Statistical Model and Benchmarking Procedure for Energy Use by U.S. Public Water Systems*

Robert B. Sowby, Ph.D., P.E., M.ASCE[1]; and Steven J. Burian, Ph.D., P.E., M.ASCE[2]

ABSTRACT

Modern public water systems require energy to provide reliable, high-quality water, and two particular data problems exist. The first is the quantification of their energy requirements, for which literature is sparse and primary data collection is difficult. The second is the equitable comparison of energy use among systems that differ in size, water supply, and geographic setting. Using a recent survey of U.S. water systems' energy requirements, this study presents a statistical model that can estimate water system energy use as a function of a few accessible system-specific and geographic variables. The model offers improvements in accuracy and ease of use over previous ones. It advances the solution of the two problems by offering reasonable estimation of water system energy use in the absence of actual observations and by considering characteristics that explain much of the variation in energy use, enabling better energy benchmarking and fairer comparisons among diverse systems.

INTRODUCTION

Modern public water systems require energy to extract, treat, and deliver reliable, high-quality water to built environments. In the United States, this varies from 0.07 to 3.04 kWh/m^3 with an average of 0.48 to 0.53 kWh/m^3 (Sowby and Burian 2017a; DOE 2014; Twomey and Webber 2011; EPRI 2013). This energy-for-water relationship is one facet of the water–energy nexus, a broad research area that explores the interdependencies of water and energy resources. Water utilities' energy footprints carry financial, environmental, and social impacts that need to be understood and managed sustainably.

Despite the importance, little work has quantified or analyzed the energy requirements of public water supply, and many research needs are well documented. One recurring theme is the paucity of data, analysis, and models related to the water sector and its energy use (Healy et al. 2015; National Academies 2013; Water in the West 2013; Bazilian et al. 2011; Carlson and Wallburger 2007; Pate et al. 2007). In most cases, utility-level data are being collected but are not publicly accessible since few reporting systems or policies have been established—a curious deficiency in an era of "big data" (Chini and Stillwell 2017). Many of the available data are isolated observations, aggregated averages, or calculations that do not satisfy the resolution and quality needed by data users in government, research, engineering,

* This is a pre-publication manuscript of the following published paper: Sowby, Robert B., and Burian, Steven J. 2018. "Statistical model and benchmarking procedure for energy use by US public water systems." *Journal of Sustainable Water in the Built Environment*, 4(4): 04018010. https://doi.org/10.1061/JSWBAY.0000864.

[1] Water Resources Engineer, Hansen, Allen & Luce, Inc., 859 W. South Jordan Pkwy. Ste. 200, South Jordan, UT 84095 USA (corresponding author). Email: rob.sowby@hansenallenluce.com

[2] Professor, Dept. of Civil and Environmental Engineering, Univ. of Utah, 110 Central Campus Drive Ste. 2000, Salt Lake City, UT 84112 USA.

and professional associations who have repeatedly acknowledged these data gaps. The identification of similar needs by these diverse stakeholders testifies of their broad significance.

This study addresses two specific data problems in the United States. The first is the difficulty of quantifying water system energy requirements. Unlike water use, energy use is not typically reported by water utilities, so the data must be collected individually, as in Sowby and Burian's (2017a, 2017b) survey. If the data even exist, this primary data collection can be time consuming since most water utilities are unaccustomed to such requests. For these reasons the literature on this subject has been sparse to date. Past efforts include average or categorized energy intensities as reported by Twomey and Webber (2011), Young (2015), EPRI (2013), DOE (2012), and a more complete U.S. survey by Sowby and Burian (2017a, 2017b). Until regular reporting is mandated and/or the process is streamlined, collecting energy data from a large number of water systems is difficult. An improvement would be the development of an accurate statistical model, based on the observations already available, that could estimate water system energy use as a function of a few key variables and therefore substitute for otherwise infeasible primary data collection.

The second problem is the difficulty of fairly comparing energy use among diverse water systems. Even if a water system's energy use is known, judging its performance and comparing it to other water systems is another matter. Water systems and their energy consumption vary greatly for several reasons, some of which are beyond the systems' control. A small system with clean, abundant surface water will have a markedly different energy footprint than a large system that pumps deep groundwater; comparing the two directly is inappropriate. Even within the same system the energy intensity can change over time, further complicating fair comparisons across several water utilities (Sowby and Burian 2017a). There is therefore a need to identify and quantify the factors that influence water systems' energy use. Researchers have supposed that water source type, utility size, conveyance distance, climate, treatment technology, infrastructure age, and topography are important factors, but little research has quantified these (Wakeel et al. 2016; Young 2015; Tidwell et al. 2014; Water in the West 2013; Twomey and Webber 2011; Klein 2005). Cheng and Karney (2017), in determining the scaling properties of water supply networks, identified a strong relationship between energy use and water use. One notable study is that of Carlson and Wallburger (2007), which, in developing an energy benchmarking procedure, explored relationships between a water utility's energy use and its water use, pumping horsepower, pipe length, and elevation difference. Once identified, such factors should be developed into a benchmarking tool that enables a more equitable comparison for the water system and its peers.

This study's objective was to produce an empirical model that can estimate water systems' energy use as a function of a few system-specific and geographic variables and to apply the model for energy benchmarking of water systems, thereby advancing the solution of the two aforementioned problems.

METHODS

Definitions. Here, as in Sowby and Burian's (2017a) study, a public water system or water utility is an entity that delivers potable water to the public (more than 25 people). This definition follows that of the Safe Drinking Water Act. The entity may be publicly or privately owned. Most of the systems in this study are municipal drinking water systems.

"Energy" as used here means electricity delivered to the customer. Losses in electricity generation, transmission, and distribution are not included since the measurement is made at the customer's electric meter(s). Other fuels (e.g., natural gas) are occasionally used, but since water utilities consume energy predominantly as electricity (Twomey and Webber 2011; Carlson and Wallburger 2007), the energy data used in this study are limited to electricity. Common electrical uses in water facilities include pumping, treatment, monitoring, controls, lighting, and ventilation. For an entire water system, the electricity use is summed across all facilities and divided by the volume of water delivered to obtain energy intensity in kilowatt-hours per cubic meter (Wilkinson 2000).

Data sources. Annual water and energy use data originated from a survey of 108 U.S. water utilities in 36 states by Sowby and Burian (2017a, 2017b) which represents water services for about 14% of the U.S. population. Primary data were collected from a sample of water utilities with diverse sizes, locations, and water sources. In the survey, energy intensities were traced from the natural water source to the customer meter, including the energy intensity of any imports, on a whole-system basis (Wilkinson 2000). The observation years range from 2006 to 2015; over half of the observations are from 2015. While the years differ and may affect energy use because of drought, population growth, and other time-dependent factors, they represent the most consistent and recent data available. Energy intensities from the survey agree with aggregated values reported by others (DOE 2014; Twomey and Webber 2011; Young 2015; EPRI 2013), suggesting that the sample does not differ relevantly from the population. The survey also indicated the primary water source type and service population.

Literature review, logical suppositions, and a few unique theories informed the selection of additional explanatory variables. The following data, most of which are publicly available, were added to the survey dataset:

- Further categorization of water source types into pumped surface water, gravity-fed surface water, groundwater, and imported water based on follow-up inquiries with the survey respondents. Gravity-fed surface water may be the least energy-intensive because it requires only treatment and little or no pumping.
- Average annual precipitation (PRISM 2016a, 2016b). Wetter locations may correspond to greater water availability, particularly gravity-fed surface water, leading to lower energy intensities when compared to more arid conditions.
- Average annual air temperature (PRISM 2016a, 2016b). Warmer locations may correspond to greater seasonal variation in water use, especially for outdoor demand, leading to higher energy intensities.
- Minimum vapor pressure deficit (PRISM 2016a, 2016b). Related to the previous two variables, vapor pressure deficit describes hot, dry regions and may correspond to higher energy intensities.
- The county's population density (USCB 2012). Where water services are compact among higher population densities, energy intensities may be lower than in places with more spread-out infrastructure that require more conveyance.
- The county's population growth between the 2000 and 2010 censuses (USCB 2012). Rapid growth may correspond to more energy-intensive water to quickly meet new demands.
- The state's average electricity price in 2014 (EIA 2016). Where electricity is more expensive, there may be motivation for water systems to better manage it.
- The water system's approximate elevation (USGS 2016). The basic topographic position may affect water availability and/or gravity-fed status.
- The standard deviation of elevation within each water system's estimated service area, as computed from the National Elevation Dataset (USGS 2016). A measure of topographic

variation, this is a surrogate for a water system's elevation difference or hydraulic head between source and delivery.

- The city's presence on a list of 50 "greenest" U.S. cities by *Popular Science* (Svoboda 2008). Water systems in such cities may have environmentally motivated policies, attitudes, and programs leading to more efficient energy use.
- Results from the 2012 U.S. presidential general election (Guardian 2012). Similar to the point above, democratic cities tend to be more environmentally minded and this may motivate their water systems to optimize energy use.

For spatial data, values were extracted by water system location; for tabular data, the values were linked directly. Note that the foregoing variables fall into internal (e.g., system-specific) and external (e.g., climate) variables. Table 1 presents summary statistics.

Data partitions. The observations were randomly partitioned into a training dataset (80%) and validation dataset (20%). Model development used only the training dataset; the validation dataset was hidden until after the model was developed and then tested against the model. This process helps avoid overstating the accuracy of the predictions and provides an independent measure of error which previous models have not reported.

Transformation. As in Carlson and Wallburger's (2007) and Cheng and Karney's (2017) work, the range of water utility sizes prompted a logarithmic transformation of both water use and energy use. The transformation produces a strong linear relationship between the two that serves as the basis for further specification. By linearizing the data, the transformation overcomes some of the weaknesses of the previous efforts described below. Since the relationship is nearly linear, ordinary least squares (OLS) regression was selected for modeling using the cross-section for the most recent year in the panel dataset.

Figure 1 shows the transformed datasets of Carlson and Wallburger (2007) and Sowby and Burian (2017a, 2017b) in the same units. The range, slope, intercept, and R^2 are very similar, demonstrating that the two studies, in which all data were self-reported and otherwise could not be verified, corroborate each other.

Model specification. Specification followed three minimum criteria:

1. The absolute value of all individual test statistics exceeds 2.0. This corresponds to a Type I error probability of about 0.05 or less (probability of false positive) and follows Carlson and Wallburger's (2007) stepwise method.
2. The adjusted R^2 exceeds 0.87, offering an improvement in fit over Carlson and Wallburger's (2007) model. This metric describes the overall fit on a scale from 0 to 1.
3. The model's root mean square error (RMSE) does not exceed 1.79, offering an improvement in accuracy over Carlson and Wallburger's (2007) model when converted to kilowatt-hour basis (see Appendix for conversion). This metric describes the overall accuracy of the predictions relative to the range of observed values.

All model criteria had to be balanced. For example, models without an intercept produced a near-perfect adjusted R^2 (> 0.99) but yielded unacceptably high errors (RMSE > 2.5). A desirable but optional feature was the similarity of RMSE between the training and validation datasets when the model was applied, which would indicate that the model is not prone to overfitting.

Previous efforts. While stepwise OLS regression was ultimately selected because of linearization and consistency with Carlson and Wallburger (2007), the authors previously attempted three modeling approaches which ultimately did not meet the study objectives but nonetheless informed the approach. Since the main dataset was an unbalanced panel (describing water and energy use in the same water systems over multiple years), panel models with fixed and random effects were tried first (Hsaio 2003). The fixed-effects panel model was rejected since it did not consider time-invariant explanatory variables like many of those in Table 1. The random-effects panel model was inconsistent since its composite errors were correlated with the explanatory variables (failed Hausman test). A third model, spatial interpolation by kriging, was discarded because the underlying semivariogram was weak, showing little correlation between a water system's energy intensity and that of its neighbors, somewhat defying the first law of geography (Stein 1999; Tobler 1970). These three attempts and further literature review led to the final linear model presented here.

RESULTS AND DISCUSSION

Model results. The model, presented in Table 2, includes five statistically significant and largely independent variables—water system size, gravity-fed surface water status, imported water status, average annual precipitation, and average annual air temperature—plus an intercept. All test statistics exceed an absolute value of 2.0, the adjusted R^2 value is 0.9447, and the RMSE is 0.4989, satisfying all three minimum requirements. Figure 2 shows model residuals for each variable. Applying the model to the validation dataset yields an RMSE of 0.5183, which closely matches the model dataset's RMSE of 0.4989, demonstrating that the model performs well on an independent sample.

Magnitudes and signs of coefficients. The size of the water system, expressed here as the natural logarithm of its water use, is the most influential factor and correlates positively with energy use. This finding matches that of Carlson and Wallburger (2007).

The two extreme water supply types, gravity and imported, have negative and positive coefficients, respectively. Gravity-fed water requires less pumping and therefore less energy, while imported water requires more energy for its conveyance over greater distances and elevations. This variable explains the low energy intensities observed in the water systems of Portland, Denver, New York, and Boston—all which have high-head surface water sources—and the high energy intensities observed in southern California, where water is conveyed hundreds of kilometers over hundreds of meters of elevation gain before arriving at the point of use.

Precipitation, which shows a negative coefficient, could indicate the wetness of a location and its tendency to have abundant surface water, which is generally more accessible and less energy intensive than groundwater or imported water. Temperature exhibits a positive coefficient, suggesting that warmer regions require more energy for water supply, perhaps because of greater water demand relative to cooler areas and the need for more marginal water sources.

Together, these findings support the aforementioned theories about what factors influence a water system's energy footprint.

Residuals. The residuals in Figure 2 show random, symmetric dispersion without clear correlation to the parameter values, indicating that linear regression is appropriate.

Fit and accuracy. The adjusted R^2 value of 0.9447 indicates good overall fit. The model predicts the natural logarithm of the water system's energy use within −7% to +10% when using the training dataset, and within −6% to +5% when using the validation dataset. The RMSEs for the training and validation datasets are nearly equal—0.4989 and 0.5183, respectively—demonstrating that the model performs equally well on an independent sample. The measures of adjusted R^2 and RMSE and the provision of an independent error estimate are improvements over previous models.

Limitations and future work. Examining the observations that differ most from their predictions tells where the model still does not perform well. The differences tend to decrease with system size, but the correlation is weak. No common characteristics were found that would immediately suggest an additional variable. Future work may refine this model and/or produce new models with greater power and accuracy to explain how much energy a water system consumes. Further, many water systems' energy footprints are shrinking as a result of improved energy management practices (Sowby 2016); some may be increasing due to drought, treatment of emerging contaminants, or increased reliance on imported water. Both trends reinforce the need to continually collect data on this subject.

APPLICATIONS

Energy use estimation. Since energy data in the water sector are difficult to obtain (Chini and Stillwell 2017; Sowby and Burian 2017a), the model offers an alternative to resource-intensive primary data collection as a means to estimate water system energy footprints. The only multivariable model found in the literature review was that of Carlson and Wallburger (2007), which relied on detailed water system characteristics and/or other obscure data that limit its applicability, especially to studies of numerous systems. The model presented here helps overcome these challenges by using only basic water system characteristics and publicly available climate data. It may be used, at least initially, to estimate a water system's energy use until firm data become available.

Energy benchmarking. As discussed earlier, one of the difficulties in developing energy benchmarks and related sustainability metrics for water systems is the great variability in system characteristics that influence their energy use and complicate fair comparisons. The model overcomes much of this difficulty by quantifying the effects of a few pertinent characteristics that would otherwise render the comparison inappropriate if not impossible.

The natural logarithms of energy use are nearly normally distributed with mean 15.59 and standard deviation 2.16. Figure 3 shows the observations and a cumulative normal distribution curve. (Both are reversed so a low energy use corresponds to a higher percentile.) For the purposes of benchmarking, the distribution is considered to be exactly normal.

Consider a water system of given characteristics and an observed energy use whose natural logarithm is y. Given these characteristics, the model will predict the natural logarithm of energy use for a theoretical water system with the same characteristics, called $\hat{y}$. The ratio $y/\hat{y}$ indicates how much higher or lower the actual energy use is relative to the predicted value. There also exists a sample mean, $\hat{y}_{\text{mean}}$, which may be used to scale the ratio to the sample.

Using these three numbers, one may compare the water system to its theoretical peers:

$$E_{\text{adj}} = \hat{y}_{\text{mean}} \frac{y}{\hat{y}}$$

where E_{adj} is the natural logarithm of adjusted energy use, $\hat{y}_{\text{mean}}$ is the sample mean (here, 15.59), y is the natural logarithm of the observed energy use, and $\hat{y}$ is the natural logarithm of the predicted energy use (or, more precisely, the expected energy use of a theoretical water system with the same characteristics). The value of E_{adj} corresponds to a percentile ranking on the curve of Figure 3, assuming that the normal distribution applies to the theoretical peers as well as to the overall sample. This method follows that of Carlson and Wallburger (2007).

As an example, consider the water system described in Table 3 that delivered 103,000,000 m^3 of water in one year. The natural logarithm of its observed energy use (20,400,000 kWh) is 16.8. Using the same characteristics, the model predicts a value of 17.2. The ratio of these two is 0.978, indicating that the observed energy use is somewhat less than predicted. Multiplying by the sample mean 15.6, the adjusted natural logarithm of energy use is 15.3. On the curve of Figure 4, this value corresponds to the 57th percentile, or a score of 57 out of 100 among its peers, slightly above average.

Applying the benchmark procedure to Sowby and Burian's (2017a, 2017b) dataset, several dissimilar water utilities actually have the same score. For example, the water systems serving Boise, Denver, and Tampa—although they differ in size, topography, water supply, and climate—all score 55. Likewise, water systems serving Houston, Milwaukee, and St. Louis all score 46. Both cases demonstrate how the model normalizes great differences and identifies groups of similar performers.

This method can also be used to evaluate what magnitude of energy reductions are needed to achieve a higher score (thus defining an energy savings goal) or what impact proposed energy management projects will have. The water system may then pursue the energy savings using power company programs, qualified consultants, and/or published energy management guidance (AWWA 2016; Jones and Sowby 2014; UDDW 2014; Liu et al. 2012; NYSERDA 2010; EPA 2008). Energy management in the water sector is a major sustainability opportunity and many water systems have already achieved significant energy savings (Sowby 2016), while new research and resources will continue to promote energy reductions.

Jordan Valley Water Conservancy District, for example, serves the greater Salt Lake City area. Using this benchmarking procedure and data provided by the District (Todd Schultz, pers. comm.), its 2013 score would have been 37. In 2014 the District began an energy management program that delivered verified energy savings (Sowby et al., forthcoming). The District's 2014 and 2015 scores increased to 41 and 42, respectively, illustrating the incremental improvement likely attributed to the program.

SUMMARY AND CONCLUSIONS

Based on a recent survey and public datasets, a model of U.S. water systems' energy use was developed that offers improvements in accuracy and ease of use over previous models. While more sophisticated methods and more exact models may follow, this study identified a few important internal and external characteristics—water use, water source type, precipitation, and temperature—that are easily obtained. The model can provide reasonable estimates of energy use where primary data collection is infeasible. Since it explains much of the variation in energy use among water systems, the model is conducive to energy benchmarking, peer comparisons, and energy management planning.

APPENDIX

The units of root mean square error (RMSE) in both models are unusual since they are based on natural logarithms and different measures of energy consumed by the water utility. Carlson and Wallburger's (2007) RMSE is 0.56 when the base unit inside the logarithm is 1,000 British Thermal Units (kBTU). This is to be called ln(A). For comparison with this study, one seeks the equivalent RMSE when the base unit inside the logarithm is kilowatt-hours (kWh). This is to be called ln(B). The conversion is as follows and considers only end-use electricity consumed by the water utility (not counting losses during power generation, transmission, and distribution).

Recall 1 kBTU = 0.293 kWh. Then $A = 0.293B$ and $\ln(A) = \ln(0.293B)$. Separating the product within the logarithm, $\ln(A) = \ln(0.293) + \ln(B)$. Therefore, $\ln(B) = \ln(A) - \ln(0.293)$. Replacing with the value of ln(A) and solving, one obtains $\ln(B) = 0.56 - \ln(0.293) = 0.56 - (-1.23) = 1.79$. This conversion enables the RMSEs from both models to be compared.

ACKNOWLEDGEMENT

The authors thank professors Michael Barber, Tom Cova, Jeff Horsburgh, Christine Pomeroy, R. J. Porter, and Zain Al-Houri for their insights on this study. Support and comments from members of the Urban Water Group at the University of Utah and from colleagues at Hansen, Allen & Luce, particularly Tyler Allen, are also appreciated. The authors also appreciate comments received from two anonymous reviewers.

REFERENCES

AWWA (American Water Works Association). (2016). *Energy management for water utilities.* AWWA, Denver.

Bazilian, M., Rogner, H., Howells, M., Hermann, S., Arent, D., Gielen, D., Steduto, P., Mueller, A., Komor, P., Tol, R. S. J., and Yumkella, K. K. (2011). "Considering the energy, water and food nexus: Towards an integrated modelling approach." *Energy Policy*, 39(12), 7896–7906.

Carlson, S. W., and Wallburger, A. (2007). *Energy index development for benchmarking water and wastewater utilities*. Project no. 3009. Awwa Research Foundation, Denver.

Cheng, L., and Karney, B. W. (2017). "Organization and scaling in water supply networks." *Physical Review E,* 96, 062317.

Chini, C. M., and Stillwell, A. S. 2017. "Where Are All the Data? The Case for a Comprehensive Water and Wastewater Utility Database." *Journal of Water Resources Planning and Management* 143(3).

DOE (U.S. Dept. of Energy). (2012). Table 8.1.2: Average Energy Intensity of Public Water Supplies by Location. *2011 Buildings energy data book*.

DOE (U.S. Department of Energy). (2014). *The water–energy nexus: Challenges and opportunities*. DOE/EPSA-002.

EIA (U.S. Energy Information Agency). (2016). State electricity profiles. http://www.eia.gov/electricity/state/archive/2014/ (Mar. 24, 2016).

EPA (U.S. Environmental Protection Agency). (2008). *Ensuring a sustainable future: An energy management guidebook for wastewater and water utilities*.

EPRI (Electric Power Research Institute). (2013). *Electricity use and management in the municipal water supply and wastewater industries.* Technical report 3002001433. EPRI, Palo Alto.

Guardian, The. (2012). Full U.S. 2012 election county-level results. *The Guardian*, Nov. 14. https://www.theguardian.com/news/datablog/2012/nov/07/us-2012-election-county-results-download.

Healy, R. W., Alley, W. M., Engle, M. A., McMahon, P. B., and Bales, J. D. (2015). *The water–energy nexus—An earth science perspective*. U.S. Geological Survey Circular 1407. U.S. Dept. of the Interior, Reston.

Hsiao, C. (2003). *Analysis of panel data*. Cambridge University Press.

Jones, S. C., and Sowby, R. B. (2014). "Water system optimization: Aligning energy efficiency, system performance, and water quality." *J. Am. Water Works Assoc.*, 106(6), 66–71.

Klein, G. (2005). *California's water–energy relationship.* Final staff report CEC-700-2005-011-SF. California Energy Commission.

Liu, F., Ouedraogo, A., Manghee, S., and Danilenko, A. (2012). *A primer on energy efficiency for municipal water and wastewater utilities*. Energy Sector Management Assistance Program (ESMAP) technical report 001/12. The World Bank, Washington.

National Academies, The. (2013). "Sustainable energy and materials: addressing the energy–water nexus." Addressing the energy–water nexus: roundtable on science and technology for sustainability. Science and Technology for Sustainability Program, Board on Energy and Environmental Systems, Water Science and Technology Board, June 6.

NYSERDA (New York State Energy Research & Development Authority). (2010). *Water & wastewater energy management best practices handbook*.

Pate, R., Hightower, M., Cameron, C., and Einfeld, W. 2007. *Overview of energy–water interdependencies and the emerging energy demands on water resources.* SAND 2007-1349C. Sandia National Laboratories.

PRISM (PRISM Climate Group). (2016a). 30-year normals. Oregon State University. http://prism.oregonstate.edu/normals/.

PRISM (PRISM Climate Group). (2016b). "Descriptions of PRISM spatial climate datasets for the conterminous United States." http://prism.oregonstate.edu/documents/PRISM_datasets.pdf (Aug. 2016).

PSCW (Public Service Commission of Wisconsin). (2016). Water Statewide Statistical Benchmarks. http://psc.wi.gov/utilityInfo/water/benchmark.htm (accessed June 2, 2016).

Sowby, R. B. (2016). "Energy management in the water sector: A major sustainability opportunity." *Proc. 1st International Electronic Conf. on Water Sciences*, Nov. 15–29.

Sowby, R. B., and Burian, S. J. (2017a). "Survey of energy requirements for public water supply in the United States." *J. Am. Water Works Assoc.*, 109(7), E320–E330.

Sowby, R. B., and Burian, S. J. (2017b). "Energy Intensity Data for Public Water Supply in the United States." Dataset. <http://doi.org/10.5281/zenodo.1048275>.

Sowby, R. B., Jones, S. C., Packard, A. E., and Schultz, T. R. (2017). "Jordan Valley Water redefines sustainable water supply through energy management." *J. Am. Water Works Assoc.* 109(10), 38–45.

Stein, M. L. (1999). *Interpolation of spatial data: Some theory for kriging.* Springer, New York.

Svoboda, E. (2008). "America's top 50 green cities." *Popular Science*, Feb. 8. http://www.popsci.com/environment/article/2008-02/americas-50-greenest-cities.

Tidwell, V. C., Moreland, B., and Zemlick, K. (2014). "Geographic footprint of electricity use for water services in the western U.S." *Env. Sci. & Tech.*, 48, 8897–8904.

Tobler, W. R. 1970. "A computer movie simulating urban growth in the Detroit region." *Econ. Geography*, 46 (supplement), 234–240.

Twomey, K. M., and Webber, M. E. (2011). "Evaluating the energy intensity of the U.S. public water supply." *Proc. ASME 2011 5th International Conf. on Energy Sustainability, ES2011-54165,* 1735–1748.

UDDW (Utah Division of Drinking Water). (2014). *Drinking water energy (cost) savings handbook*. Utah Dept. of Environmental Quality, Salt Lake City.

USCB (U.S. Census Bureau). (2012). 2*010 Census of population and housing, population and housing unit counts*, CPH-2-1, United States Summary. U.S. Government Printing Office, Washington.

USGS (U.S. Geological Survey). (2016). National Elevation Dataset (NED). 1-arc-second digital elevation models (DEMs). https://ned.usgs.gov/ (Nov. 19, 2016).

Wakeel, M., Chen, B., Hayat, T., Alsaedi, A., and Ahmad, B. (2016). "Energy consumption for water use cycles in different countries: A review." *Applied Energy*, 178, 868–885.

Water in the West. (2013). *Water and energy nexus: A literature review*. Stanford University, Stanford.

Wilkinson, R. (2000). "Methodology for analysis of the energy intensity of California's water systems, and an assessment of multiple potential benefits through integrated water–energy efficiency measures." Lawrence Berkeley Laboratory and California Institute for Energy Efficiency.

Young, R. (2015). "A survey of energy use in water companies." ACEEE white paper. American Council for an Energy-Efficient Economy, Washington.

Table 1. Summary Statistics

Variable	Units	Average	Median	Standard Deviation	Minimum	Maximum
Internal Variables						
Indicator of pumped surface water supply (1 if pumped surface water constitutes >50% of supply; 0 otherwise)		0.37	0	0.49	0	1
Indicator of gravity-fed surface water supply (1 if gravity-fed surface water constitutes >50% of supply; 0 otherwise)		0.11	0	0.32	0	1
Indicator of groundwater supply (1 if groundwater constitutes >50% of supply; 0 otherwise)		0.48	0	0.50	0	1
Indicator of imported water supply (1 if imported water constitutes >50% of supply; 0 otherwise)		0.04	0	0.19	0	1
Service population	People	429,000	74,500	1,070,000	100	8,270,000
Annual energy use	Kilowatt-hours (kWh)	36,000,000	6,800,000	110,000,000	13,000	1,100,000,000
Annual water use	Cubic meters (m^3)	75,000,000	11,000,000	180,000,000	11,000	1,400,000,000
External Variables						
Average annual precipitation, 1981–2010	Centimeters (cm)	60.7	62.1	29.5	12.3	113.3
Average annual air temperature, 1981–2010	Degrees Celsius (°C)	11.6	10.8	−13.7	3.02	22.7
Minimum vapor pressure deficit, 1981–2010	Millibars (mbar)	1.81	1.59	1.32	0.25	9.86
County population density, 2010	Persons per square mile	1,380	373	3,880	2	35,400
County population growth, 2000–2010	Fraction	0.14	0.11	0.14	−0.08	0.58
State average price of electricity, 2014	U.S. dollars per kilowatt-hour	0.10	0.09	0.03	0.07	0.16
Elevation	Meters (m) above sea level	606	269	693	4	2,900
Standard deviation of elevation within service area	Meters (m)	109	38	145	3	584
Indicator of presence on list of 50 greenest cities (1 if on list; 0 otherwise)		0.16	0	0.37	0	1
Indicator of democratic vote in 2012 U.S. presidential election (1 if democratic vote; 0 otherwise)		0.49	0	0.50	0	1

Table 2. Model Results for Natural Logarithm of Water System Energy Use

Variable	Coefficient	Standard Error	Test Statistic
Natural logarithm of annual water use in cubic meters	0.8934	0.0287	31.10
Indicator of gravity-fed water supply (>50%)	−0.9494	0.2140	−4.44
Indicator of imported water supply (>50%)	1.2759	0.2726	4.68
Average annual precipitation in centimeters	−0.0054	0.0021	−2.62
Average annual air temperature in degrees Celsius	0.0360	0.0164	2.20
Intercept	0.9713	0.3991	36.66
R^2	0.9480		
Adjusted R^2	0.9447		
Root Mean Square Error	0.4989		
Number of Observations	86		

Table 3. Water System Example

Variable	Value	Coefficient	Product
Natural logarithm of annual water use in cubic meters	18.5	0.8934	16.5
Indicator of gravity-fed water supply	1	−0.9494	−0.950
Indicator of imported water supply	0	1.2759	0.0000
Average annual precipitation in centimeters	18.3	−0.0054	−0.0985
Average annual temperature in degrees Celsius	22.3	0.0360	0.804
Intercept	1	0.9713	0.971
			Sum = 17.2
Natural logarithm of observed energy use (y)	16.8		
Natural logarithm of predicted energy use ($\hat{y}$)	17.2		
Sample mean ($\hat{y}_{\mathrm{mean}}$)	15.6		
Natural logarithm of adjusted energy use (E_{adj})	15.3		
Percentile score	57/100		

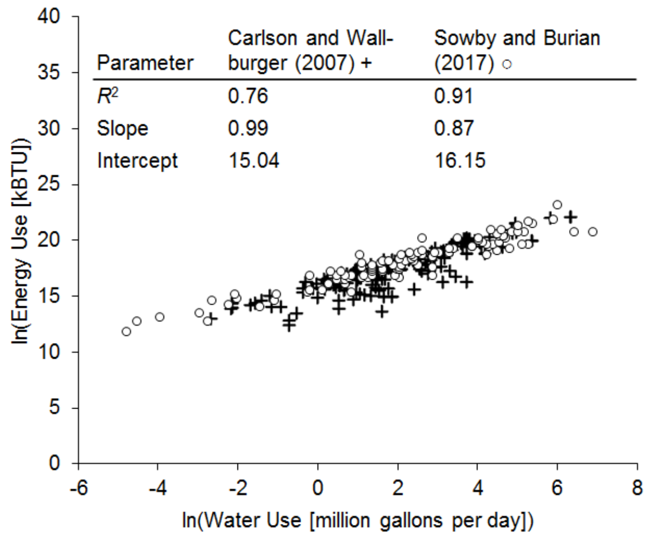
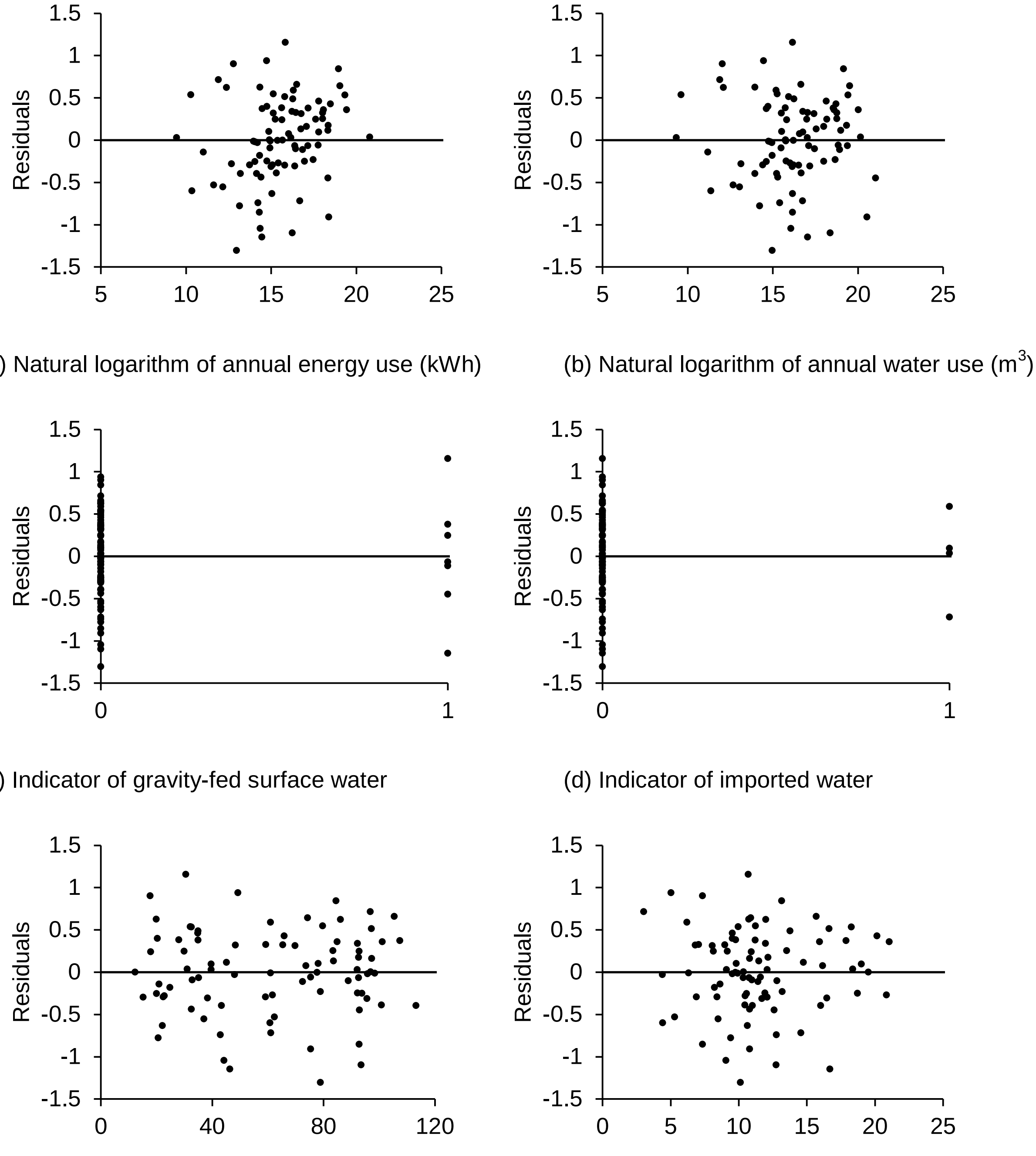


Fig. 1. Comparison of two datasets on water system energy use. Axes show natural logarithms of water and energy use. Points represent annual observations for individual water systems. (Carlson and Wallburger [2007] figure reprinted with permission. © Water Research Foundation.)



(a) Natural logarithm of annual energy use (kWh)

(b) Natural logarithm of annual water use (m^3)

(c) Indicator of gravity-fed surface water

(d) Indicator of imported water

(e) Average annual precipitation (cm)

(f) Average annual temperature (°C)

Fig. 2. (a–f) Model residuals. Horizontal axes show variable value and vertical axis shows residual, the difference between the observed and predicted value of natural logarithm of the water system's annual energy use.

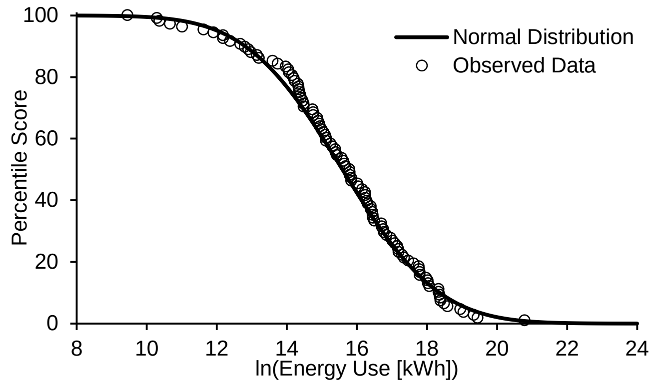


Fig. 3. Water system energy use distribution. The natural logarithms of water system energy use are nearly normally distributed. The curve is reversed so lower energy uses correspond to higher scores.

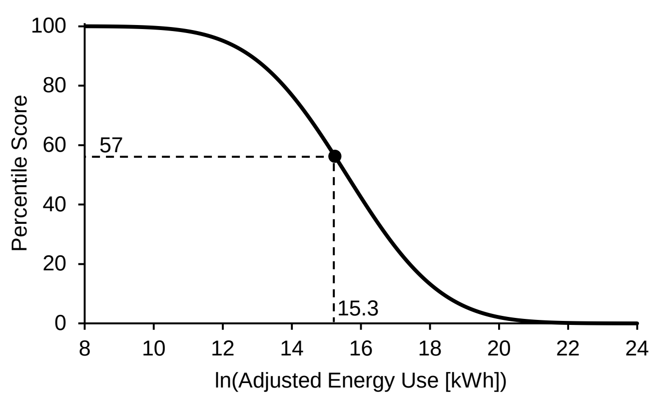


Fig. 4. Benchmark application example. The graph shows how a water system's actual energy use, relative to its expected value, corresponds to a percentile score based on the distribution.