CSCE Annual Conference
Growing with youth – Croître avec les jeunes

Laval (Greater Montreal)
June 12 - 15, 2019

UNDERSTANDING THE WATER-ENERGY NEXUS IN URBAN AREAS: A CLUSTER ANALYSIS OF URBAN WATER AND ENERGY CONSUMPTION

Lufan Wang[1,2] and Nora M. El-Gohary[1]
[1] Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[2] lwang105@illinois.edu

Abstract: The rapid urbanization and population growth are changing the global water- and energy-use patterns. A better understanding of the complex interactions between the urban water and energy systems is crucial in developing healthy, resilient, and sustainable cities. Although the water-energy nexus has attracted much research attention in recent years, substantial knowledge gaps still exist. There is still a limited understanding of the interrelationships between urban water and energy usage. First, most of the existing efforts analyzed water and energy usage separately. Second, they used a limited number of cases in their analyses. Towards addressing these knowledge gaps, this paper proposes a data-driven methodology for identifying and understanding the different water and energy consumption patterns of cities. A cluster analysis of 89 U.S. cities was conducted, taking socioeconomic factors (e.g., population, median household income, and total number of housing units), local climate conditions, and water and energy consumption data into account. The results show that the cluster analysis can help identify and characterize urban water-energy consumption patterns. These patterns could provide insights to support water-and-energy decision and policy making.

1 INTRODUCTION

The rapid urbanization and expansion of metropolitan areas have resulted in severe demands on water and energy resources, which have intensified the pressure on the limited resources, thereby threatening the sustainability of the urban environment. Water scarcity already affects four out of every ten people (United Nations 2017). Predictions forecast that almost half of the global population will face water scarcity by 2030 (OECD 2012), and the total world energy consumption will increase by 28% by 2040 (EIA 2017a). In the United States, about 355 billion gallons of water are withdrawn for use each day, among which about 38% withdrawals are for thermoelectric power and 14% for public supply (Maupin et al. 2014). About 80% of the total energy consumption comes from fossil fuels (EIA 2017a), and about 4% of the nation's electricity generation is water-related. As water and energy are two intertwined and closely related critical resources, understanding the complex interactions, synergies, and dependencies between the urban water and energy systems is crucial in developing healthy, resilient, and sustainable cities.

The increasing attention to the water-energy nexus has attracted a boom of research efforts on this subject. Existing studies have focused on various geographical scales, such as household (e.g., Hussien et al. 2017), city (e.g., Chini and Stillwell 2018a), regional (e.g., Marie Marsh 2008), national (e.g., Howells and Rogner 2014), and global (e.g., Endo et al. 2017) scales. Despite the importance of these efforts, substantial knowledge gaps still exist. There is still a limited understanding of the interrelationships between urban water and energy usage. First, most of the existing efforts analyzed water and energy usage separately (Dai et al. 2018), and existing water and energy policies are developed largely in isolation

(Hussey and Pittock 2012). For example, Novia et al. (2016) conducted a clustering analysis focusing on urban water demand and supply, and Lee et al. (2017) conducted a lifecycle analysis focusing on energy intensity and greenhouse gas emission quantification. Second, most of the existing studies used a limited number of cases for their analyses. For example, Kontokosta and Jain (2015) analyzed the determinants and spatial patterns of water consumption for only 2,300 multi-family buildings within New York City.

To address the aforementioned knowledge gaps, this paper proposes a data-driven clustering-based methodology for identifying the different water and energy consumption patterns of cities. Specifically, 89 major cities across the United States were analyzed using spectral clustering. Water and energy consumption features, local climate features, and socioeconomic features (e.g., population, median household income, and total number of housing units) were used in the analysis.

2 BACKGROUND

Cluster analysis aims to find structures in the data by identifying natural groups and similar objects. One object in a given cluster is similar to another in the same cluster, but different from the objects in other clusters. The main clustering algorithms include distance-based algorithms (e.g., k-means and hierarchical clustering), probabilistic and generative model-based algorithms [e.g., expectation maximization (EM) algorithm], density-based algorithms [e.g., density-based spatial clustering of applications with noise (DBSCAN)], grid-based algorithms [e.g., statistical information grid (STING)], and graph-based algorithms (e.g., spectral clustering). Spectral clustering aims to cluster data that are connected but not necessarily compact or clustered within convex boundaries (Von Luxburg 2007). It creates a similarity graph based on the eigenvalue matrix and uses the graph Laplacian to project the data to a lower-dimensional feature space with an easily separable shape, which can be clustered using simple methods such as k-means (Von Luxburg 2007).

Cluster analysis has been widely used in urban studies for characterizing city typologies, identifying geographical areas with similar use, studying human activity patterns, and understanding city dynamics. For example, using a hierarchical clustering methodology, Mikelbank (2004) analyzed the population, place, economy, and government-related characteristics for 3,567 U.S. suburban places, and identified ten distinct suburb types. Louf and Barthelemy (2014) clustered 131 cities worldwide into four clusters based on the distribution of sizes and shapes of their city block forms and urban street patterns. Becker et al. (2011) clustered city residents and visitors based on the anonymized call detail records (CDRs), for better understanding the flow of people into and out of the city. And, Cranshaw et al. (2012) studied the composition, structure, and character of a city based on a cluster analysis of the social media check-ins generated by the city residents, which revealed the current dynamics of the local urban areas.

More recently, a limited number of efforts used cluster analysis in the water-energy nexus domain. For example, Noiva et al. (2016) identified six city clusters by analyzing 142 cities around the world based on city size, per capita water consumption, and net annual water balance. Haben et al. (2016) used a finite mixture model-based clustering method to identify the peak energy demand behaviors of residential customers. Other studies (e.g., Candelieri 2017 and Tang et al. 2014) used clustering of time series data for forecasting water and energy demands.

3 RESEARCH METHODOLOGY

The proposed data-driven methodology for clustering cities based on their urban water and energy consumption includes four steps: (1) data collection; (2) data preprocessing; (3) spectral clustering; and (4) city characterization.

3.1 Data Collection

Four main types of data were collected for this research: water consumption data, energy consumption data, socioeconomic data, and local climate data. Socioeconomic and local climate features were considered in this water-energy analysis, because they were selected as energy indicators by the Energy

Information Administration (EIA) for conducting energy assessments and measuring state energy performance (EIA 2017b); and they have been found to be correlated with water-use patterns (Kontokosta and Jain 2015).

The water consumption data were collected from a database published by Chini and Stillwell (2018b). This database was developed by sending open-record requests to utilities in 127 cities across all 50 states and the District of Columbia. All the 127 cities have a population greater than 100,000, or are the largest cities in their states. Data about drinking water volume, wastewater volume, their corresponding embedded energy consumption, and the service population for each utility were included in the database. For this paper, cities with missing drinking water volume data were excluded, resulting in 89 cities. The total water consumption amount was assumed to be equal to the total volume of drinking water supplied by the utilities. A map of the selected 89 cities is shown in Fig. 1.

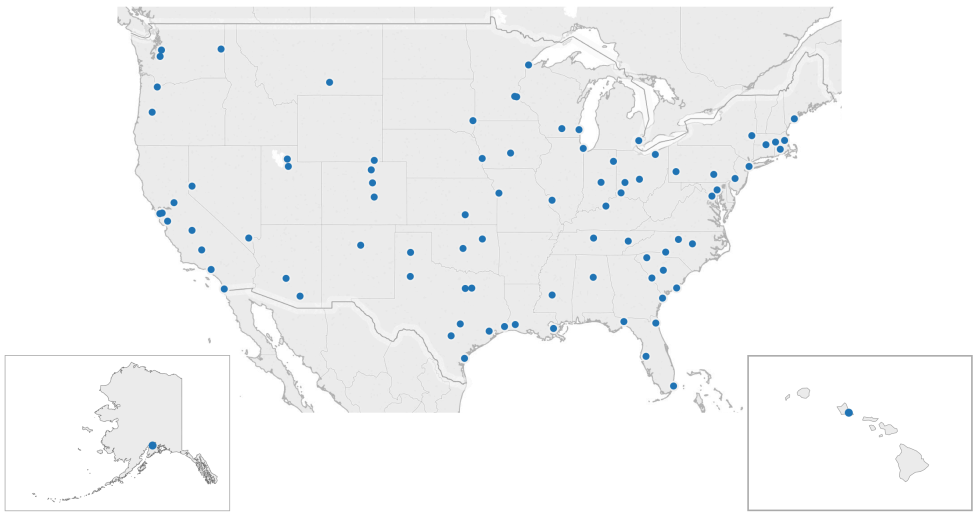

Fig. 1: Map of the selected 89 cities

The energy consumption data were collected from the State Energy Data System (SEDS) by EIA. This database provides comprehensive estimates of energy production, consumption, prices, and expenditures every year by energy source and sector (EIA 2017b). However, for analysis and forecasting purposes, the state level is the smallest unit in the EIA estimates, due to the possible inconsistency of the boundaries for the incorporated places or the census designated places (CDPs) (i.e., cities, towns, or villages). Therefore, in this study, the energy consumption data for each city were substituted by its corresponding state-level data.

The socioeconomic data were collected from the American Community Survey (ACS), which is conducted by the U.S. Census Bureau, and provides detailed social, economic, and housing statistics for every community every year (US Census Bureau 2018). The data included in this research are the features about population, median household income, total number of housing units, and total travel time to work.

The local climate data were collected from the 1981-2010 Climate Normals, which are the latest three-decade averages of the climatological variables that were released by the National Climatic Data Center (NCDC) of the National Oceanic and Atmospheric Administration (NOAA) (NCDC 2018). The database includes hourly, daily, monthly, seasonal, and annual temperature and precipitation climate normals for over 9,800 stations across the United States. The data included in this study are the annual temperature and precipitation normals.

A brief description of the features is presented in Table 1.

Table 1: Description of selected features

#	Feature Name	Description
1	Per capita water consumption (Gallon)	Total annual drinking water volume per capita
2	Per capita energy consumption (Million Btu)	Total annual energy consumption per capita
3	Population	Total population
4	Median household income ($)	Median household income in the past 12 months (in 2017 inflation-adjusted dollars)
5	Housing units (#)	Total number of unweighted sample housing units
6	Total travel time to work (<10 min)	Total population who travel to work in less than 10 minutes
7	Total travel time to work (10-29 min)	Total population who travel to work in 10 to 29 minutes
8	Total travel time to work (30-59 min)	Total population who travel to work in 30 to 59 minutes
9	Total travel time to work (>=60 min)	Total population who travel to work in 60 minutes or more
10	Total precipitation (Inches)	Annual total precipitation normals
11	Average temperature (°F)	Annual average temperature normals

3.2 Data Preprocessing

The data, from the aforementioned multiple sources, were fused and preprocessed. First, the raw water consumption data, which were in different temporal scales (i.e., monthly or annually), were aggregated to annual scales for consistency. Second, the total annual water consumption data were transformed to water consumption per capita, which was calculated by dividing the total volume of the annual drinking water by the service population. Third, the data values were normalized to the same magnitude between zero and one, by applying the min-max normalization method, as shown in Eq. (1), where x_i' and x_i are the normalized and original data, respectively, and $\max(x)$ and $\min(x)$ are the maximum and minimum values in x_i.

[1] $x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)}$

3.3 Spectral Clustering

This step aimed to use spectral clustering to cluster the 89 cities based on the aforementioned features, including the selection of the optimal number of clusters. The Python Scikit-learn module (Pedregosa et al. 2011) was used for implementing the clustering algorithm. The number of clusters was selected by running the algorithm several times, calculating the R-squared (RS) index, and using the elbow plot. The RS index indicates the extent of dissimilarities between the different clusters. It calculates the percentage of variance, i.e., the ratio of the between-group variance to the total variance. The elbow plot is a graph that shows the RS index results against the number of clusters, which takes an elbow-like shape. After the “elbow” point, the addition of more clusters will not significantly improve the percentage of variance explained by the clusters. Therefore, that “elbow” point corresponds to the optimal number of clusters.

3.4 City Characterization

This step aimed to analyze the features in each cluster, both qualitatively and quantitatively, and identify the different types of cities corresponding to the clusters. The t-SNE (t-Distributed Stochastic Neighbor Embedding) method (Van Der Maaten and Hinton 2008), as well as the box plots and kernel density estimation (KDE) plots for the distribution of each feature were used for the analysis. The t-SNE method is a technique for dimensionality reduction, which is particularly developed for enhancing the visualization of high-dimensional datasets in a two-dimensional space. It preserves the local structure in a dataset, such that the points that are close in the high-dimensional space remain close in the new, low-dimensional space.

The box plots and KDE plots are non-parametric methods to display the variations and estimate the probability density functions of the data samples.

4 RESULTS AND DISCUSSION

The dataset was clustered using the spectral clustering algorithm, with the number of clusters ranging from two to ten. The RS index was computed and the elbow plot was developed, as shown in Fig. 2. Six clusters were found to be the optimal number of clusters. The map in Fig. 3 displays how these clusters distribute spatially and how they relate to the geographic regions designated by the U.S. Census Bureau (e.g., west south central region, south Atlantic region, mountain region, and Pacific region) and the climate regions by the Building America Program of the Department of Energy (e.g., cold, mixed humid, and marine). Although the clusters do not follow the geographic or climate regions (e.g., Cluster 2 spans over four geographic regions and three climate regions), they are somewhat related (e.g., the majority of the cities in Cluster 1 is in the Pacific region). Fig. 4 shows the scatter plot of the 89 cities in a two-dimensional space produced by t-SNE. It conveys a clearer visualization of the relative similarities between the cities and clusters. Figs. 5 and 6 show the box plots and KDE plots for the distribution of each feature.

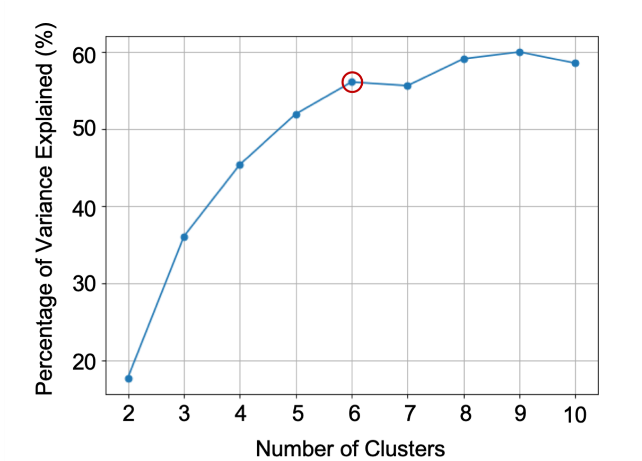

Fig. 2: Elbow plot: percentage of variance explained by the clusters against the number of clusters

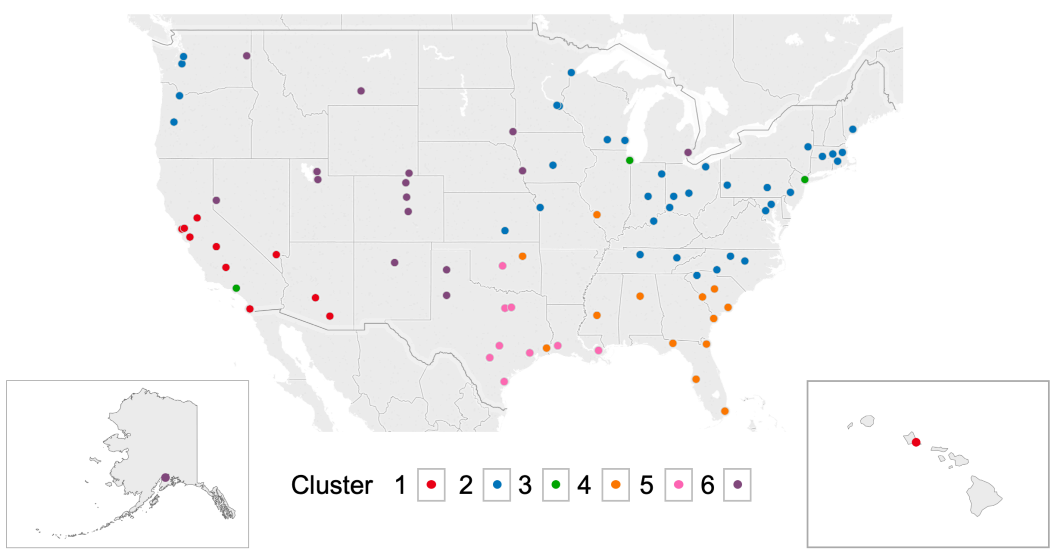

Fig. 3: Map of the selected 89 cities after clustering

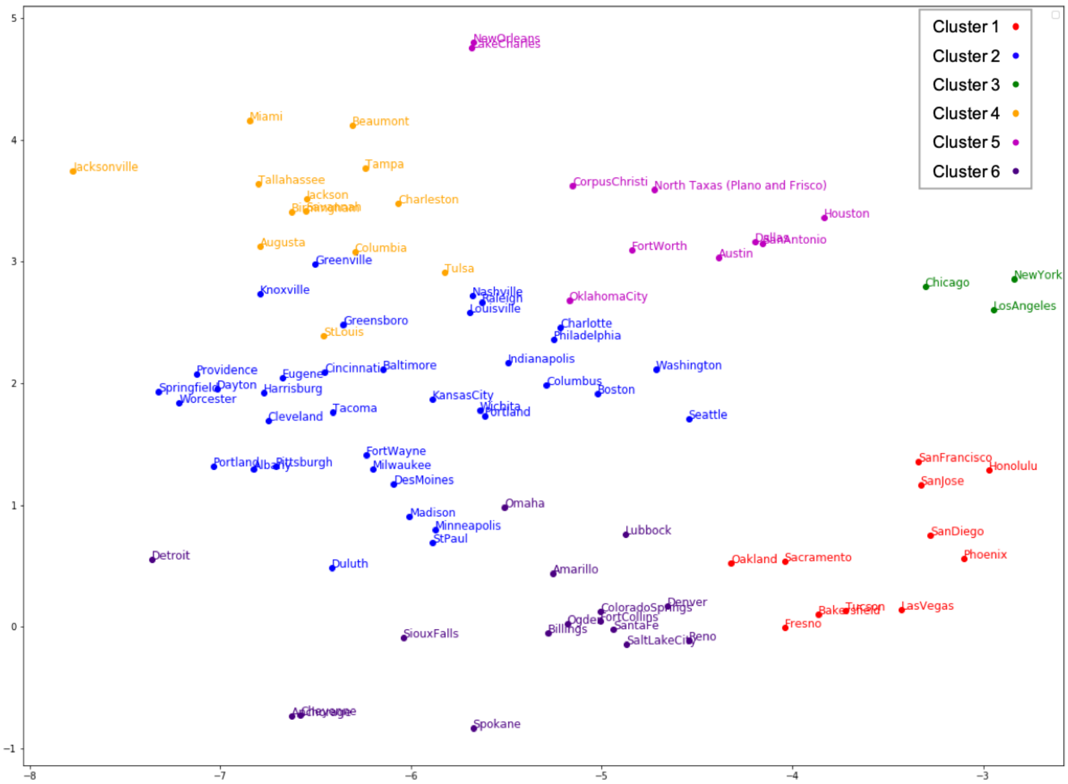

Fig. 4: Scatterplot of the 89 cities in two-dimensional space by t-SNE

Based on the clustering results, six types of cities were identified and characterized.

- Type 1 (Cluster 1)

Cluster 1 includes the major cities in California, such as San Francisco, Sacramento, San Jose, and San Diego. In addition, Las Vegas in Nevada, Phoenix and Tucson in Arizona, and Honolulu in Hawaii were also included in Cluster 1. This cluster has a medium level of water consumption and the lowest per capita state energy consumption. The cities in this cluster are mostly medium-to-large in size and have a dry and hot climate. On average, the cities in this cluster have the largest median household income, the second largest population and number of housing units, the lowest precipitation, and a relatively high average temperature compared to other clusters. This cluster is already showing relatively good energy performance, due to the high average city temperatures. According to EIA, space heating is the largest household energy-consuming end use (EIA 2018). Without much need for space heating, the cities in this cluster have potential for additional energy savings. More efforts are needed to improve their water consumption performance, which is only medium, especially with the majority of cities in this cluster suffering from serious shortages of water.

- Type 2 (Cluster 2)

Cluster 2 has the largest number of cities. The representative cities are Seattle, Portland, Cleveland, Baltimore, and Milwaukee. In general, these cities are medium in size and have a humid but cold climate. Although this group of cities has a low water consumption and a medium-level state energy consumption, their box plots tend to have long tails, which indicates that the consumption levels varied within the cluster. For example, the water consumption in Seattle is about 82 kilogallons per capita, compared to the median, 45 kilogallons per capita.

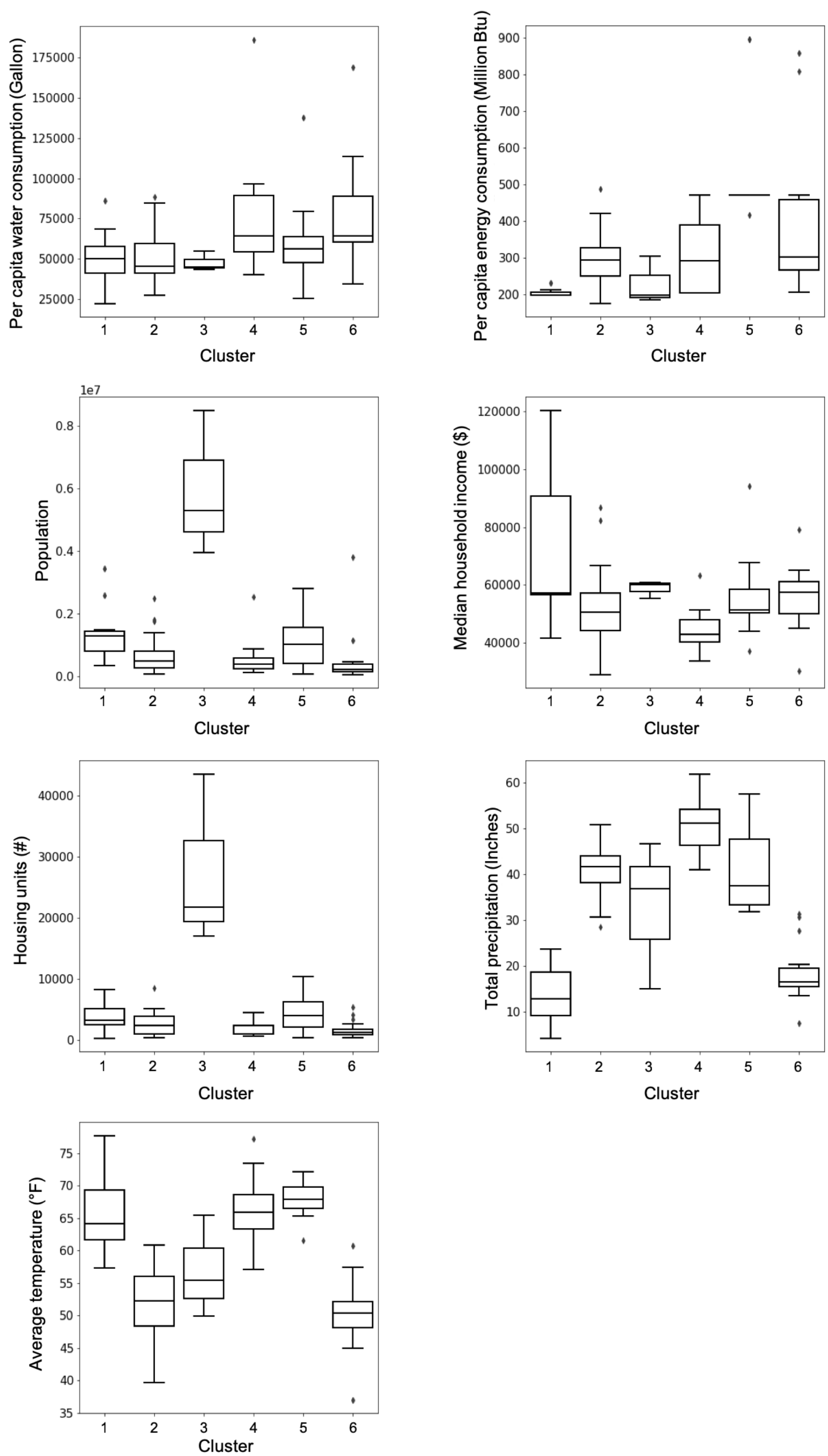

Fig. 5: Distribution of Feature 1-5 and 10-11 by cluster

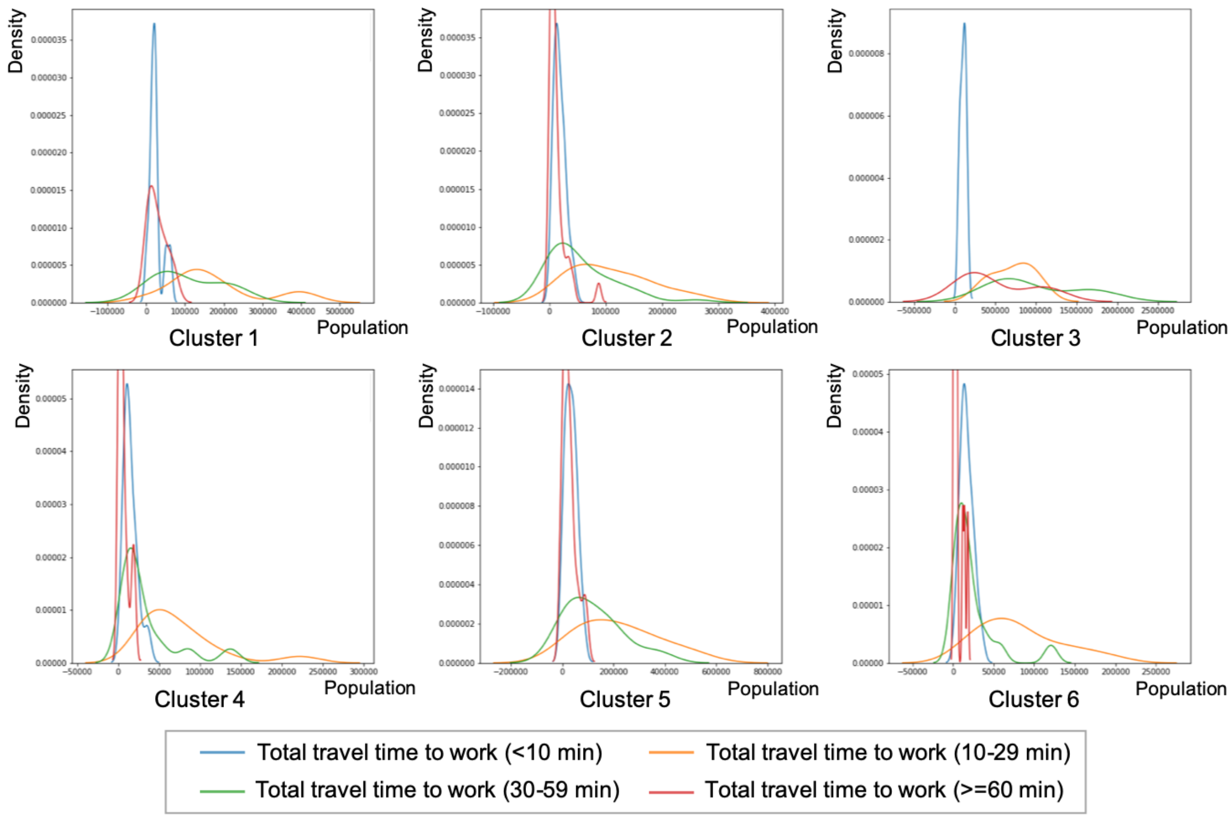

Fig. 6: KDE plot of Feature 6-9 by cluster

- Type 3 (Cluster 3)

The three megacities, New York, Los Angeles, and Chicago, form Cluster 3. Compared to the cities in the other clusters, these three largest U.S. cities have a significantly larger population and number of housing units. But interestingly, their average water consumption levels are low, even though they are located in relatively different climate regions. This indicates that there are other factors for this relatively good water performance, other than climate. For example, these megacities often have energy-efficient programs, green initiatives, and regular benchmarking efforts to save water and energy, which could be a factor. Similarly, they might have more water- and energy-saving buildings (e.g., LEED-certified buildings) or more frequent building renovations, which could improve the cities' overall water and energy performance.

- Type 4 (Cluster 4)

The majority of the cities in Cluster 4 are located in the south Atlantic region, such as Miami, Charleston, Birmingham, and Savannah. On average, the cities in this cluster have the highest total precipitation and a relatively high average temperature. With abundant water and energy resources, these cities have high water consumption and medium-to-high state energy consumption levels. With these high consumption rates, the cities in this cluster need to pay more attention to water and energy savings.

- Type 5 (Cluster 5)

The cities in Cluster 5 are mainly located in the west south central and the east south central regions. The majority of Type 5 cities comes from Texas. Examples include Houston, Dallas, Austin, and San Antonio. In addition, Oklahoma City, New Orleans, and Lake Charles are also included in Type 5. These cities are medium-to-large in size, and are located in hot climate regions. This cluster has a large variance in population size and number of housing units. But, on average, the cities in this cluster have the highest average temperature, a medium level of precipitation, a medium level of water consumption, and the highest level of state energy consumption. Texas, Louisiana, and Oklahoma are the top energy-producing states in the United States. Specifically, Texas is the largest energy-producing state and the largest energy-consuming state in the nation. Although these cities have abundant natural resources, they need to consider possible energy-saving initiatives or more efficient energy distribution systems for saving energy.

- Type 6 (Cluster 6)

The cities in Cluster 6 are mainly spanned in the mountain region, such as Denver, Salt Lake City, Omaha, and Santa Fe. The majority of Type 6 cities are small in size and are located in a relatively cold and dry climate zone. On average, the cities in this cluster have the second lowest annual precipitation and the lowest average temperature. The water and state energy consumption for Type 6 cities are both high. The water, energy, and socioeconomic characteristics of Type 6 cities are, in general, similar to those of Type 4. For example, they both have similar population sizes, median number of housing units, and travel-time-to-work patterns. However, they are located in two distinct climate regions – Type 6 is cold and dry, while Type 4 is hot and humid. This result indicates that regardless of the city location and climate regions, other factors such as neighborhood characteristics, building characteristics, and individual behaviors could have significant influence on the water and energy performance. Further analysis will be conducted in future work to study the reasons behind these differences.

5 CONCLUSIONS AND FUTURE WORK

In this paper, a cluster-based approach for analyzing the urban water and energy consumption patterns was proposed. A total of 89 cities across the United States were analyzed. In addition to water and energy consumption features, seven socioeconomic features (e.g., population, median household income, and total number of housing units) and two local climate features (i.e., annual temperature and precipitation normals) were considered in the analysis. Spectral clustering was conducted, and six meaningful city clusters were identified. The cities were classified into six city types according to the clustering results. This research shows the potential to use clustering for characterizing urban water and energy consumption patterns. The proposed approach could be used to better understand the water-energy consumption patterns and saving potentials of the U.S. cities, which could provide insights to support water-and-energy decision and policy making. The aforementioned analysis of the 89 cities is only preliminary. The next steps of this work are to consider more relevant features in the cluster analysis, introduce temporal variations in the analysis, and further validate the proposed approach.

ACKNOWLEDGEMENTS

This research is based upon work supported by the Strategic Research Initiatives (SRI) Program by the College of Engineering at the University of Illinois at Urbana-Champaign.

REFERENCES

Becker, Richard A, Ramon Caceres, Karrie Hanson, Ji Meng Loh, Simon Urbanek, Alexander Varshavsky, and Chris Volinsky. 2011. “A Tale of One City: Using Cellular Network Data for Urban Planning.” *IEEE Pervasive Computing* 10 (4):18–26.

Candelieri, Antonio. 2017. “Clustering and Support Vector Regression for Water Demand Forecasting and Anomaly Detection.” *Water* 9 (3): 224.

Chini, Christopher M., and Ashlynn S. Stillwell. 2018a. “The State of U.S. Urban Water: Data and the Energy-Water Nexus.” *Water Resources Research* 54 (3):1796–1811.

Chini, Christopher M., and Ashlynn S. Stillwell. 2018b. “The Urban Energy-Water Nexus: Utility-Level Water Flows and Embedded Energy.” HydroShare.

Cranshaw, Justin, Raz Schwartz, Jason I Hong, and Norman Sadeh. 2012. “The Livehoods Project: Understanding Collective Activity Patterns of a City from Social Media.” In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*: 58–65.

Dai, Jiangyu, Shiqiang Wu, Guoyi Han, Josh Weinberg, Xinghua Xie, Xiufeng Wu, Xingqiang Song, Benyou Jia, Wanyun Xue, and Qianqian Yang. 2018. “Water-Energy Nexus: A Review of Methods and Tools for Macro-Assessment.” *Applied Energy* 210: 393–408.

Endo, Aiko, Izumi Tsurita, Kimberly Burnett, and Pedcris M Orencio. 2017. “A Review of the Current State of Research on the Water, Energy, and Food Nexus.” *J. of Hydrology: Regional Studies.* 11: 20–30.

Haben, Stephen, Colin Singleton, and Peter Grindrod. 2016. “Analysis and Clustering of Residential Customers Energy Behavioral Demand Using Smart Meter Data.” In *IEEE Transactions on Smart Grid*: 136–44.

Howells, Mark, and H-Holger Rogner. 2014. “Assessing Integrated Systems.” *Nature Climate Change* 4 (4): 246–47.

Hussey, Karen, and Jamie Pittock. 2012. “The Energy-Water Nexus: Managing the Links between Energy and Water for a Sustainable Future.” *Ecology and Society* 17 (1).

Hussien, Wa’el A., Fayyaz A. Memon, and Dragan A. Savic. 2017. “An Integrated Model to Evaluate Water-Energy-Food Nexus at a Household Scale.” *Env. Modelling and Software* 93: 366–380

Kontokosta, Constantine E., and Rishee K. Jain. 2015. “Modeling the Determinants of Large-Scale Building Water Use: Implications for Data-Driven Urban Sustainability Policy.” *Sustainable Cities and Society* 18: 44–55.

Lee, Mengshan, Arturo A Keller, Pen-Chi Chiang, Walter Den, Hongtao Wang, Chia-Hung Hou, Jiang Wu, Xin Wang, and Jinyue Yan. 2017. “Water-Energy Nexus for Urban Water Systems: A Comparative Review on Energy Intensity and Environmental Impacts in Relation to Global Water Risks.” *Applied Energy* 205: 589–601.

Louf, Rémi, and Marc Barthelemy. 2014. “A Typology of Street Patterns.” *J. R. Soc. Interface* 11(101).

Luxburg, Ulrike Von. 2007. “A Tutorial on Spectral Clustering.” *Statistics and Computing* 17(4): 395–416.

Maaten, Laurens Van Der, and Geoffrey Hinton. 2008. “Visualizing Data Using T-SNE.” *Journal of Machine Learning Research*. 9: 2579-2605.

Marie Marsh, Debborah. 2008. “The Water-Energy Nexus: A Comprehensive Analysis in the Context of New South Wales.” PhD Diss.

Maupin, M.A., J.F. Kenny, S.S. Hutson, J.K. Lovelace, N.L. Barber, and K.S. Linsey. 2014. *Estimated Use of Water in the United States in 2010: U.S. Geological Survey Circular 1405. US Geological Survey.*

Mikelbank, Brian A. 2004. “A Typology of U.S. Suburban Places.” *Housing Policy Debate* 15 (4):935–64.

National Climatic Data Center (NCDC). 2018. “Data Tools: 1981-2010 Normals.” https://www.ncdc.noaa.gov/cdo-web/datatools/normals.

Noiva, Karen, John E. Fernández, and James L. Jr Wescoat. 2016. “Cluster Analysis of Urban Water Supply and Demand: Toward Large-Scale Comparative Sustainability Planning.” *Sustainable Cities and Society* 27: 484–496.

OECD. 2012. *OECD Environmental Outlook to 2050*. Organization for Economic Co-operation and Development, 8, 397–413.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and et al. 2011. “Scikit-Learn: Machine Leraning in Python.” *J. Mach. Learn. Res.* 12 (10):2825–2830.

Tang, Fan, Andrew Kusiak, and Xiupeng Wei. 2014. “Modeling and Short-Term Prediction of HVAC System with a Clustering Algorithm.” *Energy and Buildings* 82 (October). Elsevier:310–321.

U.S. Energy Information Administration (EIA). 2017a. “International Energy Outlook 2017.”

U.S. Energy Information Administration (EIA). 2017b. “State Profiles and Energy Estimates for U.S. States.” https://www.eia.gov/state/?sid=US.

U.S. Energy Information Administration (EIA). 2018. “Today in Energy: Space Heating and Water Heating Account for Nearly Two Thirds of U.S. Home Energy Use.” https://www.eia.gov/todayinenergy/detail.php?id=37433.

United Nations. 2017. “Global Issues: Water.” http://www.un.org/en/sections/issues-depth/water/.

US Census Bureau. 2018. “Introduction to the American Community Survey.” https://www.census.gov/programs-surveys/acs/guidance/training-presentations/acs-intro.html.