Statistical Analysis of Automatic Meter Reading in the Multifamily Sector

John P. McCary

John McCary, P.E., is potable and reclaimed water planning team leader for Hillsborough County Public Utilities Department in Tampa.

This article presents the results of a high-frequency water use evaluation using one-minute data for two multifamily residential complexes that are customers of Hillsborough County Public Utilities Department (Utility). Automatic meter reading (AMR) data loggers are used with short-range wireless communication, which allow for ease of data collection by driving by and downloading the data from the data loggers. Starting in September 2013, over 1,400,000 data points have been collected and stored in a database for the two study areas, and the database provides easy access to water use data aggregated to any combination of time of day and day of the week. Both study areas are supplied by 8-in. master meters with AMR data loggers. The data loggers record in 10-gal increments, and the data storage is limited to 16,000 data points, which requires downloading every 11 days in order to avoid gaps in the data. Study Area 1 has 440 residential units, with an estimated population of 893 residents; study Area 2 has 257 residential units, with an estimated population of 447 residents.

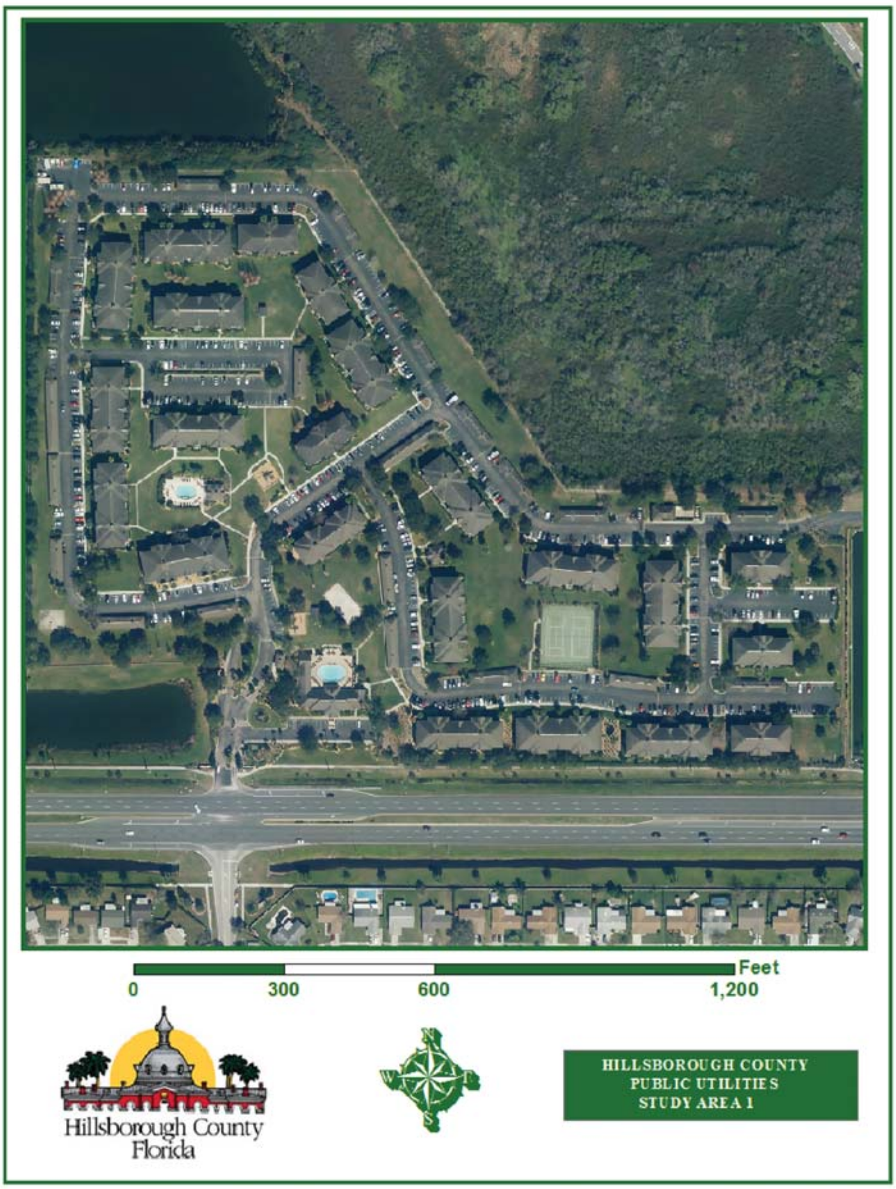

Figure 1. Aerial View of Study Area 1

The purpose of the analysis was to evaluate the value of the high-frequency water use "big data": How can the data be used to improve service by making better design and operating decisions? Specifically, focus was on comparing the measured results with the normal distribution to see if peak flows could be accurately estimated by traditionally collected billing data containing average use over monthly readings. Applying the normal distribution approximation using only the mean flow value, with the assumption that the standard deviation is half the mean flow, results in a distribution that visually resembles the measured distribution and adequately estimates peak flows at different levels of aggregation. This conclusion is subjective, as it is up to the individual, depending on application, to determine how close of an approximation is needed.

Future applications of the data and additional collection efforts will evaluate water use distributions at varying temporal scales with applications in design and operations optimization. It is unlikely that large data collection efforts are necessary to predict flow distributions and peak flows; however, future research will evaluate how much data collection is necessary to accurately forecast demand patterns and account for seasonal variations.

Background

Beginning in August 2013, an AMR data collection and analysis case study began for the

Utility. The entire study group consisted of one large single-family residential (SFR) neighborhood, two multifamily residential (MFR) complexes, one commercial big-box retail store, and one hospital. This analysis focused 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 the study was 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 (DeOreo et al, 1996; Buchberger and Wells, 1996; Mayer et al, 1999; Blokker et al., 2010; Buchberger et al, 2003).

Study Area 1

Shown in Figure 1, the MFR complex has 440 units on a parcel classified by the Department of Revenue (DOR) Code 0310 (Multifamily Residential > Nine 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 five-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-in. master meter, with an AMR data logger with recording capability in 10-gal 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, 744,785 data points have been collected. The average flow during the period of record is approximately 52,000 gal per day (gpd) or 36.1 gal per minute (gpm). The resulting gal per capita per day (gpcd) is 58.

Study Area 2

Shown in Figure 2, the MFR complex has 257 multifamily 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 five-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-in. master meter, with an AMR data logger with the same recording capability as Study Area 1. Over the period of record, 700,628 data points have been collected. The average flow during the period of record is approximately 28,300 gpd or 19.6 gpm, and the resulting gpcd is 63.

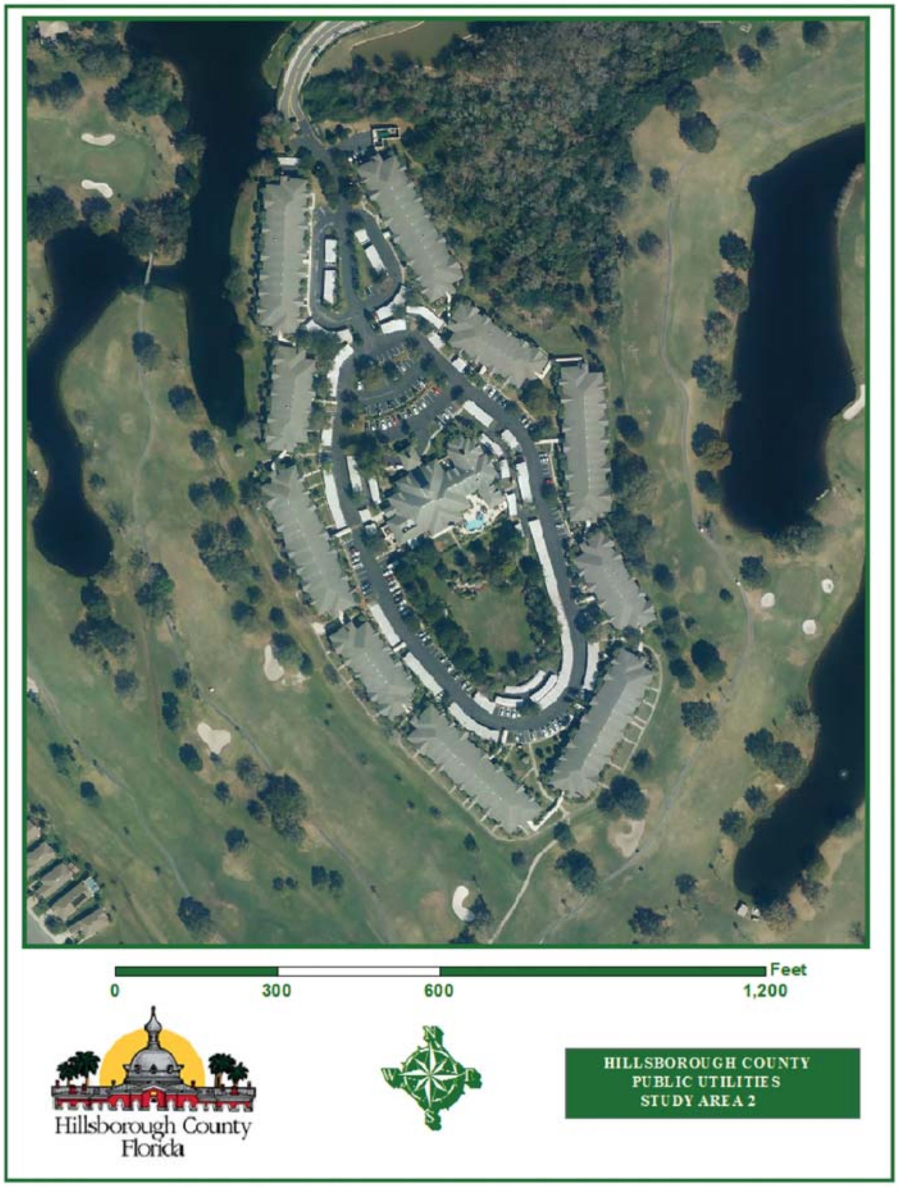

Figure 2. Aerial View of Study Area 2

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 one-minute water use data to be evaluated at higher frequencies and up to the aggregated, more commonly collected monthly billing data.

Monthly and Daily Average

Figure 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 readings can be seen in Figure 3, with wide variations in reported water use prior to the meters being replaced. For both meters,

Continued on page 34

Continued from page 33

the data has been more consistent once replaced with a meter and an AMR data logger.

Figure 4 shows the daily average water use for both Study Area 1 and 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 the end of summer. While not the subject of this article, future research will include investigating seasonality and how it impacts demand patterns over the year.

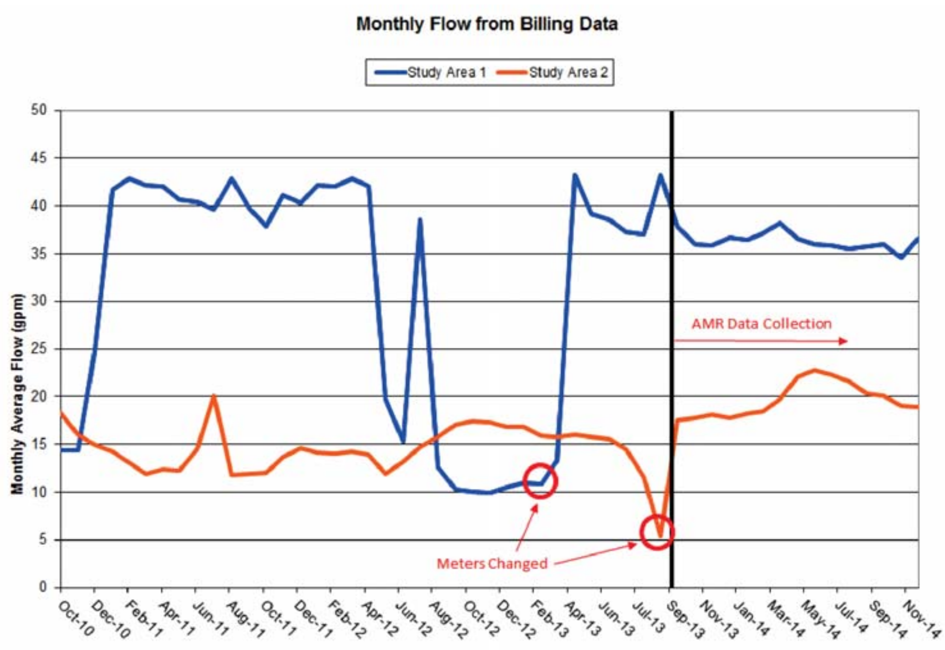

Figure 3. Average Monthly Flow From Billing Data

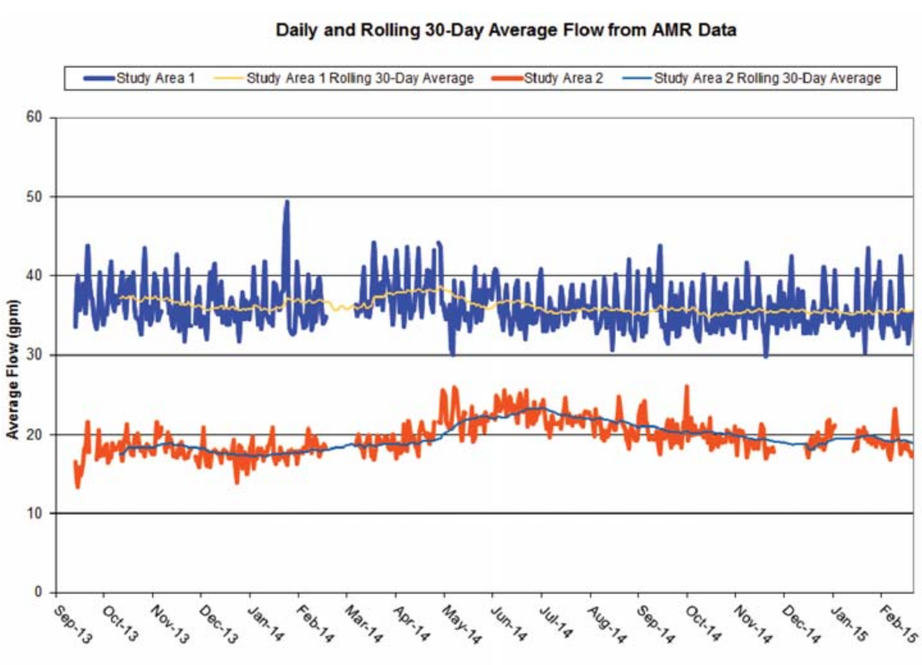

Figure 4. Daily and Monthly Average Flow From Automatic Meter Reading Data

Demand Patterns

Figures 5 and 6 show the average values for both study areas reflected in a weekly time series. Each point on the graph represents all of the data available for that time of day and the day of the week averaged together. For one year of data, each one-minute value on the graph represents the mean of each of the individual 52 weekly one-minute data points for that minute and day of the week. When aggregated up to the one-hour time step, each one-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 one-minute to one-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.

Comparison of Measured Data With Normal Distribution

Previous work on a limited dataset indicated that of the more common probability distributions, the normal (also known as Gaussian) distribution had the best fit. Conceptually, this makes sense because of the Central Limit Theorem, which basically states that when many random variables are combined, each having independent distributions, the combined distribution approaches a normal distribution. Rather than performing a detailed analysis using various distribution fitting tests and confidence intervals, a simple question was analyzed: Based on knowing only one value, the mean water use, how accurate would the normal distribution be at estimating minimum and peak flow rates at

Continued on page 36

Continued from page 34
one-minute, five-minute, 15-minute, and one-hour time steps?

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 distributing values around the actual mean flow, with an assumed standard deviation equal to half the mean flow. 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 7 shows the distribution of flows 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 of the total period 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 on-site 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. Also of note is that, during the period of record, flow was recorded 98.3 percent of the time, with the remaining 1.7 percent of the time resulting in zero flow.

Because these particular meter registers record the data in discrete 10-gal increments, the data columns in Figure 7 are displaying the actual data reported by the data logger and are not the result of binning by the database. The reported value for each one-minute interval carries the remainder of the value forward from the previous time step if it didn't result in a discrete 10-gal increment.

The following example illustrates this concept. Assume that for three consecutive minutes, the actual flow values are 1 gal, 21 gal, and 8 gal, respectively; the data logger would report the flow values as 0 gal, 20 gal, and 10 gal, 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 gal; however, the maximum cumulative error over any period of record is -10 gal.

Figure 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 gal around the discrete 10-gal increment. As an example, the data point used for graphing the probability at 10 gal used the difference between the cumulative probability at 15 gal and 5 gal. This affects the display of the results only; it doesn't have any impact on the normal distribution calculations.

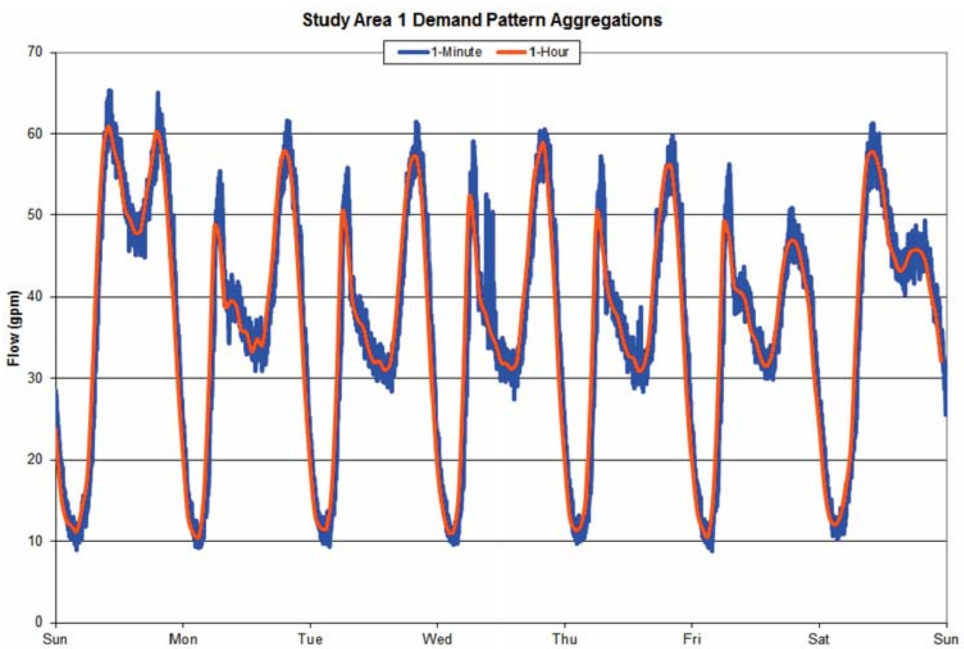

Figure 5. Aggregated Demand Patterns for Study Area 1

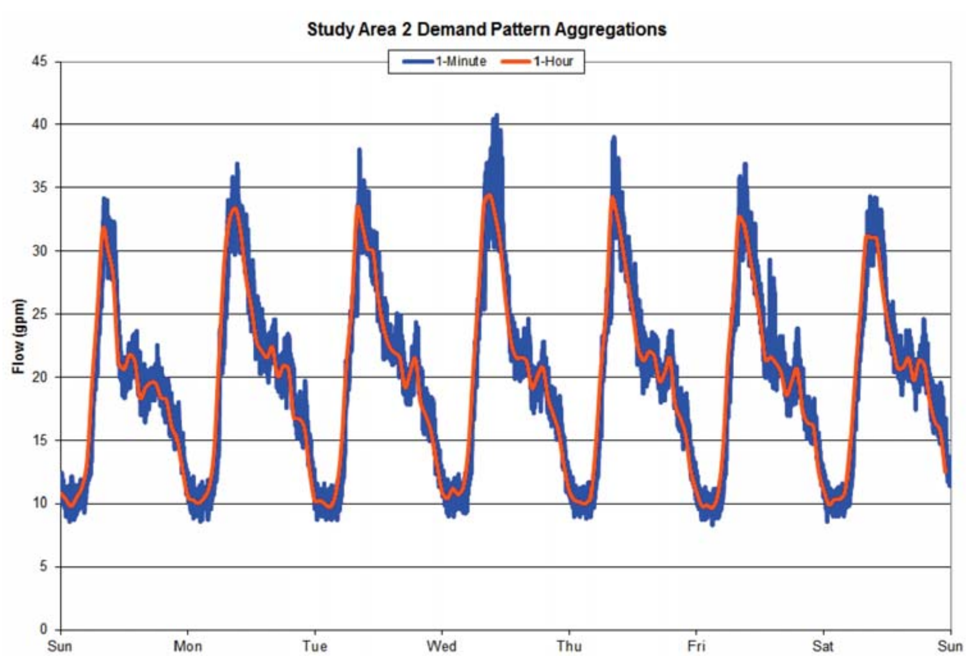

Figure 6. Aggregated Demand Patterns for Study Area 2

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. In order to do this, a simple question was asked: What flow would result for a given time-period statistic, like peak hour, assuming the probability of occurrence is consistent with the actual percentage of time that the period of interest occurs? 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 half the mean flow.

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 one-minute flow values during the week were assumed to occur over exactly one minute, which would equate to a frequency of 0.01 percent of time during the week. So, the minimum one-minute value for each week was selected from the cumulative distribution whose flow value corresponded to 0.01 percent, and the peak one-minute flow value was selected from the corresponding value at 99.99 percent. For five-minute, 15-minute, and one-hour flow values, the corresponding time periods occur at 0.05, 0.15, and 0.60 percent of the time, respectively.

Referring to Figure 8, the expected peak flow values are not visually evident because of the "flattened" curve above the 99 percent 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. While Figure 8 is representative of the entire dataset, this is consistent with the individual weekly distributions as well. Therefore, Table 1 doesn't summarize the minimum values, but it is important to note that the actual data recorded a zero value every week for the one- and five-minute levels of aggregation for both study areas. At the 15-minute and one-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 1 shows the weekly summarization of all 76 weeks, with peak flows at one-minute, five-minute, 15-minute, and one-hour levels of aggregation. The "percent difference" values in the table reflect the summarization 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 percent reported under the "Peak One-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 percent greater than the predicted peak flow.

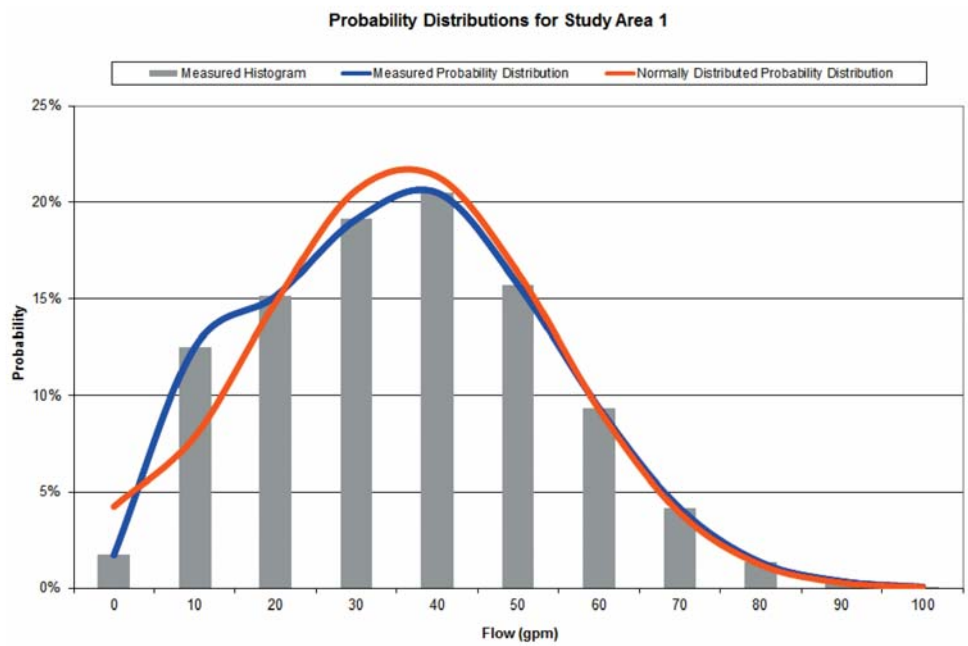

Figure 7. Probability Distributions for Study Area 1

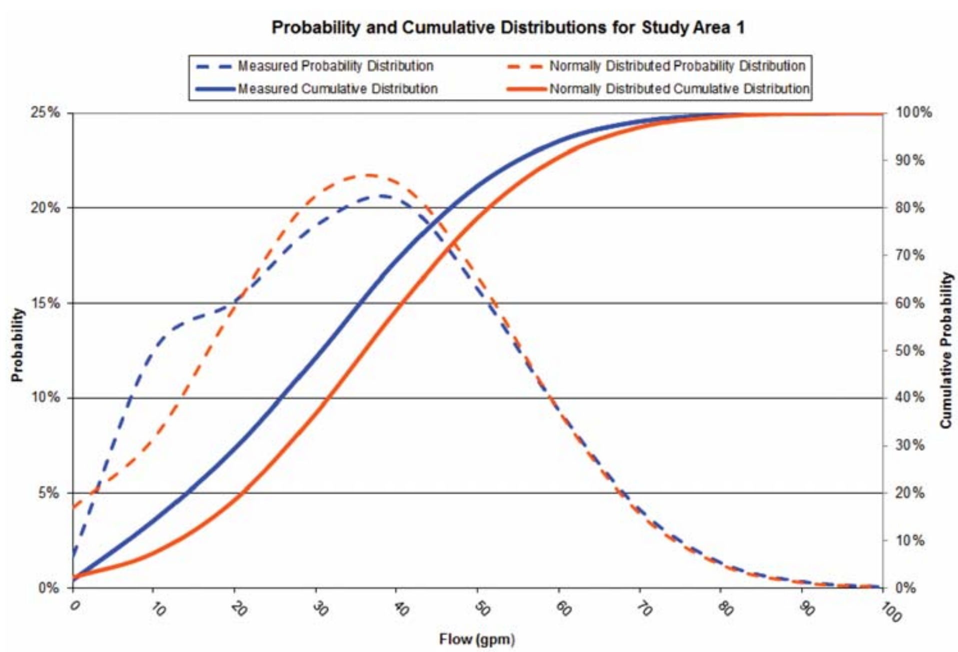

Figure 8. Probability and Cumulative Distributions for Study Area 1

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 data were assumed to be 100 percent correct, and these data were applied to the meter accuracy curves published for the 23 meters currently approved for use by the Utility at the sizes of 4, 6, and 8 in. For each flow value recorded for the two study areas, the meter accuracy error for each of the 23 meters was individually applied and the cumulative error for each meter type was calculated.

Figure 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-in. 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

Continued on page 38

Continued from page 37

green line). In this case, both the meters with highest negative and positive cumulative errors are compound meters. As can be seen in Figure 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-in. meter used resulted in a -0.2 percent error, and the meters with the highest negative and positive cumulative errors resulted in -2.3 and +0.4 percent, respectively. While not graphed, Study Area 2 had similar results with the actual 8-in. meter resulting in 0 percent error, and the meters with the highest negative and positive cumulative errors resulting in -1.8 and +0.6 percent, respectively.

Conclusions

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 MRF complexes, representing a combined population estimated at 1,340 residents. Applying the normal distribution approximation, using only the mean flow value with the assumption that the standard deviation is half the mean flow, results in a distribution that visually resembles the measured distribution. Using the normal distribution approximation provides an adequate estimate of the expected peak flows at different levels of aggregation. 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 evaulating 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 percent.

Table 1. Summarization of Weekly Measured and Predicted Values for 76 Weeks

Summarization of Individual Weekly Values:			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	-	76.6	84.4	89.8	97.1
		Average Flow	-	81.4	89.6	95.4	103.1
		Peak Flow	-	87.6	96.4	102.7	111.0
	% Difference	Minimum	-	-35%	-20%	-20%	-13%
		Average	-	-15%	-1%	0%	3%
		Maximum	-	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	-	38.1	41.9	44.6	48.3
		Average Flow	-	44.4	48.9	52.1	56.3
		Peak Flow	-	54.0	59.4	63.3	68.4
	% Difference	Minimum	-	-34%	-20%	-21%	-11%
		Average	-	-10%	8%	12%	16%
		Maximum	-	20%	28%	26%	33%

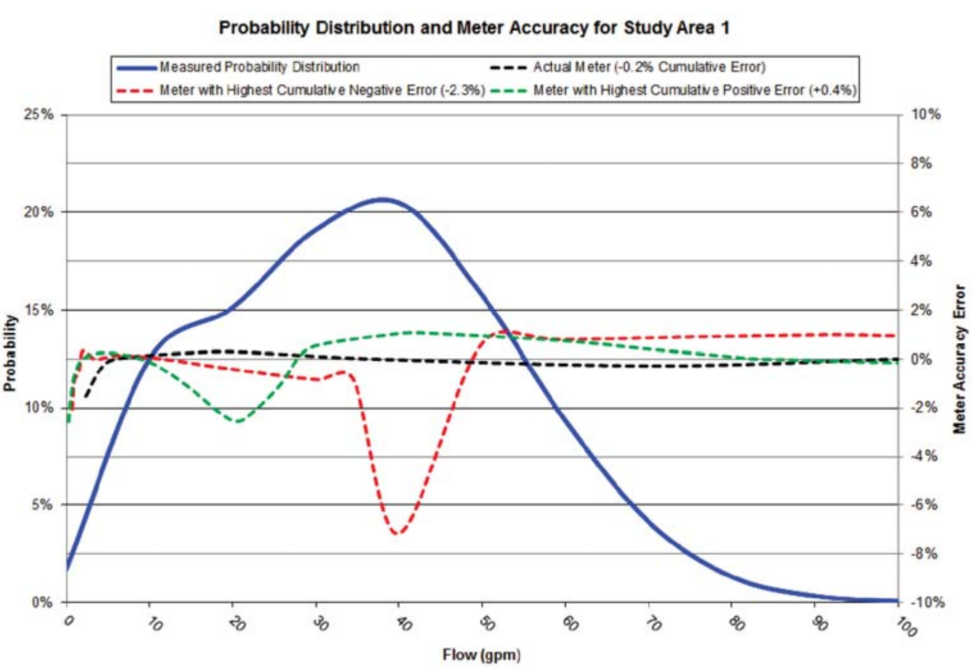

Figure 9. Measured Probability Distribution Versus Meter Accuracy

Acknowledgments

The author would like to acknowledge the staff members at Hillsborough County Public Utilities Department who have been involved in the AMR program. While this list doesn't encompass everyone involved, special thanks are in order for Daniel Hammon, Kenny Carlisle, Gary Hunter, David Butierries, Kimani Thomas, Norman Vik, and Paul Schaedler.

References

- Blokker, E., J. Vreeburg, and J. van Dijk, 2010. Simulating Residential Water Demand with a Stochastic End-Use Model. *Journal Water Resources Planning and Management*, 136(1), 19-26.
- Buchberger, S.G., and G.J. Wells, 1996. Intensity, Duration, and Frequency of Residential Water Demands. *Journal Water Resources Planning and Management*, 122(1), 11-19.
- Buchberger, S.G., J.T. Carter, Y. Lee, and T.G. Schade, 2003. *Random Demands, Travel Times, and Water Quality in Deadends.* Denver, Colo.: AwwaRF.
- DeOreo, W.B., J.P. Heaney, and P. Mayer, 1996. Flow Trace Analysis to Assess Water Use. *Journal American Water Works Association*, Vol. 88, No. 1.
- Friedman, K., Heaney, J.P., Morales, M. and R. Switt, 2010. Water Use and Demand Management Options for the Multifamily Residential Sector. FSAWWA 2010 Fall Conference. Orlando, Fla.
- Mayer, P., W.B. DeOreo, E.M. Opitz, J.C. Kiefer, W.Y. Davis, B. Dziegielewski, and J.O. Nelson. 1999. *Residential End Uses of Water.* Denver, Colo.: AwwaRF.