

LETTER • OPEN ACCESS

Diverse paradigms of residential development inform water use and drought-related conservation behavior

To cite this article: Kimberly J Quesnel *et al* 2020 *Environ. Res. Lett.* **15** 124009

View the [article online](#) for updates and enhancements.

Environmental Research Letters



LETTER

Diverse paradigms of residential development inform water use and drought-related conservation behavior

OPEN ACCESS

RECEIVED

21 February 2020

REVISED

7 August 2020

ACCEPTED FOR PUBLICATION

11 September 2020

PUBLISHED

18 November 2020

Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Kimberly J Quesnel^{1,2,3} , Saahil Agrawal⁴ and Newsha K Ajami^{2,3}

¹ Blue Forest Conservation, 2716 6th Ave, Sacramento, CA 95818, United States of America

² Stanford Woods Institute for the Environment, Stanford University, 473 Via Ortega, Stanford, CA 94305, United States of America

³ Bill Lane Center for the American West, Stanford University, 473 Via Ortega, Stanford, CA 94305, United States of America

⁴ Department of Management Science and Engineering, Stanford University, 475 Via Ortega, Stanford, CA 94305, United States of America

E-mail: newsha@stanford.edu

Keywords: urbanization, water demand, water conservation, clustering

Supplementary material for this article is available [online](#)

Abstract

Widespread urbanization has led to diverse patterns of residential development, which are linked to different resource consumption patterns, including water demand. Classifying neighborhoods based on urban form and sociodemographic features can provide an avenue for understanding community water use behaviors associated with housing alternatives and different residential populations. In this study, we leveraged built environment data from the online real estate aggregator Zillow to develop neighborhood typologies and community clusters via a sequence of unsupervised learning methods. Five distinct clusters, spatially segregated despite no geospatial inputs, were associated with unique single-family residential water use and conservation patterns and trends. The two highest-income clusters had divergent behavior, especially during and after a historic drought, thus unraveling conventional income–water use and income–water conservation relationships. These clustering results highlight evolving water use regimes as traditional patterns of development are replaced with compact, water-efficient urban form. Defining communities based on built environment and sociodemographic characteristics, instead of sociodemographic features alone, led to 3% to 30% improvements in cluster water use and conservation cohesion. These analyses demonstrate the importance of smart development across rapidly urbanizing areas in water-scarce regions across the globe.

1. Introduction

Today, 55% of the world's population lives in urban areas with the percentage expected to grow to 68% by 2050 (United Nations Department of Economic and Social Affairs Population Division 2019). As residents move away from rural areas and into cities, suburbs, and towns, many different patterns of residential development and urban form are emerging, in turn affecting urban system sustainability (Jabareen 2006, Pandit *et al* 2017, Pickett and Zhou 2015). Variations in population density, housing structures, and neighborhood configurations are linked to different energy use patterns, transportation behaviors, and public health outcomes (Berrigan and Troiano 2002, Ewing and Rong 2008, Güneralp *et al* 2017, Stokes and Seto 2019). Housing features including urban

versus suburban communities (Breyer and Chang 2014), infill development (Sanchez *et al* 2020), changing building and landscaping codes (Brelsford and Abbott 2017, Garcia and Islam 2019), and nontraditional housing arrangements (Barnett *et al* 2019) can also lead to heterogeneous residential water use behavior, which has direct implications for water resources management.

Simultaneously, sociodemographic characteristics like income and education levels often explain variations in residential water use (Schleich and Hillenbrand 2009, Shandas and Parandvash 2010, House-Peters and Chang 2011, Brelsford and Abbott 2017, Quesnel and Ajami 2017, Fan *et al* 2017). Together, built environment and social features provide the building blocks for defining neighborhoods and communities for urban water demand

assessment and planning (Jackson-Smith *et al* 2016, Stoker *et al* 2019), which in turn dictate water supply and infrastructure investment decisions at city to neighborhood scales (House-Peters and Chang 2011, Stoker and Rothfeder 2014, Gurung *et al* 2016). These characteristics are also critical for understanding water conservation behavior during drought (Fielding *et al* 2012, Polebitski and Palmer 2013, Mini *et al* 2015), when strategic resource management is particularly critical.

Grouping residential customers for water resources planning and management requires linking bottom-up, household-level data about the built environment with top-down Census block-group or tract scale features. Built environment and social data can also be acquired through customer surveys (Randolph and Troy 2008, Harlan *et al* 2009, Willis *et al* 2013, Hannibal *et al* 2018), but these can be expensive, time-intensive, and limited in scope. Some researchers have successfully accessed assessor records on homes and parcels (for example: Chang *et al* 2017, Brelsford and Abbott 2017), but this data is generally challenging to obtain in bulk, digitized formats and must be acquired on a city-by-city or county-by-county basis. Around the US, a few agencies have moved to publicly available digital records like New York's Primary Land Use Tax Lot Output (PLUTO) database, which can be used for building-level water demand studies (Kontokosta and Jain 2015), although these are not yet common.

New websites like Zillow, Redfin, or Trulia that aggregate and digitize records from multiple sources, including public agencies, offer the possibility to overcome this historic obstacle, but these sources have yet to be leveraged as a tool for analyzing urban water use. Together with Census information, this housing data can be used as an alternative to traditional sources to develop a holistic depiction of communities within a city, a spatial-scale which provides the granularity of within-city information while being more practical than customer-level analyses. This research aims to demonstrate the value of combining data on single-family residential housing features and urban form from Zillow with Census data to identify residential community groupings. In turn, these classifications can help water resource decision-makers better understand their customers' water use behavior, design optimal conservation policies, and plan for future resources needs and allocation.

We performed our analysis within a single utility—exploring high-resolution data within a small area to gain insights into behavior not possible at more aggregated scales while discovering information that can be applied in a broader context. Our study area, the City of Redwood City (Redwood City) is situated on the San Francisco Bay peninsula in California and represents a microcosm of diversity—in 2017, block-group level median household income within the service area ranged from

less than \$40 000 (including several block-groups classified as Disadvantaged Communities) to over \$220 000. The San Francisco Bay Area represents a region with varied and evolving water supply and demand regimes, making it a particularly valuable place to study urban water with lessons for other growing, semi-arid and arid regions across the Western U.S. and the world (Gonzales and Ajami 2017b). In our study, we focused specifically on single-family residential water use, which accounts for 65% of California's urban water use (California Department of Water Resources 2016) and about half of Redwood City's potable use (City of Redwood City 2015). Redwood City demand is seasonal, with higher use in the summer due to landscape irrigation.

Our research takes place over a 10-year period from 2008–2017 which included two historic droughts and an economic recession. In particular, the 2012–2016 drought was one of the most severe in California's history (U.S. Geological Survey: California Water Science Center 2018). The drought was not only exceptional hydrologically, but also in terms of state and local political actions, public awareness, and news media coverage that led to high drought saliency and has been associated with high conservation rates (Quesnel and Ajami 2017, Gonzales and Ajami 2017a, Bolorinos *et al* 2020). This historic drought provides an important setting for examining not only the drivers of water conservation, but also rebound once mandatory restrictions were lifted and the drought was declared over (Gonzales and Ajami 2017a, Bolorinos *et al* 2020). Evaluating water conservation during the 2012–2016 drought, an extreme event more likely to occur in the future, provides an opportunity to evaluate changing residential water use behavior across customers under escalating climatic and policy regimes.

2. Methods

2.1. Data and data integration

The first step in our analysis was to gather, process, and integrate multiple data sources using text-matching algorithms, geocoding, and common identifiers. We aggregated and averaged (1) parcel-level housing information from Zillow (Zillow 2017) to define the built environment and (2) Census block-group level demographic information from the U.S. Census and American Community Survey (Simply-Analytics 2017) to quantify social structure at the block-group level, the smallest spatial scale of Census information. Our final database included 14 features (7 from Zillow and 7 from the U.S. Census), centered and scaled, averaged at the block-group level ($n = 46$) which were spatially distributed throughout the service area. We calculated average monthly customer-level water use for each block-group from 2008–2017. See the supplemental information for more detailed information on the input data, our cleaning, filtering,

and combining procedures, and the final characteristics of the database.

2.2. Principal component analysis and hierarchical clustering

From our final database, we generated community clusters through a sequence of unsupervised learning methods. The built environment and sociodemographic features were highly correlated which prompted us to employ a dimensionality reduction technique. We performed principal component analysis (PCA) to transform the features into a set of linearly uncorrelated variables, thereby compressing our data into a smaller number of variables that explain most of the variation in the feature space and making the clustering simpler and more intuitive. Transforming a set of correlated features into principal components (PCs) to better understand customer profiles has been used across resource sectors, including in water demand analyses both for load shape clustering (Cominola *et al* 2019) and in regression models (Polebitski and Palmer 2013). We made a scree plot to determine the optimal number of PCs to retain for clustering and identified an elbow at the 5th component. These five PCs together represented 89.6% of the variance in the data (see supplemental information (<https://stacks.iop.org/ERL/15/124009/mmedia>)).

The PCs were then used as inputs into our clustering algorithm to organize the block-groups into communities. We used the hierarchical clustering on principal components (HCPC) method (Husson *et al* 2010), which is a combination of partitional and hierarchical clustering. First, a hierarchical clustering procedure is performed using Ward's criterion based on the Euclidean distance. The tree is manually or optimally cut to form an initial set of clusters. Then a partitional clustering method, in this case k-means, is applied to improve the initial groupings obtained from the hierarchical clustering. We applied HCPC to the 46 block-groups and their values for the first five PCs. Based on the intra-cluster inertial plot and values, we determined that five clusters was the optimal solution (see supplemental information). We performed PCA and HCPC using the Factominer package in R (Lê *et al* 2008). We classified a feature as important to a principal component (PC) based on its contribution: if each variable uniformly contributed to each PC, the expected value would be $(1 \text{ feature}) / (14 \text{ total features}) * 100 = 7\%$, and we used this 7% cutoff for a feature to be considered important for a PC (Kassambara 2017).

We repeated our principal component and clustering procedure without the Zillow data, using only the 7 Census features to generate community clusters. By creating a 'counterfactual' scenario in which urban environment information is not available, we could compare water use distributions to determine the

added value of incorporating built environment features in neighborhood analyses. In this case, we identified an elbow at 3 PCs which were then used for clustering. Based on the intra-cluster inertial plot and values, we determined that five clusters was again the optimal solution (see supplemental information).

2.3. Water use and conservation comparative analyses

We evaluated cluster water use and conservation quantities and trends to evaluate how different residential neighborhood typologies are linked to water demand. Monthly block-group water use and summer monthly block-group water use distributions within each cluster were not normal (Shapiro–Wilks test, all 10 tests $p < 0.05$), thus we used the nonparametric Kruskal–Wallis test to statistically compare water use distributions *between* community clusters. Tests for (1) all months and (2) summer months were significant ($p < 0.05$) indicating that water use distributions across groups were not statistically similar. To tease out individual group differences, we used Dunn's post-hoc test with the Benjamini–Hochberg method of adjusting p-values to minimize false discovery rates.

We examined block-group water conservation and rebound between 2014 and 2017, a drought stress-test, which provides insight into how communities react to an external system shock, to capture changes in water use related to extreme drought, political actions, and local policies within Redwood City (table 1). In January 2014, the California Governor declared a drought state of emergency and called for voluntary water reductions across the state (California State Water Resources Control Board 2014). As the drought progressed, the state issued a resolution calling for conserving potable outdoor water use (California State Water Resources Control Board 2014) which coincided with Redwood City implementing outdoor water use restrictions for single-family residential customers. Customers were limited to 2-day a week outdoor watering where residential addresses ending with odd numbers were allowed to water on Monday and Thursday and with even numbers water on Tuesday and Friday. Following a statewide drought declaration in April 2015, June 2015 brought state-level mandatory water use restrictions where Redwood City was mandated to achieve 8% overall savings (California State Water Resources Control Board 2018); however, the city did not change their single-family residential irrigation policy, only the level of enforcement. The outdoor water use restriction ended in June 2016 when the mandatory conservation regulations ended and a state-wide call for 'self-certified conservation goals' at which time Redwood City set their goal to 0% (California State Water Resources Control Board 2018). The drought was declared over in April 2017.

Table 1. Drought policy timeline.

Starting month	Ending month	Local policy for single-family residential customers	Statewide policies and sentiment
January 2014	July 2014		Call for voluntary conservation
August 2014	May 2015		Outdoor water use restrictions
June 2015	May 2016	2 d a week watering restriction	First mandatory urban water use restrictions
June 2016	April 2017		'Self-certified' conservation goals
May 2017	December 2017		Drought declared over

We calculated monthly block-group absolute and relative (percentage) water conservation rates compared to the same month in 2013 in order to follow California's statewide drought mandate requirements (California State Water Resources Control Board 2015). The absolute and relative block-group level conservation rate distributions had similar features as the water use distributions (were not normal, Shapiro–Wilks test, all 10 tests $p < 0.05$; were not statistically similar, Kruskal–Wallis test, both tests, $p < 0.05$) so we deployed the same statistical testing procedures for the group comparisons. We compared conservation rate distributions between clusters for each of the five distinct policy periods.

To investigate the importance of the built environment in grouping customers for water use analyses, we also created clusters trained only using Census data. We compared cluster cohesion, or within cluster sum of squares (WCSS), between the two solutions for four different water-related variables: water use, water use with seasonality removed, absolute water conservation, and percent water conservation, which were not inputs into the clustering algorithms:

$$\sum_i \sum_{x \in C_j} (x - \bar{x}_i)^2 \quad (1)$$

where the Euclidean distance is calculated between each point x and the mean \bar{x} in cluster i and then squared and summed across all clusters. A smaller WCSS indicates tighter cohesion and thus a clustering result that is more useful for water use and conservation analyses.

3. Results

3.1. Neighborhoods, communities and their characteristics

Five residential community clusters emerged from our analyses. One important outcome is that although no geospatial inputs i.e. location coordinates were introduced into our model, the clusters were separated into spatially distinct groups (see supplemental information). This unprompted spatial

segregation provides evidence for both geographical sorting of people into similar networks and community-scale urban planning while showing the effectiveness of clustering by built environment and social features. This kind of spatial grouping also paves the way for neighborhood-level distributed water system integration, for example decentralized water recycling infrastructure and geospatially tailored customer outreach (Gurung *et al* 2016).

The features of each cluster can be examined by spider plots of feature value z-scores (figure 1) and average feature values (table 2). Cluster A is comprised of 11 block-groups and formed by block-groups/neighborhoods with generally low-income, high-density families with many people per household. Cluster B is made up of 12 block-groups and is the most similar to Cluster A in that neighborhoods in this community represent lower income customers. However, unlike Cluster A, block-groups are linked to nonfamily renters with bigger household sizes. Cluster C is made up of 11 block-groups and represents the average of Redwood City across the social and built environment.

An especially interesting result is the division of high-income customers into two distinct groups. Cluster D and Cluster E both contain block-groups with high positive values in PC1, representing affluent, highly educated neighborhoods with bigger houses. Cluster D, containing 4 block-groups, however, represents an area with younger residents living in newer houses on smaller lots with more people per household. These characteristics are in contrast to the community of Cluster E, containing 8 block-groups, which is dominated by older houses on large lots with less people per household.

Aerial imagery provides evidence for these neighborhood typologies (figure 1). For example, these pictures show dense development for Clusters A, B, and D, although the development in Cluster D shows a planned community with shared instead of individual lawns. Imagery from Clusters C and E both show spaced out single-family residential properties. The lots in Cluster E are substantially larger than those in Cluster C, however, and most include lawns.

Table 2. Mean cluster block-group built environment and sociodemographic features.

		Cluster A	Cluster B	Cluster C	Cluster D	Cluster E
Built environment features (Zillow)	Lot size in sq. ft.	5589	6230	6901	4744	14 646
	# of units > 1	0.2	0.4	0.0	0.0	0.0
	Year built	1942	1949	1958	1993	1961
	# of rooms	6.0	6.7	6.6	8.4	7.5
	# of bathrooms	1.5	2.0	2.1	2.7	2.5
	Property value	\$ 195 134	\$ 326 532	\$ 369 167	\$ 407 987	\$ 555 136
	Home value	\$ 200 919	\$ 266 207	\$ 313 947	\$ 392 823	\$ 450 988
Sociodemographic features (US Census)	Median household size	3.8	2.7	2.8	3.0	2.9
	% renter occupied	67%	69%	21%	36%	12%
	Population density per sq. mile	16 373	14 990	7382	8805	4035
	% nonfamilies	12%	21%	15%	15%	12%
	% population older than 65	7%	11%	18%	10%	20%
	% college educated or higher	16%	31%	50%	71%	56%
	Median household income (\$2017)	\$ 67 020	\$ 74 365	\$ 120 134	\$ 163 456	\$ 159 980

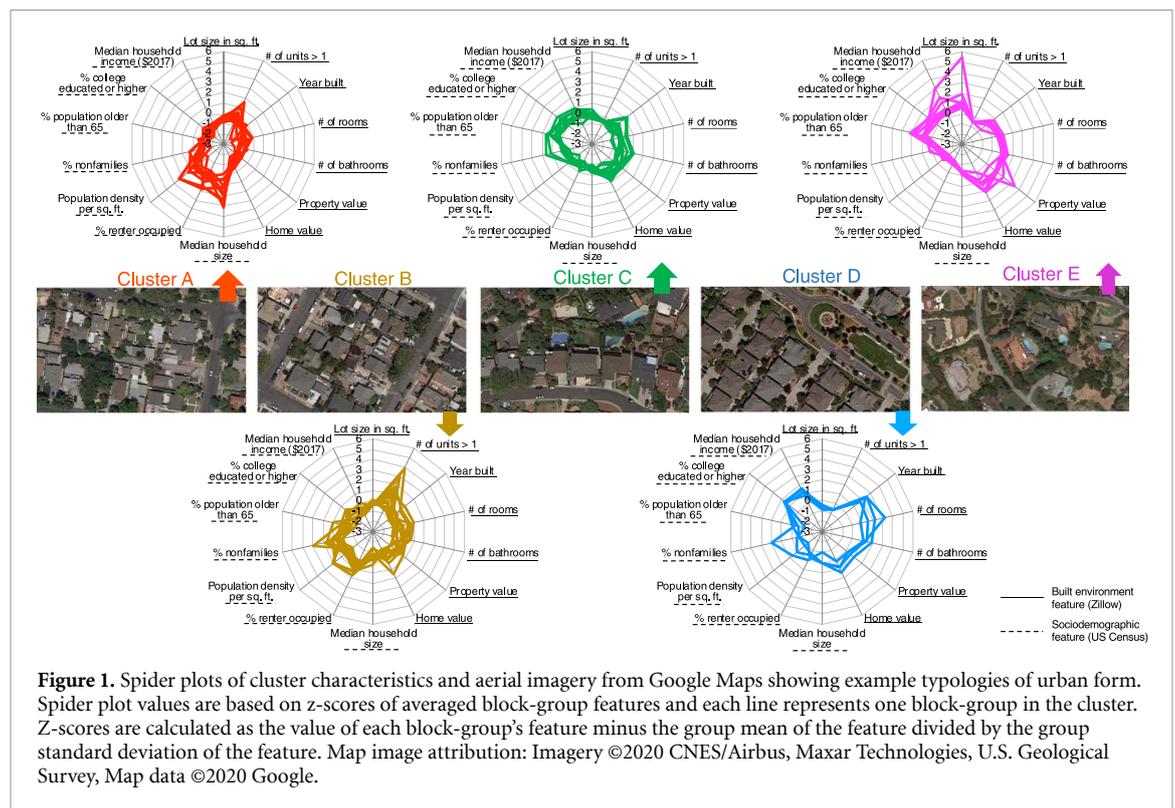


Figure 1. Spider plots of cluster characteristics and aerial imagery from Google Maps showing example typologies of urban form. Spider plot values are based on z-scores of averaged block-group features and each line represents one block-group in the cluster. Z-scores are calculated as the value of each block-group’s feature minus the group mean of the feature divided by the group standard deviation of the feature. Map image attribution: Imagery ©2020 CNES/Airbus, Maxar Technologies, U.S. Geological Survey, Map data ©2020 Google.

3.2. Community clusters and their water use

What do these neighborhood clusters mean for water use patterns and trends? A cycle plot shows average monthly water use within each cluster for the entire 2008–2017 time period (figure 2(A)) with average water use and summer water use values presented in table 3. To better understand the seasonality shown in the cycle plot, we calculated peaking factors

(Polebitski and Palmer 2010) for each block-group as the ratio of average monthly water use in the summer months divided by average monthly water use in the winter months (figure 2(B), table 3). All block-groups have average peaking factors greater than 1, indicating at least some outdoor water use within all single-family neighborhoods throughout the city. Additionally, block-group peaking factors within each cluster were not widely distributed, demonstrating

the link between neighborhood typologies and outdoor water use.

We present boxplots within violin plots, which are boxplots with rotated kernel density shapes, to visualize average monthly and average monthly summer water use by customers in block groups within each cluster (figures 2(C) and (D)). Statistically, eight of the ten water use distribution comparisons were significantly different (supplemental information) while all ten summer water distribution comparisons were significantly different (supplemental information), indicating overall different water use among the clusters. The two comparisons that were not statistically different were between Clusters A and C and between Clusters B and D. The cycle plot sheds light on these connections. Clusters B and D do in fact show similar seasonal use, in contrast to Clusters A and C, where Cluster A, with older houses and more residents per household, has higher winter use and Cluster C, with more homeowners (compared to renters) and bigger lot sizes, has higher summer use.

Block-groups in Cluster A had medium monthly and summer monthly water use compared to the other clusters and the lowest average peaking factor of 1.15 (a value shared by Cluster D). Cluster A is also the lowest income community, with high density housing, the most people per household, and mostly renters. The other low-income, majority renter community, Cluster B, had a similar mean summer water use but lower winter water use and therefore a slightly higher peaking factor of 1.21. This divergence could be due to a higher percentage of families, more people per household, and more multi-unit (although still considered single-family residential) buildings in neighborhoods that define Cluster A compared to Cluster B. Cluster C, the average of the service area across most features, exhibits the second highest monthly water use, summer monthly water use, and peaking factor of 1.28, indicating room for increased outdoor conservation and efficiency. Interestingly, Clusters C and D have similarly low winter water use to that of Cluster B, pointing towards indoor efficiency within these higher-income groups.

Cluster D had the lowest summer water use and also the lowest peaking factor of 1.15. If we only examined block-group demographics, this would be a surprising result as neighborhoods in this community are characterized by high-income, highly-educated homeowners, with larger household sizes, and these features are typically associated with higher water use (Harlan *et al* 2009, Polebitski and Palmer 2010, House-Peters *et al* 2010, Mini *et al* 2014, Breyer *et al* 2018). However, by incorporating features of the built environment, extracted from Zillow, we can understand that low water use is linked to new urban form, built in the 1990s compared to the 1940s–1960s like the rest of the city. Additionally, through

aerial imagery (figure 1), we found that within the high-income neighborhoods of Cluster D, many areas have substituted individual lawns for community or homeowner's association lawns. While water was still being used for outdoor irrigation around these houses, water use for shared lawns is not included in this study. Water use decisions made at the individual homeowner (or renter) level are fundamentally different than those made by homeowner associations and landscape professionals. In addition, many of these shared lawns are being irrigated with recycled water, making them one avenue for more efficient green spaces in urban settings (Quesnel and Ajami 2019).

Finally, customers in Cluster E had the highest average water use and highest average summer water use, including the highest mean peaking factor of 1.35, compared to the other cluster communities, which can at least be partially explained by the large lots and large houses. Notably, Cluster E contains one outlier block-group which has substantially higher income, water use, and lot sizes compared to the rest of the service area and the other block-groups in the cluster. The other six block-groups within Cluster E do, however, have the next highest water use profiles, thus still representing a cohesive high water-use group.

3.3. Water conservation and rebound

Customers were able to achieve high conservation rates, especially as the drought progressed (table 4, figure 3). During the voluntary conservation period in the first half of 2014, customers in each cluster conserved at similar relative rates, with average cluster monthly block-group savings between 6%–12% compared to 2013 across the service area. As the drought progressed and statewide outdoor water use restrictions were coupled with local watering policies, average cluster monthly block-group savings increased to 14%–23% compared to 2013. Throughout these first two policy periods, the three lower water use clusters A, B, and D had similar absolute savings while the two higher water use clusters C and E conserved in parallel and had the highest relative savings.

In spring 2015, mandatory water use restrictions were implemented across the state which led to peak drought awareness during the 2014–2017 period (Bolorinos *et al* 2020). This mandate, coupled with the continuation of local watering restrictions, resulted in the policy period with the highest relative and absolute conservation rates within each cluster during the drought, with average cluster monthly block-group savings of 23%–37% compared to 2013.

When the statewide and local mandatory water restrictions were lifted, the 'self-certified' goal policy period went into effect. Customers maintained high average cluster monthly block-group conservation rates of 20%–24% compared to 2013 despite the

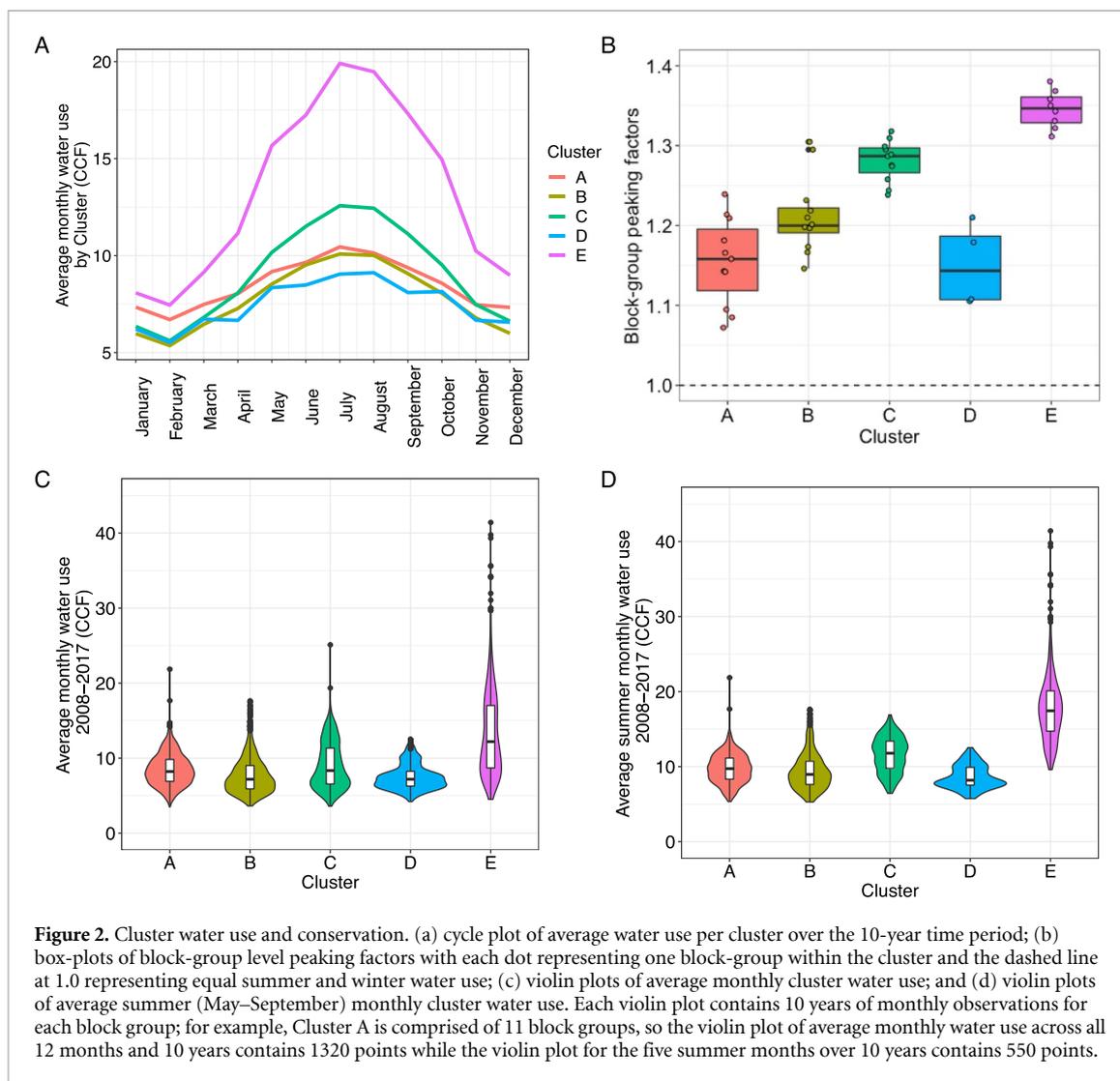


Figure 2. Cluster water use and conservation. (a) cycle plot of average water use per cluster over the 10-year time period; (b) box-plots of block-group level peaking factors with each dot representing one block-group within the cluster and the dashed line at 1.0 representing equal summer and winter water use; (c) violin plots of average monthly cluster water use; and (d) violin plots of average summer (May–September) monthly cluster water use. Each violin plot contains 10 years of monthly observations for each block group; for example, Cluster A is comprised of 11 block groups, so the violin plot of average monthly water use across all 12 months and 10 years contains 1320 points while the violin plot for the five summer months over 10 years contains 550 points.

Table 3. Cluster descriptions, water use, and peaking factors.

	Mean monthly block-group water use (CCF) [sd]	Mean monthly summer (May–September) block-group water use (CCF) [sd]	Mean block-group peaking factor [range]
Cluster A	8.47 [2.11]	9.76 [1.96]	1.15 [1.07–1.24]
Cluster B	7.76 [2.51]	9.44 [2.51]	1.21 [1.15–1.30]
Cluster C	9.02 [3.04]	11.56 [2.36]	1.28 [1.24–1.32]
Cluster D	7.47 [1.64]	8.62 [1.54]	1.15 [1.11–1.21]
Cluster E	13.30 [5.71]	17.92 [4.89]	1.35 [1.31–1.38]

absence of restrictions. However, in the spring of 2017 when the drought was declared over, conservation rates lessened to 11%–25% compared to 2013.

From the beginning of the drought until the self-certified goals were lifted, Clusters A and D had similar relative and absolute conservation rates, which were the lowest in the service area. These two groups have dramatically different demographic and housing features: Cluster A is comprised of low-income renters in small, older houses compared to Cluster D which contains block-groups defined by high-income homeowners in large, newer houses. However, both have small-lot sizes, the lowest

peaking factors and therefore low household outdoor water use. One notable difference occurs when the drought ends. Clusters A and B, with similar demographic and built environment profiles, have similar water use rebound responses while Cluster D exhibits the highest rebound (lowest conservation rates) in the service area, indicating different community responses to restrictions being lifted.

During every policy period, customers in the two communities with the highest overall water use, summer water use, and highest peaking factors, Clusters C and E, had the largest absolute savings, signaling

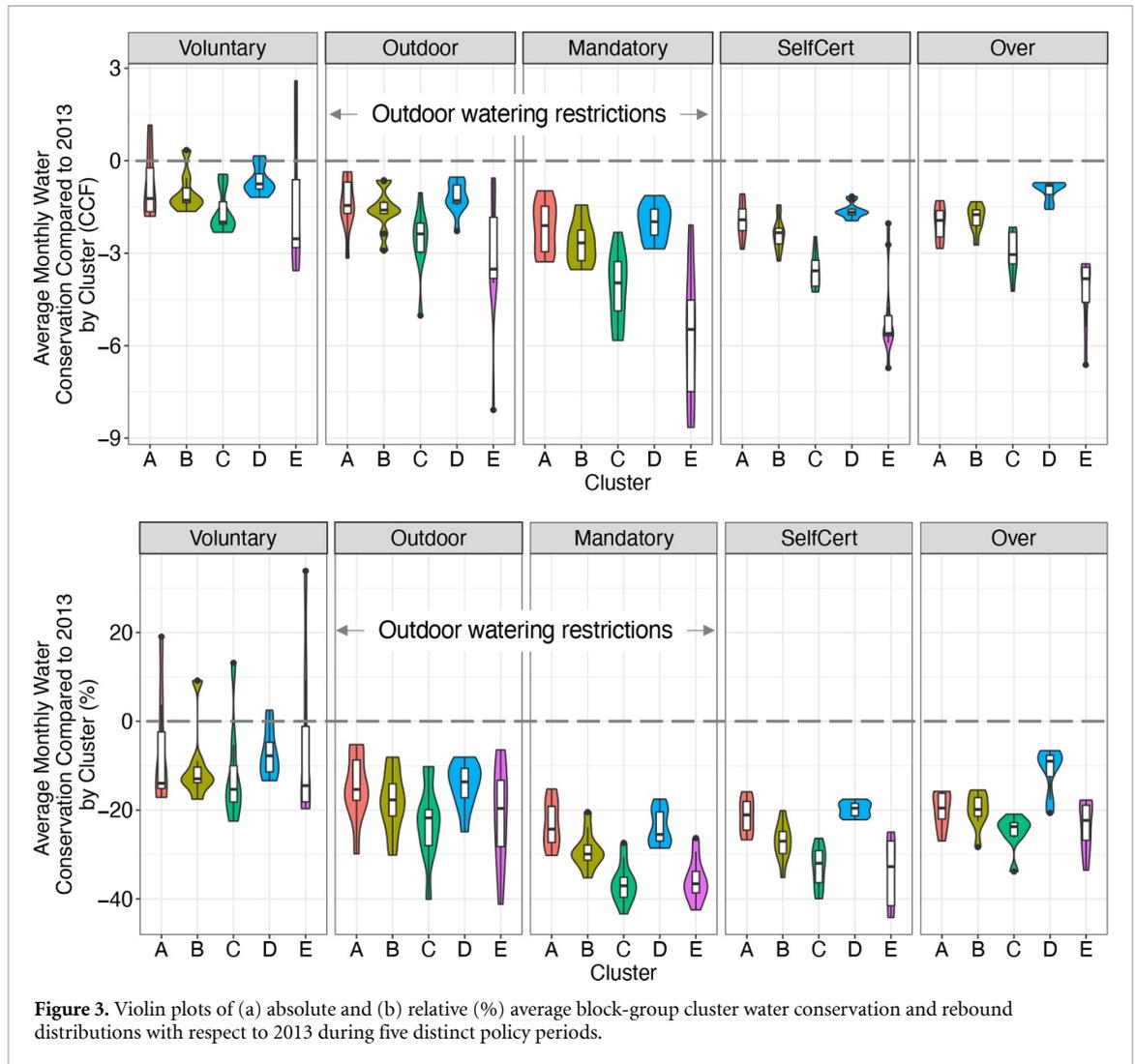


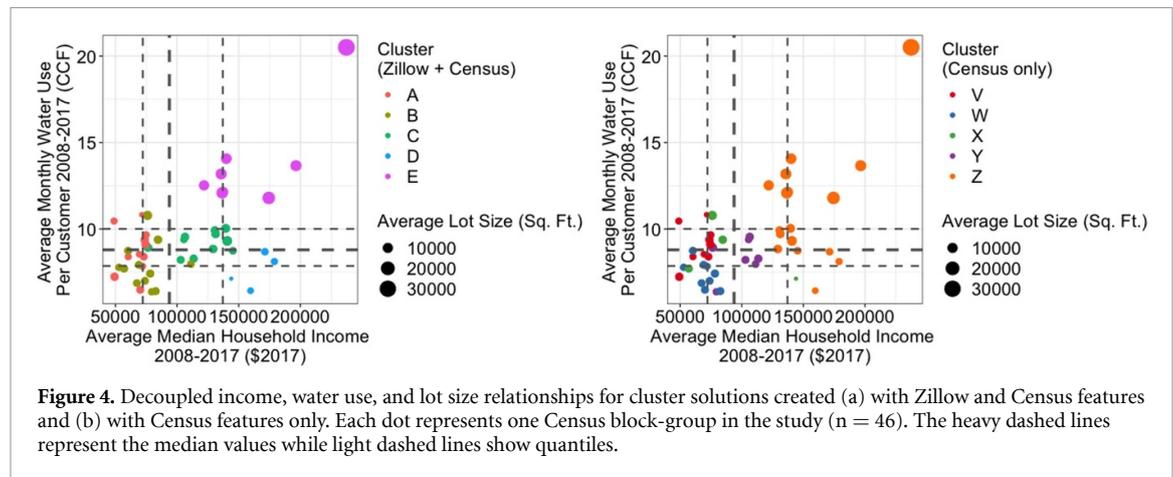
Figure 3. Violin plots of (a) absolute and (b) relative (%) average block-group cluster water conservation and rebound distributions with respect to 2013 during five distinct policy periods.

Table 4. Mean cluster absolute and relative conservation during each policy period. Letters in () under each number indicate that the distribution of block-group conservation rates in that cluster was not significantly different than the distributions within those clusters. For example, during the Mandatory policy period, the distributions of average block-group absolute conservation rates within Clusters A and D were not statistically different.

Absolute Conservation					
Average block-group monthly conservation w.r.t. 2013 (CCF)					
	Policy Period				
	Voluntary	Outdoor	Mandatory	SelfCertified	Over
Cluster A	-0.8 (B, D)	-1.4 (B, D)	-2.2 (D)	-2.0 (D)	-2.0 (B)
Cluster B	-1.0 (A,D)	-1.6 (A,D)	-2.6	-2.4	-1.9 (A)
Cluster C	-1.6 (E)	-2.6 (E)	-4.1	-3.6	-3.0
Cluster D	-0.6 (A,B)	-1.2 (A,B)	-2.0 (A)	-1.6 (A)	-1.0
Cluster E	-1.5 (C)	-3.2 (C)	-5.7	-5.0	-4.3
Relative Conservation					
Average block-group monthly conservation w.r.t. 2013 (%)					
	Policy Period				
	Voluntary	Outdoor	Mandatory	SelfCertified	Over
Cluster A	-7% (B,C,D,E)	-15% (D)	-23% (D)	-21% (D)	-20% (B)
Cluster B	-10% (A,C,D,E)	-18% (E)	-29%	-27%	-20% (A)
Cluster C	-12% (A,B,E)	-23% (E)	-37% (E)	-33% (E)	-25% (E)
Cluster D	-7% (A,B,E)	-14% (A)	-24% (A)	-20% (A)	-11%
Cluster E	-6% (A,B,C,D)	-21% (B,C)	-36% (C)	-34% (C)	-24% (C)

Table 5. Cluster water use and conservation cohesion (within cluster sum of squares, WCSS) and % improvement of including Zillow data with Census data in the clustering algorithm compared to clustering based on Census data only.

	Cluster solution Zillow + Census	Cluster solution Census only	Percent Difference
Water use (CCF) Monthly, 01/2008–12/2017	59 647	70 309	–15%
Trend + remainder (CCF) Monthly, 01/2008–12/2017	24 893	35 555	–30%
Absolute Conservation (CCF) Monthly w.r.t. 2013 baseline, 01/2014–12/2017	7306	8222	–11%
Percent Conservation (%) Monthly w.r.t. 2013 baseline, 01/2014–12/2017	391 817	402 629	–3%

**Figure 4.** Decoupled income, water use, and lot size relationships for cluster solutions created (a) with Zillow and Census features and (b) with Census features only. Each dot represents one Census block-group in the study ($n = 46$). The heavy dashed lines represent the median values while light dashed lines show quartiles.

the high potential for outdoor water conservation. During later drought periods, customers in Cluster E were able to achieve higher absolute savings, however, likely linked to their higher summer and overall water use. As urban landscapes can often recover after drought periods, even with decreased irrigation (Quesnel *et al* 2019), there is a strong case for limiting outdoor water use during water supply shortages.

Researchers have consistently reported that affluent communities are responsible for the greatest savings during drought because they generally have the largest amount of outdoor space, and thus outdoor water use (Kenney *et al* 2008, Mini *et al* 2015, Breyer *et al* 2018). Yet in this study, we found that Clusters D and E were both defined by high-income, highly educated customers but had dramatically different water use behavior during and after drought due to their built environment factors. This outcome demonstrates the importance of taking into account differences in urban form when assessing and forecasting water use, including when developing drought plans.

3.4. Counterfactual scenario: comparing cluster solutions with and without the built environment

We found that across all four cluster comparisons, the clusters created using both Zillow and Census features had lower WCSS (table 5). The clusters that

included Zillow showed a 15% improvement in overall water use cohesion and a 30% improvement in de-seasoned water use cohesion, pointing to the importance of including urban form not only when assessing total water quantities but also for examining water use trends and behavior over time. Both water conservation WCSS metrics also improved for the cluster solution that included Zillow, although to a lesser extent at 11% and 3% performance increases.

The cluster solution that takes into account both built environment and sociodemographic features splits the high-income customers into two distinct clusters (figure 1) leading to a logical grouping of their water use patterns (figure 2) and conservation and rebound behavior (figure 3). Comparing block-group water use and income relationships for both cluster solutions (figure 4) further highlights the important decoupling between income and water use, which comes through in the cluster solution only when built environment features are included.

The two cluster solutions diverge the most specifically when examining the higher-income block-groups. In the first cluster solution that includes features of the built environment (figure 4(A)), the higher-income customers are stratified into three groups that also correspond to three different levels of water use, despite no water use characteristics being used to generate the clusters themselves.

This is especially important for the highest-income customers, where the block-groups are separated into three groups—Cluster E with the highest water use, Cluster C with high to medium water use, and Cluster D with lower water use. However, in the cluster solution with sociodemographic features only (figure 4(B)), the high-income customers are all grouped into one neighborhood typology, Cluster Z, but water use spans the range of the entire city. The traditional method of classifying customers and neighborhood based only on affluence will not be sufficient for understanding water use trends and tailoring various demand management strategies as new types of urban development are integrated into cities. This distinction has important consequences for demand forecasting and supply planning. If water agencies over-predict water use from high-income customers, they will be unable to optimally invest in diverse water supplies and will mis-allocate resources throughout the community.

4. Discussion and conclusion

This research demonstrates the coupled role of urban form and sociodemographic characteristics in water use and drought-related conservation behavior. A crucial finding of this study is the importance of the built environment when classifying urban water customers and creating residential neighborhood typologies useful for water use and conservation analyses and developing demand management strategies. Traditionally, many researchers have stratified customers by income, finding that affluence is associated with high water use, high levels of drought-related conservation, and subsequently high conservation backslide post-drought. These water use characteristics are typically linked to more wealthy customers having bigger lots and therefore using more water outdoors. However, this study proves that income does not always correlate to high water use, and that affluent residents choosing to live in dense, new developments have very different water use patterns and water use behavior.

We bypassed previous data constraints faced by researchers, utilities, municipalities and other public agencies trying to better understand how built environment factors are linked to water use by utilizing data from an online real estate aggregator. By obtaining customer-level information from the Zillow ZTRAX database, this research for the first time demonstrates the possibility for this kind of platform to serve as a new tool for incorporating housing features into water use analyses. City and county assessors are legally required to fulfill public record requests and provide data in response to inquiries, but acquiring these individual datasets for comparative regional, state, or national assessments can be prohibitive. Thus, using aggregated data generated by Zillow and similar websites can provide a gateway for widespread analyses, for example opening

the door for multi-city or even multi-state studies. These new websites offer a way to access built environment data uniformly in one place instead of filing separate records request for each service area of interest. While this research presents an alternative to traditional methods for obtaining information about the built environment, water use data remains sparse and challenging to acquire (Chini and Stillwell 2016, Josset *et al* 2019).

As this is the first research to use Zillow in water demand research, there are a host of avenues for future research. For example, our study was set within one utility, but the potential for Zillow truly lies in its ability to cross administrative borders. For example, where cross-city and multi-city comparisons in different counties previously required obtaining data from each individual jurisdiction, Zillow or other online aggregators store this data in one central location. Having one central database also provides the benefit of data consistency.

Additionally, here we focused primarily on neighborhood clustering to inform residential water use modeling at the monthly scale, but future studies could explore how emerging databases, online platforms, and data aggregators with high spatial and/or temporal resolution can be coupled with higher temporal resolution water use data from smart meters to better understand and predict water use and conservation. These insights could then be used to develop short-term and long-term demand management strategies. Another avenue for future research would be to investigate how different emerging urban form paradigms including infill development, peri-urban, and suburban designs shift both short- and long-term water use behavior. Land managers and water managers often do not coordinate or even talk to each other (Gober *et al* 2013), but with these new developments and unprecedented urbanization, there is increasing pressure for integrated planning and management.

This research lays the framework for future big data-driven urban water research and provides evidence for the implications evolving urban form on water use. Our results also point out future paths of more water-efficient urban development. Cities all over the world are expanding at rapid rates, and there are many different ways for them to develop. Here, we showed that dense housing patterns and new houses, regardless of the size, can result in lower water use than traditional sprawl. However, these low water use communities also have lower conservation rates and sometimes faster post-drought water demand rebound, and actions must be taken to account for this reduced water system flexibility. City planners and water managers must work together to develop cities of the future that house an increasingly large portion of the population and meet the wide-range of sustainability-oriented goals critical to addressing 21st century urban challenges.

Acknowledgments

We would like to thank the City of Redwood City and in particular Justin Chapel, Debbie Ivazes, and Sindy Mulyono-Danre for their time and effort in providing data and valuable feedback. Research assistance by Tim Hsu was critical for initial data processing. We appreciate feedback from Jordyn Wolfand and Jose Bolorinos during manuscript development. We also thank three anonymous reviewers for their feedback which improved the manuscript. This research was developed under STAR Fellowship Assistance Agreement no. FP-91778101-0 awarded to K J Q by the U.S. Environmental Protection Agency (EPA). It has not been formally reviewed by EPA. The views expressed are solely those of the authors, and EPA does not endorse any products or commercial services mentioned in this poster. Funding was also provided, in part, by the National Science Foundation Engineering Research Center for Re-inventing the Nation's Urban Water Infrastructure (ReNUWIt) (Award No. EEC-1028968), the Bill Lane Center for the American West, and Stanford Woods Institute for the Environment.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Kimberly J Quesnel  <https://orcid.org/0000-0002-4933-4218>

Newsha K Ajami  <https://orcid.org/0000-0003-4421-3764>

References

- Barnett M J, Jackson-smith D and Endter-wada J 2019 Implications of nontraditional housing arrangements for urban water management in the United States intermountain west *Soc. Nat. Resour.* **32** 508–29
- Berrigan D and Troiano R P 2002 The association between urban form and physical activity in U.S. adults *Am. J. Prev. Med.* **23** 74–79
- Bolorinos J, Ajami N K and Rajagopal R 2020 Consumption change detection for urban planning: monitoring and segmenting water customers during drought *Water Resour. Res.* **56** 1–16
- Brelsford C and Abbott J K 2017 Growing into water conservation? decomposing the drivers of reduced water consumption in Las Vegas, NV *Ecol. Econ.* **133** 99–110
- Breyer B and Chang H 2014 Urban water consumption and weather variation in the Portland, Oregon metropolitan area *Urban Clim.* **9** 1–18
- Breyer B, Zipper S C and Qiu J 2018 Sociohydrological impacts of water conservation under anthropogenic drought in Austin, Texas (USA) *Water Resour. Res.* **54** 3062–80
- California Department of Water Resource 2016 Urban Water Use Efficiency: A Resource Management Strategy of the California Water Plan 1–36 (www.water.ca.gov/LegacyFiles/waterplan/docs/rms/2016/02_Urban_Water_Efficiency_July2016.pdf)
- California State Water Resources Control Board 2014 *Resolution No. 2014-0038* (Sacramento, CA: California State Water Resources Control Board) (www.waterboards.ca.gov/board_decisions/adopted_orders/resolutions/2014/rs2014_0038_regs.pdf)
- California State Water Resources Control Board 2015 *Resolution No. 2015-0032* (Sacramento, CA: California State Water Resources Control Board) (www.waterboards.ca.gov/board_decisions/adopted_orders/resolutions/2015/rs2015_0032.pdf)
- California State Water Resources Control Board 2018 Water Conservation and Production Reports (Sacramento, CA: California State Water Resources Control Board) (www.waterboards.ca.gov/water_issues/programs/conservation_portal/conservation_reporting.shtml)
- Chang H, Bonnette M R, Stoker P, Crow-Miller B and Wentz E 2017 Determinants of single family residential water use across scales in four western US cities *Sci. Total Environ.* **596–597** 451–64
- Chini C M and Stillwell A S 2016 Where are all the data? the case for a comprehensive water and wastewater utility database *J. Water Resour. Plan. Manag.* **143** 01816005
- City of Redwood City 2015 *2015 Urban Water Management Plan* (Redwood City, CA: City of Redwood City) (www.redwoodcity.org/home/showdocument?id=8091)
- Cominola A, Nguyen K, Giuliani M, Stewart R A, Maier H R and Castelletti A 2019 Data mining to uncover heterogeneous water use behaviors from smart meter data *Water Resour. Res.* **55** 9315–9333
- Ewing R and Rong F 2008 The impact of urban form on U.S. residential energy use *Hous. Policy Debate* **19** 1–30
- Fan L, Gai L, Tong Y and Li R 2017 Urban water consumption and its influencing factors in China: evidence from 286 cities *J. Cleaner Prod.* **166** 124–33
- Fielding K S, Russell S, Spinks A and Mankad A 2012 Determinants of household water conservation: the role of demographic, infrastructure, behavior, and psychosocial variables *Water Resour. Res.* **48** W10510
- Garcia M and Islam S 2019 The role of external and emergent drivers of water use change in Las Vegas *Urban Water J.* **15** 888–98
- Gober P, Larson K L, Quay R, Polsky C, Chang H and Shandas V 2013 Why land planners and water managers don't talk to one another and why they should! *Soc. Nat. Resour.* **26** 356–64
- Gonzales P and Ajami N K 2017b The changing water cycle: impacts of an evolving supply and demand landscape on urban water reliability in the Bay area *Wiley Interdiscip. Rev. Water* **4** e1240
- Gonzales P and Ajami N 2017a Social and structural patterns of drought-related water conservation and rebound *Water Resour. Res.* **53** 1–38
- Güneralp B, Zhou Y, Üрге-Vorsatz D, Gupta M, Yu S, Patel P L, Fragkias M, Li X and Seto K C 2017 Global scenarios of urban density and its impacts on building energy use through 2050 *Proc. Natl Acad. Sci.* **114** 8945–50
- Gurung T R, Stewart R A, Beal C D and Sharma A K 2016 Smart meter enabled informatics for economically efficient diversified water supply infrastructure planning *J. Cleaner Prod.* **135** 1023–33
- Hannibal B, Sansom L and Portney K E 2018 The effect of local water scarcity and drought on water conservation behaviors *Environ. Sociol.* **5** 294–307
- Harlan S L, Yabiku S T, Larsen L and Brazel A J 2009 Household water consumption in an Arid City : affluence, affordance, and attitudes *Soc. Nat. Resour.* **22** 691–709
- House-Peters L A and Chang H 2011 Urban water demand modeling: review of concepts, methods, and organizing principles *Water Resour. Res.* **47**
- House-Peters L, Pratt B and Chang H 2010 Effects of urban spatial structure, sociodemographics, and climate on residential

- water consumption in Hillsboro, Oregon *JAWRA J. Am. Water Resour. Assoc.* **46** 461–72
- Husson F, Josse J and Pagès J 2010 Principal component methods - hierarchical clustering - partitional clustering: why would we need to choose for visualizing data? *Tech. Rep. Appl. Math. Dep.* 1–17 (http://factominer.free.fr/docs/HCPC_husson_josse.pdf)
- Jabareen Y R 2006 Sustainable urban forms: their typologies, models, and concepts *J. Plan. Educ. Res.* **26** 38–52
- Jackson-Smith D, Stoker P A, Buchert M, Endter-Wada J, Licon C, Cannon M S and Li S 2016 Differentiating urban forms: A neighborhood typology for understanding urban water systems *Cities Environ.* **9** 5
- Josset L, Allaire M, Hayek C, Rising J, Thomas C and Lall U 2019 The U.S. water data gap—a survey of state-level water data platforms to inform the development of a national water portal *Earth's Future* **7** 433–49
- Kassambara A 2017 *Practical guide to principal component methods in R: PCA, M (CA), FAMD, MFA, HCPC, factoextra* (STHDA)
- Kenney D S, Goemans C, Klein R, Lowrey J and Reidy K 2008 Residential water demand management: lessons from Aurora, Colorado *JAWRA J. Am. Water Resour. Assoc.* **44** 192–207
- Kontokosta C E and Jain R K 2015 Modeling the determinants of large-scale building water use: implications for data-driven urban sustainability policy *Sustain. Cities Soc.* **18** 44–55
- Lê S, Josse J and Husson F 2008 FactoMineR: an R package for multivariate analysis *J. Stat. Softw.* **25** 1–18
- Mini C, Hogue T S and Pincetl S 2014 Patterns and controlling factors of residential water use in Los Angeles, California *Water Policy* **16** 1054–69
- Mini C, Hogue T S and Pincetl S 2015 The effectiveness of water conservation measures on summer residential water use in Los Angeles, California *Resour. Conserv. Recycl.* **94** 136–45
- Pandit A et al 2017 Infrastructure ecology: an evolving paradigm for sustainable urban development *J. Cleaner Prod.* **163** S19–S27
- Pickett S T A and Zhou W 2015 Global urbanization as a shifting context for applying ecological science toward the sustainable city *Ecosyst. Health Sustain.* **1** 1–15
- Polebitski A S and Palmer R N 2010 Seasonal residential water demand forecasting for census tracts *J. Water Resour. Plan. Manag.* **136** 27–36
- Polebitski A S and Palmer R N 2013 Analysis and predictive models of single-family customer response to water curtailments during drought *JAWRA J. Am. Water Resour. Assoc.* **49** 40–51
- Quesnel K J and Ajami N K 2017 Changes in water consumption linked to heavy news media coverage of extreme climatic events *Sci. Adv.* **3** e1700784
- Quesnel K J and Ajami N K 2019 Large landscape urban irrigation: a data-driven approach to evaluate conservation behavior *Water Resour. Res.* **55** 771–86
- Quesnel K J, Ajami N and Marx A 2019 Shifting landscapes: decoupled urban irrigation and greenness patterns during severe drought *Environ. Res. Lett.* **14** 064012
- Randolph B and Troy P 2008 Attitudes to conservation and water consumption *Environ. Sci. Policy* **11** 441–55
- Sanchez G M, Terando A, Smith J W, Garcia A M, Wagner C R and Meentemeyer R K 2020 Forecasting water demand across a rapidly urbanizing region *Sci. Total Environ.* **730** 139050
- Schleich J and Hillenbrand T 2009 Determinants of residential water demand in Germany *Ecol. Econ.* **68** 1756–69
- Shandas V and Parandvash G H 2010 Integrating urban form and demographics in water-demand management: an empirical case study of Portland, Oregon *Environ. Plan. B* **37** 112–28
- SimplyAnalytics 2017 US Census and American Community Survey Data (www.simplyanalytics.com)
- Stoker P, Hinners S, Jackson-Smith D, Buchert M and Levine Z 2019 Neighborhood effects on parcel-level water use *Sustain. Water Resour. Manag.* **5** 1303–13
- Stoker P and Rothfeder R 2014 Drivers of urban water use *Sustain. Cities Soc.* **12** 1–8
- Stokes E C and Seto K C 2019 Characterizing and measuring urban landscapes for sustainability *Environ. Res. Lett.* **14** 045002
- U.S. Geological Survey: California Water Science Center 2018 2012–2016 California Drought: Historical Perspective (<https://ca.water.usgs.gov/california-drought/california-drought-comparisons.html>)
- United Nations Department of Economic and Social Affairs Population Division 2019 *World Urbanization Prospects: The 2018 Revision. Key Facts* (New York: United Nations) (<https://population.un.org/wup/Publications/Files/WUP2018-KeyFacts.pdf>)
- Willis R M, Stewart R A, Giurco D P, Talebpour M R and Mousavinejad A 2013 End use water consumption in households: impact of socio-demographic factors and efficient devices *J. Cleaner Prod.* **60** 107–15
- Zillow 2017 ZTRAX: Zillow Transaction and Assessor Dataset, 2017-Q4 (Zillow Group, Inc.) (www.zillow.com/ztrax/)