Classification, benchmarking, and hydroeconomic modeling of nonresidential water users

MIGUEL A. MORALES[1] AND JAMES P. HEANEY[1]

[1]Department of Environmental Engineering Sciences, University of Florida

The US Environmental Protection Agency and the Water Research Foundation have identified the development of a standardized method for classifying and benchmarking nonresidential water users as a major research need. The methodology proposed in this article uses water utility billing data spatially linked to property-appraisal and business data to arrive at a detailed description of how nonresidential customers use water. Property-appraisal and business databases are available nationwide and provide an extensive, standardized classification scheme through the North American Industry Classification System, along with data on building area, number of employees, and annual sales as measures of size that can be used to develop water use benchmarks. Additionally, this methodology allows coefficients for water use per dollar of economic activity to be developed and incorporated into hydroeconomic models and other tools used to model the interaction between water use and the economy. For this analysis, data on 4,622 nonresidential parcels in Austin, Texas, were used.

Keywords: *benchmarking, commercial water use, hydroeconomic modeling, nonresidential water use, water conservation, water demand planning, water use classification*

Nonresidential—or commercial, industrial, and institutional (CII)—water users account for at least 29% of public water supplied in the United States (Solley et al, 1998). Despite the significant contribution to water demand by these sectors, research on CII water use has lagged behind studies of residential water use. A US Environmental Protection Agency (USEPA, 2009) white paper titled "Water Efficiency in the Commercial and Institutional Sector" summarizes many of the information and research needs related to water use by the commercial and institutional sectors. These findings would also apply to the industrial water-use sector.

The primary challenge in evaluating CII water use is that usage patterns vary widely as a result of the diversity of the CII sectors, which range from small convenience stores to regional shopping malls. The USEPA white paper cites a lack of subsector-specific data, such as facility water use and end use, as well as the lack of benchmarks with which to set targets. Application of subsector information is complicated by the variability in utility classification of CII customers. For example, some utilities might classify restaurants in a separate group from hotels and motels, whereas others might combine both groups in a hospitality category. Other utilities might exclude anything beyond a simple commercial sector on the basis of meter size.

Dziegielewski et al (2000) also highlighted the problem associated with the lack of a standardized classification scheme for CII customers. CII water-use estimates can be expected to vary widely depending on the demographics of the utility and how CII sectors are defined. Thus, comparisons of CII water use across agencies and water utilities are complicated by these dissimilar schemes of classifying customers.

Because of the difficulty of obtaining normalization measures, water use–efficiency benchmarking systems for CII sectors have largely been limited to simple normalizations of water use. Dziegielewski et al (2000) presented efficiency benchmarks for schools, hotels and motels, office buildings, restaurants, and food stores on the basis of the 25th percentile of water-use intensity ratios normalized by a number of variables. Benchmark normalization variables were chosen on the basis of their ability to denote the intensity of activity within given establishments (e.g., meals served by restaurants, occupancies at hotels and motels, transactions at food stores), but these measures of normalization were obtained through costly survey and audit data and are thus difficult to readily apply elsewhere.

Using a variety of normalization variables, Brendle Group Inc. (2007) calculated water use benchmarks based on average values for restaurants, schools, hotels and motels, and nursing and assisted-living facilities in Colorado. Endter-Wada et al (2008) used airborne multispectral imagery, water billing data, and estimates of irrigation need to determine over-irrigation for households and businesses in Layton, Utah, thus demonstrating a need-based approach to benchmarking. The percentage of total water use and other water uses besides irrigation were not addressed, however, and business attributes were arrived at through surveys. Morales et al (2011) joined property-appraisal and water billing databases to arrive at average water use and percentile ratios normalized by building area for the top 16 CII water-use sectors in the state of Florida. Other researchers have also used property-appraisal information to model water use elsewhere, though these studies considered only broad sectors: the commercial sector was modeled in Salt Lake City, Utah, by Stoker and Rothfeder (2014), and the commercial and industrial sectors were modeled by Shandas and Parandvash (2010) in Portland, Ore.

To provide further insight into a possible nationwide CII classification scheme and to improve water use modeling and efficiency benchmarking, this research explored the water use patterns within 4,622 CII parcels in Austin, Texas. Water use in these parcels was linked to county property-appraisal and business databases, which provide physical and economic attributes of interest as well as land use and business classifications. This article compares three classification schemes provided by Austin's water billing database, the city's property appraiser, and six-digit North American Industry Classification System (NAICS) categories from the business database. The NAICS offers a standardized, nationwide taxonomy and is available from various vendors that provide business data across the United States. The links between water billing, property, and business databases also allow for the normalization of water use according to building area, number of employees, and annual sales. This article shows how statistics from these normalized measures of water use can be applied to develop benchmarks with which to target customers and compare water use across utilities.

Additionally, by normalizing water use according to annual sales and by classifying CII customers according to NAICS categories, water use can be readily incorporated into hydroeconomic models (Harou et al, 2009) and other tools used to model the interaction between water use and the economy. For example, application of the input–output approach for resource allocation and environmental analysis dates back to the late 1960s (Miller & Blair, 2009). The approach described in this article builds on the input–output framework by incorporating an additional vector or matrix of activity coefficients that describe the intensity of resource use (in this case, water) per dollar of output within a given industry (Hendrickson et al, 2006).

AUSTIN'S WATER BILLING DATABASE

Austin, Texas, was selected as a case study to demonstrate the data-driven methodology proposed in this article. The city of Austin provided average annual water use data for September 2010 through August 2011 for each of the 17,187 CII premises in its service area. A premise denotes a single address in the billing database, and each premise is described by a code designating the type of premise. A review of the premise types suggests that Austin uses considerable flexibility in designating CII customer classes. Daily CII water demand in Austin averaged 53.4 mgd over the period for which data were provided, a significant 36% of total water use. Though outdoor water use is understood to generally be a fairly small fraction of total CII water use (Morales et al, 2011), this water billing period coincided with stage 1 watering restrictions in Austin, which limited CII irrigation to twice a week at night. The climate in Austin is characterized as humid subtropical, with hot summers and mild winters.

TRAVIS CENTRAL APPRAISAL DISTRICT DATABASE

The city of Austin's property appraiser, Travis Central Appraisal District (TCAD), provides physical and economic attributes for the properties within Travis County, Texas. For this study, the 2010 TCAD tax-roll database, including 289,943 properties (or parcels) in the district, was obtained. The TCAD data contain a series of tables, four of which were manipulated, aggregated to the parcel level, and joined in preparation for linking with Austin's water billing database. For example, TCAD improvement type codes identify 102 land use classifications. A given property or parcel can have multiple land use classifications based on various improvements on that parcel. To create one-to-one matches between the data in the TCAD tables and the water billing database, the land use classification associated with the parcel's improvement of greatest value was used to define that parcel. Thus, the parcel served as the basic unit of analysis. Improvement value was used because it was the only variable available for differentiating among parcel improvements.

Of the 289,943 parcels in the TCAD database, 12,764, or 4.4%, were determined to be classified as CII parcels according to TCAD improvement type codes. In addition to its land-use classification scheme, the TCAD database includes useful attributes for characterizing and normalizing water use, such as main building area, climate-controlled area, parcel area, and year that improvements were built. Unfortunately, the climate-controlled area was seldom populated with data for the CII parcels.

BUSINESS DATABASE

Business databases, which provide beneficial information for characterizing CII water users, are available from a number of vendors. For this analysis, data from a private company[1] were used. This company provides records for nearly 12 million businesses in the United States. Though specific data can be purchased directly from the company, the Environmental Systems Research Institute (ESRI) also provides this company's business data at a cost of $2,400 for a county, $4,800 for a state, and $12,000 for the entire United States.

The private company's data for Travis County, Texas, were obtained through ESRI and were current as of January 2013. These data provided several attributes of potential interest, including four-digit categories from the Standard Industrial Classification system and six-digit NAICS categories, plus information on number of employees, annual sales, building area, and latitude and longitude coordinates as geocoded by ESRI. Data on number of employees, annual sales, and building area can be used as measures of size with which to normalize and compare CII water users, as well as drivers of water use to explain water use patterns. Also, annual sales (an output) can be evaluated as a function of number of employees, building area, or both (two inputs). The latitude and longitude coordinates allow for simple spatial joining to other geodatabases such as that of the TCAD.

Geocoding of addresses is an imperfect operation. Addresses can be assigned to the wrong spatial location and thus be associated with the wrong parcel. The geocoded data provided by ESRI is filtered to include only businesses with a high geocoding score. For Travis County business data, the minimum match score is 85%, with an average match score of 99.5%. In maintaining and adding to its business database, the private company references several sources, including directory listings such as Yellow Pages and business white pages; annual reports; 10-K forms and Securities and Exchange Commission information; federal, state, and municipal government data; business magazines; newsletters and newspapers; and information from the US Postal Service. To ensure accurate and complete information, the company conducts annual telephone verifications with each business listed in its database.

Other academic studies, principally within the field of public health, have examined the validity and uncertainty associated with using business data such as these. Fleischhacker et al (2013) carried out a literature review of 19 studies that looked at the use of business data to identify retail food outlets. These authors concluded that the company whose data were used in the current study had higher validity than Dun & Bradstreet, the only other provider of business data studied. In terms of validity of the data, undercounting of businesses appeared to be the primary concern. Liese et al (2013) reported that the private company undercounted supermarkets and grocery stores by 29%. When the accuracy of the classification assignment was considered, the undercounting increased to 39%.

DATABASE JOINS

The city of Austin provided a geodatabase[2] that maps Austin Water's premise identification (ID) and TCAD's property ID. This mapping platform was used to join Austin's water use database with TCAD's property attribute database. Table 1 shows the number of premises matched in the various databases. The geodatabase join that provided the TCAD's property ID link reduced the premise count matches to 87% and the water use matches to 85% of the premises available from the water billing database. The data loss in this join was mostly attributable to incomplete geocoding of water meters. The subsequent join with the TCAD database resulted in matches of 70% of CII premises, accounting for 58% of total CII water use. The significant reduction of premise count matches resulting from the TCAD join was largely attributed, through correspondence with TCAD personnel, to the exclusion of tax-exempt parcels from the TCAD database. The private company's business data were linked with the property appraiser's parcels by assigning the geocoded business points from the private company's database to the nearest geographic information system's polygon parcels from the property appraiser's data by means of a spatial join. Joining the Austin Water–TCAD database with the private company's database reduced the number of linked premises to 48% from 70% and the water use matches to 49% from 58% (Table 1).

The loss of data associated with the various database joins highlights a practical limitation of this data-driven approach in that it generally provides only a nonrandom sample of all the CII customers of a given utility. This issue is of concern in overall coverage of a utility's customers and at the parcel level. For example, in the water billing database, all of the utility's water meters might not be geocoded and linked to parcels, and thus all the water meters on any given parcel might not be accounted for. Incomplete data at the parcel level are of particular concern because parcel attributes from the property appraisal and business databases are thus associated with only a fraction of actual water use on that parcel, and this affects the development of metrics and benchmarks. Similarly, at the utility level, the TCAD database is largely missing attributes for tax-exempt parcels.

The private company's database provided only a nonrandom sample of all of the businesses served by the utility, so the business attributes of all the businesses on a parcel might not have been accounted for. Still, the commonality of building area, a primary measure of size, in both the TCAD and the private company's databases permitted validation or scaling up of the private company's data. When the TCAD database join was used as a baseline, the overall undercounting shown by the private company's data was 32%, which is similar to the 29% reported by Liese et al (2013). The water use and building area statistics provided in Table 1 hint at the representativeness of the data samples available through each database join. The overall trend was that as databases were joined and the sample size was reduced, the data were skewed slightly toward CII customers with greater water use and larger building areas.

TABLE 1 Comparison of CII premises, water use, and building areas associated with sequential joins of water billing, property ID–GIS linkage, TCAD, and private company* databases for Austin, Texas

Premise and Parcel Count Matches	Water Billing Database	Property ID Link	TCAD Database Join	Private Company* Database Join
Premise count	17,187	14,921	12,086	8,208
Percent matched	100%	87%	70%	48%
Water use—*mgd*	53.40	45.64	31.13	26.24
Percent matched	100%	85%	58%	49%
Average premise water use—*kgal/month*†	256 (730)	269 (744)	311 (778)	439 (930)
Parcel count	N/A	N/A	8,418	5,230
Average parcel building area—*sq ft*†	N/A	N/A	37,413 (51,963)	57,427 (56,992)

CII—commercial, industrial, and institutional, GIS—geographic information system, ID—identification, N/A—not applicable, TCAD—Travis Central Appraisal District
* Infogroup, Papillion, Neb.
†Standard deviations shown in parentheses

ANALYSIS OF CII CLASSIFICATIONS

The results of joining the water billing, TCAD property appraisal, and private company business databases were discussed in the preceding section. This section focuses on the use of these joined databases to analyze the classification of Austin's CII water users. Three distinct classification schemes were evaluated. Austin's water billing database provided 43 categories of premise type; the TCAD database offered 82 improvement type codes identifying land use classifications, and the private company's data included all 24 two-digit NAICS categories. Because this analysis was carried out at the parcel level and multiple premises can be associated with a single parcel, joining to the TCAD classifications required that premises be aggregated up to the level of the parcel, the basic unit of this analysis.

To assign a primary classification to a given parcel, primary premise types were associated with the type of premise exhibiting the greatest water use on that parcel. Similarly, primary TCAD improvement type codes were assigned to parcels on the basis of highest appraised value, the only available metric for differentiation. In the private company's classification scheme, six-digit NAICS categories were presented at the business level. The hierarchical structure of NAICS codes allowed the classifications to be truncated down to the two- and three-digit level, where primary classifications were assigned on the basis of highest number of employees. If a tie occurred when the classifications were based on number of employees, the largest building area was subsequently used to determine primary classification, followed by the greatest annual sales. A minimum resolution of parcel-level analysis was required, given the uncertainty associated with geocoding meters in the water billing database and businesses in the private company's database and given that the parcel is generally the smallest geographic unit provided by property appraisal databases.

The CII classifications were evaluated using the functional relationship between water use and building area. This approach was determined to be more appropriate than using measures of homogeneity (e.g., coefficients of variation) because of the heterogeneous nature of CII customers. The primary concern in evaluating CII classifications for use by utilities is the relationship between water use and size, not necessarily how homogeneous the building areas within a given sector are.

Prior to this evaluation, outliers were identified by means of the interquartile range method, in which any value more than 1.5 times the interquartile range below the first quartile or above the third quartile is considered an outlier (Navidi, 2010). Because the interquartile range method is designed for normal distributions and because water use, property, and business attributes were found to follow lognormal distributions, the natural logarithm of each variable was taken to convert to normal distributions so that the method could be effectively applied. Though outliers were identified at the sector level, on average 12% of parcels were removed as outliers. Following the removal of outliers, linear and power functions were fit for each sector by means of regression. Linear fits were forced through the origin to be compared with the power fits, which also pass through the origin. Table 2 provides the respective R^2 values of these fits for the top 10 water use sectors in each of the three classification schemes. As a measure of importance, the total water use per sector is also shown in Table 2.

The top 10 CII water use sectors in Austin's premise type classifications, TCAD's improvement type classifications, and NAICS's two-digit categories used 89, 61, and 87% of total CII water use in Austin, respectively. TCAD's improvement type classifications reasonably accounted for less water use in its top 10 sectors, given that it consists of 82 sectors compared with Austin's 43 premise type sectors and NAICS's 24 two-digit sectors. Austin's Manufacturing and Industrial Building sector used the most water, accounting for nearly 28% of total CII water use. Likewise, the counterpart manufacturing sectors from TCAD and NAICS were the largest water-using sectors within those classification schemes. To a large extent, TCAD's sectors are more specific than those in the other two classification schemes, leading to drastically lower parcel counts. For example, TCAD's Major Industrial Manufacturing sector comprised only five parcels, compared with Austin's 61 manufacturing parcels and the 118 parcels in NAICS's two-digit manufacturing sectors.

Overall, the relationship between water use and building area had an R^2 value of 0.32 using a linear fit through the origin and an R^2 value of 0.35 using the power function fit. Certain R^2 values were negative because R^2 compares the fitted model with the average of the data, as shown in Eq 1. If the model yields a worse fit than simply taking the average of the data (the regression sum of squares is greater than the total sum of squares), the R^2 value is negative. The heterogeneity of CII sectors is apparent in Table 2 with regard to the appropriateness of a given functional relationship. A negative R^2 value for the linear fit proves the inappropriateness of that functional relationship when an average value provides a better fit. The power-function relationships, by definition, provide better fits compared with the linear functional relationships forced through the origin.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \tag{1}$$

in which R^2 = the coefficient of determination, SS_{res} = the regression sum of squares or explained sum of squares, SS_{tot} = the total sum of squares, y_i = the data set of values, f_i = the modeled values, and $\bar{y}$ = the mean of the dataset values.

To facilitate comparing the three classification schemes analyzed, Table 3 provides summary power-function R^2 statistics for the relationship between water use and TCAD building area. To ensure that small sectors do not skew the results, the statistics in this table include only sectors with a minimum sample size of 30 parcels. Austin's premise type classification scheme had 16 sectors that met this criterion and an average R^2 value of 0.37. The two-digit NAICS categories had 20 sectors that met the minimum sample criterion, an average R^2 value of 0.45, and a reduced standard deviation compared with Austin's premise type scheme. Despite TCAD's larger number of sectors, and thus finer resolution among classifications, it provided the poorest average R^2 value at 0.23. As a measure of variability, the range of R^2 values is also shown in Table 3.

TABLE 2 Linear and power R^2 values for the relationship between water use and building area in the top ten CII water use sectors within the three classification schemes evaluated for Austin, Texas

Classification System	Sector Description	Parcel Count*	Water Use Versus TCAD Building Area: Linear R^2	Water Use Versus TCAD Building Area: Power R^2	Total Water Use *mgd*	Total Water Use *%*
Austin's Premise Type	Manufacturing/Industrial Building	61	0.21	0.22	7.21	27.5
	Public Works/Utilities	17	– 0.19	0.0008	3.32	12.7
	Commercial Irrigation/Sprinklers	372	0.40	0.43	2.78	10.6
	Office Building	985	0.67	0.70	2.6	9.9
	Reclaim Water	2	N/A	N/A	1.3	5.0
	Retail/Wholesale	662	0.21	0.24	1.29	4.9
	Hotel/Motel	66	0.66	0.69	1.28	4.9
	Other Meter Usage Apartment	290	0.36	0.38	1.23	4.7
	Restaurant/Bar/Lounge	387	– 0.08	0.14	1.18	4.5
	Service Establishments	636	0.08	0.11	1.04	4.0
	Other 33 premise types	1,108	0.05	0.13	2.99	11.4
TCAD's Improvement Code	Major Industrial–Manufacturing	5	0.90	0.90	3.46	13.2
	Office Building (> 35,000 sq ft)	146	0.65	0.66	3.05	11.6
	Major Industrial–Engineering	3	0.16	0.30	2.47	9.4
	Single-Family Dwelling	287	0.02	0.02	2.4	9.1
	Industrial (> 20,000 sq ft), < 25% FO	145	0.28	0.28	0.94	3.6
	Major Industrial–Office	4	0.79	0.81	0.92	3.5
	Office Building (10,000–35,000 sq ft)	178	0.15	0.17	0.8	3.0
	Commercial Strip Center (> 10,000 sq ft)	148	0.17	0.21	0.75	2.9
	Office Building Hi-rise (≥ 6 stories)	35	0.32	0.33	0.65	2.5
	Restaurant	152	0.22	0.26	0.61	2.3
	Other 72 improvement codes	3,424	0.25	0.27	10.18	38.8
Two-Digit NAICS Code	33: Manufacturing	118	0.75	0.75	4.66	17.8
	56: Administrative and Support Services	281	0.30	0.38	3.89	14.8
	72: Accommodation and Food Services	482	0.35	0.43	3.05	11.6
	54: Professional, Scientific, and Technical Services	499	0.66	0.69	2.8	10.7
	44: Retail Trade	586	0.43	0.43	1.75	6.7
	81: Other Services	640	0.10	0.19	1.71	6.5
	23: Construction	236	0.41	0.43	1.65	6.3
	62: Health Care and Social Assistance	363	0.22	0.31	1.59	6.1
	42: Wholesale Trade	290	0.59	0.59	0.83	3.2
	45: Retail Trade	186	0.17	0.17	0.8	3.0
	Other 14 two-digit NAICS codes	893	0.33	0.34	3.51	13.4
	Total	4,622	0.32	0.34	26.24	100.0

CII—commercial, industrial, and institutional, FO—finished out, NAICS—North American Industry Classification System, TCAD—Travis Central Appraisal District

*Parcel sample count after the removal of outliers

TABLE 3 Summary power fit R^2 statistics for sectors with a sample size of at least 30 parcels in classification schemes analyzed for Austin, Texas

Sectors With Sample Sizes of ≥ 30 Parcels	Austin's Premise Type	TCAD Improvement Code	Two-Digit NAICS
Number of sectors	16 of 43	27 of 82	19 of 24
Average R^2 value*	0.37 (0.30)	0.23 (0.20)	0.45 (0.20)
R^2 range	0.04–0.97	0.00–0.66	0.12–0.82

NAICS—North American Industry Classification System, TCAD—Travis Central Appraisal District

*Standard deviations shown in parentheses

As shown in Tables 2 and 3, the NAICS compares favorably with the two other classification schemes analyzed. In addition to being the only classification system that is standardized throughout the United States, the NAICS offers the additional advantage of allowing finer resolution of classifications through its hierarchical coding system, which includes codes of up to six digits. Water use statistics for the top 15 three-digit NAICS sectors in Austin, which account for 70% of total parcels and 81% of total CII water use, are provided in Table 4. Because three-digit NAICS categories are more specific than two-digit NAICS categories, one would expect the relationship between CII building area and CII water use sectors to generally strengthen. This is demonstrated in Table 4, in which the R^2 value of the power fit relationship between water use and building area is shown to increase in 11 of the top 15 three-digit NAICS sectors compared with their two-digit counterparts. Sectors in which the R^2 value decreased with their three-digit classification likely had strengthening water use relationships elsewhere within their broader two-digit classifications. For example, category 722—Food Services and Drinking Places shows a sharp decrease in R^2, though category 721—Accommodation exhibits an even greater increase in R^2.

BENCHMARKING

To characterize the top 15 three-digit NAICS sectors in Austin, average building areas obtained from the TCAD database and average water use data (with standard deviations in parentheses) are shown in Table 5. Even at this level of disaggregation, the large standard deviations compared with the mean values indicate considerable heterogeneity among the customers within the various sectors. Across sectors, this heterogeneity appears to be greater in terms of water use compared with building area. Weighted average water use coefficients, developed by summing the average water use of all parcels within a given sector and dividing by their total building area (Eq 2), are also shown in Table 5. Weighted water use coefficients are often used by researchers (Morales et al, 2011; Dziegielewski et al, 2000) to account for the wide range in size of CII customers. Building area is the chosen measure of normalization, given that Morales et al (2011) showed that it is the variable most highly correlated with total water use as opposed to parcel area or year built.

$$\bar{q}_j = \sum_{i=1}^{n} Q_{ij} / \sum_{i=1}^{n} BA_{ij} \tag{2}$$

in which $\bar{q}_j$ = the coefficient for weighted water use per building area in subsector j (in gallons per square foot per day), n = number of parcels in subsector j, Q_{ij} = the average water use of parcel i in subsector j (in gallons per day), and BA_{ij} = the square footage of all buildings on parcel i in subsector j.

TABLE 4 R^2 power fit comparisons of water use versus building area for the top 15 three-digit NAICS sectors in Austin, Texas, and their counterpart two-digit NAICS sectors

Three-Digit NAICS Code	Three-Digit NAICS Sector Description	Parcel Count*	Water Use Versus TCAD Building Area—Power Fit R^2	R^2 of Two-Digit Counterpart	Change %	Total Water Use *mgd*	Total Water Use %
334	Computer and Electronic Product Manufacturing	40	0.96	0.75	26.9	4.5	17.1
561	Administrative and Support Services	275	0.4	0.38	7.3	3.87	14.7
541	Professional, Scientific, and Technical Services	503	0.7	0.69	2.0	2.84	10.8
722	Food Services and Drinking Places	394	0.22	0.43	– 49.2	2.11	8.0
238	Specialty Trade Contractors	159	0.6	0.43	39.2	1.52	5.8
621	Ambulatory Health Care Services	204	0.45	0.31	46.2	1.11	4.2
721	Accommodation	68	0.75	0.43	76.1	0.96	3.7
423	Merchant Wholesalers, Durable Goods	193	0.72	0.59	21.4	0.66	2.5
445	Food and Beverage Stores	183	0.15	0.43	– 65.9	0.64	2.5
811	Repair and Maintenance	280	0.21	0.19	11.7	0.61	2.3
812	Personal and Laundry Services	171	0.28	0.19	52.0	0.57	2.2
813	Religious, Grantmaking, Civic, Professional, and Similar Organizations	170	0.4	0.19	113.7	0.55	2.1
452	General Merchandise Stores	24	0.27	0.17	52.5	0.46	1.8
522	Credit Intermediation and Related Activities	108	0.54	0.55	– 2.4	0.44	1.7
531	Real Estate	164	0.11	0.12	– 10.5	0.44	1.7
	Other NAICS classifications	1,593	0.44	0.34	26.8	4.97	18.9
	Total	4,622	0.34	0.34	0.0	26.24	100.0

NAICS—North American Industry Classification System, TCAD—Travis Central Appraisal District

*Parcel sample count after removal of outliers

Average water use coefficients normalized by TCAD building area can also be calculated at the parcel level, allowing the determination of percentiles that can be used for benchmarking. For example, if a CII customer's water use per building area is above the 75th percentile in its respective sector, this customer would be a prime candidate for further investigation because it uses water more intensely than the majority of its peers. Thus, a utility can use the percentiles for water use per building area shown in Table 5 to target customers within the upper percentiles for water conservation or as a means of identifying customers in the lower percentiles for meters that might be under-recording water use. On the other hand, CII customers with high water use rates may be more successful enterprises that have their end-use devices used more often.

Table 5 also provides the regression-fitted linear and power functions between water use and TCAD building area for the top

TABLE 5 Building area and water use statistics for the top three-digit NAICS sectors with the greatest water use in Austin, Texas

Three-Digit NAICS Code	Parcel Count*	TCAD Average Building Area *sq ft*†	Average Water Use *gpd*†	Average Weighted Water Use Coefficient *gal/sq ft/d*	Average Water Use Coefficient (*gal/sq ft/d*) *Percentiles*			Linear and Power Fits
					25th	50th	75th	Water Use (gpd) Versus TCAD Building Area (sq ft)
334	40	126,318 (311,227)	23,730 (106,638)	0.188	0.0215	0.0598	0.0888	L: m = 0.291; R^2 = 0.79 P: a = 9.42e-7: b = 1.89; R^2 = 0.96
561	275	26,460 (55,035)	2,099 (4,533)	0.0793	0.0188	0.0528	0.154	L: m = 0.0552; R^2 = 0.34 P: a = 3.57: b = 0.657; R^2 = 0.40
541	503	31,355 (67,399)	2,159 (4,284)	0.0689	0.031	0.0668	0.133	L: m = 0.0556; R^2 = 0.68 P: a = 0.769: b = 0.788; R^2 = 0.70
722	394	7,470 (9,613)	2,687 (3,535)	0.36	0.193	0.36	0.657	L: m = 0.236; R^2 = 0.08 P: a = 28.9: b = 0.524; R^2 = 0.22
238	159	18,908 (30,091)	756 (1,291)	0.04	0.0146	0.0302	0.0663	L: m = 0.0351; R^2 = 0.59 P: a = 0.241: b = 0.83; R^2 = 0.60
621	204	15,700 (23,991)	1,944 (4,178)	0.124	0.0507	0.0888	0.155	L: m = 0.119; R^2 = 0.45 P: a = 0.23: b = 0.941; R^2 = 0.45
721	68	60,445 (64,344)	12,811 (14,665)	0.212	0.145	0.193	0.259	L: m = 0.204; R^2 = 0.74 P: a = 0.726: b = 0.894; R^2 = 0.75
423	193	36,120 (72,983)	2,008 (5,088)	0.0556	0.0136	0.0327	0.0775	L: m = 0.0584; R^2 = 0.72 P: a = 0.0663: b = 0.99; R^2 = 0.72
445	183	5,679 (7,953)	1,286 (1,889)	0.226	0.0851	0.184	0.337	L: m = 0.135; R^2 = 0.02 P: a = 20.2: b = 0.497; R^2 = 0.15
811	280	6,754 (7,642)	381 (586)	0.0565	0.0183	0.0372	0.0783	L: m = 0.044; R^2 = 0.16 P: a = 0.981: b = 0.687; R^2 = 0.21
812	171	10,204 (17,821)	1,386 (2,018)	0.136	0.0709	0.125	0.238	L: m = 0.0688; R^2 = 0.02 P: a = 23.5: b = 0.469; R^2 = 0.28
813	170	12,554 (20,175)	1,163 (1,746)	0.0926	0.0341	0.0846	0.167	L: m = 0.0646; R^2 = 0.33 P: a = 2.84: b = 0.656; R^2 = 0.40
452	24	85,079 (74,733)	12,505 (24,295)	0.147	0.0407	0.0833	0.161	L: m = 0.153; R^2 = 0.24 P: a = 7.71e-7: b = 2.01; R^2 = 0.27
522	108	10,584 (14,712)	1,785 (2,899)	0.169	0.0456	0.123	0.311	L: m = 0.153; R^2 = 0.54 P: a = 0.235: b = 0.96; R^2 = 0.54
531	164	71,913 (100,422)	2,197 (4,658)	0.0305	0.00699	0.0252	0.118	L: m = 0.0226; R^2 = 0.02 P: a = 55.8: b = 0.355; R^2 = 0.11
Other 74	1,593	25,648 (51,980)	2,085 (4,875)	0.0813	0.0248	0.0612	0.157	L: m = 0.0657; R^2 = 0.43 P: a = 0.372: b = 0.858; R^2 = 0.44
Total	4,622	23,065 (45,973)	2,245 (5,478)	0.0973	0.0281	0.0755	0.193	L: m = 0.0747; R^2 = 0.32 P: a = 0.824: b = 0.8; R^2 = 0.34

NAICS—North American Industry Classification System, TCAD—Travis Central Appraisal District

*Parcel sample count after removal of outliers
†Standard deviations shown in parentheses

15 three-digit NAICS sectors in Austin. The general linear and power functions used are shown in Eqs 3 and 4. Because the linear fits are forced through the origin, the power fits always provide a better fit, given the added exponent parameter. Of the top 15 water use sectors, 13 also display power fits with exponents < 1, indicating the presence of diminishing marginal returns within most CII sectors because the rate of water use per building area decreases as building areas increase. Despite the persistent heterogeneity of building area and water use within sectors, the modeled relationships of these two variables remain strong. For the top three-digit NAICS sectors, the power fit R^2 values ranged between 0.96 for Computer and Electronic Product Manufacturing and 0.11 for Real Estate.

$$f_i = mx_i \text{ (Linear)} \tag{3}$$

$$f_i = ax_i^b \text{ (Power)} \tag{4}$$

in which f_i = the modeled values, x_i = the dataset's independent values, m = the slope parameter for linear fit forced though the origin, a = the factor parameter for power fit, and b = the exponent parameter for power fit.

Similar characterization and benchmarking statistics, derived from the business data provided through the join with the private company's data, are provided in Table 6. As a means of ensuring that all businesses within a given parcel were accounted for in the statistics, the analysis included only parcels in which the building areas in the databases of the TCAD and the private company matched within 10% (an assumed level of allowable variability). This further reduced the sample sizes within the sectors analyzed, skewing samples toward smaller building areas with lower water use, compared with the statistics shown in Table 5.

Average number of employees and annual sales (with standard deviations in parentheses) provide an additional means of characterizing CII water users and one that is more closely tied to economic activity. For example, this information helps identify which customers might have high water use per building area as a result of increased business as opposed to inefficient water use fixtures. A comparison of the R^2 statistics in Table 6 with those in Table 5 shows that the relationship between water use and building area is generally stronger than the relationship between water use and number of employees. A few key exceptions are in the service sectors of Food Services and Drinking Places, Ambulatory Health Care Services, and Personal and Laundry Services. Diminishing marginal returns are prevalent in the relationship between water use and number of employees, with 12 of the top 15 three-digit NAICS sectors in Austin showing power function fits with an exponent < 1.

APPLICATION TO HYDROECONOMIC MODELS

The statistics in Table 6 can also be used outside of traditional water use planning and conservation analysis. The availability of NAICS classifications and annual sales from the private company's data allows these statistics to be applied to macro-level economic models, such as those based on the input–output approach. Input–output analysis was first developed in the 1930s by economist Wassily Leontief, who used it to model the entire economy of the United States. The approach divides an economy into distinct sectors and relies on large matrixes that describe the monetary interactions among sectors by means of sales.

Two general schools of thought follow for incorporating resource allocation into the input–output framework. Leontief (1970) suggested an additional row and column be appended to the A matrix, through which resources and environmental factors are essentially treated as additional sectors. Hendrickson et al (2006) argue that externally augmented models, which simply use the output vector from an input–output analysis, yield results that are equivalent to Leontief's 1970 proposal while reducing computational requirements. The externally augmented model formulation is shown in Eq 5, in which R_i is a matrix with diagonal elements representing resource use per dollar of output, and the result, b_i, is a vector of resource use. The numbers for average water use and annual sales shown in Table 6 can be divided to arrive at coefficients of weighted water use per dollar, and these coefficients can be used to populate the R_i matrix for any specific area of study. Additionally, coefficients based on number of employees can be calculated for models used to examine the effect of water resource allocation on employment.

$$b_i = R_i\,[I - A]^{-1}\,y \tag{5}$$

in which b_i = the vector for resource use (in gallons per day); R_i = the matrix with diagonal elements representing resource use per dollar of output (in gallons per dollar per day); I = the n by n identity matrix; A = the matrix of a_{ij} normalized, intersector interaction coefficients, in which a_{ij} terms equal the input from sector i to sector j, divided by the total output of sector j; and y = the vector for final demand (in dollars).

This methodology, which uses county property appraisal and business data, also relies on a geographic framework that facilitates incorporation of the method into hydroeconomic models where the spatial distribution of supply and demand is an essential component (Harou et al, 2009). Parcel-level allocation of water demand estimates afforded by this approach uses a fine spatial resolution, which can be readily aggregated up to coarser geographic units of other, more macro-level, hydroeconomic models. The link to hydroeconomic models allows for solution-oriented tools that can help decision-makers develop institutional policies to plan for future infrastructure expansion and water allocation, achieve environmental and economic goals, and better evaluate the economic impact of these policies.

SUMMARY, CONCLUSIONS, AND FUTURE WORK

The linking of water billing, property appraisal, and business databases was evaluated in this research, with Austin, Texas, used as a case study. Overall, joining these three databases resulted in a match of 48% of premises and 49% of water use within Austin's CII water use sectors. On the basis of this sample, the various classification schemes provided by the three data sources were evaluated using water use statistics. The classification scheme of the NAICS, provided by the business database, provided results that were at least comparable to those of the two other classification schemes analyzed. NAICS is the only classification scheme that is standardized throughout the United States, and it permits varying degrees of class aggregation on the basis of a hierarchical coding structure in which the codes range from two to six digits. For this case study, business data were obtained from a private company that provides data throughout the United States, lending broad applicability to this approach.

This study also investigated useful applications for the joining of these databases. The availability of statistics on water use, building area, employment, and annual sales allows the development of benchmarking metrics that can be used to target

TABLE 6 Business, building area, and water use statistics for the top three-digit NAICS sectors with the greatest water use in Austin, Texas

Three-Digit NAICS Code	Parcel Count*	TCAD Average Building Area *sq ft*†	Average Number of Employees†	Average Annual Sales *$1,000*†	Average Water Use *gpd*†	Average Weighted Water Use Coefficient *gal/emp/d*	Average Water Use Coefficient (*gal/employee/d*) *Percentiles*			Linear and Power Fits
							25th	50th	75th	Water Use (gpd) Versus Number of Employees
334	7	40,320 (52,651)	83.0 (116.2)	60,152 (99,653)	1,889 (2,242)	22.8	17.4	30.6	84.1	L: m = 14.6; R^2 = 0.08 P: a = 505: b = 0.362; R^2 = 0.39
561	34	7,229 (8,783)	22.9 (42.6)	1,711 (3,239)	1,304 (2,254)	56.8	7.96	26.8	114	L: m = 29.8; R^2 = 0.07 P: a = 416: b = 0.465; R^2 = 0.26
541	63	11,595 (17,686)	28.2 (57.5)	4,130 (10,044)	796 (1,266)	28.3	18.2	30.1	48	L: m = 19.2; R^2 = 0.54 P: a = 129: b = 0.641; R^2 = 0.68
722	101	4,475 (3,564)	25.9 (19.9)	1,297 (1,689)	1,957 (1,715)	75.4	46.7	74.4	112	L: m = 72.1; R^2 = 0.58 P: a = 86.2: b = 0.954; R^2 = 0.58
238	13	6,779 (6,769)	26.8 (24.5)	3,394 (4,976)	271 (300)	10.1	3.75	9.32	11	L: m = 8.75; R^2 = 0.29 P: a = 26: b = 0.728; R^2 = 0.32
621	27	4,684 (4,385)	7.5 (7.4)	$655 (1,261)	496 (658)	66.3	29.2	52.2	81.5	L: m = 66.1; R^2 = 0.54 P: a = 78: b = 0.943; R^2 = 0.54
721	16	61,103 (41,370)	27.3 (26.9)	1,839 (2,553)	13,230 (13,531)	484	325	473	636	L: m = 406; R^2 = 0.35 P: a = 1530: b = 0.681; R^2 = 0.44
423	28	23,935 (35,476)	24.0 (26.7)	9,164 (25,873)	1,266 (2,833)	52.8	11	24.7	52	L: m = 68.7; R^2 = 0.57 P: a = 2.26: b = 1.78; R^2 = 0.65
445	24	3,019 (1,164)	6.3 (4.0)	1,258 (2,216)	1,212 (1,300)	193	62	143	323	L: m = 161; R^2 = -0.03 P: a = 571: b = 0.435; R^2 = 0.069
811	66	5,191 (4,161)	7.1 (7.3)	555 (761)	336 (587)	47.3	13.9	29.5	78.9	L: m = 34.3; R^2 = 0.02 P: a = 126: b = 0.545; R^2 = 0.08
812	26	8,327 (9,049)	12.4 (19.4)	1,228 (3,649)	2,651 (7,376)	213	43	93.9	280	L: m = 304; R^2 = 0.78 P: a = 1.07: b = 2.29; R^2 = 0.93
813	36	8,562 (9,581)	8.0 (9.3)	1,787 (3,512)	703 (785)	87.5	34	68.7	174	L: m = 56.8; R^2 = -0.03 P: a = 339: b = 0.411; R^2 = 0.16
452	1	118,694	150	31,955	6,159	41.1	N/A	N/A	N/A	N/A N/A
522	18	5,450 (2,894)	12.1 (10.5)	4,111 (3,476)	816 (887)	67.4	16.9	51.5	97.5	L: m = 49.6; R^2 = -0.07 P: a = 278: b = 0.467; R^2 = 0.15
531	9	14,398 (19,889)	12.4 (8.5)	846 (559)	915 (752)	73.5	41.7	74.2	97.7	L: m = 67; R^2 = 0.3 P: a = 193: b = 0.642; R^2 = 0.36
Other 74	293	16,945 (23,655)	25.2 (41.0)	6,935 (22,248)	1,509 (2,835)	59.9	15	36	109	L: m = 41.2; R^2 = 0.2 P: a = 241: b = 0.634; R^2 = 0.27
Total	821	13,120 (21,296)	23.3 (40.4)	3,305 (7,989)	1,601 (3,504)	68.6	18.5	47.8	113	L: m = 43.2; R^2 = 0.12 P: a = 316: b = 0.595; R^2 = 0.2

NAICS—North American Industry Classification System, N/A—not applicable, TCAD—Travis Central Appraisal District

*Parcel sample count after removal of outliers; includes only parcels in which the building area from the TCAD database and the private company' from which business data were obtained were within 10% of each other
†Standard deviations shown in parentheses

customers for water conservation or meter replacement and to compare water use across utilities. This information can also be applied in the development of coefficients for water use per dollar of economic activity. Furthermore, these coefficients can be incorporated into macro-level economic models, such as those based on the input–output approach, or hydroeconomic models, which evaluate water resource management options in a manner that is integrated with economic values.

This article describes a data-driven methodology and its applications. Though outdoor water use has been shown to constitute a fairly small fraction of CII water use (Morales et al, 2011), its significance varies by sector and region. The water use statistics provided in this article should be applied with caution outside of the Austin, Texas, area because of a level of uncertainty that cannot be quantified until more utilities across the country apply this approach. The intent of the authors is for this methodology to be applied elsewhere, where site-specific water use statistics can be developed, applied in utility or regional modeling or benchmarking efforts, and compared with the values presented for Austin.

Future work should address how this data-driven approach applies to water-demand modeling and should add to the understanding of how coefficients of water use intensity change over time. Multivariate regression should be used to analyze the effect of multiple predictor variables, and stepwise regression or other statistical tools should be used to determine which variables are significant predictors of water use within a given sector. Additionally, many variables that could be significant drivers of CII water use were not addressed in this work (e.g., price, climate, customer attitudes), and their effects should be evaluated. Total water use was not disaggregated in this study, but future work using submetering along with estimates of irrigated area and cooling requirements would provide greater insight into CII end uses. The use of submetering is suggested because the separation of indoor and outdoor water use is complicated in the CII sectors, given the influence of other seasonal drivers besides climate (e.g., vacation time, holidays, and shopping seasons). The method of removing data outliers could be improved, especially when it is applied at the utility level to identify customers to target. It should also be understood that outliers removed for benchmarking are obvious targets for further investigation. Though the authors believe the approach described in this article is applicable nationwide, the availability of parcel-level, county property-appraisal databases that can be spatially joined to water billing and proprietary business databases should be assessed because there might be regional differences in the availability and quality of these data.

In general, the data-driven approach described in this article advances the field of modeling urban water demand by providing a means of gaining greater insight into the highly variable potable water use of CII sectors. The use of business data offers a nationwide classification scheme that can explain much of this variability across sectors. Because the methodology described in this article contributes to a better understanding of urban water use and its link to the economy, it should improve water demand and hydroeconomic modeling.

ACKNOWLEDGMENT

The authors appreciate the support of the Water Research Foundation and the advice of Jack Kiefer and Lisa Krentz of Hazen and Sawyer. Neither the foundation nor Hazen and Sawyer has reviewed or endorsed the contents of this article; only the authors are responsible for its content. The cooperation of Austin Water and Travis Central Appraisal District in providing access to their databases is also appreciated.

ABOUT THE AUTHORS

Miguel A. Morales *(to whom correspondence should be addressed) is a postdoctoral associate in the Department of Environmental Engineering Sciences at the University of Florida, 424 A.P. Black Hall, Gainesville, FL 32611; miguel22@ufl.edu or miguel@wateraim.com. Morales has seven years of experience in researching methods for estimating water use and targeting customers in relation to water demand management practices. He is also the president and founder of WaterAIM Inc., a consulting firm that provides data-driven analyses and planning services related to water demand management and water conservation to utilities, regional water supply authorities, and state and federal agencies that manage or regulate water supplies. He has BS, ME, and PhD degrees from the University of Florida, Gainesville.* ***James P. Heaney*** *is a professor in the Department of Environmental Engineering Sciences at the University of Florida, Gainesville.*

FOOTNOTES

[1]InfoGroup, Papillion, Neb.
[2]ArcGIS, ESRI, Redlands, Calif.

PEER REVIEW

Date of submission: 03/26/2014
Date of acceptance: 08/19/2014

REFERENCES

Brendle Group Inc., 2007. Benchmarking Task Force Collaboration for Industrial, Commercial, and Institutional Water Conservation. Colorado WaterWise Council, Denver.

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

Endter-Wada, J.; Kurtzman, J.; Keenan, S.P.; Kjelgren, R.; & Neale, C., 2008. Situational Waste in Landscape Watering: Residential and Business Water Use in an Urban Utah Community. *Journal of the American Water Resources Association,* 44:4:902.

Fleischhacker, S.E.; Evenson, K.R.; Sharkey, J.; Pitts, S.B.; & Rodriguez, D.A., 2013. Validity of Secondary Retail Food Outlet Data: A Systematic Review. *American Journal of Preventive Medicine,* 45:4:462.

Harou, J.J.; Pulido-Velazquez, M.; Rosenberg, D.E.; Medellín-Azuara, J.; Lund, J.R.; & Howitt, R.E, 2009. Hydro-Economic Models: Concepts, Design, Applications, and Future Prospects. *Journal of Hydrology,* 375:3-4:627.

Hendrickson, C.T.; Lave, L.B.; & Matthews, H.S., 2006. *Environmental Life Cycle Assessment of Goods and Services: An Input-Output Approach.* Resources for the Future, Washington.

Leontief, W., 1970. Environmental Repercussions and The Economic Structure: An Input-Output Approach. *Review of Economics and Statistics,* 52:3:262.

Liese, A.D.; Barnes, T.L.; Lamichhane, A.P.; Hibbert, J.D.; Colabianchi, N.; & Lawson, A.B., 2013. Characterizing the Food Retail Environment: Impact of Count, Type, and Geospatial Error in Two Secondary Data Sources. *Journal of Nutrition Education and Behavior,* 45:5:435. http://dx.doi.org/10.1016/j.jneb.2013.01.021.

Miller, R.E. & Blair, P.D., 2009 (2nd ed.). *Input-Output Analysis: Foundations and Extensions*. Cambridge University Press, Cambridge, U.K. http://dx.doi.org/10.1017/CBO9780511626982.

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 AWWA,* 103:6:84.

Navidi, W., 2010 (3rd ed.). *Statistics for Engineers and Scientists.* McGraw-Hill, New York.

Shandas, V. & Parandvash, G.H., 2010. Integrating Urban Form and Demographics in Water-Demand Management: An Empirical Case Study of Portland, Oregon. *Environment and Planning B: Planning and Design,* 37:1:112.

Solley, W.; Pierce, R.; & Perlman, H., 1998. Estimated Use of Water in the United States in 1995. US Geological Survey Circular 1200. US Department of the Interior, Washington.

Stoker, P. & Rothfeder, R., 2014. Drivers of Urban Water Use. *Sustainable Cities and Society,* 12:1:1.

US Environmental Protection Agency, 2009. Water Efficiency in the Commercial and Institutional Sector: Considerations for a WaterSense Program. USEPA WaterSense Program, Washington.