HYDROGRAPH SEPARATION OF INDOOR AND OUTDOOR BILLED WATER USE IN FLORIDA'S SINGLE FAMILY RESIDENTIAL SECTOR

John Palenchar, Kenneth Friedman, and James P. Heaney
University of Florida, Conserver Florida Water Clearinghouse

Gainesville, Florida

To Appear, Proc. Fall 2009 FSAWWA Fall Conference, Orlando

BACKGROUND

Use of potable water for irrigation can be reduced through conservation or may be replaced by a lower quality source such as reclaimed wastewater, surface water, or private well water. Ideally, the utility has information on customers who are using reuse water or have private wells. Often reuse customers are not metered and customers who use wells or surface sources for their irrigation water are difficult to track. Private residential irrigation wells, even if a permit is on record with a county health department, have no central data source to identify their prevalence or location. Whitcomb (2005) surveyed a cross section of 3,521 homes in 16 cities in Florida regarding their water use patterns including their irrigation source. The percentage of irrigation that is provided by the utility ranges from as low as 21% for Melbourne and St. Petersburg to a high of 100% in Tallahassee with an average of 64%. The potential for source substitution in a given area is proportional to the accounts that draw or will draw irrigation water from the potable water system. Thus, a critical value which must be determined is the proportion of accounts that use the potable supply for irrigation and the quantity of that use (Palenchar et al. 2009).

The vast majority of single family residential (SFR) water customers are served by a single meter that records total use, typically on a monthly basis. However, some, often larger, SFR customers, have two meters so that the regular metered "indoor" and irrigation metered "outdoor" uses are recorded separately. If dual-metering is not used, then the minimum month method is a popular way to perform single metered hydrograph separation. The minimum month method assumes that outdoor water use ceases in the winter because irrigation water is not needed (Dziegielewski et al. 1993; Vickers 2001). Unfortunately, this approach is less valid for warmer climates like Florida where year round irrigation is practiced (DeOreo et al. 2008; Mayer et al. 2009).

Aside from dual-meter accounts, it is difficult to find direct data in Florida on the proportion of total water use that is associated with indoor purposes or, by subtraction, outdoor purposes. This information has been collected historically by: surveys, special metering of some or all of the customers, or by using dual-metered billing records when available (Mayer and DeOreo 1999; Whitcomb et al. 2005). In this paper, a direct way to estimate the offline customers has been developed using one or two years of billing data to infer from the time series signatures of the water use patterns the proportion of accounts using potable water only for indoor purposes and the volume of that use. Outdoor flow is equal to total flow minus indoor flow or:

$$Q_{irrigation} = Q_{total} - Q_{indoor} \tag{1}$$

Customer level billing data were combined with parcel level customer attributes databases from the Florida Department of Revenue (FDOR) and the Alachua County Property Appraisers (ACPA). U.S. Census data were used to assign a population per house estimate for each of the residences. This household size estimate is the average population for the Census block(s) in which the parcel is located. The combined relational database provides an excellent profile of each of the customers by providing new insights into the nature of residential water use in Florida which should prove invaluable in conservation evaluations.

K-MEANS CLUSTERING

A popular method to group a data set into categories that are of interest is using the k-means clustering algorithm. XLSTAT© Version 2009.4.03 (Addinsoft 2009) was used for this analysis. The algorithm optimizes clustering by minimizing the objective function shown in equation 2.

$$\min E = \sum_{i=1}^{k} \sum_{p \in C_i} \left\| \hat{p} - \hat{m}_i \right\|^2 \tag{2}$$

Where

k = number of subsets in the global set

$\hat{m}_i$ = multidimensional centroid or mean point of cluster set C_i

$\hat{p}$ = multidimensional data point in question, and

E = sum of squares of errors in the Euclidean distances to the iteration subset means

The results are *k* clusters that are as compact and separate as possible (Everitt et al. 2001; Han and Kamber 2006, Xu and Wunsch 2009). The k-means cluster algorithm uses the following basic steps to determine appropriate assignment to clusters:

- Input the number of clusters.
- Assign each data point randomly to a cluster.
- The centroid of each cluster is determined, and then each point is assigned to the cluster whose centroid is closest to it by Euclidian distance.
- This process is repeated until no points switch clusters (convergence) indicating that the minimum value of *E* in equation 2 has been found (Han and Kamber 2006).

Each data point is the mean (x_1) and standard deviation(x_2) of the monthly water use for the customer. The mean gives a direct measure of the size of the water use whereas the standard deviation measures the monthly variability in water use for each customer. Thus, offline irrigation users would be expected to have a monthly water use pattern with relatively low variability. For these indoor users, the mean water use is directly proportional to the persons per house. The users were divided into the following three clusters:

1) smaller users with lower variation.
2) medium users with medium variation, and
3) larger users with higher variation.

The hypothesis is that cluster 1 members correspond to customers who are using other sources of irrigation water or have no or minimal demand for irrigation water. This has been presented as a viable means to estimate the potable water offset of substitute irrigation sources in the single family sector (Palenchar et al. 2009).

STUDY AREA PATTERNS OF WATER USE

With the cooperation of Gainesville Regional Utilities (GRU), one year of monthly water use data was collected for 1,403 dual metered customers and 29,507 single metered residential customers from October 2007 to September 2008. These customer level billing data were combined with parcel level customer attribute databases from the Florida Department of Revenue (FDOR) and the Alachua County Property Appraiser (ACPA). U.S. Census data were used to assign a population per house estimate for each of the residences. This household size estimate is the average population for the Census block(s) in which the parcel is located. Significant effort may be required to link the parcel level data from the utility, U.S. Census, the ACPA, and FDOR if the utility has not already done this work.

The monthly use of water, for single and dual-meter accounts over the period of observation is shown in Figure 1. From this figure, it can be observed that indoor water use is relatively constant and averages less than 200 gallons per account per day (gpad). This value of indoor use appears to be relatively constant independent of single or dual metering. Gainesville Regional Utilities imposes a residential wastewater charge which is calculated as $4.10 per thousand gallons (kgal.) times the maximum of January or February's regular metered (non-irrigation) usage. This charge provides a strong incentive to minimize water use during those months in the dataset. As a result of this financial inducement, the minimum month method may be a good approximation of indoor use during that period.

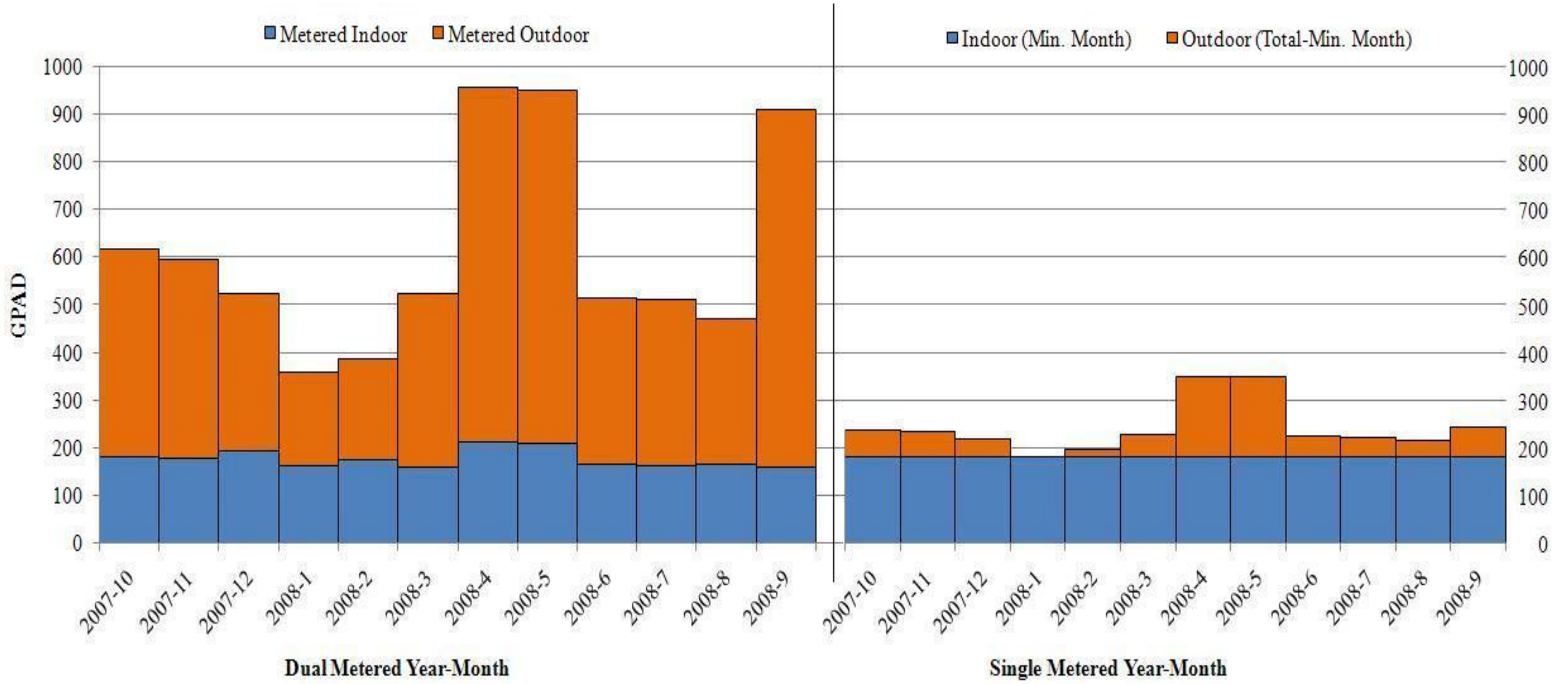

Figure 1. Average indoor and outdoor water use for 1,403 dual metered (left figure) and 29,507 single metered (right figure) residential accounts in the Alachua County Study area.

The minimum calculated average flow for all the regular meters in the month of January 2008 is 184 gpad. The 2000 U.S. Census block data yielded an average household size (HHS_{ave}) of 2.53 people per house. At an HHS_{ave} of 2.53, this gives the service area a minimum month usage rate of 72.7 gallons per capita per day (gpcd).

Dual meters ideally measure irrigation and non-irrigation uses separately. Thus, the indoor use can be assumed to be the regular (non-irrigation) metered flow. The total water use of dual-metered accounts is much larger than that of single-metered accounts. It is hypothesized that these accounts do not represent the majority of accounts in the utility due to higher than usual irrigation use. From Figure 2, the median use for dual-metered accounts is 532 gpad, with irrigation accounting for 355 gpad (532-2.53*70) or 67%. In contrast the single-meter group has a median use of 192 gpad, with irrigation accounting for 15 gpad (192-2.53*70) or 8%.

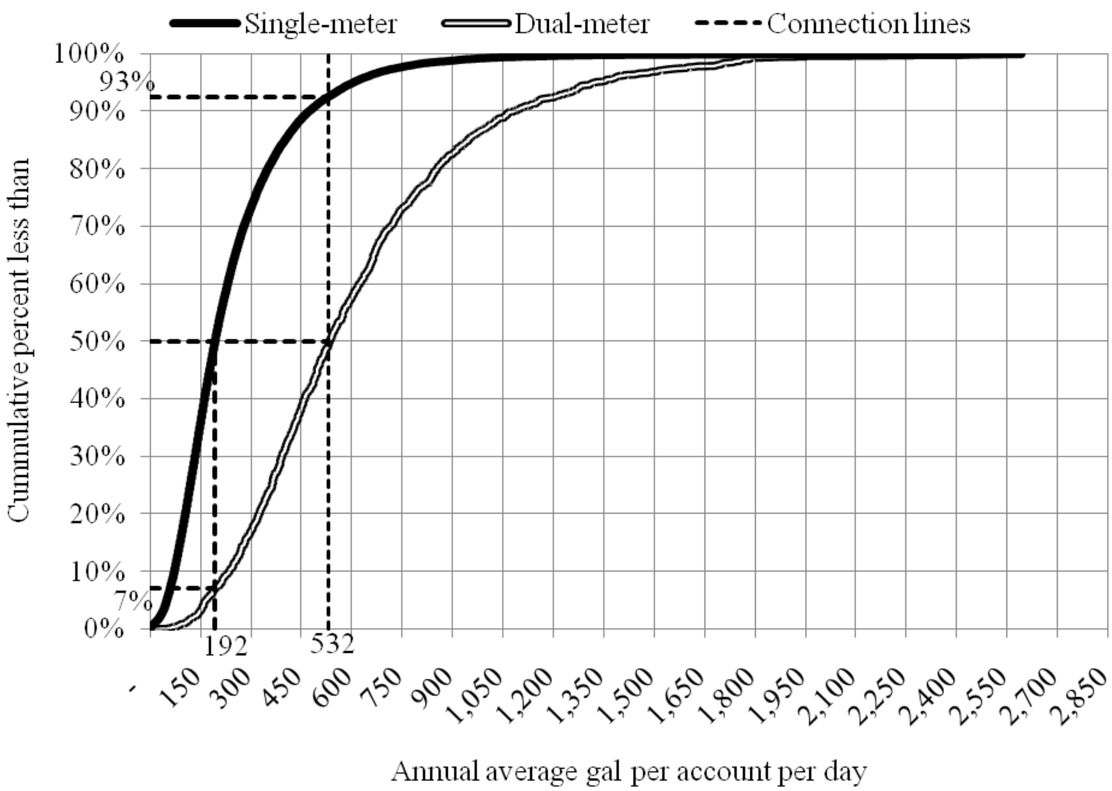

Figure 2. Cumulative frequency distributions of total water use for 29,507 single and 1,403 dual metered SFR accounts in the GRU service area.

By comparing the median annual average flow for the 1,403 dual-meter accounts to the 29,507 single-meter accounts, the dual-meter accounts can be shown to be unrepresentative of the majority of users. Single-meter accounts in general are minimal users of potable irrigation water.

The regular metered flow for dual-meter accounts shows an arithmetic mean of 178 gpad and a standard deviation of 68 gpad. Using Census block level data, it was determined that the group had an HHS_{ave} of 2.59 people per house. Thus, the average indoor per capita use for the area by this method was 68.7 gallons per person per day (gpcd). However, by looking at the data in Figure 3, it can be observed that 19 of the indoor values are above 560 gpad. Using 70 gpcd as an expected value, the result is an average household size of over eight people, an unlikely event (Friedman 2009). The values above 560 gpad can be considered erroneous in determining expected indoor use. By looking at a scatter plot of the annual average and monthly standard deviation for each account, the outlier data becomes quite evident. Thus to capture a quality assured (QA'd) sample a bounded dataset is used as shown by the exploded portion of Figure 3.

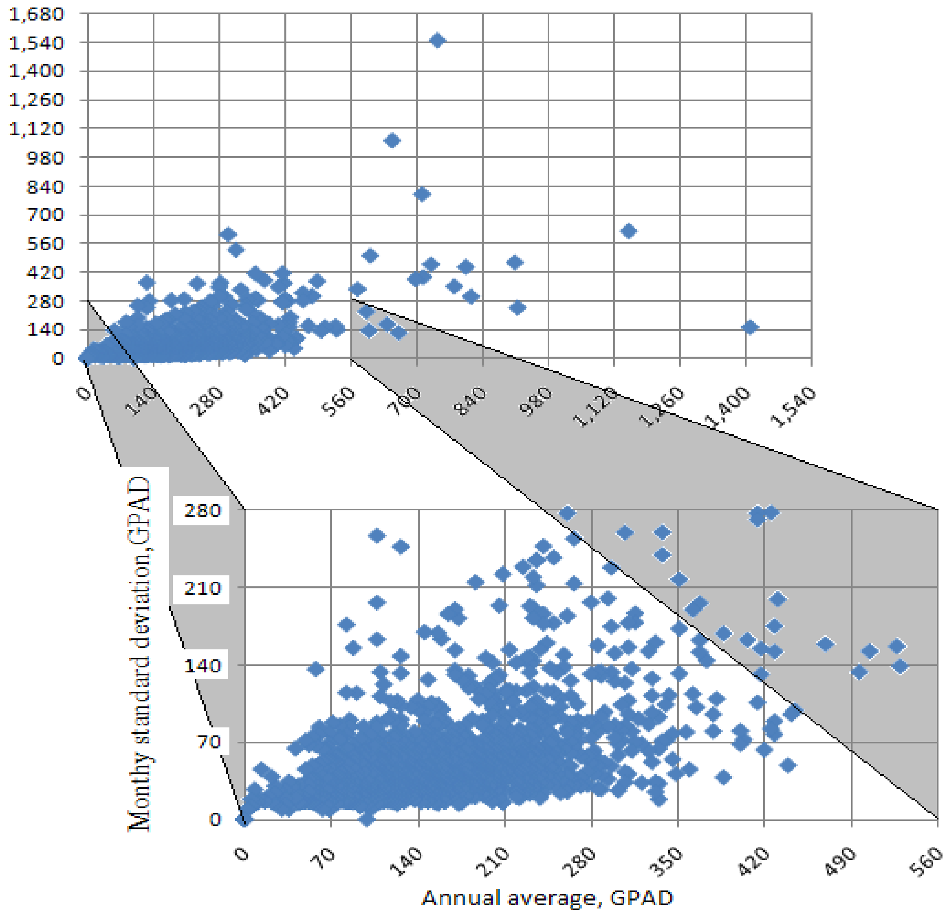

Figure 3. Scatter plot of the regular metered flow from the dual-metered GRU accounts. An extruded view details the accounts that are expected to represent the distribution of actual indoor use in the area

The resulting QA'd dataset has a mean of 165 gpad or 63.7 gpcd and a standard deviation of 54 gpad. These values are only slightly lower than the mean of 178 gpad and standard deviation of 68 gpad calculated previously. National studies performed by Aquacraft Inc. (Aquacraft 2005; DeOreo et al. 2008; Mayer and DeOreo 1999) have shown that indoor water use is relatively constant and averages around 70 gpcd.

Thus, several calculations of indoor use can be produced for the study area, as shown in Table 2. These results give a clear *a priori* value for indoor use in the study area of about 70 gpcd. This value will be used as a check of the cluster assumed to have no or minimal demand for irrigation water. In absence of a direct data, 70 gpcd should be used to estimate indoor water use for SFR customer groups in Florida (Friedman 2009).

Table 2. Results of current methods for calculating the indoor water use in the study area.

Method	GPCD
Minimum month	72.7
Regular metered from dual-meter accounts	68.7
Referenced approximation (Friedman 2009)	70.0
QA'd regular metered from dual-meter accounts	63.7

CLUSTERING OF Q_{TOTAL}

To cluster indoor and outdoor use, an overall dataset of Q_{total} was created form the single and dual-meter accounts. The regular and irrigation metered flow for each of the dual-metered accounts in the complete (not QA'd) dataset were summed to give Q_{total} for each customer in the group. The single-meter accounts already report Q_{total}. The annual average and the monthly standard deviation were calculated for the overall dataset. The data points for each customer's annual average gpad (x_1) and monthly standard deviation (x_2) were binned using widths calculated using the Scott (1979) method. The binned data was then put into the three dimensional histogram, shown in Figure 4.

The k-means clustering algorithm was applied to form three clusters based on x_1 and x_2. The numerical results are summarized in Table 3. The percent of customers in each group illustrates the densities of x_1 and x_2 shown in Figure 4. The results in Table 3 indicate that over 70% of the use in the service area is characteristic of indoor use with a mean of 156 gpad or 62.4 gpcd only slightly lower than the 63.7 gpcd taken from the QA'd metered indoor usage. Thus, only 29.4% of accounts appear to use potable water for irrigation but these accounts use 57.5% or the water.

Table 3. The k-means cluster centroids and water use percentages.

Irrigation Group	Mean GPAD	Standard Deviation GPAD	% of SFR Customers	% of SFR Water Use
1.) Minimal/Offline	156	65	70.6%	42.5%
2.) Mid-range	434	233	25.3%	42.4%
3.) Upper	949	598	4.1%	15.1%
Overall/Total	259	130	100.0%	100.0%

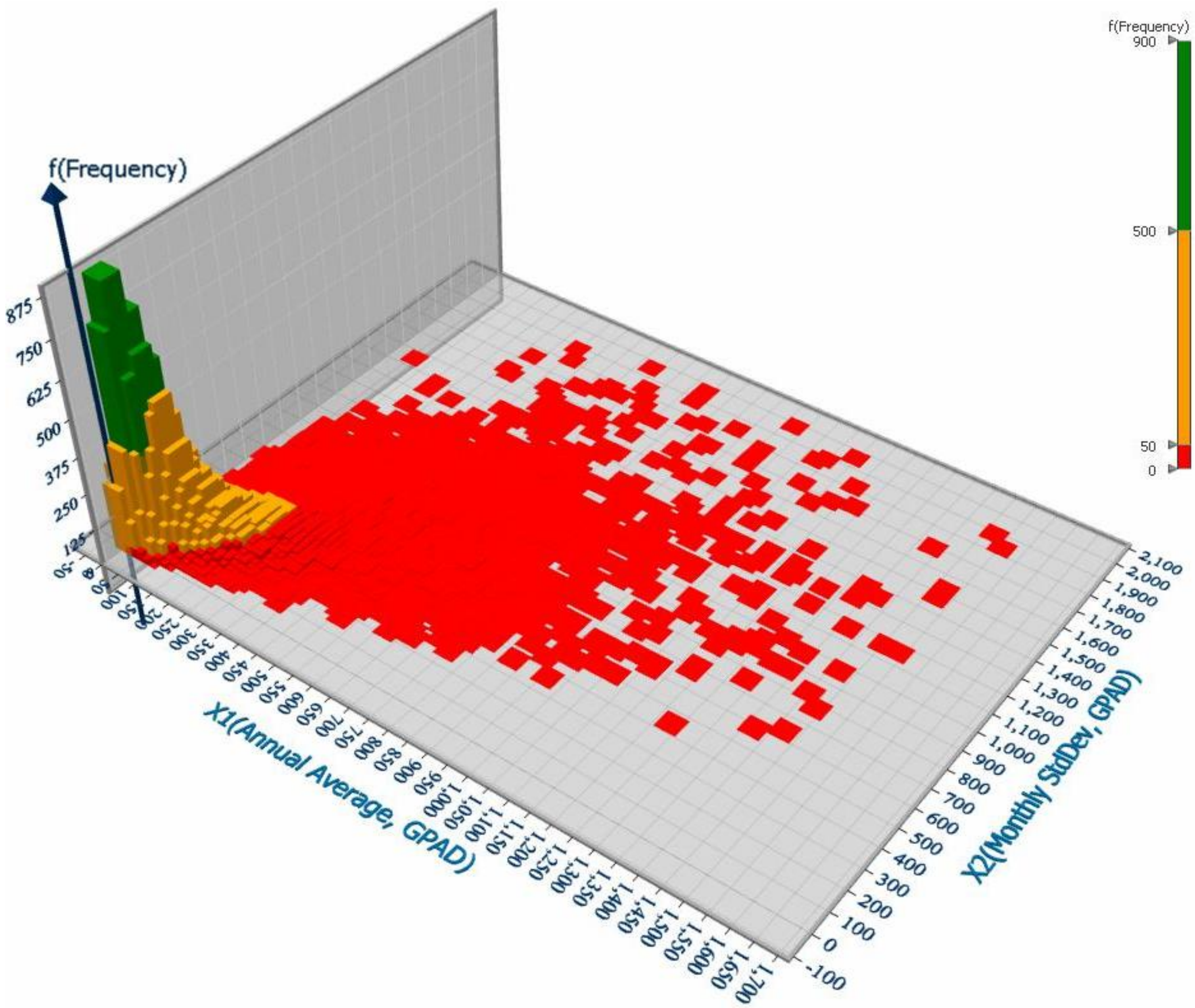

Figure 4. Frequency histogram of the means (x_1) and standard deviations (x_2) of the total monthly flow for all accounts in the Alachua County study area. Red indicates frequencies between 0 and 49; orange indicates frequencies between 50 and 499; green indicates frequencies 500 and over.

Lognormal distributions were fit to each cluster. The marginal probability density functions are shown by the lines in Figure 5. The histogram of the cluster marginal density functions were binned using widths calculated by the Scott (1979) method in Matlab©. Overlaps in the clusters can be observed in Figure 5. By this visualization, the annual average upper bound on minimal/offline use is about 275 gpad (109 gpcd) and the upper bound on monthly standard deviation is about 125 gpad (49 gpcd). These observations do not say that all water use above these bounds is for irrigation. These bounds simply state that as accounts go further past them the use is more likely to be for irrigation. The overall (all clusters) annual average was 259 gpad and the monthly standard deviation was 130 gpad. This indicates that the majority of customers on the system do not use potable water for irrigation as evidenced by the frequency histogram (Figure 4).

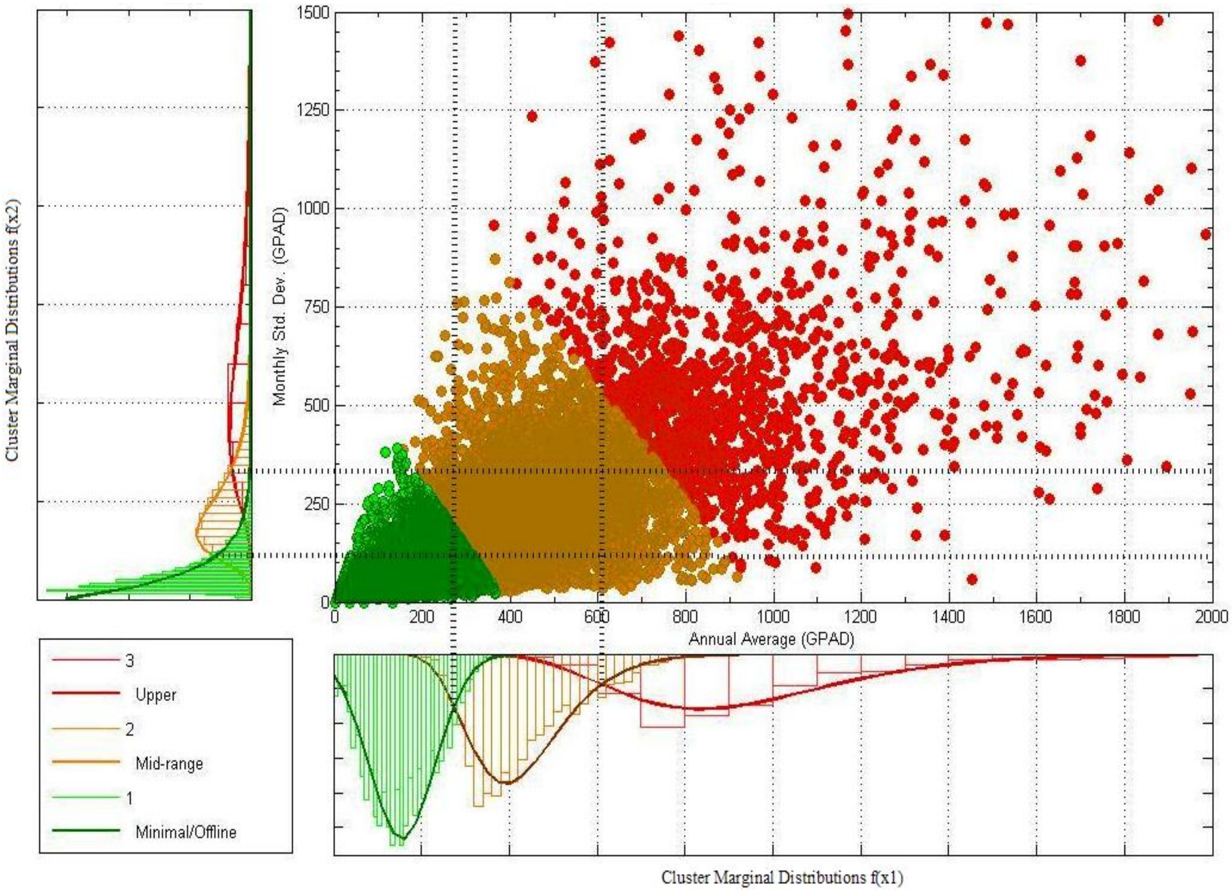

Figure 5. K-means clusters for the overall dataset showing scatter and lognormal marginal distributions of each data cluster: 1) Minimal/Offline, 2) Mid-range, and 3) Upper.

Combining the ACPA with FDOR databases gives extensive data on property and structural attributes for each customer. The ACPA database has several important fields for analysis including:

- Presence of a sprinkler system,
- effective year built,
- stories,
- baths,
- gross area of structure,
- area of drive/walkways,
- patios,
- screened enclosures, etc...

No differentiation is made between automatic or manual systems in this database. The effective year built is the actual year of construction or the year of major renovation. The area data in the ACPA database can be used in conjunction with FDOR data to determine the irrigable area for each parcel. Adding the census data allows the estimation of HHS_{ave} for the clusters. The median of the irrigable area ($A_{irrigable}$) was used to calculate the application rate (AR) in in./yr.

because the mean was skewed by outliers, e.g., the maximum value being 2,015,766 square feet, returning unreasonable irrigable area for SFR parcels. The *AR* was calculated as

$$Q_{irrigation}\left(\frac{gal.}{day}\right)*365.25\left(\frac{day}{yr.}\right)*0.1337\left(\frac{ft^3}{gal.}\right)*A_{irrigable}\left(ft^2\right)*12\left(\frac{in.}{ft.}\right)=AR\left(\frac{in.}{yr.}\right) \quad (3)$$

Where

$Q_{irrigation}$ is calculated using equation 1.

The cluster indoor gpad (Q_{indoor}, gpad) was calculated as the product of the cluster HHS_{ave} and the Cluster 1 gpcd.

Indoor use, Q_{indoor},was assumed using Cluster 1 gpcd as this value was close to the *a priori* assumed value.. It is recommended that a minimum of three clusters be used but more should be added to individual analyses such that the annual average of the minimal cluster is as close to 70 gpcd as possible.

Table 3. Summary of selected cluster attributes form the Alachua County study area

	Cluster 1	Cluster 2	Cluster 3	Overall
Count	21,821	7,811	1,278	30,910
Q_{total}, gpad	156	434	947	259
HHS_{ave}	2.50	2.59	2.63	2.53
Q_{indoor}, gpcd	62.2	62.2	62.2	62.2
Q_{indoor}, gpad	156	161	164	157
$Q_{irrigation}$, gpad	-	273	783	101
% dual-meter accounts	1%	9%	34%	5%
% accounts with sprinklers	17%	47%	76%	27%
$A_{irrigable}$, median ft^2	8,067	9,434	15,594	8,447
AR, in./yr.	-	17.0	29.4	7.0
Effective Area, ft^2	1,889	2,485	3,342	2,100
Effective Year Built	1982	1988	1992	1984
Just Value	$171,186	$241,073	$370,192	$197,074

Cluster 3, comprises newer, larger, and more expensive homes on larger lots. Cluster 3 homes contain the majority of dual-meter accounts and sprinkler systems and have the highest water use and application rate. Somewhere less than about 44% (21,821*17% / 30,910*27%) of homes with in-ground irrigation do not appear to be using potable water for irrigation as evidenced by the percent of accounts with sprinklers in Cluster 1. Overall only about one quater of the accounts appear to use the potable water system for irrigation.

CONCLUSIONS

Total potable water use for the single family residential sector can be separated into its respective indoor and outdoor components using a simple point estimation of the indoor portion.

Additionally, a more disaggregated k-means clustering algorithm may be used to differentiate between levels of outdoor use. A point estimate of indoor use of 70 gpcd and average household size from the U.S. Census can be use to estimate indoor water use. Replacing this value for Q_{indoor} in equation 1, the equation for irrigation use becomes:

$$Q_{irrigation}\text{ (gpad)} = Q_{total}\text{ (gpad)} - 70\text{ gpcd}*HHS_{ave} \tag{4}$$

Outdoor use consists primarily of irrigation and can be broken down into three groups using the k-means clustering algorithm, as described in this paper. These groups are:

1) Minimal/Offline,

2) Mid-range, and

3) Upper.

It is recommended that a minimum of three clusters be used. More clusters may be added to individual analyses such that the annual average of the minimal cluster is as close to the *a priori* value of 70 gpcd as possible.

The first cluster represents those accounts that use little or no irrigation water because they do not irrigate or they irrigate using alternative sources, i.e., reuse water and private wells. These are essentially indoor only customers. The second cluster indicates the average irrigation users. The irrigation use may be calculated using equation 4 or by replacing the indoor 70 gpcd in equation 4 with the gpcd of the first cluster or:

$$Q_{irrigation}\text{ (gpad)} = Q_{total}\text{ (gpad)} - (\text{Cluster 1 gpcd})*HHS_{ave} \tag{5}$$

This result would be the annual average irrigation use in gpad. The third cluster represents the above average users of irrigation water. Their irrigation use can be calculated using equation 4 or equation 5. This group represents the largest potential for savings in irrigation use. The difference between Cluster 3 and Cluster 2 is the saving that may be expected by improving the system efficiency (landscape or irrigation) of this group to the average efficiencies of the area. The potable water savings that can be realized by removal of Cluster 2 and Cluster 3 accounts from the potable water system may be much more than this but perhaps not as much as once thought (Palenchar et al. 2009).

ACKNOWLEDGEMENTS

The guidance and support of the Conserve Florida Water Clearinghouse Advisory Committee is appreciated. The financial support provided by the South Florida, Southwest Florida, and St. Johns River Water Management Districts and the Florida Department of Environmental Protection has been essential. Hillsborough County Water Utility, Tampa Bay Water, and the Gainesville Regional Utility provided essential data to make this study possible.

REFERENCES

Addinsoft. (2009). "XLSTAT, Statistical software for MS Excel: k-means clustering application." http://www.xlstat.com/en/support/tutorials/cluster2.htm.

Aquacraft. (2005). "Water and Energy Savings from High Efficiency Fixtures and Appliances in Single Family Homes." Boulder, CO.

DeOreo, W., Whitcomb, J., Mayer, P., Henderson, J., Gleick, P., Wolfe, G., Nelson, J. and Gauley, B. (2008). "California Single Family Water Use Efficiency Study: Draft Report." Aquacraft, Inc., Water Engineering and Management, Boulder, CO.

Dziegielewski, B., Opitz, E., Kiefer, J., and Baumann, D. (1993). "Evaluation of Urban Water Conservation Programs: A Procedures Manual." California Urban Water Agencies, Sacramento, CA

Everitt, B., Landau, S., and Leese, M. 2001. *Cluster Analysis*, 4th Edition, Arnold Press, London.

Friedman, K. (2009). "Evaluation of Indoor Uraban Water Use and Water Loss Management as Conservation Options in Florida," Thesis (ME), University of Florida, Gainesville, FL.

Han, J., and Kamber, M. (2006). *Data Mining : Concepts and Techniques*, Elsevier, Morgan Kaufmann, Boston, MA

Mayer, P., DeOreo, W., Hayden, M., Davis, R., Caldwell, E., Miller, T., and Bickel, P. (2009). "Evaluation of California Weather-based "Smart" Irrigation Controller Programs." Proposition 13 Urban Water Conservation Outlay Grant Agreements 4600003098 and 4600003099, CUWCC, ed., Aquacraft, Inc., Water Engineering and Management, Boulder, CO.

Mayer, P., and DeOreo, W. (1999). *Residential End Uses of Water*, American Water Works Association Research Foundation, Denver, CO.

Palenchar, J., Friedman, K., and Heaney, J. (2009). "Reuse and Private Wells to Offset Irrigation with Potable Water in Urban Water Systems." *Florida Watershed Journal*, to appear.

Scott, D. W. (1979). "On Optimal and Data-Based Histograms." *Biometrika*, 66(3), 605-610.

Vickers, A. (2001). *Handbook Of Water Use and Conservation.*Waterplow Press, Amherst, MA.

Whitcomb, J. (2005) "Florida Water Rates Evaluation Of Single-Family Homes". South Florida Water Management District, Southwest Florida Water Management District , and St. Johns River Water Management District.

Xu, R. and Wunsch, D. 2009. *Clustering.* J. Wiley and Sons, Hoboken, NJ