his master's degree. His research was focused on integrating surface water and groundwater modeling, and he graduated with his master's degree in 2005.

John started working for the Hillsborough County Public Utilities Department in 2003 while finishing his master's degree. His work at Hillsborough County involved planning for the future of the distribution system, which included demand analysis, hydraulic analysis, and managing large datasets. While working on demand evaluations and conservation, he was introduced to Dr. James Heaney who was leading a research team on improving bottom-up demand evaluations at the University of Florida's Department of Environmental Engineering

Sciences. John developed a working relationship with Dr. Heaney that led to his pursuit of a Ph.D. Fortunately, the opportunity allowed him to stay employed full time while pursuing his studies. In September 2011, John and Lorrie had their second son, Jamason. The balance between family, work, and academia made the journey challenging but rewarding. John received his Ph.D. in Environmental Engineering Sciences in December 2017.