

Leading-Digit Patterns from Smart Water Meters

Robert B. Sowby, Ph.D., P.E., and Daniel R. Jones, P.E.
Hansen, Allen & Luce, Inc.
November 16, 2020

ABSTRACT

As infrastructure develops and adoption of smart water meters increases, new techniques are needed to validate and learn from the large datasets they produce. Patterns of leading digits (i.e., first non-zero digits, 1 through 9) can support this task. This study examines leading digits in hourly smart meter readings from a western U.S. water utility with over 5,000 customer connections. Benford analysis, power law analysis, and leading-digit-frequency analysis all indicate that the readings tend toward values that start with 1. The findings suggest that readings from smart water meters—and, by extension, water use by individual customers—could be expected to follow a particular nonuniform pattern of leading digits and that deviation from the pattern may indicate data errors or abnormal water use. Applications are suggested for validating water use data, comparing multiple datasets, checking projections, and assessing meter performance. Additional work is needed to further explore the beneficial uses of leading-digit patterns and other data signatures in water use data from diverse datasets.

INTRODUCTION

In an era of big data and sustainability, water use data are more plentiful and important than ever. Decision makers need accurate data to maintain the vital service of providing clean, reliable water. The data must be fine enough in both space and time to characterize individual water use profiles, justify water conservation measures, prioritize leak detection, and inform planning activities.

The Promise of Smart Meters

As part of advanced metering infrastructure (AMI) that is now becoming mainstream, smart water meters that record customer water usage at hourly (or smaller) time scales promise to help solve these problems. An estimated 17% of North American water utilities use smart meters and the share is growing (AWWA 2018; Russell 2019). The resulting large datasets may be mined for insights into water use habits of customers and the planning, design, and operation of the water utility, with two-way communication between both parties.

One driver of smart meter deployment is the desire for better data on the timing and type of end use. The data help improve meter accuracy (Boyle et al. 2013; Beal and Flynn 2015), models and projections (Beal and Stewart 2014; Gurung et al. 2014; Romano and Kapelan 2014; Creaco et al. 2016; Candelieri et al. 2017; Pesantez et al. 2020; Xenochristou et al. 2020; Shafiee et al. 2020), and dynamic pricing (Rougé et al. 2018).

When combined with customer feedback, the high-resolution data can help reduce water use and alert users and operators of problems. Davies et al. (2014) and Sønnderlund et al. (2016) both reported measurable reductions in water use when combining smart meters and near-real-time consumption feedback. Others have shown how the data can help customize water management activities (Cardell-Oliver et al. 2016; Cahn et al. 2020). During a contamination incident, one Utah water utility examined smart meter data from the affected area to determine which residents had flushed their plumbing; water utility staff then visited and helped those who had not (Hansen, Allen & Luce 2020).

In addition to security, hardware, software, logistics, funding, and politics (Mutchek and Williams 2014; Gurung et al. 2017; Casteletti et al. 2018; Monks et al. 2019), one of the most prominent challenges for continued smart meter adoption is data management (Cominola et al. 2015, 2018; Sowby et al. 2019). Specifically, Savić et al. (2014) stated that the main scientific challenges “are the management and extraction of useful information from vast amounts of high-resolution consumption data” and turning them into decision-support tools. Some examples include algorithms to detect meter fraud by identifying trends of gradually decreasing consumption, suddenly decreased consumption, and/or abnormally low consumption in the meter profiles (Monedero et al. 2016) and distinctive time-series patterns that indicate leakage, backflow, pump failures, and pipe bursts (Britton et al. 2013; Mix et al. 2020; Wu and Liu 2020). Additionally, Meng et al. (2017) reports that the perception of smart infrastructure by water management professionals, particularly as it relates to performance, user-friendliness, and cost effectiveness, remains a challenge.

Despite the challenges, smart water meters and their data continue to drive advancements in hydraulic modeling, leak detection, water conservation, and customer service (Russell 2019), “[ushering] in a new era in design, operation, and management of urban water infrastructure” (Rasekh et al. 2016).

The Value of Leading-Digit Patterns

One area of smart meter analysis yet to be explored is the pattern of leading digits (i.e., first non-zero digits, 1 through 9) in the readings and how these patterns might be exploited. Leading-digit patterns have been found in diverse datasets, including populations (Sandron and Hayford 2002), financial transactions (Nigrini and Mittermaier 1997), integer sequences (Washington 1981), and streamflow (Nigrini and Miller 2007). When expected leading-digit patterns are established, deviation from the pattern can indicate problems. In accounting, leading-digit analysis is used to detect fraud since fabricated or randomly generated data do not follow a natural pattern (Durtschi et al. 2004; Nigrini and Mittermaier 1997). Sambridge et al. (2010) applied leading-digit analysis to validate observations of exoplanet masses, seismic wave speeds, and global temperature anomalies. Sowby (2018) found a leading-digit pattern in annual potable water use data aggregated by U.S. public water suppliers, counties, and states and suggested that the pattern could help validate water use observations. The study recommended examination of patterns in higher-resolution data and water end-uses. Assigned numbers (e.g., phone numbers, identification numbers, and prices) and values with limited ranges do not follow the same patterns.

The most well-known and well-documented leading-digit pattern is Benford’s law, in which most numbers start with 1, regardless of the decimal place (Fig. 1). Benford (1938), like Newcomb (1881) before him, noted how the first pages of logarithm books were more worn out than the last, suggesting that users were looking up the logarithms of numbers starting with low first digits more frequently than those with high first digits.

Benford theorized that in a dataset of values spanning several orders of magnitude in any consistent units, a leading digit d (where $d = 1, 2, \dots, 9$) would occur with frequency f according to

$$f(d) = \log_{10} \left(1 + \frac{1}{d} \right) \quad (1)$$

resulting in the distribution shown in Fig. 1. While similar expressions exist for second and third digits, the distribution of first digits is the most distinctive.

In data that follow Benford's law, the digits are not uniformly distributed 1/9 or 11.1% of the time as one might expect. Instead, 1 appears as the leading digit about 30% of the time and the frequency decreases thereafter for digits 2 through 9. This is apparent on any logarithmic scale, where the distance, on paper, between 1 and 2 is much greater than that between 8 and 9. Benford's law applies best when the data range covers an integer order of magnitude; otherwise, the pattern may deviate in the tails. In such cases the data may be curtailed so the log range becomes a whole number and Benford's law can be observed.

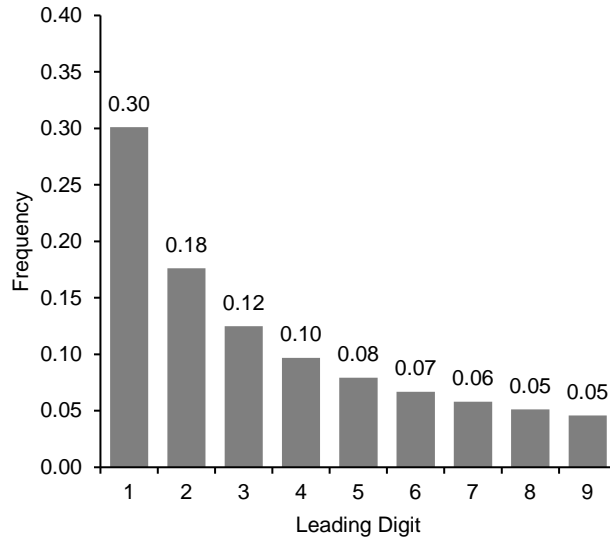


Fig. 1. Frequency of Leading Digits according to Benford's Law

Other patterns of leading digits have also been observed, such as power law distributions, which are even more extreme than Benford's law. Nigrini and Miller (2007) analyzed hydrologic data and found that streamflow measurements follow Benford's law, but that lake areas follow a steep power law.

In a power law distribution, a leading digit d (where $d = 1, 2, \dots, 9$) would occur with frequency f according to

$$f(d) = \frac{d^{-m} \left(1 - \left(\frac{d+1}{d} \right)^{-m} \right)}{1 - 10^{-m}} \quad (2)$$

where m is determined from analysis of the observations by Newman's (2005) method:

$$m = n \left[\sum_{i=1}^n \ln \left(\frac{x_i}{x_{\min}} \right) \right]^{-1} \quad (3)$$

Where n is the sample size, x_i (where $i = 1, 2, \dots, n$) are the observations, and x_{\min} is the minimum observation. Note in Eq. (2) that when $m = 0$, the equation simplifies to Benford's law in Eq. (1), after applying L'Hospital's Rule to handle indeterminacy (Nigrini and Miller 2007). When $m > 0$, the pattern is steeper than Benford's law and favors low leading digits even more. When $-1 < m < 0$, the pattern is less steep than Benford's law but still favors low leading digits. When $m = -1$, pattern of leading digits is uniform (flat). When $m < -1$, the pattern favors high leading digits.

Research Objectives

Do water use data exhibit any distinctive leading-digit patterns? If so, how might such patterns be beneficially used? This study is the first to examine leading-digit patterns from smart water meters and intends to provide a valuable technique for extracting insights from these new, large datasets. Using a sample dataset, the authors identify patterns of the first leading digit in the meter reads. Potential applications to data validation, data quality monitoring, water use projections, and meter performance are then recommended.

DATA AND METHODS

Data Source and Selection

Xylem Inc., a large American water technology provider, provided anonymized smart meter data from a Western U.S. water utility. The original dataset consists of hourly reads for 5,765 meters from 2016 to 2018.

For this study, the dataset was limited to the year 2018 (for reasons of data quantity), to nonzero readings (since 0 cannot be a leading digit), and to meters with resolution of 3.785 L (1 gallon) (the finest resolution of the provided data, so as to provide consistent resolution and avoid rounding). The resulting dataset contains 13,836,912 observations of meter ID, timestamp, and readings across 5,243 unique meters. The data range from 1 to 10,000,000 gallons (3.78 L to 37,800,000 L), so the log range is equal to an integer (7) and no curtailing was necessary. The statistical tests described below were carried out for each meter each month (a “meter-month”). The sample size in each meter-month ranged from 0 (no nonzero reads) to 744 (24 hours over 31 days). Meter-months with a sample size of zero were excluded. Of 62,916 possible meter-months (5,243 meters across 12 months), 58,234 remained in the final analysis.

Units

The native meter readings, as a product of both meter hardware and software, were in discrete gallons and remained so throughout the analysis. Leading-digit patterns generally persist regardless of the units, as long as the units are consistent and the data are not dominated by a few values (Hill 1995b). However, this particular dataset consists mostly of small values that would bias the leading-digit distribution if converted. To avoid such bias, the underlying data were not converted to SI units.

Statistical Tests

Benford's law. Conformance to Benford's law was evaluated with the Kolmogorov–Smirnov (KS) test (Massey 1951). The test is a common choice for Benford analysis, including water-related applications by Sowby (2018) and Nigrini and Miller (2007). The null hypothesis is that the sample conforms to Benford's law, i.e., that the cumulative distribution of leading digits in the empirical dataset is the same as that in Benford's law. The null hypothesis is rejected if the KS test statistic exceeds the critical value at a significance level of $\alpha = 0.05$, indicating a statistically significant departure from Benford's law. The critical value decreases with sample size, effectively tolerating less departure in larger samples. The test was performed for the hourly reads in each meter-month, as well as for the entire dataset of hourly reads. Since the number of reads differs each month (especially after filtering the data as described above), it is important that any statistical test consider the sample size, which the KS test does.

Steep power law. A power law distribution was evaluated by computing the power law exponent as described by Newman (2005) and Nigrini and Miller (2007) for each meter-month's hourly reads, as well as

for the entire dataset of hourly reads. Readings were determined to follow a steep power law if and only if $m > 0$ as described in (3) above (i.e., a pattern steeper than Benford’s law).

Most frequent leading digit. Regardless of conformance to Benford’s law or a power law, it is also of interest simply to know what the most frequent leading digit is in any particular dataset. For this reason, a leading-digit-frequency analysis was performed on hourly reads from each meter-month and the most frequent leading digit was reported. Of particular interest is what proportion of meter-months have 1 as the most frequent leading digit from the hourly reads.

Because of differences in statistical significance in each test, it is possible for readings to satisfy multiple patterns (e.g., Benford’s law *and* power law).

Data Processing

The data were processed with R (R Core Team 2019), combining published R packages with custom scripts and loops coded specifically to address the format of the provided data.

RESULTS AND DISCUSSION

According to the KS tests, hourly reads in 16,668 meter-months (28.6%) followed Benford’s law. According to the power law tests, hourly reads in all 58,234 meter-months (100.0%) followed a steep power law. In 57,202 meter-months (98.3%), 1 was the most frequent leading digit in the hourly reads. Some meter-months passed multiple tests. See Table 1. Each of the three tests indicates that the smart meter readings in the sample dataset tend toward values beginning with 1.

Table 1. Leading-Digit Patterns from Sample Smart Water Meter Dataset

Leading-digit pattern	Number of meter-months satisfying pattern ^a
Benford’s law	16,668 (28.6%)
Steep power law (leading with 1)	58,234 (100.0%)
1 is most frequent leading digit	57,202 (98.3%)

a. Of total 58,234 meter-months

The findings for this particular dataset suggest that readings from smart water meters—and, by extension, water use by individual customers—could be expected to follow a particular nonuniform pattern of leading digits, whether Benford’s law, a steep power law, or simply the pattern where 1 is the most frequent leading digit (regardless of the frequencies of subsequent digits). The results accord with those of Sowby (2018) for coarser water use data and prompt further exploration of other datasets and how this pattern might be exploited.

Taken as a whole, the dataset does not conform to Benford’s law. The sample size (over 13,800,000) is so great that the tolerance for departure from the Benford distribution (the critical KS value) is very small and the null hypothesis of conformance must be rejected, even though data in individual meter-months (in smaller samples) still fit the pattern. The entire dataset more closely resembles a steep power law. See Fig. 2. The distribution is clearly steeper than Benford’s law. The computed best-fit power law exponent [Eq. (3)] for the dataset is $m = 0.43$, which is much greater than Benford’s law, in which $m = 0$. In other words, the smart meter data show even more favor toward 1 as the leading digit.

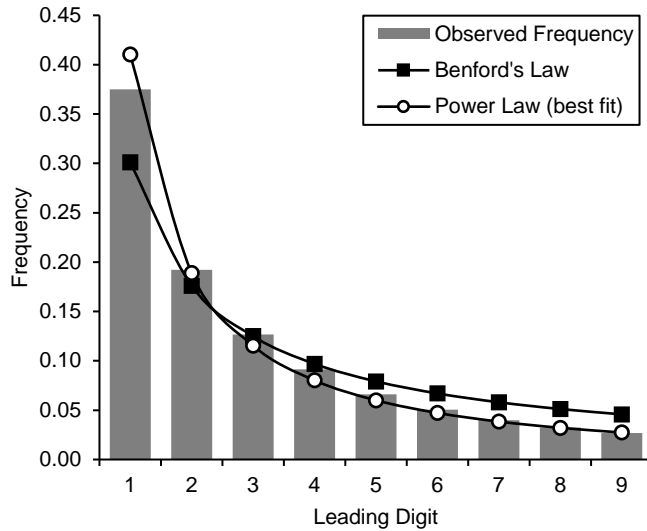


Fig. 2. Leading-Digit Frequencies for Entire Sample Dataset

APPLICATIONS

The distinctive pattern of leading digits among smart water meter readings suggests several useful applications and possible future regulatory implications to be explored in future work. A few examples are given here.

Validating Water Use Data

To complement existing validation methods, water use data could be screened by a leading-digit analysis before being finalized. As in accounting (Durtschi et al. 2004; Nigrini and Mittermaier 1997), water use data samples that deviate from the expected pattern could be flagged for further attention since they may be incomplete, be rounded too much, contain errors, or indicate abnormal water use (Sowby 2018; Nigrini and Miller 2007). Such functions could be programmed as quality control measures into smart meter software backends and water utility databases, or stand alone for ad-hoc processing. Where most water uses are measured only once and cannot otherwise be corroborated, leading-digit analysis can indicate general conformance or nonconformance to expected patterns. Leading-digit analysis could be particularly useful as data are aggregated to larger and larger scales, as in state and federal reporting: each receiving party could examine leading-digit patterns in the submitted datasets before accepting them. In nonconforming datasets, investigation may focus on the subsets of leading digits that are particularly under- or overrepresented.

For example, the Utah Division of Water Resources publishes data on non-potable water use for irrigation in municipal settings. Four years of annual totals, 2015–2018, were provided for several hundred water suppliers. In 2015, 2016, and 2018, the data conform to Benford's law by the KS test, but in 2017 they do not, showing unusually high frequencies of leading digits 6, 7, and 8. This might prompt a reevaluation of the 2017 data to uncover errors, unusual water use, or other reasons for departure from the expected pattern.

Comparing Multiple Datasets

Leading-digit patterns may also help one compare the accuracy of two or more datasets on the same water uses and inform a choice about which dataset to use. Sowby (2018) illustrated such a case where, given two datasets on public water suppliers' use, the dataset known to be more accurate also more closely followed Benford's law. The same may be true of smart meters and individual water users. A related application is comparing the accuracy of water use datasets over time (e.g., years), where increasing conformance to a predictable pattern could also indicate increasing data accuracy.

Continuing the example from the Utah Division of Water Resources above, the 2018 data show the best overall fit to Benford's law (smallest KS value relative to critical value) of the four years in question. It was in this same year that the Division improved its analysis methods after recommendations from a third-party audit (Bowen Collins & Associates and Hansen, Allen & Luce 2018). The better fit is one indicator that the improvements were effective. Such a test may belong among ongoing efforts to monitor data quality.

Checking Models and Projections

If historic water use for a certain water group follows leading-digit patterns, so should future water use. Water use models and projections (such as those by Romano and Kapelan [2017] and Candelieri [2017]) therefore could undergo leading-digit analysis to see if the results are reasonable (Hill 1995a). When results do not follow the expected pattern, the assumptions, methods, and base data may be revisited.

For example, Hejazi et al. (2014, Table 6) published projections of 2050 and 2095 global water withdrawals for 32 economic sectors and 6 scenarios (384 values total). The data conform to Benford's law by the KS test. The finding of a leading-pattern in these water use projections is consistent with this research and with Sowby (2018).

Assessing Meter Performance

Water meter performance can deteriorate over many years as the components wear out and the calibration wanders. Meters are also subject to accidental damage and deliberate tampering (meter fraud) that affect the readings. Such changes may affect the leading-digit patterns. Similar to what Monedero et al. (2016) did for actual water usage, a time series of test statistics on leading digits (such as the KS test statistic) might show trends of decreasing performance or permanent changes at the meter that prompt further investigation. The effect of meter wear on leading-digit patterns is not known, so this is an area for further research. Many years of smart meter records, which were not immediately available to the authors, may be necessary to determine if such analysis is useful.

LIMITATIONS

One of the main limitations in this study, as well as in future applications, is the computational demand of assessing leading-digit patterns in large water use datasets. Here, the analysis required several hours of custom programming and processing on a desktop computer. Other datasets are even larger, with more meters and more frequent reads. If every meter must be analyzed, then new computational methods, optimized algorithms, and more efficient data formats are needed to streamline the process and make continued analysis more efficient, as others have already noted (Savić et al. 2014; Cominola et al. 2015). Otherwise, random sampling might be more practical.

It is possible that the leading-digit patterns found here are specific to this water utility; other sites with other water-use profiles due to land use, climate, pricing, and demographics may exhibit different patterns. While Benford's law is expected to apply whenever the above criteria are met, these patterns need to be tested for more datasets and sites to reach a general conclusion that individual water use follows these patterns.

The patterns discovered in this study may be a product of the particular meters used. In the dataset examined here, 1 is the most frequent leading digit among reads in gallons, and 3.785 L (1 gallon) is the resolution of the selected meters. This may or may not be coincidence. However, 37.85 L (10 gallons), not 3.785 L (1 gallon), is not the most frequent reading, so direct correlation with the resolution of 3.785 L (1 gallon) can be ruled out. Still, this prompts questions about relationships between meter brands, hardware, resolution, and frequency and the resulting leading-digit patterns.

As in any leading-digit analysis, zeros must always be excluded since 0 is not a non-zero leading digit (0 is never a significant leading digit). Because of low discrete volumetric resolution and/or intervals with little or no water use, zero most often indicates "no activity." As such, many smart meter readings may be zero, so large portions of the original data may not qualify for analysis. If zeros (and other actual values) are to be analyzed, other methods must be employed. In future metering technologies, "no activity" might be more accurately recorded as a null value instead of a zero.

CONCLUSION

This study examined hourly readings from over 5,000 smart water meters in a 12-month period. Benford analysis, power law analysis, and lead-digit-frequency analysis all indicated the readings tend toward values that start with 1. The findings suggest that readings from smart water meters—and, by extension, individual customer water use—might be expected to follow a particular nonuniform pattern of leading digits and that deviation from the pattern may indicate data errors or abnormal water use. As the implementation of smart water meters and related technologies grows, leading-digit analysis, where appropriate, could complement existing methods to enhance the insights one can extract from these large datasets. Applications are suggested for validating water use data, comparing multiple datasets, checking projections, and assessing meter performance. Further work should improve computational ability and explore leading-digit patterns from meters in different geographic locations and from meters of different resolution.

ACKNOWLEDGMENTS

The authors thank Joseph Dryer and Travis Smith of Xylem Inc. for providing the data used in this study.

REFERENCES

- AWWA (American Water Works Association). 2018. "2018 state of the water industry report." Denver: AWWA.
- Beal, C. D., and J. Flynn. 2015. "Toward the digital water age: Survey and case studies of Australian water utility smart-metering programs." *Utilities Policy* 32: 29–37. <https://doi.org/10.1016/j.jup.2014.12.006>.
- Beal, C. D., and R. A. Stewart. 2014. "Identifying residential water end uses underpinning peak day and peak hour demand." *J. Water Resour. Plann. Manage.* 140 (7): 04014008. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000357](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000357).
- Benford, F. 1938. "The law of anomalous numbers." *Proc. Am. Phil. Soc.* 78 (4).
- Bowen Collins & Associates and Hansen, Allen & Luce. 2018. *State of Utah Water Use Data Collection Program*. Draper, UT: Bowen Collins & Associates.
- Boyle, T., D. Giurco, P. Mukheibir, A. Liu, C. Moy, S. White, and R. Stewart. 2013. "Intelligent metering for urban water: A review." *Water* 5 (3): 1052–1081. <https://doi.org/10.3390/w5031052>.

- Britton, T. C., R. A. Stewart, and K. R. O'Halloran. 2013. "Smart metering: enabler for rapid and effective post meter leakage identification and water loss management." *J. Cleaner Prod.* 54: 166–176. <https://doi.org/10.1016/j.jclepro.2013.05.018>.
- Cahn, A., D. Katz, and A. Ghermandi. 2020. "Analyzing water customer preferences for online feedback technologies in Israel: A prototype study." *J. Water Resour. Plann. Manage.* 146 (4): 06020002. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001179](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001179).
- Candelieri, A. 2017. "Clustering and support vector regression for water demand forecasting and anomaly detection." *Water* 9 (3): 224. <https://doi.org/10.3390/w9030224>.
- Cardell-Oliver, R., J. Wang, and H. Gigneyh. 2016. "Smart meter analytics to pinpoint opportunities for reducing household water use." *J. Water Resour. Plann. Manage.* 142 (6): 04016007. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000634](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000634).
- Castelletti, A., A. Cominola, A. Facchini, M. Giuliani, P. Fraternali, S. Herrera, M. Melenhorst, I. Micheel, J. Novak, C. Pasini, A. E. Rizzoli, and C. Rottondi. 2018. "Chapter 7: Gamified approaches for water management systems: An overview." *Smart Water Grids: A Cyber-Physical Systems Approach*. Boca Raton: CRC Press. <https://www.doi.org/10.1201/b21948-7>.
- Cominola A., M. Giuliani, A. Castelletti, D. E. Rosenberg, and A. M. Abdallah. 2018. "Implications of data sampling resolution on water use simulation, end-use disaggregation, and demand management." *Env. Modelling & Software* 102: 199–212. <https://doi.org/10.1016/j.envsoft.2017.11.022>.
- Cominola, A., M. Giuliani, D. Piga, A. Castelletti, and A. E. Rizzoli. 2015. "Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review." *Env. Modelling & Software* 72: 198–214. <https://doi.org/10.1016/j.envsoft.2015.07.012>.
- Creaco E., P. Kossieris, L. Vamvakieridou-Lyroudia, C. Makropoulos, Z. Kapelan, and D. Savić. 2016. "Parameterizing residential water demand pulse models through smart meter readings." *Env. Modelling & Software* 80: 33–40. <https://doi.org/10.1016/j.envsoft.2016.02.019>.
- Davies, K., C. Doolan, E. Van Den Honert, and R. Shi. 2014. "Water-saving impacts of smart meter technology: An empirical 5 year, whole-of-community study in Sydney, Australia." *Water Resour. Research* 50 (9): 7348–7358. <https://doi.org/10.1002/2014WR015812>.
- Durtschi, C., W. Hillison, and C. Pacini. 2004. "The effective use of Benford's law to assist in detecting fraud in accounting data." *J. Forensic Accounting* V: 17–34.
- Gurung, T. R., R. A. Stewart, A. K. Sharma, and C. D. Beal. 2014. "Smart meters for enhanced water supply network modelling and infrastructure planning." *Resources, Conservation and Recycling* 90: 34–50. <https://doi.org/10.1016/j.resconrec.2014.06.005>.
- Gurung, T. R., R. A. Stewart, C. D. Beal, and A. K. Sharma. 2017. "Smart meter enabled informatics for economically efficient diversified water supply infrastructure planning" *J. Cleaner Prod.* 135: 1023–1033. <https://doi.org/10.1016/j.jclepro.2016.07.017>.
- Hansen, Allen & Luce. 2020. *Investigation of Paradise Valley Well fluoride overfeed [Sandy City, Utah]*. South Jordan, UT: Hansen, Allen & Luce.
- Hejazi, M., J. Edmonds, L. Clarke, P. Kyle, E. Davies, V. Chaturvedi, M. Wise, P. Patel, J. Eom, K. Calvin, R. Moss, and S. Kim. 2014. "Long-term global water projections using six socioeconomic scenarios in an integrated assessment modeling framework." *Technological Forecasting and Social Change* 81: 205–226. <https://doi.org/10.1016/j.techfore.2013.05.006>.
- Hill, T. P. 1995a. "The significant-digit phenomenon." *Am. Mathematical Monthly* 102 (4): 322–327.
- Hill, T. P. 1995b. "Base-invariance implies Benford's law." *Proc. Am. Mathematical Soc.* 123 (3): 887–895.
- Massey, F. J. 1951. "The Kolmogorov-Smirnov test for goodness of fit." *J. Am. Stat. Assoc.* 46 (253): 68–78. <https://doi.org/10.2307/2280095>.
- Meng, T., D. Hsu, and B. Wadzuk. 2017. "Green and smart: Perspectives of city and water agency officials in Pennsylvania toward adopting new infrastructure technologies for stormwater management." *J. Sust. Water Built Environ.* 3 (2). <https://doi.org/10.1061/JSWBAY.0000824>.
- Mix, N., A. Lai, K. Thompson, and S. C. Seachrist. 2020. "Advanced metering infrastructure: Reducing water loss, improving security, and enhancing resiliency." *J. Am. Water Works Assoc.* 112 (2): 38–49. <https://doi.org/10.1002/awwa.1447>.
- Monedero, I., F. Biscarri, J. I. Guerrero, and M. Peña. 2016. "Detection of water meter under-registration using statistical algorithms." *J. Water Resour. Plann. Manage.* 142 (1): 04015036. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000562](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000562).

- Monks, I., R. Stewart, O. Sahin, and R. Keller. 2019. "Revealing unreported benefits of digital water metering: literature review and expert opinions." *Water* 11 (4): 838. <https://doi.org/10.3390/w11040838>.
- Mutchek, M., and E. Williams. 2014. "Moving towards sustainable and resilient smart water grids." *Challenges* 5 (1): 123–137. <https://doi.org/10.3390/challe5010123>.
- Newcomb, S. 1881. "Note on the frequency of use of the different digits in natural numbers." *Am. J. Mathematics* 4 (1): 39–40. <https://www.jstor.org/stable/2369148>.
- Newman, M. E. J. 2005. "Power laws, Pareto distributions and Zipf's law." *Contemp. Physics* 46 (5): 323–351. <https://doi.org/10.1080/00107510500052444>.
- Nigrini, M. J., and L. J. Mittermaier. 1997. "The use of Benford's law as an aid in analytical procedures." *Auditing* 16 (2).
- Nigrini, M. J., and S. J. Miller. 2007. "Benford's law applied to hydrology data—Results and relevance to other geophysical data." *Mathematical Geology* 39: 469–490. <https://doi.org/10.1007/s11004-007-9109-5>.
- Pesantez, J. E., E. Z. Berglund, and N. Kaza. 2020. "Smart meters data for modeling and forecasting water demand at the user-level." *Env. Modelling & Software* 125: 104633. <https://doi.org/10.1016/j.envsoft.2020.104633>.
- R Core Team. 2019. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rasekh, A., A. Hassanzadeh, S. Mulchandani, S. Modi, and M. K. Banks. 2016. "Smart water networks and cyber security." *J. Water Resour. Plann. Manage.* 142 (7): 01816004. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000646](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000646).
- Romano, M., and Z. Kapelan. 2014. "Adaptive water demand forecasting for near real-time management of smart water distribution systems." *Env. Modelling & Software* 60: 265–276. <https://doi.org/10.1016/j.envsoft.2014.06.016>.
- Rougé, C., J. J. Harou, M. Pulido-Velazquez, E. S. Matrosov, P. Garrone, R. Marzano, A. Lopez-Nicolas, A. Castelletti, and A. E. Rizzoli. 2018. "Assessment of smart-meter-enabled dynamic pricing at utility and river basin scale." *J. Water Resour. Plann. Manage.* 145 (5): 04018019. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000888](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000888).
- Russell, P. R. 2019. "The internet of water." *Engineering News-Record*, September 2/9, 28–33.
- Sambridge, M., H. Tkalčić., and A. Jackson. 2010. "Benford's law in the natural sciences." *Geophysical Research Letters* 37: L2230. <https://doi.org/10.1029/2010GL044830>.
- Sandron, F., and S. R. Hayford. 2002. "Do populations conform to the law of anomalous numbers?" *Population*, 57 (4), 755–761. <https://doi.org/10.2307/3246664>.
- Savić, D., L. Vamvakeridou-Lyroudia, and Z. Kapelan. 2014. "Smart meters, smart water, smart societies: The iWIDGET project." *Proc. Engineering* 89: 1105–1112. <https://doi.org/10.1016/j.proeng.2014.11.231>.
- Shafiee, M. E., A. Rsaekh, L. Sela, and A. Preis. 2020. "Streaming smart meter data integration to enable dynamic demand assignment for real-time hydraulic simulation." *J. Water Resour. Plann. Manage.* 146 (6): 06020008. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001221](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001221).
- Sønderlund, A. L., J. R. Smith, C. J. Hutton, Z. Kapelan, and D. Savić. 2016. "Effectiveness of smart meter-based consumption feedback in curbing household water use: Knowns and unknowns." *J. Water Resour. Plann. Manage.* 142 (12): 04016060. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000703](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000703).
- Sowby, R. B. 2018. "Conformance of public water use data to Benford's law." *J. Am. Water Works Assoc.* 110 (12) E52–E59. <https://doi.org/10.1002/awwa.1161>.
- Sowby, R. B., S. J. Burian, C. M. Chini, and A. S. Stillwell. 2019. "Data challenges and solutions in energy-for-water: Experience from two recent studies." *J. Am. Water Works Assoc.* 111 (2): 28–33. <https://doi.org/10.1002/awwa.1233>.
- Washington, L. C. 1981. "Benford's law for Fibonacci and Lucas numbers." *Fibonacci Quarterly*, 19 (2).
- Wu, Y., and S. Liu. 2020. "Burst detection by analyzing shape similarity of time series subsequences in district metering areas." *J. Water Resour. Plann. Manage.* 146 (1): 04019068. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001141](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001141).
- Xenochristou, M., Z. Kapelan, and C. Hutton. 2020. "Using smart demand-metering data and customer characteristics to investigate influence of weather on water consumption in the UK." *J. Water Resour. Plann. Manage.* 146 (2): 04019073. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001148](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001148).