DISSERTATION

A DATA-DRIVEN CHARACTERIZATION OF MUNICIPAL WATER USES IN THE CONTIGUOUS UNITED STATES OF AMERICA

Submitted by

Cibi Vishnu Chinnasamy

Department of Civil and Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2024

Doctoral Committee:

Advisor: Mazdak Arabi

Sybil Sharvelle
Travis Warziniack
Christopher Goemans

ABSTRACT

A DATA-DRIVEN CHARACTERIZATION OF MUNICIPAL WATER USES IN THE CONTIGUOUS UNITED STATES OF AMERICA

Municipal water systems in the United States (U.S.) are facing increasing challenges due to changing urban population dynamics and socio-economic conditions as well as from the impacts of weather extremities on water availability and quality. These challenges pose a serious risk to the municipal water providers by hindering their ability to continue providing safe drinking water to residents while also securing adequate supply for economic growth. A data-driven approach has been developed in this study to characterize the trends, patterns, and urban scaling relationships in municipal water consumption across the Contiguous United States. Then using sophisticated and robust statistical methods, water consumption patterns are modeled, identifying key climatic, socio-economic, and regional factors.

The first chapter of this data-driven study looked at municipal water uses of 126 cities and towns across the U.S. from 2005 to 2017, analyzing the temporal trends and spatial patterns in water consumption and identifying the influencing factors. Water usage in gallons per person per day, ratio of commercial, industrial, and institutional (CII) to Residential water use, and percent outdoor water consumption were statistically calculated using aggregated monthly and annual water use data. The end goal was to statistically relate the variations in CII to Residential water use ratio across the municipalities with their local climatic, socio-economic, and regional factors. The results indicate an overall decreasing trend in municipal water use, 2.6 gallons per person annually, with greater reductions achieved in the residential sector. Both Residential and

CII water use exhibit significant seasonality over an average year. Large cities, particularly in the southern and western parts of the U.S. with arid climates, had the highest demand for water but also showed the largest annual reductions in their per capita water consumption. This study also revealed that outdoor water use varied significantly from 3 to 64 percent of the Total water consumption across the U.S., and it was highest in smaller cities in the western and arid regions. Factors such as April precipitation, annual vapor pressure deficit, number of employees in the manufacturing sector, total percentage of houses built before 1950, and total percentage of single-family houses explain much of the variation in CII to Residential water use ratio across the CONUS.

The second chapter leverages high-resolution, smart-metered water use data from over 900 single-family households in Arizona for the water year 2021. This part of the study characterizes the determinants or drivers of water consumption patterns, specifically in single-family households, and presents a framework of statistical methods for analyzing smart-metered water consumption data in future research. A novel approach was developed to characterize household appliance efficiency levels using clustering techniques on 5-second interval data. Integrating water consumption data with detailed spatial information of the household and building characteristics, along with local climatic factors, yielded a robust mixed-effects model that captured the variations in household water uses with high accuracy at a monthly time-step. Local air temperature, household occupancy level, presence of a swimming pool, the year the household was built, and the efficiency of indoor appliances and irrigation systems were exhibited to be the key factors influencing variations in household water use.

The third and fourth chapter of this study reanalyzed the water consumption data of those 126 municipalities. The third chapter dwelled into the estimation of the state of water consumption efficiencies or economics of scale in the municipal water systems using an econometrics framework called urban scaling theory. A parsimonious mixed-effects model that combined the effects of socio-economic, built environment, and regional factors, such as climate zones and water use type, was developed to model annual water uses. The results confirm efficiencies in water systems as cities grow and become denser, with CII water use category showing the highest efficiency gains followed by the Residential and Total water use categories. A key finding is the estimation of the unique variations in water use efficiency patterns across the U.S. These variations are influenced by factors such as population, housing characteristics, the combined effects of climate type and geographical location of the cities, and the type of water use category (Residential or CII) that dominates in each city.

The fourth or the final chapter synthesizes the lessons learned previously about the drivers of municipal water uses and explores the development of a model for predicting monthly water consumption patterns using machine learning algorithms. These algorithms demonstrated improved capabilities in predicting the Total monthly water use more accurately than the previous modeling efforts, also controlling for factors with multi-collinearity. Climatic variables (like precipitation and vapor pressure deficit), socio-economic and built environment variables (such as income level and housing characteristics), and regional factors (including climate type and water use type dominance in a city), were confirmed by the machine learning algorithms to strongly influence and cause variations in the municipal water consumption patterns.

Overall, this study showcases the power of data-driven approaches to effectively understand the nuances in municipal water uses. Integration of the lessons learned and the statistical frameworks used in this study can empower water utilities and city planners to manage municipal water demands with greater resiliency and efficiency.

ACKNOWLEDGEMENTS

Firstly, I am grateful to my Ph.D. advisor, Dr. Mazdak Arabi, for his patience, research mentorship, and consistent belief in my ability to complete this research work despite my own doubts and personal tragedies. His constant reminder to focus on the small things in life was a guiding light during my Ph.D. journey. This research work was accomplished through the financial support of the National Science Foundation (NSF) grants – 1444758, as part of the Urban Water Innovation Network (UWIN), and OAC-1931363. This research was also financially supported [in part] by the U.S. Department of Agriculture (USDA), Forest Service.

I would like to thank and acknowledge all the 126 water utility companies in the USA, who provided the monthly water use data for this research. I also thank the engineering and data management team at Flume, Inc. for providing their smart-metered water use data for my research. Special thanks to Mr. Peter Mayer for enabling the collaboration with Flume and for extending his mentorship during my Ph.D. journey.

I also thank my parents and family in India, the many friends in the U.S., and my partner, Sharayah, for their unwavering support. They taught me that a Ph.D. degree is just a part of life and not life itself. They gave me many small moments in the past 5 years to look back on and cherish that period with fondness despite the many hardships that I encountered.

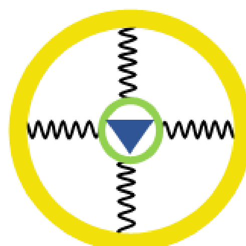
Lastly, I am grateful to this land, what is now called ‘Fort Collins’, for hosting and nourishing me. The multitude of medicine and wisdom I have received in this place has taught me the importance of being humane, humble, and heart centered, despite the number of intellectual achievements I may attain. May I continue to walk my life embodying all that.

DISCLAIMER

The findings and conclusions in this report are those of the author and should not be construed to represent any official U.S. Government or USDA determination or policy. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. government.

All the valuable and novel information presented in this report was made possible by all the ideas and support of countless numbers of people who crossed paths with me since the beginning of my Ph.D. journey at Colorado State University. All the misgivings in the report are not anyone else's but my own carelessness or naivety.

DEDICATION



To the element and spirit of Water,

and

To the grace of many who guided and supported me towards the completion of this program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEFINITION OF TERMS

Built environment -- includes all of the physical parts of where we live and work (e.g., homes, buildings, streets, open spaces, and infrastructure).

City, town, village, parish, and borough (as defined in the United States) -- refer to https://www2.census.gov/geo/pdfs/reference/GARM/Ch9GARM.pdf

CII users -- Commercial, Industrial and Institutional water users.

Climatic -- Koppen level-1 climate classifications – Arid, Continental, Temperate and Tropical.

Consumptive use -- that part of water withdrawn for supply that is evaporated, transpired by plants, incorporated into products or crops, consumed by humans or livestock, or otherwise removed from the immediate water environment. Also referred to as *"water consumed"* or *"water consumption"*.

Contiguous United States (CONUS) -- officially called the '*conterminous United States*', consists of the 48 adjoining U.S. states and the Federal District of the United States of America. The term excludes the only two non-contiguous states, Alaska and Hawaii, and all other offshore insular areas, such as American Samoa, Guam, the Northern Mariana Islands, Puerto Rico, and the U.S. Virgin Islands.

Core Based Statistical Area (CBSA) -- a U.S. geographic area defined by the Office of Management and Budget (OMB) that consists of one or more counties (or equivalents) anchored by an urban center of at least 10,000 people plus adjacent counties that are socioeconomically tied to the urban center by commuting. CBSA consists of *"metropolitan statistical areas (MSA)"* and *"micropolitan statistical areas (μSA)"*.

County -- an administrative or political subdivision of a state that consists of a geographic region with specific boundaries and usually some level of governmental authority. The term '*county*' is used in 48 states, while Louisiana and Alaska have functionally equivalent subdivisions called "*parishes*" and "*boroughs*", respectively.

Eco-hydrological -- different ecosystems where water interacts to influence living organisms and the built environment. Koppen climate regions is loosely used as eco-hydrological regions in this report. US EPA-Ecoregions are another example of eco-hydrological regions.

Metropolitan Statistical Areas (MSA) -- an urban cluster (urban area) with a population of at least 50,000 people.

Micropolitan Statistical Areas (μSA) -- an urban cluster (urban area) with a population of at least 10,000 but fewer than 50,000 people.

Municipal water system -- a water system that has at least five service connections or which regularly serves 25 individuals for 60 days. Municipal water system is also called as a *"public* *water system",* or *"public water-supply system", or "(water) utility"*. Typically, these systems supply water to residential and CII users. They systems also supply water wholesale to large industries, or to neighboring utilities or cities or counties.

Residential users -- single-family and multi-family residential water users.

Socio-economic -- differences between groups of people caused mainly by their financial situation and demographies.

Urban area -- defined by the US Census Bureau as a contiguous set of census blocks that are densely developed residential, commercial, and other nonresidential areas.

(Municipal) Water use -- water used for i) household or domestic purposes, such as drinking, food preparation, bathing, washing clothes, dishes, and dogs, flushing toilets, and watering lawns and gardens, and for ii) motels, hotels, restaurants, office buildings, other commercial and facilities, industries and institutions. If the word '*municipal*' is used in the front of water use, the water is supplied by a "*municipal water system"* or a "*public water-supply system"*. (Municipal) Water use is also referred to as *"(municipal) water demand"*.

CHAPTER 1 – MOTIVATION FOR THIS RESEARCH WORK

Today's facts are tomorrow's allegories. Our world is constantly evolving; the ways in which we grow our food, manage our natural resources and build our cities are in a constant state of flux. Changing climatic patterns and varying human behaviors and attitudes on water use add more fluidity and ambiguity to the way water resources are managed. Fluxes in climate, socio-economic, political, and environmental values, technological innovations, and community health are ever more rapid and uncertain in the contemporary world. Therefore, it is imperative to constantly study these forces for sustainably relying on natural resources to resiliently organize human communities and steward local ecosystems.

Since the early 20th century, water demand (or, water use) management practices have significantly relied on past trends in population growth, water availability, and water use trends to make projections for future water needs. This led to the boom in infrastructure projects in the mid-twentieth century which saw water being moved across the mountains and deserts, and huge dams and reservoirs being built with a naïve thinking that these systems will be used for a long time (Magionni, 2014; Niemczynowicz, 1999; Yigzaw & Hossain, 2016). Sadly, in less than a hundred years from their first operation date, infrastructure like the Hoover Dam, that helps pool water in Lake Mead and Lake Powell, is facing an uncertain future with water levels dropping drastically threatening energy production and maintenance of water level to have a healthy aquatic ecosystem (Udall & Overpeck, 2017).

The problem seems not to be with the poor designing of these structures based on future hydrology, but due to the severely underestimated water demand projections of the future. Large water consumption relative to water availability results in decreased river flows, mostly during the dry period, and declining lake water and groundwater levels (Mekonnen & Hoekstra, 2016). Western United States especially is in the dire situation now where most of the water resources are over-appropriated and the water supply is being threatened due to varying weather patterns that is been exasperated by human-built systems and ways of operating those systems (Dettinger et al., 2015). Collection and interpretation of water-use data from water suppliers, particularly if sustained for a period, provides useful information that can be compared with other variables (Kenny & Juracek, 2012). Thus, it is urgent that water demand management practices are constantly updated and improved to enhance our water security and, in the process, develop the foresight and vision to steward our natural resources than look at them as objects of exploitation.

In the United States, greater attention has been given to developing water supplies and quantifying available waters than determining who uses water, how much they withdraw and consume, and how and where water use occurs (Marston et al. 2022). The need to stay current and proactive in water demand management sector is even more important in the contemporary world where the impacts of global politics, changing socio-economic conditions, pandemics, and alternate working options (like, work from home) are having drastic influences on end-water uses. As the entropy of human activities keeps on extending and expanding, it therefore becomes imperative for water managers to continuously study the ongoing trends in water uses (Wackernagel & Rees, 1998; Kenny & Juracek, 2012).

Smart-automated technologies, that measures water use in time intervals as small as few seconds and classify water use by specific end-use type inside and outside the building using machine learning tools, are being implemented by many utilities to gather high quality data that are vital in expanding their knowledge of their customers' water use patterns (DeOreo et al., 1996; Paterson et al., 2015; Yang et al., 2017). This information is used to plan for conservation programs along with planning for future needs. Therefore, relying on such end-use data will be more important to understand the influences of climatic, socio-economic and built environment attributes on water consumption and variability in the water-supply systems. However, obtaining such high-resolution end-use water data could be a challenge due to the sensitive and proprietary nature[1] of those datasets.

Typically, a water supply system that provides water to at least 25 people or 15 service connections, that are residential and non-residential, for at least 60 days a year is referred to as a "municipal water system" or "public water-supply system"[2]. Residential water users are households that are single-family and multi-family dwellings. Non-residential water users typically range from commercial, industrial, and institutional (CII) water users to thermoelectric powerplants, public parks and gardens, and even mining and irrigation companies. Most smart, high-resolution water meters for measuring end-uses are contemporarily implemented at the residential sector of the municipal water system. In this proposed research work, municipal water systems that provide water for residential and CII uses in cities, towns, and counties within the Contiguous United States (CONUS) is studied to improve the characterization of regional and temporal trends and patterns in water uses and to identify and quantify the forces influencing it.

The overall goals of this research proposal are to collect and use the most recent and extensive set of municipal water consumption data from across the Contiguous United States to improve our knowledge of the:

- spatial and temporal patterns and trends in residential and CII water uses,

[1] https://www.usgs.gov/data-management/proprietary-and-sensitive-data
[2] https://www.epa.gov/dwreginfo/information-about-public-water-systems

- climatic, socio-economic, built environment, and regional factors that influence water consumption patterns,
- state of efficiency in municipal water supply/ consumption, and
- ways to model and predict the seasonal variations in municipal water demand.

To accomplish these overall goals, the following research activities were specifically undertaken:

i) Characterization of municipal water uses across 126 cities and towns using aggregated city- and utility- level water supply data

ii) Characterization of single-family residential water uses employing high-resolution, smart metering data

iii) Development of robust urban scaling metrics on municipal water consumption patterns

iv) Sophisticated modeling of municipal water consumption patterns using machine learning toolkit

These above-mentioned activities were achieved using a combination of contemporary data that included, a) aggregated city (or, utility)-level water use data, and b) smart-metered, high-resolution household-level water use data. Table 1.1 highlights the summary of the characteristics of these datasets in detail.

The first three research activities mentioned above parallelly contribute to the overall objective of improving our knowledge of the municipal water consumption trends and patterns and the identification of key factors influencing them. The fourth research activity compiles all the learnings from the previous research activities and proposes an improved modeling/ projection framework for municipal water consumption within CONUS. Each of the above research activities are explained in order and in detail as individual chapters in the following pages of this proposal.

Table 1.1. Summary of data sets and their characteristics used in this research work.

Data Type	Spatial resolution	Source	Sample size	Time step	Categories	Time period
Municipal water use data	City and utility level	Surveys and open data access	126 cities and towns within CONUS	Monthly water uses	Aggregated residential (both single- and multi-family) and CII	2005 to 2017
High resolution single-family water use data	Household level	Flume Inc., and Flume Data Labs	Households within the state of Arizona	Daily end-water uses	Aggregated indoor and outdoor	October 2021 to September 2022

The scope of this research work includes the characterization of the municipal water uses or demand correlated with the easily relatable local climatic, socio-economic, built environment, and regional factors compiled from external, open-access sources. Characterization of municipal water consumption incorporating the influences of the price of water, human behaviors and attitudes, availability and quality of water for supply, social equity and environmental justice, climate (floods, wildfires) related risks to water shortages, and the impacts to social and civil systems by the recent COVID pandemic were outside the scope of this research work, but are highly recognized to have an impact of how water is produced, supplied, and consumed.

CHAPTER 2 – CHARACTERIZATION OF MUNICIPAL WATER USES ACROSS 126 CITIES AND TOWNS USING AGGREGATED CITY- AND UTILITY-LEVEL WATER SUPPLY DATA

2.1 Introduction

This study investigates municipal water use patterns across the Contiguous United States (CONUS). Municipal water supply systems provide water to residential users, both single-family houses and multi-family housing structures, and to commercial, industrial, and institutional (CII) users. In some regions, these systems also trade wholesale water to large energy and manufacturing industries or to neighboring urban centers (Dieter & Maupin, 2017), and are audited as “master meter” use or supply. Water demand management strategies are increasingly relied upon to address water supply demand challenges in cities (Gleick, 2010; Odai, 2009; Yigzaw & Hossain, 2016). For water conservation programs and/or alternate water source systems to be effective, it is imperative that an improved understanding of indoor and outdoor municipal water use patterns across various city clusters is achieved (Maas et al., 2017).

Water use in cities is influenced by several physiographic, social, ecological, and other factors. Rapid urbanization along with changes in population and climatic conditions alter water use patterns in cities (Fischer, 1972; Glaeser et al., 2001; Gordon & Richardson, 1997; Grimm et al., 2008; Mills, 2007; Rees & Wackernaegel, 1996; Stone, 1972). More than 82% of North America’s population resides in urban areas (United Nations, 2019), with that percentage projected to increase over the coming decades. Growing water uses within cities and diminishing water supplies intensify the competition for water amongst cities, irrigated agriculture, industries, energy production, ecological flows, and recreational uses (Gleick, 2010; Gober & Kirkwood, 2010; Pimentel et al., 1997). The vulnerability of water supply systems to shortage is exacerbated by changing climate (Christensen et al., 2004; Folger, 2017; Gober & Kirkwood, 2010; Yigzaw & Hossain, 2016). Quantifying the end-water uses in the municipal sector is increasingly important for urban planners and policymakers to develop robust and resilient planning and water supply management strategies.

Climate change contributes to the uncertainty in predicting rainfall and snowfall patterns, thereby presenting a water supply planning challenge for the sustainable growth of communities (Pimentel et al., 1997). Climate variability and change necessitate that cities adapt and change the way urban water resources are managed (Breyer et al., 2018; Gleick, 2010; Sharvelle et al., 2017), to address both excess water during intense precipitation events and intense drought years. Climatic forces, therefore, shape the way humans use water to grow food, produce energy, and build societies. This constant feedback cycle between humans and the natural forces is gaining recognition in the form of the "hydro-social cycle" (Linton & Budds, 2014). Thus, climatic factors interact with other forces that drive water use in cities (Turral et al., 2011). Another well documented aspect of variability in residential water use is the age of the building or the year the building structure was built. DeOreo et al. (2016) and Mayer et al. (1999) have shown that the year in which the structure was built is highly correlated with the plumbing codes which determine the water use efficiency of the building.

Municipal water supply systems provide water to residential users, both single family houses and multi-complex housing structures, and to commercial, industrial, and institutional (CII) users. In some regions, these systems also trade wholesale water to large energy and manufacturing industries or to neighboring urban centers (Dieter & Maupin, 2017), and are audited as 'master meter' use or supply. Traditionally, cities have addressed growing water supply needs and long-term water security by developing new water sources. In the mid-twentieth century, large water supply infrastructure systems, like diversion of rivers, construction of dams and reservoirs, and long-distance piping of water, were developed to support human settlements and rapid urban growth across the U.S. However, by the end of that century diminishing freshwater resources due to climate change, widening socio-economic divide along the rural-urban interface, funding shortfalls, and environmental regulations have constrained the capacity of urban centers to build new water supply infrastructure (Magionni, 2014; Niemczynowicz, 1999; Yigzaw & Hossain, 2016). These factors have led to alternative modes of securing raw water through mechanisms such as inter-basin water transfers, water reclamation and water reuse programs. In the U.S., transfer of water from agriculture to urban users and other inter-basin water transactions are made within the existing physical infrastructure and water laws (Breyer et al., 2018; Merrett et al., 2003; Paterson et al., 2015). These trades often involve high legal and infrastructural transaction costs (Dozier et al., 2017; Paterson et al., 2015).

Water demand management strategies are increasingly relied upon to address water supply-demand challenges in cities (Gleick, 2010; Niemczynowicz, 1999; Yigzaw & Hossain, 2016). Municipal water suppliers have already embarked upon planning and implementation of water conservation programs and use of alternate water sources (Hering et al., 2013; Ma et al., 2015; Paterson et al., 2015; Sharvelle et al., 2017). Water conservation programs aimed at the reduction of end-water uses through a combination of voluntary/ mandatory restrictions on outdoor water use, tiered water pricing, and improved efficiency in plumbing codes using low flow water fixtures have led to improved water security in cities (Gleick, 2010; Grisham & Fleming, 1989; Kennedy et al., 2012; Maidment & Miaou, 1986; Olmstead & Stavins, 2009; Saurí, 2013). Alternate local water resources, like graywater/ blackwater recycling, rainwater/

stormwater harvesting, desalination, and aquifer storage and recovery systems, are being recognized as effective solutions to minimizing the dependency on the extraction of limited freshwater resources (Gleick, 2010; Sharvelle et al., 2017; Chinnasamy et al., 2018). For water conservation programs and/or alternate water source systems to be effective, it is imperative that an improved understanding of indoor and outdoor municipal water use patterns across various city clusters is achieved (Maas et al., 2017).

Residential water use for various indoor activities, including clothes and dish washing, bathing, toilet flushing, kitchen and bathroom faucet use, pipe leakage, and other uses for cleaning and sanitation, have been extensively studied using water meters and smart sensors (DeOreo et al., 1996; Paterson et al., 2015; Yang et al., 2017). The residential end-water use study (REU2016) compiled by DeOreo et al. (2016) concludes that on average, indoor residential water use per capita has decreased by 15.4% in North America, compared to the REU1999 study (Mayer et al., 1999). Rockaway et al. (2011) also found declining residential water use in North America due to decreasing household size and improved efficiency in water appliance standards. Although significant strides have been made to understand indoor water conservation programs, water use in the CII sector is poorly characterized due to data paucity (Rockaway et al., 2011; Dziegielewski et al., 2000). Similarly, few studies have investigated outdoor water use patterns across ecohydrologic regions. Understanding patterns of outdoor water use for landscape irrigation and CII water use is vital for improving cost, reliability, and resiliency of municipal water supply systems (Goldstein et al., 2013).

Municipal water use characteristics may be explained by a range of demographic, ecologic, and economic factors. Among these factors, climatic factors have been well-studied by Stoker & Rothfeder (2014), who developed computer models to show that urban water use, primarily outdoor use, is influenced by seasonal changes in the climate, and showed that urban land use pattern also influences the total volume of water used. In recent literature, climatic variables are extensively being used in modeling residential, primarily outdoor, water use (Maas et al., 2017; Mostafavi et al., 2018; Puri & Maas, 2020). In cities with similar population size characteristics, water uses tend to vary considerably (Ahams et al., 2017). Weather influences across multiple cities and specifically in outdoor water use has been documented by Mini et al (2014), Gober et al (2016) and Opalinski et al. (2020). Various factors influence the residential water demand and water demands for economic growth within a city, and these factors have been well studied and documented separately for individual water use categories predominantly for the western part of CONUS (Decker et al., 2000; Folke, 2003; Kennedy et al., 2012; Magionni, 2014; Mini et al., 2014; Stoker & Rothfeder, 2014; Folger, 2017; Wang et al., 2018).

However, an approach to holistically understand municipal water use by analyzing residential and CII water use together across multiple cities for the entire CONUS is non-existent in the literature. CII water use is not well understood (DeOreo et al., 2016), increasing the complexity of characterizing urban water use. Understanding municipal water use under the lens of CII to residential water use (CII/Res) ratio is a new approach suggested in this paper, and it is speculated to be influenced by the general characteristics of a city such as climate, socio-economy and urban geography - a pattern of urban development driven by population changes that influences land use and management (McGill, 1964). Currently, these relationships are poorly understood and characterized. Thus, understanding and quantifying the factors that influence residential and CII water uses, within a city or a region, has the potential for use in urban water demand projection models and can facilitate sustainable and resilient water resource management (Puri & Maas, 2020).

2.2 Objectives

This study addresses a critical knowledge gap in characterization of municipal water use patterns across the CONUS using multi-city data. Specifically, the objectives of this study are to:

- conduct an extensive survey of water use in cities with varying demographic, economic, and ecohydrological characteristics
- explore spatial and temporal trends in end-water uses along with improved characterization of outdoor and CII water uses
- explain the variation in CII to residential water use across CONUS using climatic, socio-economic and urban-geographic factors

2.3 Data and Study Regions

A survey (Figure SI-2.1) of monthly municipal water uses from 126 cities and towns with varying demographic, economic, and eco-hydrological characteristics was conducted across the CONUS for the period 2005–2017 (Figure 2.1). In the survey for each municipality, monthly water use data was compiled under residential, CII and total water use categories, and information such as average population served per year and a digital map of the service area was also requested.

Three clusters were used to characterize municipal water use patterns across the 126 study cities in the CONUS. They are:

i) U.S. Census regions: Midwest, Northeast, South and West,
ii) Level-1 Koppen climate regions: Arid, Continental, Temperate, Tropical, and Highlands, and

iii) Population size clusters: Small (<100,000 people), Medium (100,000 to 1,000,000 people), and Large (>1,000,000 people).

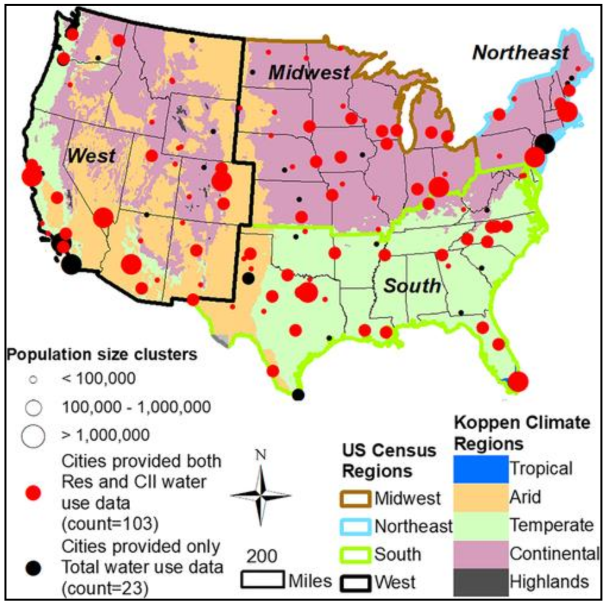


Figure 2.1. Map of the study cities clustered by U.S. Census regions and Koppen climate regions. Cities in this map are also differentiated by population size and the type of municipal end-water use data collected through survey.

Table 2.1. Groups of 81 exploratory factors used in this study.

Category	Groups of variables
Climate	Monthly max temperate, monthly max vapor pressure deficit, monthly average precipitation
Urban-geologic	Latitude, longitude, service area coverage, NLCD developed land cover area (open, low, medium, high, percent imperviousness), percentage of housing units built in different time periods, percentage of single-family houses and multi-density housing complexes, median rooms in a housing structure
Socio-economic	Population, population density, number of employees and number of establishments in different economic sectors, average family income, gross domestic product per capita

A total of 81 climatic, socio-economic, and urban-geologic variables were collected for this study at city-level (Table 2.1) for exploratory purposes. Climate variables, in terms of monthly 30-years precipitation, temperature and vapor pressure deficit normals covering the period 1981–2010, were extracted from the PRISM Climate data set (PRISM Climate Group, 2019). Land use (or, land cover) information was obtained from the USGS National Land Cover Data set (Homer et al., 2015). Demographic and housing information for each city were obtained from the U.S. Census datasets (U.S. Census Bureau, 2019) and American Community Survey. Finally, detailed economic factors such as employment information were compiled from the U.S. Bureau of Economic Analysis.

2.4 Methods

Three characteristics of municipal water use data are calculated: (a) temporal trends in gallons of water use per capita per day (GPCD); (b) the ratio of outdoor to total water use; and (c) the ratio of CII to residential water use. Monthly municipal water use data were annualized, and water use in GPCD was calculated annually for the residential, CII and total water use categories for the period 2005–2017. Then using this annual GPCD time series data set, trends in municipal water use were computed using the unbiased and non-parametric trend analysis method—Mann-Kendall Sen's slope (Sen, 1968; Hipel & McLeod, 1994).

2.4.1 Municipal water use trends – Aggregate residential, CII and total water use

Monthly municipal water use data were annualized, and water use in terms of gallons per capita per day (GPCD) was calculated annually for the residential, CII and total water use categories for the period 2005-2017. Residential GPCD estimate of a city is calculated by dividing the total annual water supplied to the residential sector by the total number of days in that year and the total population served by the city’s water utility that year. Then using this annual GPCD time series dataset, trends in municipal water use were computed using the unbiased and non-parametric trend analysis method - Mann-Kendall Sen's slope (Sen, 1968; Hipel & McLeod, 1994). Sen's slope computes the median linear rate of change of a time series dataset. Finally, the GPCD and median annual GPCD rate of the cities were clustered by U.S. Census regions, Koppen climate regions and population size clusters, to obtain the characteristics of different sub-regions across the CONUS.

2.4.2 Municipal water use trends – Outdoor to total water use

The present study characterizes percent total outdoor water use from total water use in residential and CII sector, and represents the median annual rate of reduction in outdoor water use in GPCD across the study regions.

Water use data surveyed from the study cities were provided as aggregate monthly water use under residential, CII, and master meter categories. Typically, cities in regions with distinct winter seasons do not irrigate the outdoor green spaces when snow or near frost conditions prevail, i.e., mid-November to March. Indoor municipal water use in these regions are often estimated by the average water use over the December to February period (Sharvelle et al., 2017). Other methods include billing-based methods, dual water meter methods, remote sensing methods and flow trace methods (Gleick et al. 2003; Mini et al. 2014; DeOreo et al. 2016). Although these methods can prove very effective for a small-scale study area or for a single city, they become complex and expensive to deploy for large-scale and multi-city analysis.

To support national scale analysis of water use, an approach is needed to estimate the ratio of outdoor to total water use based on readily available water use data. The approach applied for this study to estimate outdoor water use utilized two methods, one for cities with distinct irrigation seasons and one for cities with outdoor water use where seasonal variation is not observed. For cities with distinct irrigation seasons, the quantiles of monthly total water use are computed to represent indoor water use. The validity of this approach was corroborated in each study city by comparing water use quantiles in each city with the reported monthly values over the winter months. In general, the analysis shows that approximately 12 to 30 percent quantiles may be used based on the prevailing climatic conditions to separate indoor water use from total water use, under residential and CII categories, across the study cities.

A different approach was required for cities and towns in the South and West Census region, especially those in the states of Arizona, southern California, Florida, Louisiana and Texas. In these cities, separation of indoor and outdoor water use is confounded by year-round irrigation activities. For example, outdoor water uses in Miami-Dade County, Florida, remains nearly at the same level throughout the year – exhibiting lack of seasonality. In order to maintain data integrity while separating the indoor-outdoor water uses for these cities and towns, indoor GPCD results from Mayer et al. (1999), Rockaway et al. (2011), DeOreo et al. (2016) and Chang et al. (2017) were used to corroborate and adjust when warranted, the estimated outdoor water use as a fraction of the total municipal water use.

The approaches applied for estimation of indoor and outdoor demand are subject to some error that may result from regional seasonal trends in water use, e.g., car washing, running water during cold months to prevent freezing, small garden irrigation etc. However, the goal of this research is to identify trends in water use at the temporal resolution of seasonal and annual. Thus, small changes in daily water use patterns are resolved through the methodology applied that utilizes averaging.

2.4.3 CII to residential water use ratio

For the 103 cities that provided both residential and CII water use data, the CII/Res ratio was estimated for the period 2005 – 2017 on both monthly and annual time steps. From the annual CII/Res ratio timeseries dataset, the relationship between residential and CII water use in a city was explored. The median rate of change in annual CII/Res ratio was estimated from the annual CII/Res ratio timeseries dataset using the Sen's slope method (Hipel & McLeod, 1994; Sen, 1968). Finally, the generalized CII/Res ratio and the rate of change of CII/Res ratio were used to characterize the trends in CII/Res ratio across the U.S. Census regions, Koppen climate regions and population size clusters within the CONUS.

2.4.4 Identification of key factors governing the CII to residential water use ratio

A multivariate statistical analysis was used to develop a classification tree model to explain the CII/Res ratio in relation to the 81 climatic, socio-economic and urban-geologic parameters summarized in Table 2.1. First, a rigorous data multidimensionality reduction was carried out using random forest and principal component analyses to identify the most influential factors that explain the variability of CII/Res ratio in the study cities. Subsequently, these influential factors were used to build a classification tree model to classify the CII/Res ratio for cities across the CONUS.

In the random forest analysis, multiple decision trees (or, a forest, in the order of hundreds of trees) are built by selecting a random set of variables from a multidimensional dataset that is provided as input. In the present study, a random forest analysis consisting of 1000 random decision trees was built. A random forest error plot was used to reduce the size of the random forest from 1000 trees to a smaller number of trees that result in minimum prediction error and maximum variance in explaining the CII/Res ratio. All the variables used in this reduced random forest are ranked in order of their node purity, which is a loss function of the mean square error.

Node purity value of each variable represents its dominance in the random forest. The higher the node purity value of a variable, the more significant that variable is in effectively predicting the CII/Res ratio (James et al., 2013). A threshold value, based on the node purity of the ranked variables, was then arbitrarily selected to obtain a smaller subset of variables. This threshold value is randomly selected to be a low node purity value with an aim to select maximum number of variables that have high variability and relevancy in building a classification tree model for the CII/Res ratio. If a higher node purity value is used, the smaller subset of variables selected might not capture the total variability in explaining the CII/Res ratio.

Although random forest analysis helps to minimize the total number of predictors to build a classification tree model, it does not effectively prevent the selection of variables with multi-collinearity. The reason being that the principle behind random forest analysis is randomness, and with a high-dimensional dataset, the problem of inter-dependency or multi-collinearity of predictor variables is extreme (James et al., 2013). Collinearity or multi-collinearity causes redundancy, introduces complexity and impacts the efficacy of statistical analysis (Farrar & Glauber, 1967; Lafi & Kaneene, 1992). To effectively build a classification tree model of the CII/Res ratio for CONUS using unique variables that have high variance and non-collinearity, Principal Component Analysis (PCA) was applied to obtain influential variables that are not correlated (Jolliffe, 1986; Lafi & Kaneene, 1992).

The idea behind principal component analysis (PCA) is to reduce the dimensionality of a large dataset with many interrelated variables, while retaining as much of the variation present in the dataset as possible. This is achieved by transforming the original dataset of variables into a correlation and covariance matrix consisting of principal components (PCs). Each PC is uncorrelated from the others, and consists of a set of variables that retain most of the variation present in the original variables (Hotelling, 1933; Jolliffe, 1986; Lafi & Kaneene, 1992). The PCs obtained are ranked by the total variance (or, eigen value) explained by each component. With the help of a scree plot and elbow method, a PCA biplot was then built by selecting the first few PCs that can explain at least 80% of total variance in the observed variable – CII/Res ratio. A scree plot is a line plot of the eigen values of the PCs in the analysis, and it is used to retain just a handful of PCs at the point where the eigen value of PCs reach an inverted plateau resembling an elbow (Cattell, 1966).

Based on the number of PCs selected from the PCA analysis using a scree-plot, at least two variables with high correlation are selected from each PC, or from a group of variables that are clustered around a unique axis (Husson et al., 2017; Jolliffe, 1986). A classification tree model for CII/Res ratio was developed by including only one variable from each PC and by splitting the dataset into training and testing subsets. The classification tree model's adequacy was evaluated by the coefficient of determination (R^2), a non-parametric Mann-Whitney U test of the medians with a significance level of 95% and a goodness of fit tests by i) visual comparison of cumulative frequency distributions, ii) Chi-Square test and iii) Kolmogorov-Smirnov test to present a robust and parsimonious classification tree model to explain the variation in CII/Res ratio across the CONUS.

A flowchart of all methodology described above has been illustrated in Figure 2.2.

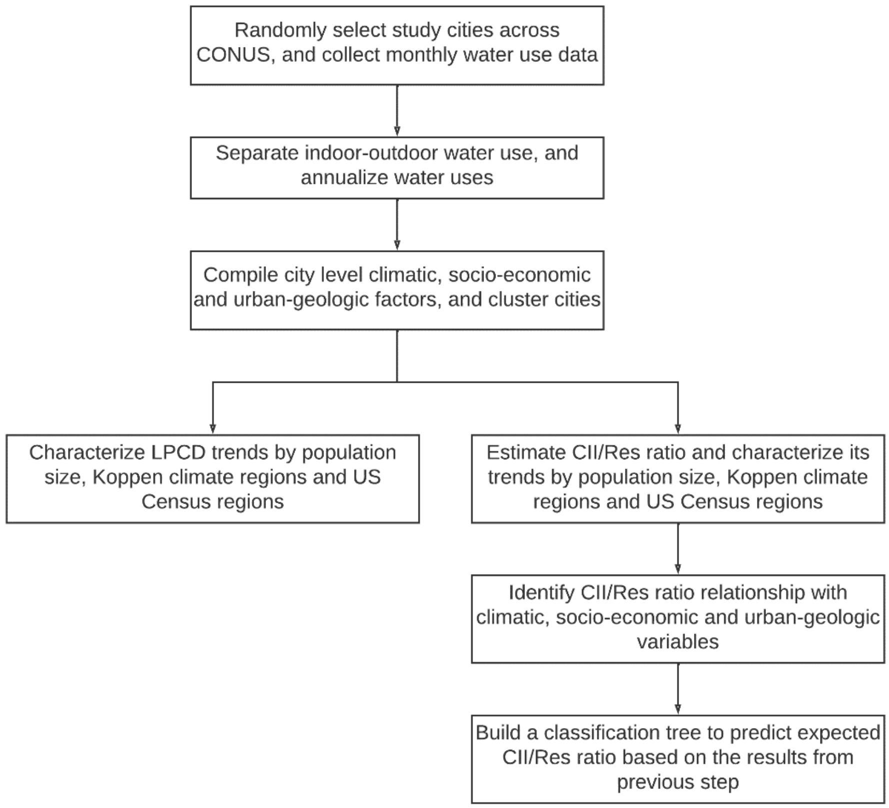


Figure 2.2. Flowchart of research analysis

2.5 Results and Discussions

The water use information compiled in this study provides one of the most comprehensive municipal water use datasets in the CONUS to date and is made available as an open access database to foster further studies on municipal water uses. The city-level monthly water use database obtained and processed in this study can improve the use of other datasets such as the U.S. Geological Survey Water Use Estimates for the Nation (Dieter et al., 2018), which is provided at county-level on a 5-year time interval and lacks the capability to explore temporal trends in municipal water supply systems as monthly or daily water use data is unavailable.

The study shows that total municipal water use in cities and towns across the CONUS has declined at a rate of approximately 2.6 gallons per capita-day on an annual basis over the last two decades, with prominent water use reductions achieved in the residential use and is similar to the findings of DeOreo et al. (2016) and Dieter & Maupin (2017). Per capita water use reduction is highest in the large population size cities, Arid climate region and in the West Census region of the CONUS. Notable differences in total water use across the study regions is attributed to outdoor water use. April precipitation, annual vapor pressure deficit, employees in manufacturing sector, number of single-family houses, and number of buildings built before 1950 best explain the variation in CII to Residential water use ratio. The following sub sections will elaborate on these results.

2.5.1 Trends in municipal water uses by population, climate and Census regions

There is an overall significant downward trend in the total water use amongst the study cities between 2005 and 2017 with a Mann-Kendall's tau of -0.44 and p-value < 0.05. Figure 2.3 provides the average value of GPCD for different categories of municipal water use within the 126 study cities across the CONUS. Residential water use dominates municipal water uses in the CONUS. The average annual residential, CII, and total water uses are 79, 58, and 143 GPCD, respectively. The median annual rate of water use reduction, measured in gallons per capita per day per year, is at 1.5 amongst residential water users, 1.2 amongst CII water users, and overall, at 2.6 for total municipal water use (Figure 2.3-Left panel). Total water use includes wholesale water trades to other utilities or large industries.

Strong seasonality is evident in municipal water uses across the CONUS (Figure 2.3-Right panel). Residential water use is notably higher between April and November, plausibly due to increased use for outdoor irrigation of lawns and green spaces. Note that, seasonality is also seen in the CII sector, since CII outdoor irrigation is cumulatively accounted with CII indoor use. Water use in December, January and February are higher than water use in March as some cities experiencing pipe freeze conditions in winter keep the faucets running and thereby, have an increased indoor water use compared to other months. Similarly, the CII/Res water use ratio shows seasonality averaging at 0.69 between April and October, while for rest of the year this ratio averages at 0.75. The "Other" water use category shown in Figure 2.3 is the master meter use which includes wholesale water transfers from a water utility to private companies, neighboring towns and/or other water utilities, and it does not show notable seasonality.

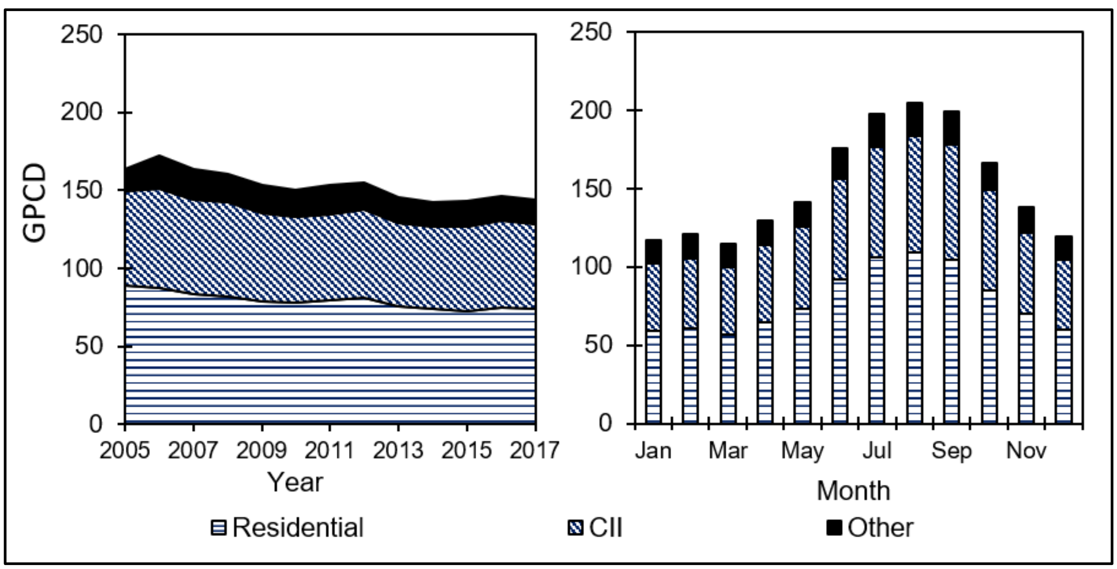


Figure 2.3. Average municipal water uses in the study cities over the 2005-2017 period: (Left panel) average annual water uses time series; (Right panel) average monthly water use.

Significant differences are depicted in municipal water use patterns among the cities differentiated by population and across the U.S. Census and level-1 Koppen climate regions (Figure 2.4). As the city population increases, the average per capita-day water use decreases (Figure 2.4-a). The lower total water use in cities with higher population is primarily associated with lower residential water use. However, large cities on average tend to have higher CII water use than small and medium size cities. Moreover, larger cities have achieved greater reduction in total water use over the analysis period (Fig 1.4-b). The median annual rate of reduction in total water use increases with increasing city population. Thus, large cities with more than a million people tend to use less water compared to small size cities, while continuing to reduce their average per capita-day water use annually. These patterns can be attributed to the fact that large cities encompass higher population and housing density compared to small and medium sized cities. These results, in congruence with previous studies, underline that multi-density housing can effectively improve the water supply efficiency and reduce overall water use in cities (Fischer, 1972; Gordon & Richardson, 1997).

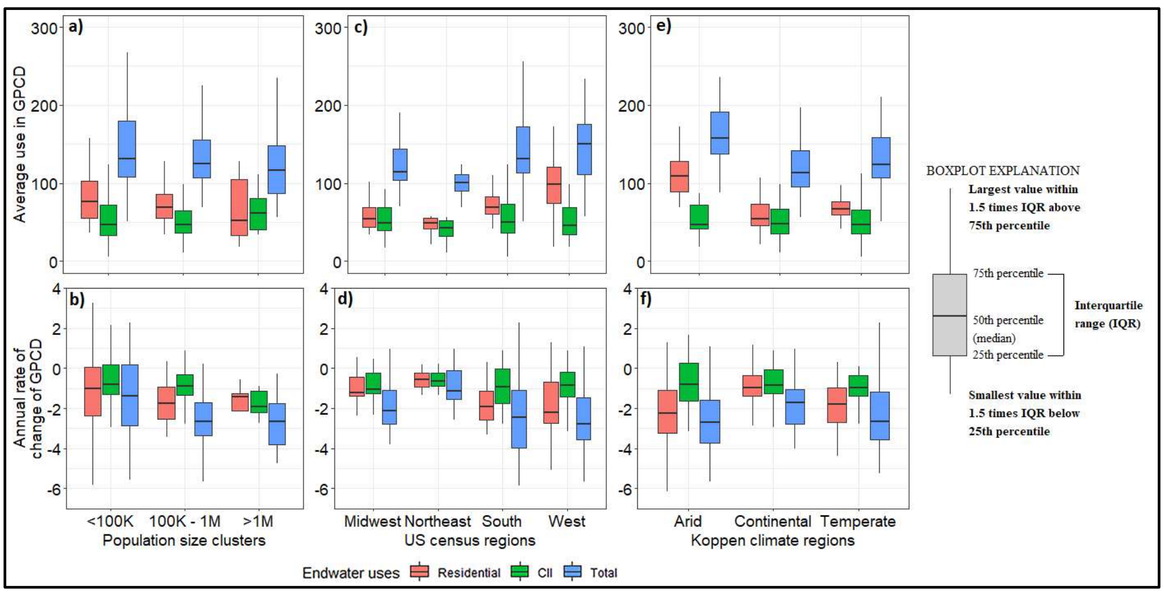


Figure 2.4. Characteristics of municipal water uses highlighting the per capita-day use and median water use reduction rate by study regions.

Contrasts in water use patterns are also evident among cities in different census regions (Figure 2.4-c, d). The Northeast region has the lowest average residential and CII per capita-day water use, while the West region has the highest average residential per capita-day water use among all the other census regions, presumably due to outdoor irrigation. However, the median annual rate of total water use reduction (Figure 2.4-d) in the Northeast census region is the lowest owing to the aging infrastructure (about 51% housing units were built before 1950) and old plumbing codes which do not incorporate water efficiency programs. The climate in the West census region is dominated by Arid and Temperate conditions (Figure 2.1), which show on average a high per capita-day total water use. The South census region, dominated by the Temperate climate type, is second to the West census region in terms of high per capita-day total water use. Also, the South and West census regions, which have higher average per capita-day total water use, also undergo highest levels of median rate of reduction in per capita-day total water use. These patterns may be explained by percentage of housing units that are built after 1950 in the regions with improved plumbing codes to promote water efficiency and water use reduction despite having the highest percentage of single-family houses.

When clustering cities by climate region, cities in the Arid climate region in CONUS have the highest average residential water use, followed by the cities in the Temperate and the Continental climate regions (Figure 2.4-e). Highest median rate of residential water use reduction has been achieved in the Arid climate region (Figure 2.4-f). Arid and Temperate climate regions have the highest total municipal water use reduction rates - above 2.5 GPCD per year.

In Figure 2.5 (Left panel), city-level GPCD estimates are compared with the county-level GPCD values which were estimated by United States Geological Survey (USGS) for the 2015 municipal water supply year. The USGS estimated GPCD data for CONUS counties is representative of "Public Supply". Overall, the city-level water use volume revealed in the present study demonstrates similar regional patterns when compared with the county-level public supply data from USGS. However, in an analysis undertaken to identify the usability of USGS data for this study (Figure SI-2.2), the USGS estimates of water withdrawn at county level are, in most cases, higher in volume than the actual amount of water supplied at city or municipal level and are also poorly correlated - correlation coefficient was 0.2 with p-value < 0.05. Hence the USGS water withdrawal estimated does not directly help in capturing the actual water use volume or trends in a city within a particular county.

The reason for this order of difference in volume between municipal level water supply data and the USGS water withdrawal estimated for Public Supply in 2015 at county-level are assumed to be, i) use of county level data where water use within municipal water supply boundaries may differ from water characterized as municipal use by USGS, ii) that there are many cities/ towns within that county boundaries with different socio-economic traits, and/ or iii) cities get their water from an outside, neighboring county through water trades. Since USGS public supply water withdrawal estimates at county-level are tallied on a bi-decadal period with estimates projected/compiled for a single year, it does not help in addressing the objectives of this study to characterize temporal trends in municipal water use other than the monthly water use data obtained from directly the cities.

Figure 2.5 (Right panel) shows the range of annual rate of change of GPCD estimates across the study regions. In some cases, water use patterns and reduction trends deviate from other cities within the same population, climate and Census regions. For example, it may be observed that cities with strong recreation and tourism economies, including Miami and Las Vegas, tend to have the highest residential, CII, and total water uses. Overall, this figure captures the general downward trend in total municipal water use across the cities in all the US Census and climate regions as indicated by the temporal trends in Figure 2.3.

Water use reduction could be attributed to factors like implementation of updated plumbing codes, water saving fixtures (e.g., low flow toilets and faucet fixtures), drought policies, water conservation in outdoor irrigation through voluntary or mandatory restrictions, and water conservation using pricing schemes, etc., (Breyer et al., 2018; DeOreo et al., 2001; DeOreo et al., 2016; Fortier & Mailhot, 2015; Magionni, 2014; Mayer, 2016; Niemczynowicz, 1999; Olmstead & Stavins, 2009). Aging infrastructure, rising cost of municipal water, and affordability are among the social and economic factors that affect municipal water use patterns and trends (Butts & Gasteyer, 2011; Etale et al., 2018; Mack & Wrase, 2017; Raj, 2005).

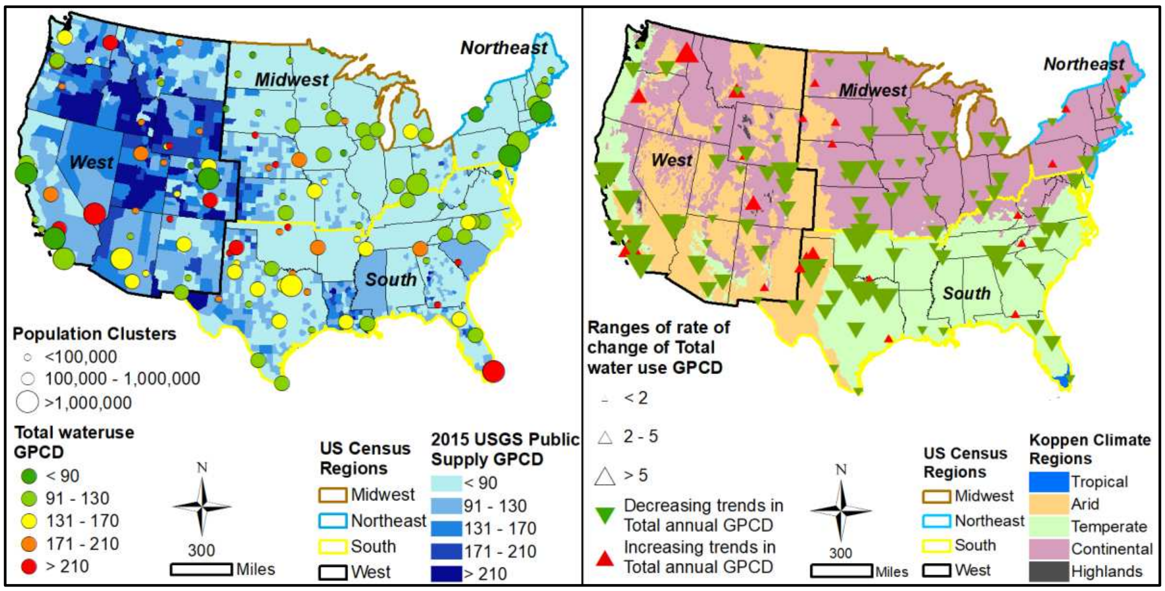


Figure 2.5. (Left panel) Map of CONUS showing the comparison between city-level total municipal water supplied in GPCD and county-level total public water supplied in GPCD. (Right panel) Map of CONUS showing the range of increasing and decreasing trends in annual rate of Total water use GPCD.

2.5.2 Outdoor water use patterns and trends

Outdoor water use makes up a significant portion of the total water use ranging between 3 and 64 percent (Figure 2.6). From Figure 2.6, it can be seen that generally cities in western part of CONUS have high outdoor water use. Cities in the Arid Koppen climate and West Census regions exhibit a higher ratio of outdoor to total water use compared to other cities across the CONUS. Small cities with a population of less than one hundred thousand residents have notably higher outdoor to total water use ratio (Figure 2.6-b). The results underline that outdoor water use is a predominant driver for varying demands in municipal water supply systems, since DeOreo et al. (2016) have established that indoor water use remains to be, on average, plateaued across the country. Similar to the total water use trends discussed earlier, the median rate of reduction in outdoor water use is higher in cities within the Arid and South climate regions, and it increases with increasing city population (Figure 2.6-c). Thus, the hypothesis assumed about outdoor water use contributing to the higher volume of total water use in cities in Arid climate and West Census regions and cities with less than 100,000 residents is verified by the trends discussed in this section. A MANOVA test of means of outdoor water demand amongst different classification within a cluster proved that there is significant difference in the means ($p < 0.05$).

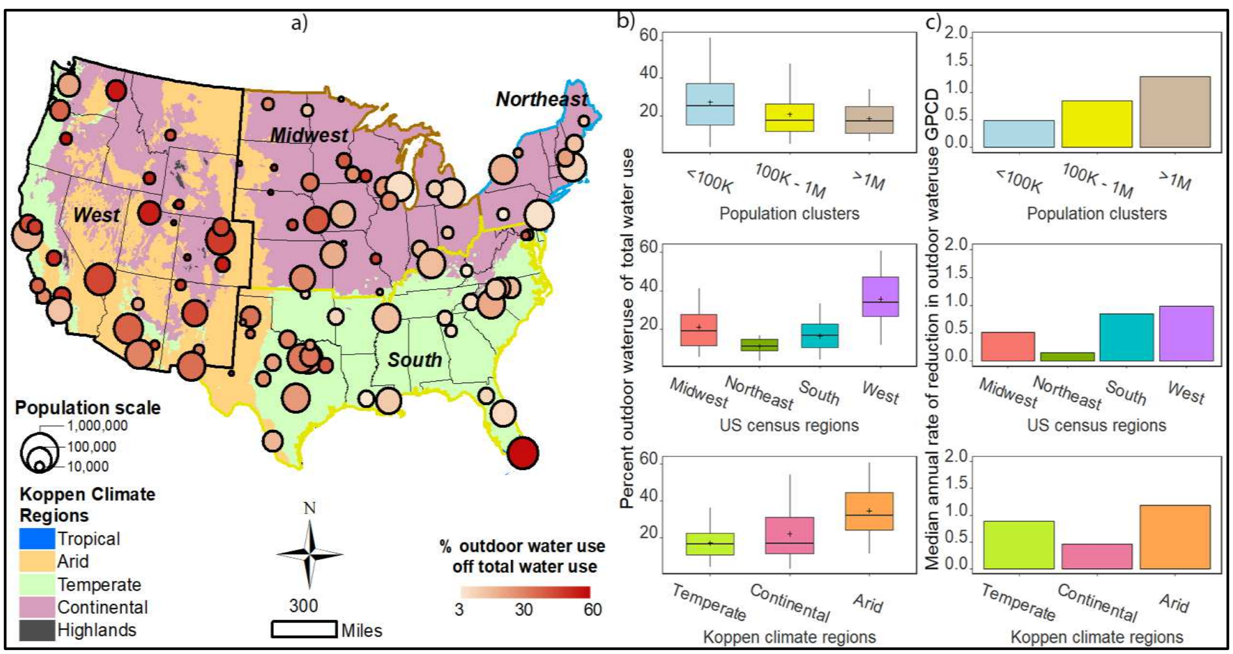


Figure 2.6. a) Map of CONUS highlighting percent outdoor water use to total water use in cities distinguished by their population size, b) Boxplot showing the characterization of percent outdoor water use to total water use by study regions, and c) level of median annual rate of reduction in outdoor water use GPCD across study regions.

2.5.3 The CII/Res municipal water use ratio by population, climate and Census regions

Using the annual CII/Res ratio estimates of the study cities, an average annual CII/Res ratio for the entire CONUS was estimated to be 0.72. This average CII/Res ratio for the CONUS does not show seasonality or variability within the study period. To reiterate, the lack of seasonality and lack of temporal change in the CII/Res ratio can be attributed to the fact that residential and CII water use exhibit similar seasonality (Figure 2.3).

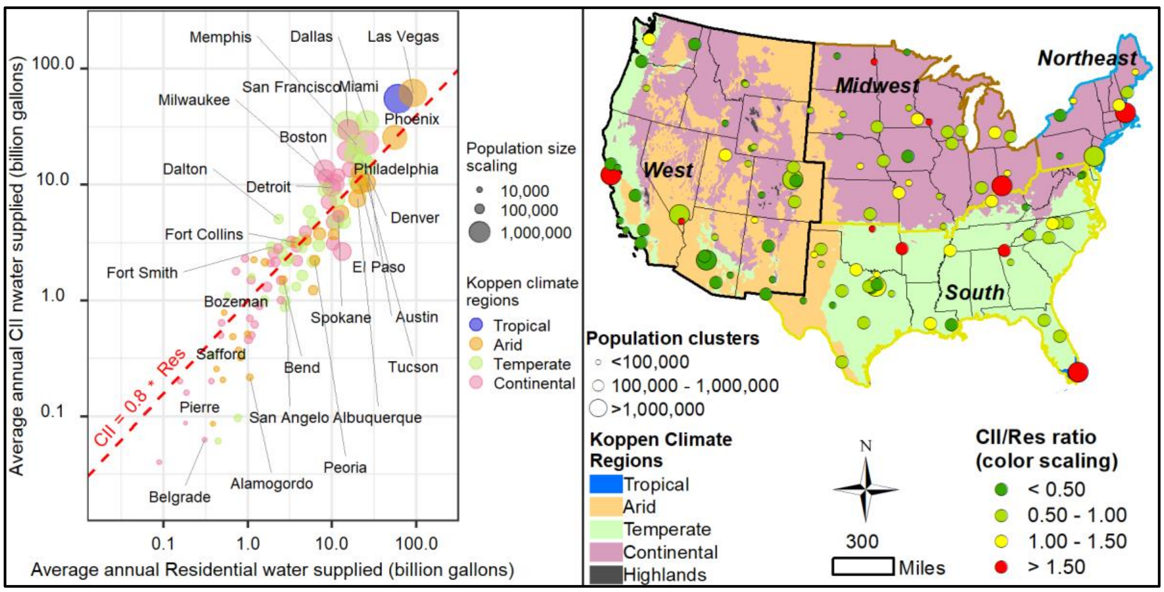


Figure 2.7. (Left panel) Bubble plot of study cities showing the variation in water use by population size and climate region type. (Right panel) Map of CONUS showing the variation in CII/Res ratio across the study cities clustered by U.S. Census regions and Koppen climate regions.

The CII/Res ratio tends to be highly variable even within the same Koppen climate and U.S. Census regions (Figure 2.7-Right panel). However, larger cities generally have a higher CII/Res ratio, above 0.8, compared to cities with lower population (Figure 2.7-Left panel). Large cities with population above 1 million people also experience the highest variation in CII/Res ratio, since they are undergoing different trajectories of population growth and economic development. While residential growth is the primary driver of changes in municipal water uses in some cities, industrial activities have increased CII water use in some other cities. These factors along with other consideration such as tourism result in higher CII/Res water use ratio in Las Vegas, Miami, Phoenix, Philadelphia, and San Francisco. An economic census data for number of establishments in Accommodation and Food Services with NAICS code 72 was used as proxy for level of tourism in a city – data is not shown in manuscript or SI.

Among U.S. Census regions, the Northeast shows the highest median level of CII/Res ratio (Figure SI-2.3). One inference for this trend is assumed to be the influence of outdoor water demand. Northeast region has the lowest outdoor water demand leading to a higher CII/Res ratio estimate (Figure 2.6-b). The West census region has the lowest average CII/Res ratio in contrast to the highest total GPCD water use. The West is also the census region with the least variation in the CII/Res ratio. This low CII/Res ratio in the West implies that residential water use dominates the municipal water use in the West as outdoor water use is highest in this region (Figure 2.6-b). This characteristic is similar in the South, which is the second highest residential water user in the CONUS, with higher average CII water use by some cities leading to high variation in its CII/Res ratio.

Examining the CII/Res ratio trends in level-1 Koppen climate regions (Figure SI-2.3), CII water use, on average, dominates residential water use (i.e., CII/Res ratio >> 1) in the Temperate and Continental climate types; residential water use is highest in the Arid climate region. The Continental climate region has highest variation in CII/Res ratio, while the Arid climate region has lowest range of CII/Res ratio which can be attributed to the highest amount of water needed for residential outdoor irrigation compared to other climate regions (Figure 2.6-b). Also, the median rate of change of CII/Res ratio for Continental climate region tends to indicate economic growth resulting in increased CII water use compared to other climate regions for the period 2005-2017.

2.5.4 Influential factors and a classification tree model for CII/Res water use ratio

Top 25 variables, illustrated in Figure 2.8 (Left Panel), were selected from the random forest analysis built using 200 trees from the initial 1000 trees (Figure SI-2.4-Left panel) with a node purity value of 0.35 as threshold. With the help of a scree plot and elbow method (Figure SI-2.4-Right panel), top five PCs were identified as optimal to build a classification tree for the CII/Res ratio, which include: "April precipitation", "Annual vapor pressure deficit", "Number of employees in manufacturing sector", "Total houses built before 1950", and "Total single-family houses" from all the other variables denoted originally in the PCA biplot (Figure SI-2.5). It should be noted that the "Number of employees in other services (except public administration)", i.e., workers in mechanic repairing, photofinishing, NGOs, personal/pet care, laundry and parking services, may be a substitute with the same fidelity for the "Number of employees in manufacturing sector". Figure 2.8 (Right Panel) represents cities clustered by population sizes and by the factors that were selected to build a classification tree.

The impact of climate variables – April precipitation and annual vapor pressure deficit, on municipal water uses can be explained by their variation among the U.S. Census and Koppen climate regions. These two climate variables primarily influence the total volume of municipal water used for outdoor irrigation based on the applied approach to partition indoor and outdoor water use in this study, and are prevalent in warm climate regions and regions with high number of single-family housing. The average total number of employees in the manufacturing sector is highest in the Northeast census region and large size cities, thereby highlighting the prevalence of CII water use. The total percentage of housing units built before 1950 is also highest in the Northeast census region and in large cities, while cities in the South and West census regions and smaller population size cities have relatively high percentage of houses that were built post 1950 with improved plumbing codes. The total percentage of single-family housing also plays a crucial role in deciding the level of residential water use in a city.

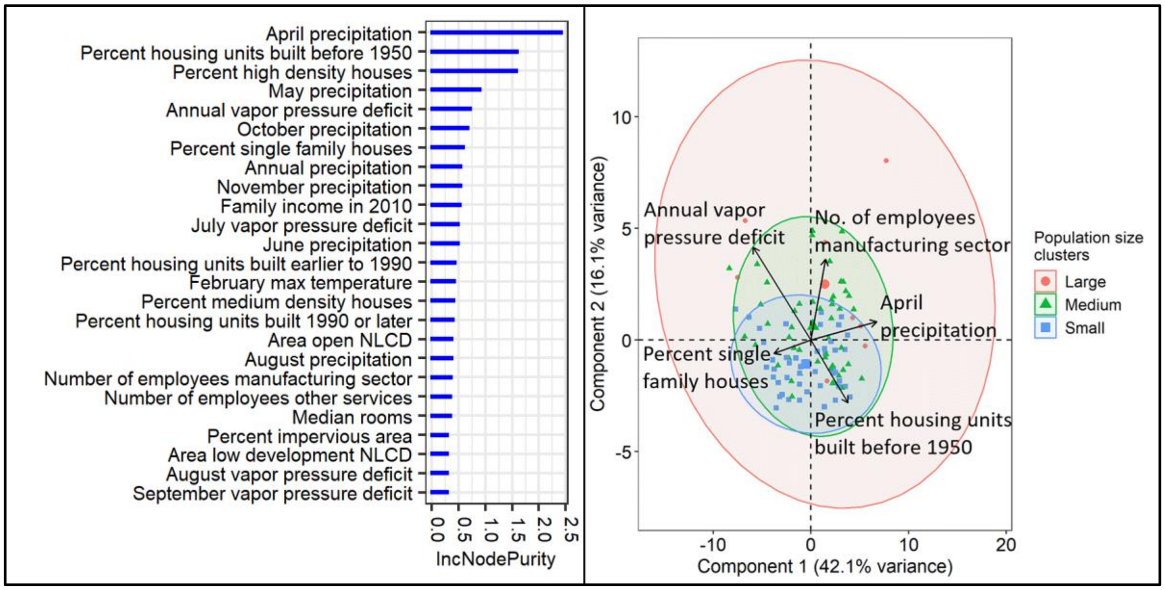


Figure 2.8. A principal component analysis biplot (right panel) of important variables from random forest analysis (left panel). The PCA biplot shows the cities clustered by their population size and their relation to the top five variables that best explain their CII/Res ratio characteristics.

Thus, these five physical and readily available variables describe a variety of municipal water use characteristics across the CONUS when summarized by the study regions. Basker et al. (2019) and Becker (2016) also highlight the importance of the bi-decadal water estimates in the manufacturing and mining sector, while Mayer et al. (1999) and DeOreo et al. (2016) have shown the importance of building plumbing codes, which are tied with the age of the building, in estimating the water efficiency of a house or an apartment complex.

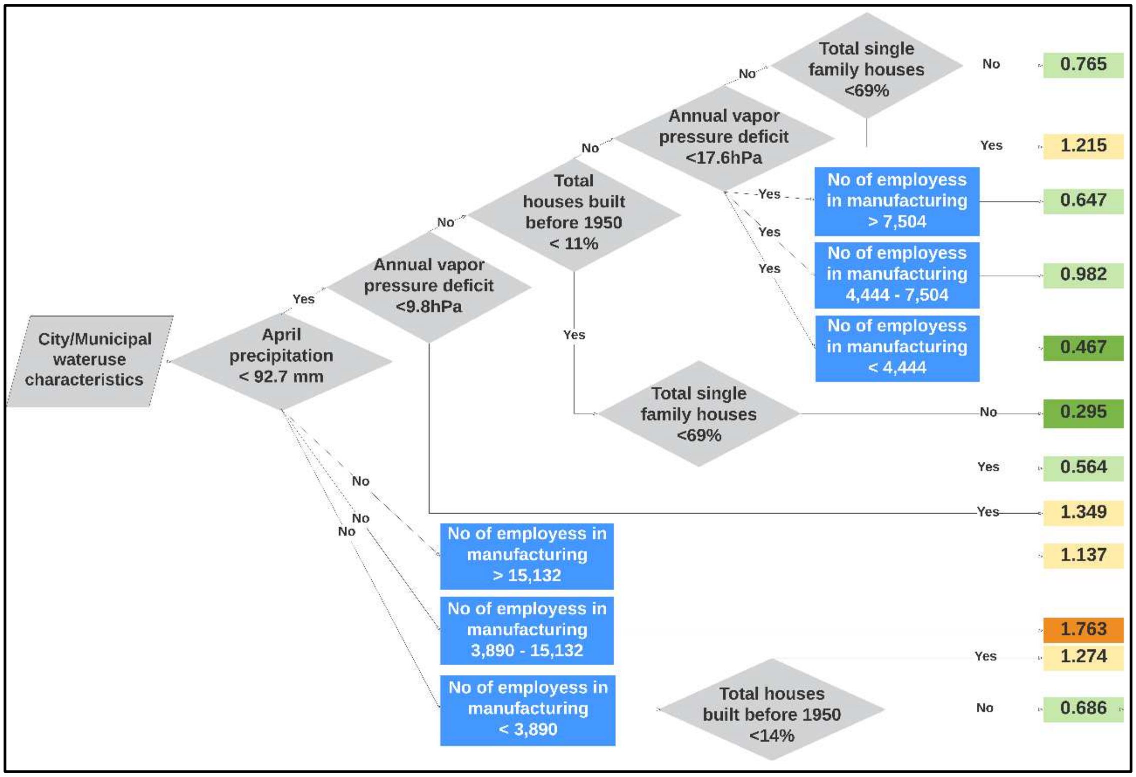


Figure 2.9. A classification tree model built to estimate the CII/Res ratio, based on the top 5 factors from the PCA analysis, for any city or town within the CONUS.

Ultimately, a classification tree model was developed using these influential factors as depicted in Figure 2.9. The classification tree's coefficient of determination (R^2) for the training and testing datasets were 0.68 and 0.76 respectively. The Mann-Whitney test on the medians of observed and decision tree classified CII/Res ratio values yielded a p value > 0.05 to validate that these two CII/Res ratio sample sets were not significantly different. Comparison of the empirical and modeled cumulative distribution functions of the CII/Res ratio indicates an adequate fit between the observed data and the fitted classification tree model – i) decent visual goodness of fit (Figure SI-2.6), ii) passed the Chi-Squared test ($p > 0.05$), but iii) failed the Kolmogorov-Smirnov goodness of fit test ($p < 0.05$). Thus, it has been observed that for a few cities, the classification tree model over-estimates the CII/Res ratio for empirical values less than 0.6 and under-estimates the ratio for empirical values above 1.7 (Figure SI-2.6).

Overall, the tree classification model provides a new approach to using readily available datasets along with the municipal water use data to be incorporated in existing urban water modeling tools, such as the that described by Sharvelle et al. (2017). Improved estimates of CII/Res ratio expands existing predictive capabilities of water demand-supply chain under varying future development scenarios, and foster capacity to assess opportunities to minimize use of traditional water supplies and maximize fresh water availability. The developed CII/Res ratios cannot be applied to identify distinct trends in each water use category, but rather to identify the expected CII/Res ratio under distinct conditions of interest (e.g., climate, socio-economic, and urban development variables).

2.6 Conclusions

This study investigates factors that influence residential and CII water use in the municipal water supply system across the CONUS. Thus, data from a comprehensive survey of monthly municipal water uses from 126 cities and towns, across ecohydrologic regions between 2005 and 2017, were compiled to explain the spatial and temporal trends in municipal water uses based on different climatic, urban-geologic, and socio-economic factors.

Notable water use reduction in the residential sector compared to the CII sector was evident over the 2005-2017 period. Cities with larger populations have lower total water use per capita-day than small and medium size cities, while those in the Arid Koppen climate region have higher per-capita water use. The CII water use per capita-day also tends to increase with population. Eventually, the trends evident in outdoor water use for landscape or lawn irrigation appear to be an important source of variability in total water demand across the population size clusters, U.S. Census regions and Koppen climate regions across CONUS.

The results also suggest that larger size cities and cities in hot, arid climate regions simultaneously achieve pronounced reductions in absolute per capita-day water use relative to smaller cities with low volume water use and cities outside arid climate regions. Although marked differences in the median rate of municipal water use reduction were characterized across different regions, strategies used by water supply providers to achieve these reductions were not investigated. Future studies focused on improving the current municipal water use data with more cites (> 126) and more recent water use data could help to unravel more trends in water use characteristics and offer better insights in modelling municipal water use.

Another major outcome of this study is the collection of 81 factors to explain the influence of climatic, socio-economic and urban-geographic characteristics of a city on the CII to residential water use ratio. This is a novel approach to unify the two major municipal water users and the characterization of CII/Res ratio has the potential for various city-level planning and governance to build a prediction model for projecting future water needs of a city under varying scenarios of development patterns. Ultimately, the amount of precipitation in April, annual vapor pressure deficit, number of employees in the manufacturing (or other services except public administration), total number of single-family houses and housing units built prior to 1950, adequately explain the variability of the CII/Res ratio. These factors serve as surrogates for climate (e.g., precipitation and temperature), infrastructural conditions such as aging water systems, economic development, and other drivers of municipal water uses in cities.

The methods provided in the present study may be used to predict the CII/Res water ratio and to separate outdoor to indoor water use ratio for other cities and towns across the CONUS. The CII/Res ratios are appropriate for estimates of water use with a monthly or annual temporal resolution. Future work related to water supply and demand planning that use projected population, climate and land use changes will benefit from the methods and results of this study. Also, municipal water use trends presented here can readily be used for model parameter estimations to enable estimation of CII water use based on residential water use.

Lastly, a noteworthy facet of this work is its potential to apply the approaches to assess impacts of economic development on municipal water use. The methodology used to build a classification tree model for the CII/Res ratio in this study could easily be translated to temporal anomalies observed in 2020, civil restrictions due to corona virus lockdowns, in order to study the forces morphing municipal water uses due to significant impacts on social and economic activities.

2.7 Data Availability

Chinnasamy, C. V., M. Arabi, S. Sharvelle, T. Warziniack, C. D. Furth, A. Dozier (2021). Data for Characterization of Municipal Water Uses in the Contiguous United States, HydroShare, https://doi.org/10.4211/hs.feb5af8990914ce2b28f18b10d65c2a2

CHAPTER 3 – CHARACTERIZATION OF SINGLE-FAMILY RESIDENTIAL WATER USES EMPLOYING HIGH-RESOLUTION, SMART METERING DATA

3.1 Introduction

This study focuses on characterizing single-family household water use patterns and drivers utilizing high-resolution, smart-metered water use data collected from more than 700 single-family households across the state of Arizona in the United States for the water year 2022 (WY-2022; October 01, 2021 to September 30, 2022). In many parts of the western U.S., especially the southwestern U.S., water is becoming a scarce resource and yet, these regions are witnessing high urban population growth (Gleick, 2010). This has huge implications for the economic productivity and the livelihood of millions of people in this region (Kyl Center, 2023; Wang & Vivoni, 2022). Research and technological advancements for monitoring water use patterns with high-resolution (in minutes or seconds), at individual building or land parcel levels, provide greater scope to uncover critical nuances in water usage drivers and to influence decision-making for a sustainable future. (Cominola et al., 2023; Gleick, 2003, 2010; Marston et al., 2018; Mekonnen & Hoekstra, 2016; Sanchez et al., 2018).

Arizona, a land of stunning desert landscapes and vibrant cities, grapples with pressing water resource challenges. In the early twentieth century, the development of large-scale irrigated agriculture attracted a surge of population growth which eventually led to large-scale urban development (Baker et al., 2004; Gober & Kirkwood, 2010). Over time, this growth trend necessitated the construction of dams, reservoirs, and the transfer of water over long distances from other watersheds (Eden et al., 2011; Wang & Vivoni, 2022). The Central Arizona Project (CAP) was designed to transport the Colorado River water over 336 miles to southern Arizona

(Figure 3.1) to protect groundwater resources in Arizona from being over-exploited (Witcher, 2022). The CAP project, completed in 1993, supports nearly a million acres of irrigated agricultural land and plays a pivotal role in the state's water supply and economy. Phoenix and Tucson, home to a combined 2.2 million people (as per 2020 Census data for those cities), rely on CAP for their municipal water supply along with several tribal nations.

Figure 3.1. Map of the Colorado River Basin with Lake Mead and Lake Powell, and the major U.S. cities receiving the Colorado River water in the lower basin. The Central Arizona Project canal is represented using a red line.

Currently, Arizona receives approximately 36 percent of its water supply from the Colorado River, with the remainder sourced from groundwater aquifers, in-state rivers, and a small fraction from reclaimed water (ADWR, 2020). Arizona is one of the seven states in the U.S. that rely on the Colorado River Basin along with Mexico and many Native American tribes (Figure 3.1). However, prolonged droughts, rapid urban developments and agricultural land expansion, and inter-state disputes over water allocation have put huge stress on the Colorado River, resulting in mandated cuts to CAP water allotments (Ault et al., 2016; Baker et al., 2004;

Bureau of Reclamation, 2023; Folger, 2017; Mekonnen & Hoekstra, 2016; Rajagopalan et al., 2009; Udall & Overpeck, 2017; Warziniack & Brown, 2019).

In the Southwest U.S., such cuts to water allocation are prophesied to reoccur in the future due to climate change impacts on water yield (Brown et al., 2013; Cosgrove & Loucks, 2015; Donkor et al., 2014; Gober et al., 2011; Salehabadi et al., 2020; Sanchez et al., 2018; Udall & Overpeck, 2017; Warziniack & Brown, 2019). The relationship between the development of Arizona's water resources and its urban growth is likely to become more controversial in the future as urban residents face increasing competition over scarce resources (Baker et al., 2004; Bolin et al., 2013; Heidari et al., 2021; Kupel, 2003; Larson et al., 2013; Wang & Vivoni, 2022). The reliance on the Colorado River water raises concerns about the long-term feasibility of maintaining such growth patterns in Arizona without jeopardizing the environment and the livelihood of future generations (Sprigg & Hinkley, 2000; Yigzaw & Hossain, 2016).

In 2019, approximately 22 percent of the total water consumption within the state of Arizona was by the municipal sector (ADWR, 2020). In particular, the issue of residential water use, especially in single-family households, is of utmost importance. It has been established that single-family households are the largest end-users and drivers of demand in the municipal water system of the U.S. (Chinnasamy et al., 2021; DeOreo et al., 2016). Phoenix metropolitan area in Arizona particularly has been reported to be single-family household dominant in its urban growth patterns with high water demands, compared to other urban areas within the US (Balling et al., 2008; Mayer et al., 1999; Ouyang et al., 2014). Private swimming pools, high water use for landscape irrigation, and sprawling development patterns with a high percentage of single-family households are the biggest challenges that Arizona faces regarding the future of water sustainability (Gleick, 2010; Gober & Kirkwood, 2010; MacDonald, 2010).

In January of 2023, a developing community's water supply was cut off from the municipal system in the city of Scottsdale, which gets most of its water from the Colorado River (Partlow, 2023). If such incidents are to be prevented from reoccurring in the future, it is imperative to recognize the importance of ongoing investigation into end water use and the forces that shape it. By continuously studying changes in residential water use, we can effectively plan for future water demand and implement policy-level changes to promote conservation and more efficient utilization of this precious resource. Therefore, urban water demand research and studies are vital to stay abreast of evolving trends, anticipate shifts in water consumption patterns, and facilitate informed decision-making that aligns with long-term sustainability goals.

With the development of advanced water metering technologies, it is now possible to take a closer look at water uses at a granular level (at the building level and appliance level) to identify leaks and different climatic, human behavior, and fixture performance components that cause variation in water uses, and use the feedback mechanism to curb excess water consumption (Schultz et al., 2018; Sønderlund et al., 2016; Yang et al., 2017). High-resolution, high-frequency monitoring, smart meters, technological advancements in Internet-of-Things (IoT), and machine learning and big data analytics have enabled improved water demand management in urban settings (Pesantez et al., 2020; Romano & Kapelan, 2014; Walker et al., 2015). Advances in flow monitoring and smart metering can empower homeowners with real-time data for better consumption management and hold immense potential for utilities. Data analytics derived from household-level monitoring enable informed decisions on water resources, infrastructure, and conservation planning. Furthermore, this technology has a tangible impact on water conservation efforts, as leaks can be swiftly identified, usage patterns comprehensively analyzed, and proactive steps taken toward a more water-efficient future.

The study provides a robust assessment of the effects of occupancy level, efficient appliances, swimming pools, and other household characteristics on water consumption patterns by controlling the effects of regional climatic divers. By utilizing advanced smart water metering data, this research advances statistical methods for improved assessment of residential end uses of water in single-family households (Mayer et al., 1999; Balling et al., 2008; Ouyang et al., 2014; DeOreo et al., 2016; and Chinnasamy et al., 2021). The large sample of metering locations in the study area comprehensively represents single-family households in Arizona. The assessments can guide development and adoption of data-driven strategies for sound water demand management in the state and across the Colorado River Basin. Similarly, the methodology developed in this study can be used for water use assessments using Flume or other advanced metering systems across the United States and worldwide.

The terms "water consumption" and "water use" are used interchangeably in this article, and both terms refer to the amount of water taken up by a household or person for various water needs. A common assumption in municipal water systems is that approximately 20 percent of the total water supplied to households is permanently lost through leaks in pipes, evapotranspiration by landscape vegetation, evaporation from swimming pools, washing cars and driveways, and irrigation system leaks.

3.2 Objectives

This study uses high-resolution, smart-metered water use data obtained at the individual consumer level to characterize the patterns and drivers of single-family household water uses.

The analysis utilizes smart-metered data to better characterize water use patterns. Specifically, the objectives of this study are to:

1) Explore the relationship between water use intensity metrics and household occupancy.
2) Assess the effects of household appliances' efficiency levels and swimming pools on indoor, outdoor, and total water consumption.
3) Develop robust statistical models to relate single-family household water consumption to important household and regional parameters to identify key drivers of water consumption.

3.3 Data and Study Region

The availability of high-resolution, smart-metered data from Flume, Inc., a smart water consumption sensor company based in San Luis Obispo, California, U.S., provides a unique opportunity to examine the study objectives. As of June 2023, Flume's smart water meter sensors are deployed in over 68,000 single-family households across the United States. The Flume system provides real-time leak alert notifications to the client and enables a better understanding of indoor and outdoor water use patterns at the household and appliance levels.

The single-family household water consumption dataset provided by Flume for this study covered the water year 2022 from October 01, 2021, to September 30, 2022 (referred to as WY-2022). The primary water consumption dataset includes 5-minute interval water metering data, categorized as indoor and outdoor water use, from over 900 households with active Flume sensors in Arizona. The water metering time series are subsequently aggregated for assessments at daily, weekly, monthly, and seasonal time steps, under the indoor- and outdoor- household water use categories. Two other datasets provided by Flume represent building/household characteristics and appliance flow characteristics.

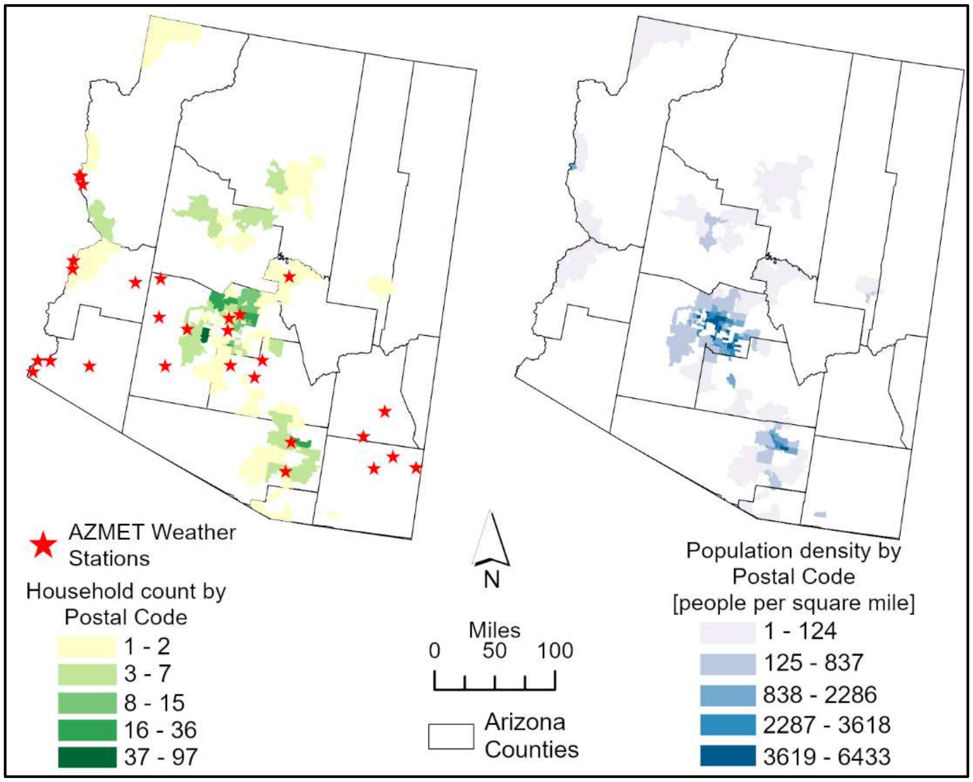


Figure 3.2. Map of Arizona showing the postal codes with number of flume households used in this study on the left panel. Daily weather information was obtained from the Arizona Meteorological Network (AZMET) weather stations highlighted using red stars. On the right panel, population density of those corresponding postal codes with flume sensors are presented using the 2020 Census information.

The households in these datasets were geographically identified only by their postal code designation in order to protect the privacy of Flume's customers. Figure 3.2 (left panel) shows the number of single-family households with active Flume sensors within Arizona by postal code polygons. The right panel illustrates population density in the study area. More household smart sensors are located in areas with higher population density.

A subset of the primary water consumption dataset was included in the analysis to improve the robustness of the statistical analyses. Households were included in the analysis only if they had water use data available for a minimum of 3 months or 90 days, and indoor and outdoor water use volumes were above 10.0 gallons per household per day (GPHD) on any given day. This process aimed to exclude low leak volumes from the analysis. However, distinguishing large leak volumes from lawn irrigation or swimming pool fill volumes proved challenging, and thus, these large leak volumes were not identified or removed during the data processing stage. After data processing, a final set of 706 households was identified for the subsequent data analysis.

The building and demographics characteristics for the selected households that were considered in the analysis include: number of bathrooms; number of occupants; irrigation type (drip, garden hose, soaker hose, sprinkler, or none); irrigation frequency (once per week, twice per week, thrice per week, more than 4 times per week, or none), swimming pool (Yes/ No), home value ($), year built, lot size (square feet), home size (square feet), and postal code.

The household-appliance characteristics for the selected households that were considered in the analysis includes per event average flow characteristics - duration (minutes), volume (gallons), and flow rate (gallons per minute), for clothes-washer, dishwasher, faucet, indoor, irrigation, leak, shower, softener, and toilet. Flume Inc. and Mayer (2022) elaborate in detail on the disaggregation process and methodology used to obtain the appliance level water consumption information from the original 5-minute interval water consumption data.

Finally, daily weather information for every household in the Flume dataset was collected from the University of Arizona's Arizona Meteorological Network (AZMET) stations (accessed January to March 2023, https://ag.arizona.edu/azmet/), which are positioned across the state (Figure 3.1). Daily precipitation (inches), daily evapotranspiration (inches) calculated using the Penman-Monteith method, and daily minimum, maximum, and mean air temperature (degrees

Fahrenheit) were gathered for all the active AZMET stations in WY-22. Using ArcGIS Pro software's ‘Find the nearest feature’ tool, all the active AZMET stations were matched by proximity with the nearest postal codes containing households with active Flume sensors (Figure 3.2).

3.4 Statistical Methods

The statistical analyses in this study include characterization of water use intensity, assessment of swimming pool’s impact on water consumption patterns, estimation of household appliance efficiencies and their impact on water consumption, and the mixed-effects modeling of household water consumption at different time steps. The statistical methods and data visualization were conducted using the R/RStudio software.

3.4.1 Characterization of water use patterns

Water use metrics in gallons per household per day (GPHD), gallons per capita per day (GPCD), and gallons per square foot (GPSF) are used in this study to characterize indoor, outdoor, and total water uses. The GPSF metric is specifically used to characterize outdoor water use. Indoor and outdoor GPCD was calculated by dividing the daily indoor and outdoor GPHD of a household by the ‘number of occupants’ specifically reported to be in that household, and the daily GPSF of a household was obtained by dividing the aggregate outdoor GPHD of a household by its ‘lot size’.

The temporal pattens in indoor and outdoor water consumption was estimated using the average indoor and outdoor GPCD of all households at a monthly time step. An average percent difference between all the monthly indoor and outdoor GPCD was then computed to generalize the average percent difference between the two water consumption categories. This representation offers insights into the total percentage of water used indoors vs. outdoors in single-family households in Arizona.

Relationships were then estimated for household water use and household occupancy level (number of residents in a household). Average annual water use characterized by GPCD and GPHD was estimated for every household, and a simple univariate regression was conducted to estimate the linear or non-linear (power, exponential, logarithmic, or polynomial) nature of the relationship between household level GPCD and GPHD versus household occupancy level for the state of AZ by taking into account all the 706 households as the sample size (*N*). Due to the proprietary nature of the datasets, only the type of relationship (linear or non-linear) and its corresponding goodness of fit, adjusted R^2 value (or, coefficient of determination) of the relationship, are presented in the results section. These relationships are presented by sub-setting the 706 households by those having a swimming pool (N=419) and those without (N=287).

3.4.2 Impact of Swimming Pools

The impact of swimming pools on household-level water use is further assessed by the end-use category – indoor, outdoor, and total. Average indoor, outdoor, and total water use in GPCD separated by two groups - households with and without a swimming pool, are estimated using the statistical box plot method or whisker plot method (Chambers et al., 1983; Becker et al., 1988; Murrell, 2005). ANOVA test of significance is then applied to test whether the patterns observed in indoor, outdoor, and total water uses for households with and without a swimming pool are significant.

3.4.3 Impact of Appliance Efficiency

Box plots are used to categorically compare indoor and outdoor water use between households. The comparison is based on the combined water efficiencies of all appliances in each home, using the gallons per capita per day (GPCD) and gallons per square foot (GPSF) metrics. Data from all the 706 households are analyzed. This process is described in the following paragraphs in detail.

Using the household appliance's average characteristics information, a statistical clustering method was applied to identify the clusters (or, efficiency levels) in the flow characteristics of each of the end-water-using appliances. The household appliances of interest are toilets, showerheads, clothes-washers, dishwashers, and faucets. The appliance flow characteristics of particular interest in this study were, 1) average gallons per event for toilets, 2) average gallons per minute per event for shower heads, 3) average gallons per event for clothes-washer, 4) average gallons per event for dishwasher, and 5) average gallons per minute per event for faucets.

These end-water-using appliances (showerheads, toilets, faucets, clothes-washers, and dishwashers) are standard fixtures in any household within the US and have been well studied in the past, especially by the 'Residential End Uses of Water' study groups (DeOreo et al., 1996, 2016; Deoreo & Mayer, 2012; Mayer et. al., 1999). Although there are other appliances used within a household, like the water softener, water heater, and different types of irrigation systems, they were not considered in the end-use appliance analytics.

Cluster analysis on the appliance flow characteristics was conducted to primarily identify water usage patterns of the individual appliances, thereby giving an estimation of whether those appliances met the U.S. plumbing code standards of efficiency. This was because Flume customers did not provide information about their household appliances (such as low-flow or high-flow toilets or faucets, top-loading or front-loading clothes-washers, etc.,) in the survey when installing the smart sensor. Finally, in determining the overall efficiency of a household, an approach was devised to look at the 'combined' efficiencies of the toilet, faucet, showerhead, clothes-washer, and dishwasher systems within a household and whether all those appliances met the industry standards or not.

The U.S. Environmental Protection Agency's guidelines for WaterSense® and the Department of Energy's ENERGY STAR® programs were considered in evaluating the efficiency of a household appliance. These programs are generally used as plumbing code standards for water use efficiencies within the U.S. The Energy Policy Act of 1992 (EPAct 1992) established the maximum flush volume for toilets at 1.6 gallons per flush (gpf), and the maximum flow rate for bathroom sink faucets, kitchen faucets, and showerheads at 2.5 gallons per minute (gpm). Subsequently, in 1998, the U.S. Department of Energy (DOE) adopted a maximum flow rate standard of 2.2 gpm for all faucets. In 2012 and 2016 respectively, the DOE issued new regulations mandating minimum water efficiency requirements for clothes washers and dishwashers (DOE, 2012; DOE, 2016; EPA, 2021). DOE (2016) regulates that dishwashers with annual energy use of 307 kWh/year must use not more than 5 gallons per cycle and systems with 222 kWh must not use more than 3.5 gallons per cycle. DOE (2012) regulates that clothes-washers which are top-loading must not use more than 14.4 gallons per load per cubic feet capacity of the system, and front-loading washers must not use more than 8.3 gallons per load per cubic feet capacity of the system. Due to the unknown information about the cubic feet capacity of the clothes washers in the Flumedatasets, 30 gallons per load is used in this study as a threshold for establishing efficiency in a clothes-washer system as determined by DeOreo & Mayer (2012).

Partitioning Around Medoids (PAM) or k-medoids clustering method (Kaufman and Rousseeuw, 1990) is used to cluster the flow characteristics of the household appliances. The PAM method was chosen over the commonly used k-means method for its capability of creating groups around the medoids which are an actual entity of the dataset that represents the group in which it is inserted, while the centroids of the k-means method are an artificial entity of the dataset (Kaufman and Rousseeuw, 1990). Kaufman and Rousseeuw (1990) elaborated that the medoid estimated using the PAM method is a data point with minimal average dissimilarity to all objects of the cluster, and they affirm that the PAM method is most suited for characterization purposes through clustering that relies on the representative objects used. Also, the k-means estimation of centroids is sensitive to the outliers, while the PAM method of estimating the medoids makes it possible to isolate the outliers while determining the medoids (Kaufmann & Rousseeuw, 1987).

In the process of performing a cluster analysis, it is essential to predetermine the optimal number of output clusters, *k*. Here, the Silhouette method is used to identify the optimal *k*. This Silhouette method proposed by Rousseeuw (1987) is a graphical display method to represent each cluster as a silhouette, which is based on the comparison of tightness and separation of the medoid clusters. The average silhouette width provides an evaluation of clustering validity and might be used to select an 'appropriate' number of clusters (Rousseeuw, 1987).

Once the flow characteristics of the appliances of every household were individually clustered using the PAM method, based on the optimal number of clusters (*k*) identified using the silhouette method, cluster medoids of each appliance were then compared with the efficiency thresholds of that particular appliance as determined by the WaterSense and ENERGY STAR programs reported earlier in this section. If a household appliance's cluster characteristics met the WaterSense and ENERGY STAR thresholds, the appliance was classified as having "High" efficiency, and if it did not meet the WaterSense and ENERGY STAR thresholds, it was classified as having "Low" efficiency. Finally, based on the efficiency levels of individual appliances, each household was then determined efficient or not by looking at the combined efficiencies of all the individual appliances – if all the appliances and fixtures (100 percent) were "High" in efficiency, then the household was classified as "High" efficiency household. If even one of the appliance or fixture categories was "Low" in efficiency, then that household was classified as a "Low" efficiency household.

3.4.4 Drivers of Water Uses

The drivers of indoor, outdoor, and total water use in single-family households in Arizona were estimated by using multivariate mixed-effects regression models (Equation 2.1) with household demographics, building characteristics, and weather response variables:

$$Y = \alpha X + \beta Z + \varepsilon \qquad (2.1)$$

where Y is a response variable or a vector of known observations (i.e., household water use volume in this study); X represents a design matrix of all the fixed effects (e.g., time-varying weather variables); α represents an unknown vector of the fixed effects coefficients; Z represents a design matrix of all the random effects (i.e., time-constant, independent household characteristic variables); β represents the vector of the random effect coefficients; and ε is an unknown vector of random errors which are assumed to be normally distributed.

Table 3.1. Variables used in the mixed-effects modeling of water use in gallons per single-family household (GPH) in Arizona.

Response variables (Y)	Fixed effect variables (X)	Random effect variables (Z)
Total GPH Indoor GPH Outdoor GPH	Month Season Year Evapotranspiration (ET) Precipitation Air Temperature (min, max, mean)	Year Built, Postal Code, Home Size, Lot Size, Home Value, Number of Bathrooms, Number of Residents, Has Pool, Irrigation Frequency, Irrigation Type, Toilet Efficiency, Showerhead Efficiency, Laundry Efficiency, Dishwasher Efficiency, Irrigation Efficiency, Faucet Efficiency

The exploratory variables used in the mixed-effects modeling process are presented in Table 3.1. The weather impacts, like temperature, precipitation, etc., have consistent effects over all the households identified close to an AZMET weather station, and therefore are considered to have fixed effects in estimating the water use for that population. However, building characteristics, like lot size, number of bathrooms, presence of a pool, etc., are independent and unique for each household, and therefore are considered to cause random effects in the modeling of household water use. This way of combinedly modeling water uses based on fixed and random effects is referred to as mixed-effects modeling (Lindstrom & Bates, 1990; Pinheiro et al., 1995; Pinheiro & Bates, 1996).

The mixed-effects modeling approach was developed using linear, log transformed, and exponential transformation at the daily, weekly, monthly, and seasonal time steps. In the log transformed mixed-effects model, Equation 2.1 is represented by transforming the response variable (Y) and the fixed and random effects matrices (X, Z) to a natural logarithmic scale. In the exponential mixed-effects model, only the response variable (Y) in Equation 2.1 is transformed to a natural logarithmic scale. In both the log transformed and exponential mixed-effects models, the fixed and random effects are used in their additive form.

Model selection was performed using evaluation criteria including the adjusted coefficient of determination (*adjusted* R^2) and the Variable Inflation Factor (*VIF*). The *VIF* is used to ensure that there is no over-fitting and multi-collinearity among the variables in the final model. For this study, 2.5 is set as the maximum *VIF* factor threshold for all variables used in the finalized mixed-effects model for different time-steps.

3.5 Results

Based on the sample of households used in this study, water consumption patterns in single-family households in Arizona are highly skewed towards outdoor uses. Strong correlations exist between household occupancy levels and water use intensity metrics. Swimming pools and household appliance efficiencies strongly influence water consumption patterns and are driving forces behind variations in household water use. The percentage outdoor to total water use varies significantly based on the conservation mindset of the householders. Overall, household water consumption can be reliably modeled using a log transformed mixed-effects model that incorporates household and regional climatic factors.

3.5.1 Single-family household water consumption patterns by indoor/outdoor category

For the 706 single-family households analyzed in this study, a strong seasonality is observed in outdoor water consumption throughout WY-2022 (Figure 3.3, top-left panel). The bottom-left panel of that figure shows that on average, outdoor water usage accounts for nearly 64 percent of the total water consumed in these households. Indoor water usage remains relatively consistent throughout the year, averaging around 50 GPCD.

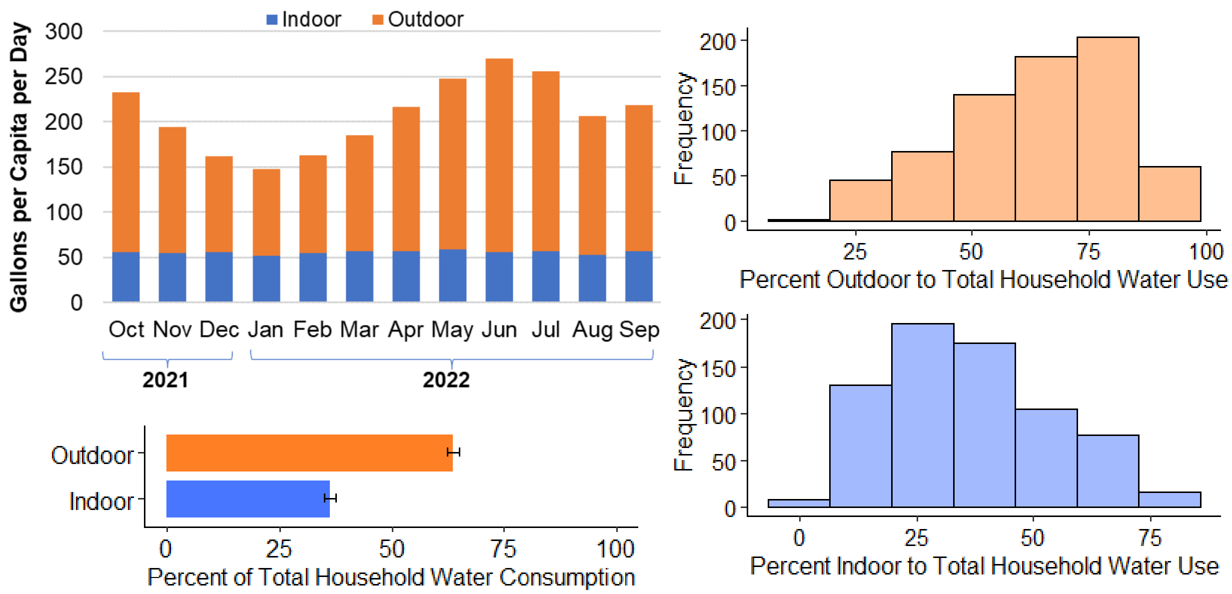


Figure 3.3. (Top-left panel) Monthly trends in GPCD of indoor and outdoor water uses in those 706 single-family households in Arizona for the water year 2022 (WY-2022). (Bottom-left panel) Percent indoor and outdoor water uses of total household water consumption with error bars. (Right panel) Histogram of percent indoor and outdoor water uses to total household water consumption.

Further statistical analysis on the percentage of indoor and outdoor water uses to total household water use revealed a normal distribution (Figure 3.3, right panel). *95% confidence interval* testing showed the expected outdoor water use ranges from 62 to 65 percent of total water use, skewing to the higher end of that range (*skewness* = -0.45). The expected indoor water use ranges from 35 to 38 percent of total water use, skewing to the lower end (*skewness* = 0.45). Both the indoor and outdoor percentages of total water use have a large *standard deviation* of 17.5 percent.

3.5.2 The effects of household occupancy

Total single-family household water consumption in Arizona is strongly related to household occupancy, and it is further influenced significantly by the presence of a swimming pool. For single-family households in Arizona, there is a statistically significant correlation between the average total water consumed across the 706 households and the household occupancy level (Figure 3.4). Model performance testing, using *adjusted* R^2, indicates an exponential relationship that best explains the variation in gallons per household per day (GPHD) based on household occupancy level. Similarly, a sub-linear power relationship most effectively captures the variation in gallons per capita per day (GPCD).

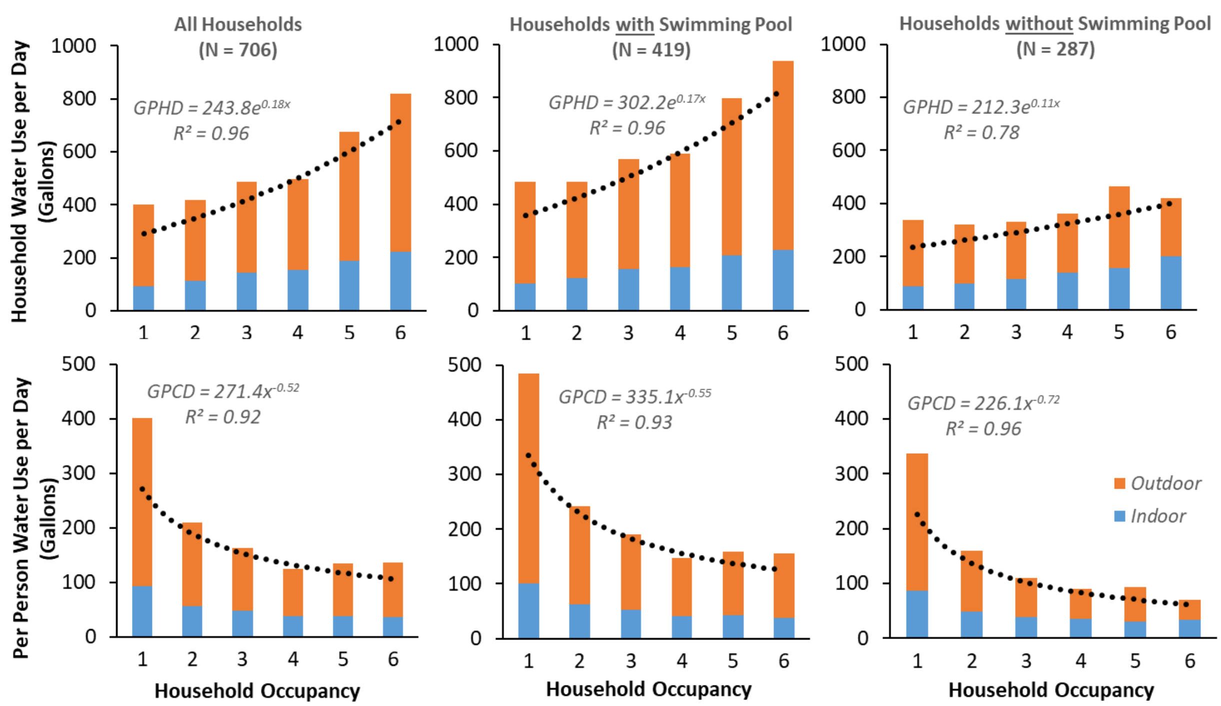


Figure 3.4. Relationships between average total water consumption in gallons per household per day (GPHD) and gallons per capita per day (GPCD) with household occupancy level (x) for single-family households in Arizona.

Figure 3.4 presents the mathematical trendline relationships, along with the goodness of fit, for the total water use category in GPHD and GPCD metrics with household occupancy levels for all the 706 households, and by grouping those households based on the presence of a swimming pool. This figure also shows the relative variability in indoor and outdoor water use in GPHD and GPCD based on the household occupancy level. On average, there is about 56 percent difference in total water used by households with a swimming pool compared to those that do not have a pool.

3.5.3 The effects of swimming pool

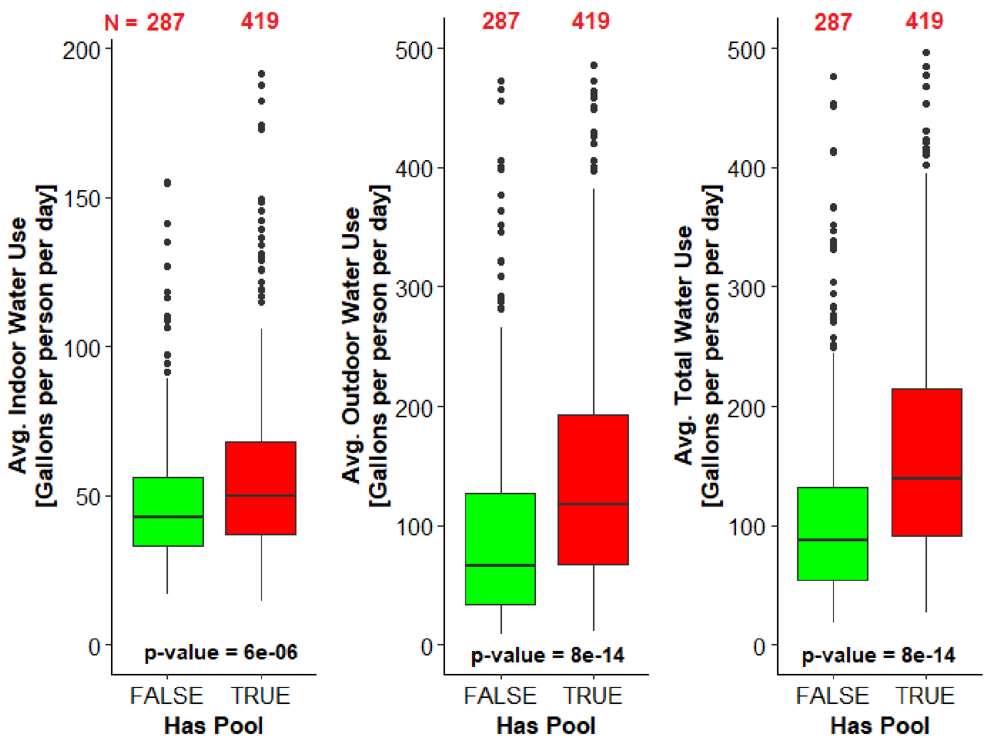


Figure 3.5. Impact of swimming pool on average household indoor, outdoor, and total water uses in GPCD. The numbers in red above the boxplots show the total number of households under that category.

Indoor water usage is substantially higher in single-family households with a swimming pool. Figure 3.4 clearly shows that the presence of a swimming pool greatly increases a household's total and outdoor water uses compared to households without a pool. Further categorical assessment of water uses in GPCD between households with and without swimming pools reveals that indoor water use is also significantly influenced (Figure 3.5). Households having a swimming pool use about 26 percent more water indoors than households without pools. The reason for this difference in indoor water was not examined as part of this study and could be a topic of future research.

3.5.4 The effects of efficient appliances

Clustering of appliance flow characteristics offers a novel way to estimate appliance efficiency and overall household efficiency using Flume's smart sensor technology.

Using the household-level appliance flow characteristics dataset, it was possible to cluster appliance characteristics into the medoid groups that were identified using the PAM method. The silhouette method results presented in Figure 3.6 show that the flow characteristics of toilets, showerheads, clothes-washers, dishwashers, and faucets for all the 706 households could be clustered into just two optimal groups, i.e., $k = 2$. Table 3.2 provides the two medoids that were estimated by the PAM method for the flow characteristics of each household appliance. Appliance characteristics of each household clustered in the Medoid-1 group were classified as "High" efficiency fixtures and those clustered in the Medoid-2 group were classified as "Low" efficiency fixtures.

Table 3.2. Medoids of individual household appliances' flow characteristics, of all the 706 sample single-family households from Arizona, estimated using the PAM method.

Appliance	Medoid-1 [High efficiency]	Medoid-2 [Low efficiency]
Toilet [avg. gallons per flush]	1.78	2.42
Showerhead [avg. gallons per minute]	1.61	2.24
Clothes washer [avg. gallons per load]	25.40	33.97
Dishwasher [avg. gallons per load]	3.28	5.42
Faucet [avg. gallons per minute]	0.015	0.070

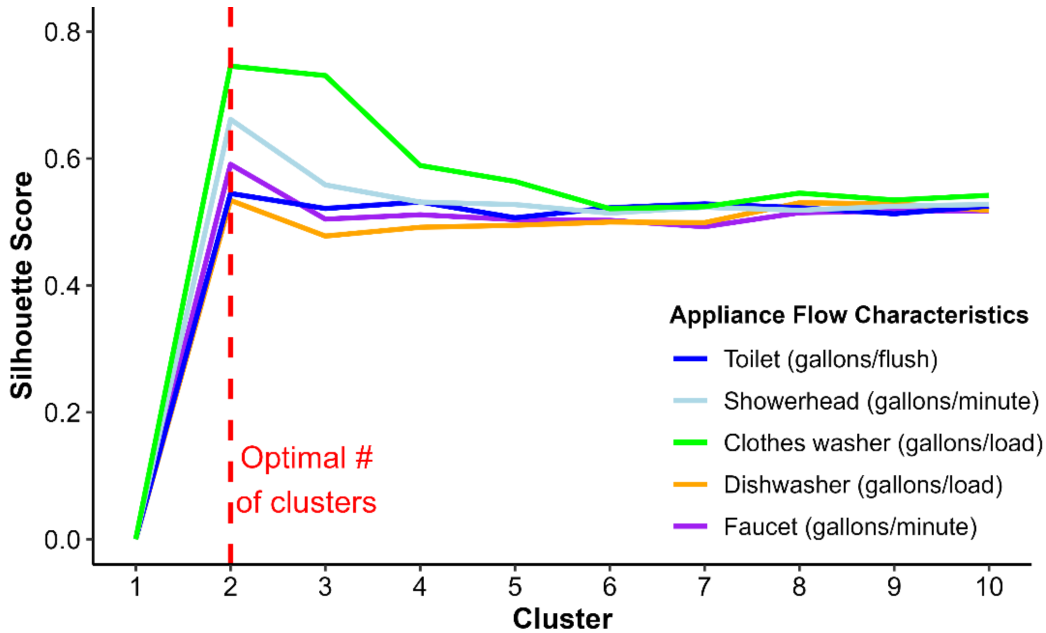


Figure 3.6. Visual representation of the optimal number of clusters identified for all the individual household appliances using the Silhouette method's scores.

Figure 3.7 shows the median value and range of the household appliance's flow characteristics clustered as "High" or "Low" efficiency fixtures based on the medoid estimates in Table 3.2. PAM method identifies "High" efficient toilets as 1.78 gallons per flush (gpf) system which is pretty close to the WaterSense standard of 1.6 gpf. Both the medoids of flow characteristic of showerheads in these sample single-family households meet the efficiency standard of 2.2 gallons per minute (Table 3.2), however, due to the presence of outliers around the medoid-2 group of the showerheads (Figure 3.7), all the household showerheads clustered around medoid-2 are considered "Low" in efficiency. Clothes-washers with "High" efficiency are clustered around the 25.4 gallons per load (gpl) medoid. Following DeOreo & Mayer's (2012) findings, all the "High" efficient clothes-washer systems use less than 30 gpl. Similarly, for dishwashers, medoid 3.28 gpl is considered a "High" efficient system, and anything around 5 gpl or more is considered to be "Low" in efficiency. The PAM method again identified the "High" efficiency dishwasher system as having 3.38 gpl and the "Low" efficiency medoid for dishwasher as 5.42 gpl, both of which are close to the DOE (2016) standards for efficient dishwashers. The faucet flow characteristic however was unusually small for both the medoid clusters (<<2.2gpm). It is assumed that there is some erroneous metric within the Flume sensor algorithm. Faucets in the households were still classified as "High" and "Low" efficient fixtures based on the medoid groups in Table 3.2.

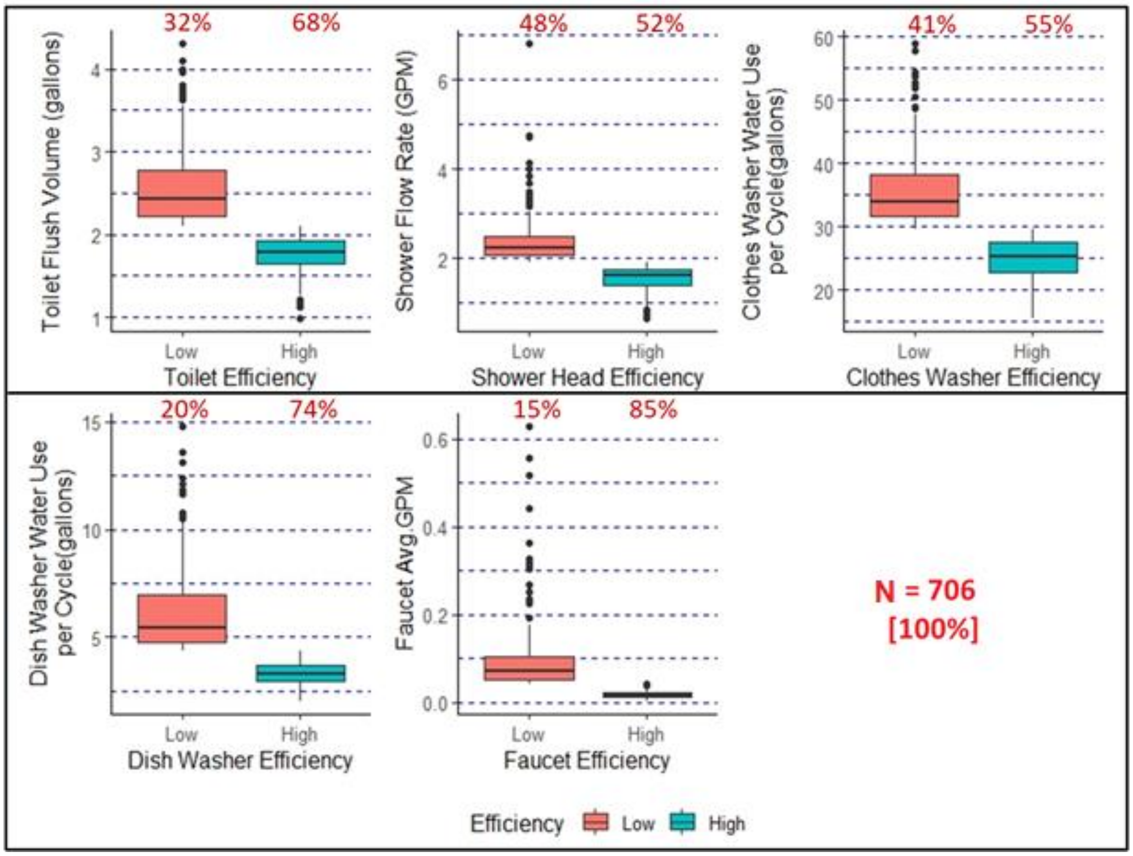


Figure 3.7. State of household appliance efficiencies were classified after clustering the appliance's flow characteristics of all the 706 sample single-family households. Percentage numbers in red color over the boxes represent the percent households within each efficiency level.

Overall, it is optimistic to note that more than 50 percent of the sample single-family households considered for this study have "High" efficiency toilets, dishwashers, and faucets, and similarly, a little over 50 percent of the households have "High" efficiency shower heads and clothes-washers.

After determining appliance efficiencies in every sample household, indoor and outdoor water uses were then characterized based on the household efficiency level. In Figure 3.8, Household Efficiency is determined by the combined efficiency classifications of toilets, showerheads, dishwashers, clothes-washers, and faucets. Per capita indoor water use was about 18.5 percent less in homes classified as "High" efficiency. Homes classified as "Low" efficiency used about 18.5 percent more water per person. Outdoor use (in GPSF) was not different between these two groups, even as indoor use varied. Indoor efficiency alone is not necessarily predictive of reduced outdoor water use.

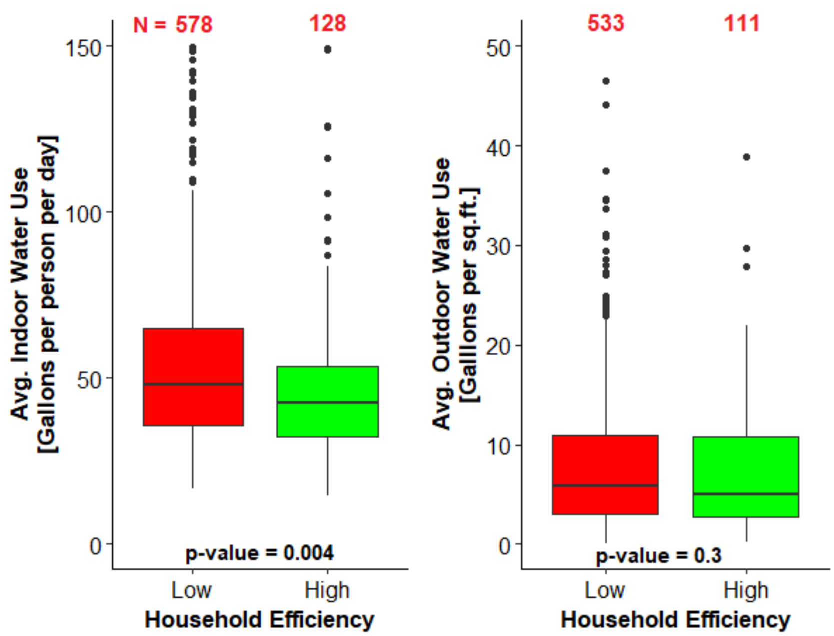


Figure 3.8. Average household-level indoor and outdoor water uses differ based on household appliance efficiency levels. The numbers in red over the boxes represent the total number of households within such an efficiency level.

3.5.5 Conservation mindset and its impact in outdoor water consumption

With the outdoor water use being highly skewed, and the presence of swimming pools and appliance efficiency levels having a significant impact on its variation across the single-family households in Arizona, outdoor water use was further analyzed using the consumption mindset framework. Households were grouped by those that i) have a swimming pool and a low overall household efficiency (*N*=377), and ii) do not have a swimming pool and have a high overall household efficiency (*N*=32). These two groups represent in volume the highest and lowest outdoor water users, respectively, among the 706 single-family households analyzed in this study. Thus, the households grouped in the latter were assumed to have a "high" conservation mindset, while the former group were assumed to have a "low" conservation mindset.

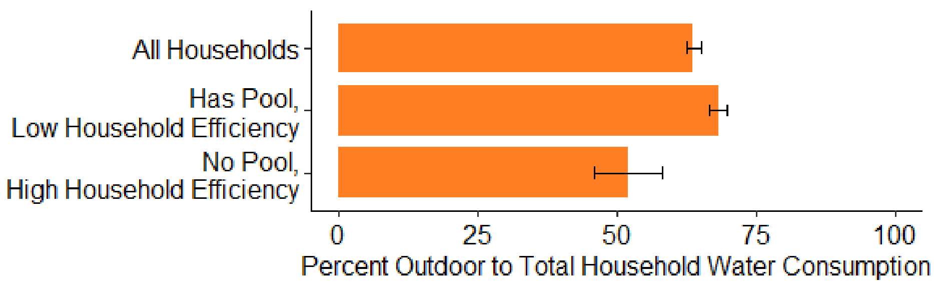


Figure 3.9. Comparison of variation in the percent outdoor to total water use among the households grouped based on their conservation mindset.

Statistical analysis of these two household groups again revealed a normal distribution in the percentage of outdoor to total water use (Figure SI-3.3). Table SI-3.2 presents their statistical characteristics with the *95% confidence interval* range of expected values. Figure 3.9 shows how the households grouped by conservation mindset compared to all households, irrespective of water conservation attitudes. Households with a low conservation mindset used up to 70 percent of total water outdoors, while those with a high mindset used as little as 46 percent outdoors. An ANOVA test confirmed a significant difference between the group means. Figure 3.9 also shows the relative comparison of these two groups to the percent outdoor water use of all 706 households grouped together.

3.5.6 Quantification of water consumption using household characteristics and regional factors

The log transformed mixed-effects models perform the best in explaining the single-family household water use variation. Figure 3.10 shows the goodness-of-fit for the log transformed mixed-effects modelling of the total household water consumption at different time-steps. Figures SI-3.1 and SI-3.2 in APPENDIX-B show the goodness-of-fit for the log transformed mixed-effects modelling of indoor and outdoor household water uses, respectively. Log transformed mixed-effects model prediction significantly improves with increasing time-steps.

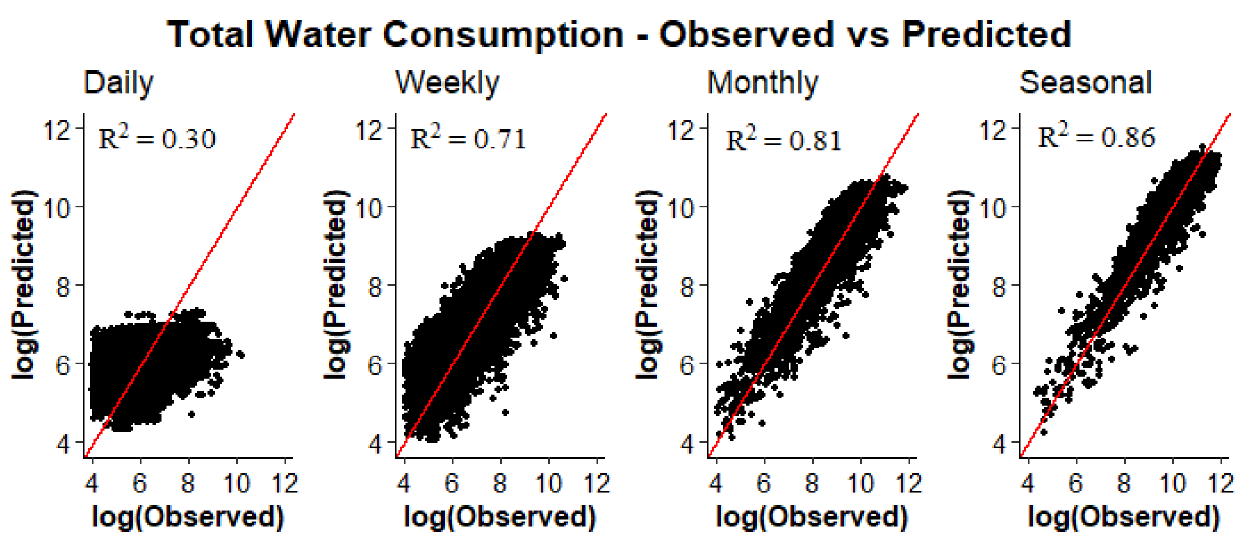


Figure 3.10. Goodness-of-fit plots of observed versus predicted household level total water consumption data at different time-steps using the log transformed mixed effects model. Model prediction improves with larger time-steps.

The performance metric for all the linear, log transformed, and exponential mixed-effects models built for different time steps of the household water uses is presented in Table 3.3.

Adjusted R^2 greater than 0.70 and *p-value* less than 0.05 were the criteria used in selecting the best performing mixed-effects models that have good prediction capabilities on the household water uses. While all the models were significant (*p-value* < 0.05) in modeling household water uses, only the log transformed mixed-effects models showed strong (*adjusted* R^2 > 0.7) predictive capabilities.

Weekly time step log transformed model produces strong (*adjusted* R^2 > 0.70) and significant (*p-value* < 0.05) results in modeling the variations in total household water consumption only, while the monthly and seasonal time-step log transformed models are strong and significant in modeling all categories of household water consumption. However, at the daily time step, the log transformed models remain significant (*p-value* < 0.05) but lack strong predictive power (*adjusted* R^2 < 0.70) and is noisy (Figure 3.10). This was also the case with modeling indoor and outdoor water uses at daily and weekly time steps using the log transformed model (Figures SI-3.1 and SI-3.2).

Table 3.3. Adjusted R^2 for the test of significance results of different time intervals of indoor, outdoor, and total - gallons of water used per household, modeling using mixed-effects models. All these models have a p-value of less than 0.05.

Time-step	Mixed-Effects Model Type	Adjusted R^2		
		Indoor GPH	Outdoor GPH	Total GPH
Daily	Linear	0.02	0.23	0.22
	Log transformed	0.12	0.27	0.30
	Exponential	0.12	0.28	0.30
Weekly	Linear	0.21	0.41	0.44
	Log transformed	0.58	0.66	0.71
	Exponential	0.47	0.59	0.63
Monthly	Linear	0.38	0.47	0.52
	Log transformed	0.74	0.77	0.81
	Exponential	0.56	0.66	0.68
Seasonal	Linear	0.48	0.48	0.53
	Log transformed	0.81	0.82	0.86
	Exponential	0.61	0.68	0.70

3.5.7 Drivers of single-family household water uses in Arizona

The fixed and random effect variables (or drivers), *X* and *Z*, (Table 3.1) that are significant in modeling the single-family household's total water consumption, *Y*, at the monthly and seasonal time-step using the mixed-effects model are presented in Table 3.4. Drivers of indoor and outdoor water uses at the monthly and seasonal time steps are presented Table SI-3.1 in APPENDIX-B. The direction and magnitude of variables in the model depended on the modeling time-step. All the variables and their estimates presented in these tables have a p-value of less than 0.05 and their variable inflation factor (VIF) of less than 2.5. Note that the model intercept estimates presented in Table 3.4 and Table SI-3.3 are independent of fixed and random effects.

Table 3.4. Important factors of single-family household's total water consumption in Arizona at the monthly and seasonal time steps, and their corresponding coefficient estimates in the log transformed mixed-effects models. "NA" denotes variable insignificance in that time-step. Astrix (*) is used under the random-effect column if that variable has a fixed-effect nature, and vice versa.

Variables	Fixed Effect Coefficients (α)		Random Effect Coefficients (β)	
	Monthly	Seasonal	Monthly	Seasonal
Air Temperature (Max)	-3.32	-2.16	*	*
Dishwasher Efficiency	*	*	0.07	0.07
Evapotranspiration	1.10	1.11	*	*
Faucet Efficiency	*	NA	-0.06	NA
Has Pool	*	*	0.13	0.12
Home Value	*	*	0.19	0.20
Irrigation Efficiency	*	*	0.84	0.84
Irrigation Frequency	*	NA	0.07	NA
Irrigation Type	*	*	0.45	0.49
Laundry Efficiency	*	*	0.08	0.07
Lot Size	*	*	0.10	0.08
Month of Year	0.32	NA	*	NA
Number of Bathrooms	*	NA	-0.07	NA
Number of Residents	*	*	0.24	0.23
Postal Code	*	*	-38.03	-32.89
Showerhead Efficiency	*	*	0.12	0.11
Toilet Efficiency	*	*	0.11	0.12
Year	NA	-1124.03	NA	*

Year Built	*	*	5.69	6.73
Intercept	492.35	8992.92	*	*

For indoor water use, the number of residents, home features like pool ownership, the number of bathrooms, and the efficiency of appliances were important drivers. Outdoor water use was driven by irrigation factors like efficiency, frequency, and type, along with weather parameters (temperature and ET), socioeconomic attributes (home value, postal code), lot size, and pool ownership. The most significant drivers of total water use aligned with those for indoor and outdoor water use. Weather parameters impact all categories of single-family household water uses in Arizona and are speculated to be due to the overall arid climate type experienced within the state.

3.6 Conclusion

This analysis highlights the pronounced impact of critical factors influencing single-family household water consumption patterns in Arizona. The factors presented in this study only are statistically tested for their influence on water use in a household, but in reality this might not entirely explain the actual variation in water use.

The analysis demonstrates a strong correlation between household occupancy and water usage, with the presence of a swimming pool further amplifying water consumption. The distinct relationship between household occupancy level and water consumption paves the way forward for policy-based and urban planning-related measures to target water conservation practices. This may involve changing housing density and associated land use patterns, and increased attention to the relationship between water use and urbanization (Gordon & Richardson, 1997; Sampson et al., 2022; Stoker et al., 2022).

Furthermore, high seasonal variation in outdoor water use emphasizes the need to manage water resources in a sustainable manner, particularly in arid regions like Arizona. The percent of outdoor water use for Arizona estimated in this study shows some efficiency gains compared to the estimates for percent outdoor water use that Balling et al., (2008) presented in their report, based on the trends and patterns they observed between 1995 and 2004 just for the Phoenix metropolitan area. Outdoor water usage remains one of the biggest end-use categories of water in Arizona's single-family homes and is particularly high for homes with swimming pools, which use about 56 percent more water than the homes without a pool. Given the role of swimming pools in mitigating the effects of extreme heat, perhaps no other use highlights the tradeoffs between reducing water use and human comfort.

The investigation of appliance flow characteristics using Flume's smart sensor technology provides insights into the state of efficiency of the household fixtures. Clustering of appliance flow characteristics unveils that a significant portion of households possess high-efficiency toilets, dishwashers, and faucets. Notably, the study shows that indoor water use is influenced and reduced by better appliance efficiency. Overall, household efficiency and the presence of a swimming pool have a huge impact on a household’s outdoor water consumption. Grouping households based on their conservation mindset shows that a wide range exists in outdoor water usage across the single-family households in Arizona.

Overall, the mixed-effects modeling efforts undertaken in this study mostly reiterate the learnings from past studies (Balling et al., 2008; Deoreo & Mayer, 2012; DeOreo et al., 2016; Ouyang et al., 2014; Stoker & Rothfeder, 2014). This study builds on previous efforts by estimating household appliance efficiency using smart-metered data to improve the predictive power of water use/demand models, thus enhancing our understanding of water demand management.

To achieve significant scientific progress and make meaningful contributions to municipal water demand management in the United States, it is crucial to expand this analysis beyond the state of Arizona or the "Arizona Sun Corridor". With the continuous growth of urban populations across the country, urban water usage is on the rise. This demographic shift has profound implications for how water is sourced, transported, and consumed, thereby necessitating a comprehensive understanding of the factors influencing water use. Smart-metered data, such as Flume's, can offer many nuanced insights into these variations in water uses, and their co-benefits for a diverse set of water management practices are yet to be realized (Monks et al., 2019).

Furthermore, the complex interplay between land use management, urban development, and climate variability impacts water supply and demand patterns across regions. Gathering nationwide, high-resolution, smart-metered data is therefore increasingly important to understand water usage dynamics. This study has thus offered a statistical framework to enable such future research using smart metering data. We hope this paves the way for more robust analysis and modeling of urban water demand across socioeconomic groups, promoting equitable and sustainable water management on a national scale.

3.7 Data Availability

Water use data for this research work was provided by Flume, Inc., a smart water sensor company that collects real-time water use data from their customers who install Flume Sensors in their households. Due to the proprietary nature of the data (owned by Flume, Inc.), those interested in obtaining data must contact Flume directly via: https://www.flumedatalabs.com/contact-us.

The weather data collected from the Arizona Meteorological Network (AZMET) stations positioned across the state can be accessed online at https://ag.arizona.edu/azmet.

CHAPTER 4 – DEVELOPMENT OF ROBUST URBAN SCALING METRICS ON MUNICIPAL WATER CONSUMPTION PATTERNS

4.1 Introduction

Within the contiguous United States, population growth patterns are becoming skewed towards urban areas with a steady migration of people from rural to urbanized areas. Based on estimates released by the USGS on the population of people connected to a water supply system within the U.S., for every 5 years, metro classified counties have seen an average of 5.5% increase in population from 1985 to 2015, while rural counties in the country have declined steadily at an average of 1.1% (Table SI-4.1 in APPENDIX-C). This trend in steady average population growth across the cities and towns within the U.S. is also captured by the municipal water use data of Chinnasamy et al. (2021) and presented below in Figure 4.1.

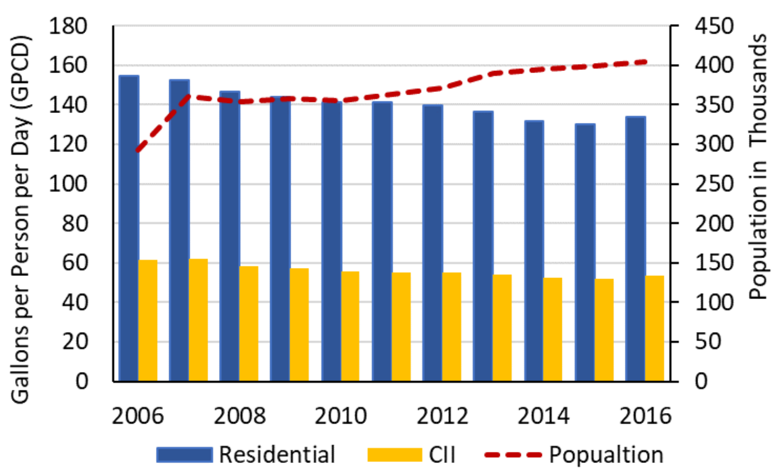


Figure 4.1. Changes in population and per capita water usage over time for 126 municipalities within CONUS that were characterized in Chapter-2.

Interestingly, Figure 4.1 presents another valuable information regarding average residential and CII water consumption using the gallons per person per day (GPCD) metric over time for those 126 municipalities. Between the study period from 2005 to 2017, despite the steady increase in population, there has been a steady decrease in water consumption or a steady increase in water savings across those municipalities, i.e., there has been about 20 percent reduction in total water use per person while the population grow by about 38 percent. Although the direct correlations for these reductions in water consumption has not been explicitly studied, advances in improving water-appliance efficiencies, water conservation efforts, smart-metering and leakage detection, and even water use restrictions in some parts of the country due to water shortages can be confidently attributed to this trend (Maas et al. 2017; Magionni, 2014; Mayer et al., 2016; Sauri, 2013; and Schultz et al. 2018).

In the past decade, studies in urban economics and complex systems have become to recognize through statistical relationships that as a place gets more urbanized, with increasing economic activity and densifying population, relationships are observed between urban attributes (especially population) and economic productivity, innovation rates, energy use and infrastructure needs (Bloom et al., 2008; Glaeser et al.,2003; Lobo et al., 2020; Nordbeck, 1971). These urban relationships are referred to as urban scaling relationships and form the basis for the urban scaling theory (Barenblatt, 2003; Bettencourt, 2013; Chave and Levin, 2003). The fundamental process at the core of this urban scaling theory is the agglomeration of social, economic, and political interactions in space and time, subject to constraints imposed by environmental conditions, technology, and institutions (Bettencourt, 2013; Schlapfer et al., 2014; Sugar and Kennedy, 2021). Figure 4.2, taken from Dewitz and U.S. Geological Survey (2021), shows the extent of the urbanized areas within the CONUS in the year 2019. This figure provides a visual representation of the land area that has been completely transformed by humans for numerous economic and social needs.

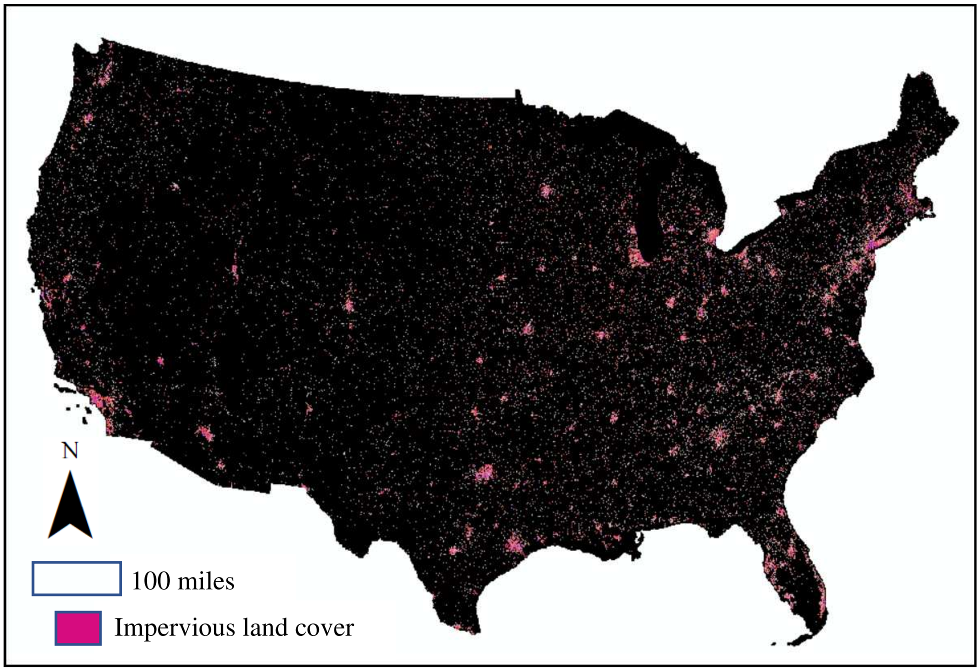


Figure 4.2. Map of CONUS in 2019 showing the extent of developed impervious land cover (non-black regions), an inherent attribute of permanent environmental changes made by humans for a variety of economic and social needs.

The nature of the scaling relationship between a city attribute (economic output, resource consumption, crime rate, innovation rate, etc.) and the city population is being recognized as metric for assessing the sustainability and livability of a city (McGill, 1964; Zhou et al., 2022). The urban scaling framework essentially is being used to debate if densification and agglomeration of people and services in a city is desirable or not in the long run. It is seen that transportation-related infrastructure development costs reduce in a sub-linear manner with increasing population size, while the amount of innovation and gross domestic product (GDP) value of a city have a super-linear relationship with increasing population size (Bettencourt, 2013). This way of understanding an urban attribute and population has many implications when it comes to taking integrated approaches to managing urban space and material, thereby enabling systems thinking.

Owing to the relative recentness of this scaling concept in urban economics, there is yet a consensus to be reached about the urban scaling relationships between municipal water uses and the corresponding population that consume it. Also, urban water use is highly complex and is influenced by a variety of technological, socio-economic, geopolitical, environmental and ecological factors. Mahjabin et al (2018), using a water footprint model, have attempted to capture all the actual and virtual uses of water within their 65 mid to large sized study cities in the U.S., and they concluded that there is a sublinear relationship between freshwater footprint and population size. They have also cautioned that while city attributes, like GDP, food production and industrial output, have a higher R^2 value (above 0.8) in their sublinear relationship with population size, freshwater footprint however has a weak sublinear relationship with population ($R^2 < 0.6$).

The criticality of this research stems from the growing recognition that water, like land, is becoming an increasingly scarce resource. Many urban regions within the U.S. are aiming to densify, yet it remains unclear if this has any impact on water consumption patterns. Therefore, the question of 'does municipal water consumption scale with population?' becomes important to answer as urban population growth trends and impacts of weather extremities on freshwater availability are adding stress to how water supplies are managed for municipal needs (Mekonnen & Hoekstra, 2016; Grigg et al, 2018; Maas et al, 2017). Moreover, the intricate relationships between water supply and land use management add to the limitations of how cities can grow and densify, while meeting the water needs for domestic and economic purposes.

This study applies the urban scaling framework to an extensive dataset of annual water supply and consumption from 2005 to 2017 across 126 municipalities within the contiguous U.S. (18). The goal of this research is to rigorously examine whether economies of scale are evident in water consumption as cities expand and economies grow. This involves a detailed estimation of the scaling relationships between municipal water use and population, along with other urban and regional characteristics. In quantifying these relationships, this study aims to contribute to a deeper understanding of dynamics to scaling in municipal water consumption, which will be a critical step towards more efficient and sustainable urban water management strategies.

4.2 Objectives

This research aims to address the overarching question: Do municipal water consumption patterns in the U.S. exhibit economies of scale? The specific objectives are:

1. To explore and identify city or regional attributes that significantly influence municipal water consumption patterns.
2. To quantify the urban scaling relationships between municipal water consumption and these identified urban and regional attributes, thereby contributing to a more sustainable and efficient approach to urban water management.

4.3 Data and Study Regions

The annual water supply/ consumption data from 126 cities and towns within CONUS from Chinnasamy et al. (2021) is used in this study. Along with it, the climatic features from the PRISM climate group put together by the Oregon State University, socio-economic factors from the 5-year estimates of American Community Survey data obtained from www.census.gov, and the national land use/land cover features from the Multi-Resolution Land Characteristics (MRLC) consortium group were compiled for those 126 study cities and towns (Table 4.1).

Table 4.1. Features collected to hypothesize the influence on water use patterns within the U.S.

Variables
US Census region, Koppen-Climate region, cii.res ratio, Elevation, Population, Population density, Precipitation, Max temperature, Max vapor pressure deficit, Total Households, Percent Single Family households, Percent Multi Family households, Single-family to multi-family household ratio, Percent Townhomes, Percent Mobile-homes, Median year household built, Median per person income, Income Gini-index, Percent total educated, Percent some education, Percent highly educated, Imperviousness acres/percent, Highly developed land acres/percent, Medium developed land acres/percent, Low developed land acres/ percent, Open space land acres/percent, Other land use acres/percent

4.4 Methods and Research Flow

Two new systems of clustering cities and towns 1) based on the combination of the U.S. Census and Koppen-Climate regions, and 2) based on the CII/Res ratio of cities are presented below. Furthermore, detailed explanation of the urban scaling theory and its modified version adopted in this study to fit the municipal water consumption data that varies significantly by the new clusters systems are presented in the following sub-sections.

4.4.1 Combining US Census and Koppen-climate region classifications.

In Chapter-2, those 126 cities and towns were clustered based on US Census regions, Koppen-Climate regions, and population size clusters (Figure 4.4-Left Panel). Population size clusters are analogous to population size, and therefore are not used in this study. The left panel of Figure 4.4 shows the evident overlap between the US Census Regions and Koppen-Climate regions in their variations in explaining water use, especially for the South region dominated by Temperate climate and the Midwest and Northeast dominated by the Continental climate. The right panel subplots of Figure 4.4 show the ranges of the ‘Total’ water use per capita variation over time, and the overlap in the ranges of GPCD between US Census and Koppen-Climate regions are again evident except that the 'Northeast' census region's variations in Total GPCD are not well described by the Koppen-Climate regions.

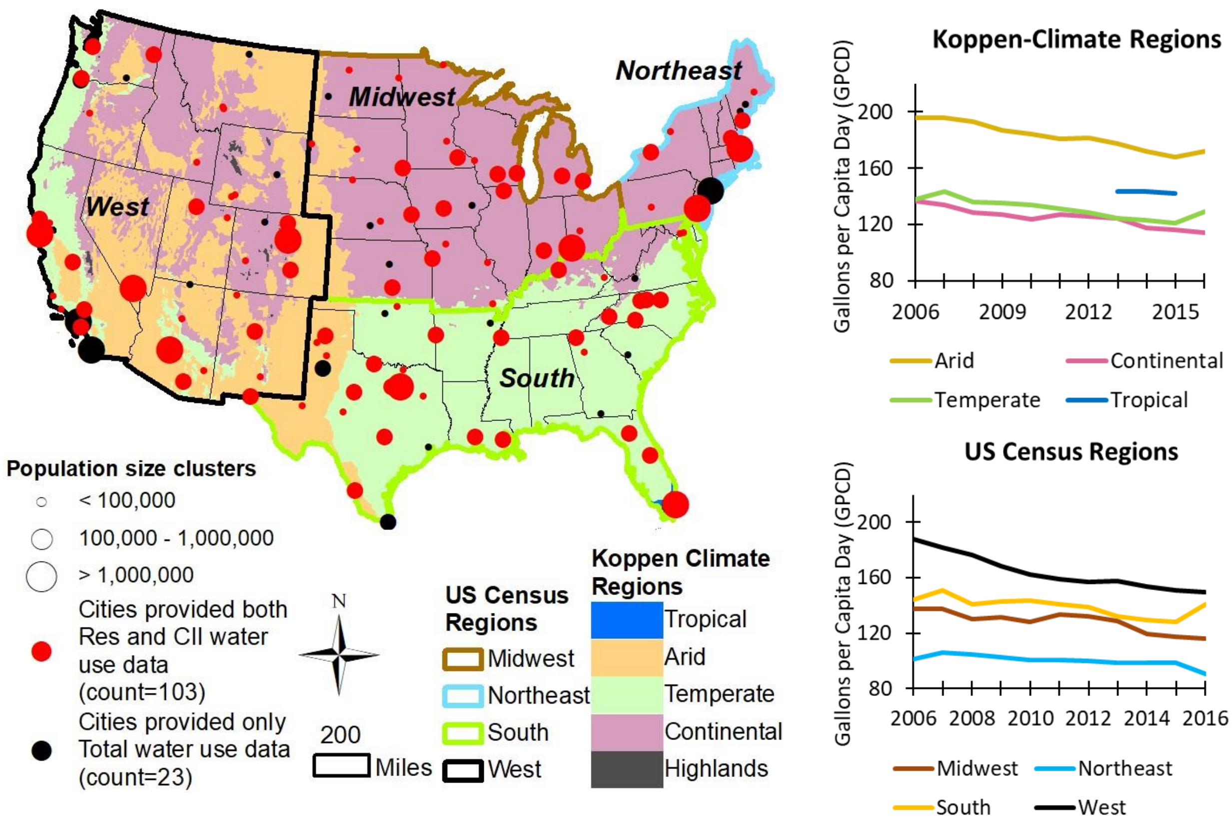


Figure 4.3. (Left panel) The map of study cities and towns clustered by US Census and Koppen-Climate regions. **(Right panel)** Subplots of total water use in GPCD over time for cities clustered by the census sand climate regions.

Therefore, a geospatial merging process was undertaken to combine the Koppen-Climate and US Census regions into a unified system of classification, called the "Climate-Census regions", which captures the unique variations in water uses across space enabling the avoidance of redundancy in regionalization.

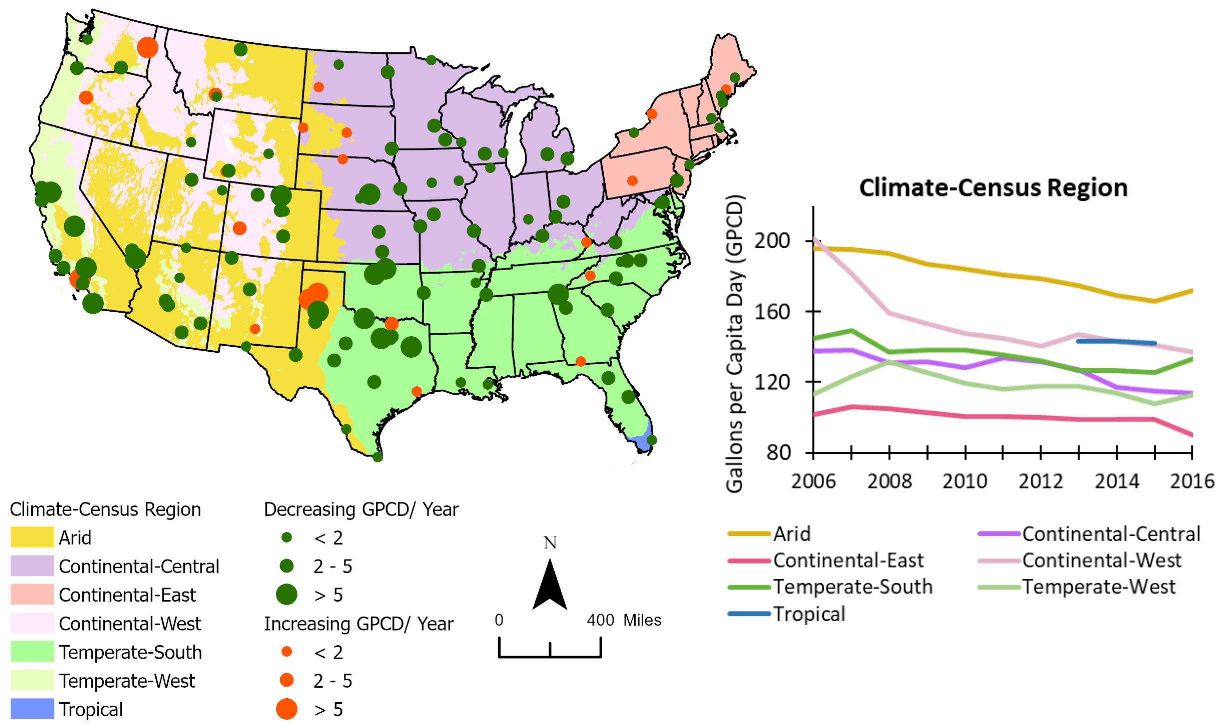


Figure 4.4. (Left panel) Map combining the Koppen-Climate and US Census regions into a new 'Climate-Census Region' classification. Also shows the annual average trends in Total water use, using GPCD metric. **(Right panel)** Shows the variations in Total water use GPCD after aggregating cities by the Climate-Census regions.

4.4.2 Clustering CII/Res ratio

The CII/Res ratio that was estimated for the cities using the water use data and the CII/Res ratio classification tree model in Chapter 2 were used to further cluster the cities and towns based on their water use type. The Partitioning Around Medoids (PAM) method was selected for clustering due to its robustness against outliers and effectiveness in creating meaningful and interpretable clusters. PAM, distinct from other clustering techniques, minimizes the sum of dissimilarities between data points and their respective medoids, making it particularly suitable for the CII/Res ratio characteristics of cities. The determination of the optimal number of clusters was guided by silhouette analysis, which suggested that two clusters were the most appropriate for our dataset (Figure 4.6-Left panel). 0.48 and 1.25 were the two medoids of the CII/Res ratio dataset estimated using the PAM method. These medoids effectively represent different water use patterns in the dataset and were used to distinctly categorize cities into residential and CII water use type dominance (Figure 4.6-Right panel). Of the 126 cities in this study, 89 of them clustered as residential water use dominant and the remaining 37 cities were clustered as CII water use type dominant.

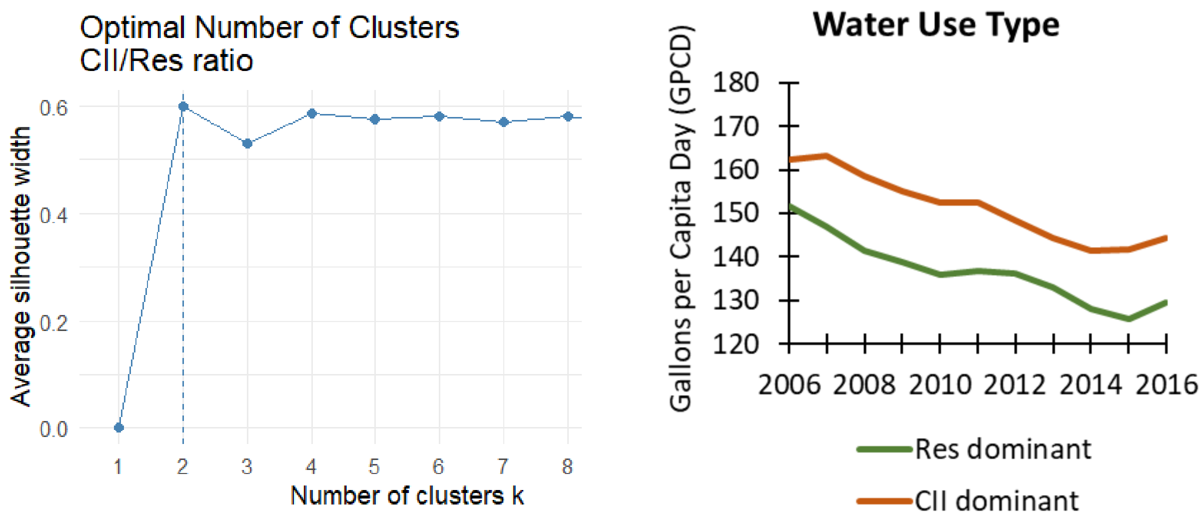


Figure 4.5. (Left panel) Optimal number of clusters identified using PAM method to cluster cities based on their CII/Res ratio. **(Right panel)** Total water use GPCD trends of cities clustered as dominated by Residential water use and CII water use.

Table SI-4.2 in APPENDIX-C presents the classification of all the 126 cities by their Climate-Census and Water Use Type classifications.

4.4.3 Mathematical Framework of Urban Scaling

Scaling theory (Bettencourt, 2013) states that most urban attributes (*Y*) vary continuously with population size(*N*) and are described mathematically using the power-law scaling relationship.

$$Y = Y_O N^{\beta} \tag{4.1}$$

where, Yo is the intercept; and β is a dimensionless scaling exponent that explains the power function relationship between the city attribute and the city population that is studied.

Equation 4.1 in its log transformed format is:

$$\ln(Y) = \ln(Y_o) + \beta * \ln(N) \tag{4.2}$$

Urban scaling theory presents this through a mathematical power function relationship presented in Figure 4.2. Bettencourt (2013) also empirically explains that as city population grows the city attribute can experience one of these three changes:

i) increase drastically in a super-linear manner ($\beta > 1$), such as the case with social quantities, like number of inventions and crime rate,

ii) decreases drastically in a sub-linear trend ($\beta < 1$), such as the urban infrastructure development costs per capita; or

iii) vary linearly in proportion to population size changes ($\beta = 1$).

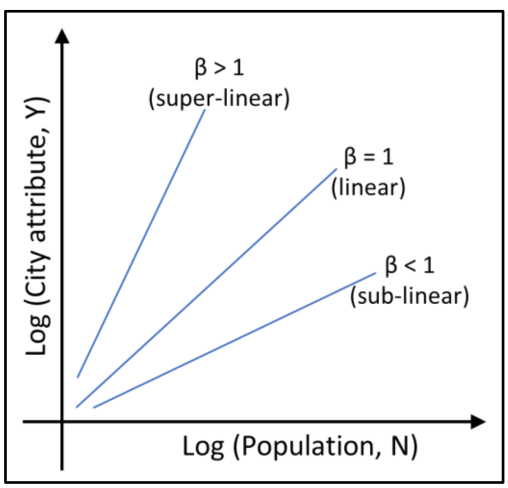


Figure 4.6. Types of scaling relationships between a city attribute and population.

When applying Equation 4.1 to the to the field of water resources management and planning, a general hypothesis is typically considered that the water consumption or the water withdrawn for supply (W) varies linearly with population (U) as shown in Equation 4.3.

$$W = \Phi * U \tag{4.3}$$

where, W is the city attribute (Y) for water consumption; U is the city population (N); Φ is the intercept; and β is 1.

Using the USGS county-level water withdrawal estimates for the U.S., Brown et al. (2013) empirically proposed the following equation to account for changes in Φ using the compound effect of periodic growth in every county(j) for a given year (Y):

$$\Phi_{j,Y} = \Phi_{j,Y-5}\,[(1 + g_{DIV})(1 + d_{DIV})^{Y-Yo}]^5 \tag{4.4}$$

where, Yo is baseline year of water withdrawal data availability; g and d are growth and decay rates respective and specific to the spatial regions (DIV) of the country and are broadly classified as eastern and western part of the U.S.

4.4.4 Modified Urban Scaling Framework for Municipal Water Uses

The evidence of regional variation in water use categories is evident from Chapter 2 and especially from Figure 2.4 and Figure 2.6. Both those figures show that clustering cities by US Census and Koppen-Climate regions have a significant variation on how cities consume water. This introduces the mixed effect component to water use modeling that is based on regional parameters along with the fixed effects to water use variation caused by factors like time and population. This is sort of in similarity to the Equation 4.3 where Brown et al. (2013) show that water use per capita is influenced differentially by the eastern and western U.S. Therefore, to incorporate the effects of regional variation, especially by the US Census and Koppen-Climate regions, Equation 4.1 is appropriately modified to combine the random (Z) effects of city regionalization that vary with time and space as:

$$W_{Z_{j_k}} = \beta_0 * U_i^{\beta_i} * Z_{j_k} \tag{4.5}$$

Where, β_0 is the model intercept incorporating the fixed effects; U_i is fixed effects from population land use characteristics, etc.; β_i is the fixed effects co-efficient; Z_{j_k} is the random effect induced by specific classifications/clusters of cities (*j*), such as Climate-Census regions and Water Use Type, with their *k* unique classifications. Equation 4.5 in its log transformed version is ultimately used in the annual water consumption data fitting process, and it is presented as:

$$\log\left(W_{Z_{j_k}}\right) = \log(\beta_0) + \beta_i * \log(U_i) + \log(Z_{j_k}) \tag{4.6}$$

For the sake of simplicity, all the variables in Table 4.1 that vary over time, including time variables, are classified as fixed effects and the regional variables that vary based on space are classified as random effects in this study.

4.4.5 Model Feature Selection and Parsimonious Model Identification

The process of selecting and finalizing model variables was meticulously conducted using R/RStudio. During this phase, several key statistical metrics were employed to evaluate various combinations of features, ensuring the selection of an optimal model. These metrics included:

1) **Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC):** Both AIC and BIC are crucial for comparing models, with a lower score in either indicating a better balance between model complexity and fit.
2) **Log-Likelihood**: This metric measures the likelihood of the model accurately representing the observed data. A higher log-likelihood value generally signifies a better model fit.

3) **Variable Inflation Factor (VIF):** VIF was used to assess multicollinearity among predictors. A threshold VIF score of less than 2.5 was used to ensure minimal collinearity, thereby enhancing the reliability and interpretability of the model.

The combination of these metrics guided the process of identifying the most suitable model. The goal was to achieve a model with minimal features while ensuring that the AIC, BIC, and log-likelihood values were optimized, indicating an effective balance between model simplicity and explanatory power. This rigorous selection process helped in formulating a statistically robust and parsimonious model, capable of yielding reliable and interpretable results.

After finalizing the model selection, diagnostic tests were conducted to ensure the validity of the model assumptions. These tests included an analysis of normality, homoscedasticity, and randomness of the residuals to verify that the model adequately captured the underlying patterns in the data without systematic biases.

4) **Normality of Residuals**: A Quantile-Quantile (Q-Q) plot was utilized to assess the normality of the residuals. This plot compares the observed distribution of residuals with a theoretical normal distribution. Any significant deviations from the reference line in the Q-Q plot would indicate a departure from normality.

5) **Homoscedasticity**: The homogeneity of variance (homoscedasticity) of the residuals was examined by plotting the residuals against the predicted values. Ideally, this plot should not exhibit any patterns and the residuals must be randomly spread across the plot. This indicates that the variance of the residuals is constant across all levels of the predicted values.

6) Randomness: The Autocorrelation Function (ACF) was employed to check for randomness in the residuals. This involved analyzing the correlation of the residuals with their lagged values. A lack of significant autocorrelations at all lag intervals would suggest that the residuals are random, confirming that the model accounts for the data's inherent patterns.

These diagnostic tests are crucial for validating the reliability and robustness of the mixed-effects models used in the study. They ensure that the models provide an accurate and unbiased representation of the relationships between the variables of interest.

4.4.6 Scaling Estimates from the Finalized Mixed-Effects Models

In the mixed effects model, a critical methodological component was the estimation of water use efficiencies in relation to population growth. This was achieved by incorporating log-transformed population data as a key predictor. The focus was on the coefficient of this log-transformed population variable [$\beta_i * \log(population)$], with particular attention to whether this β_i value was less than 1.

To ascertain the robustness of the scaling estimate of population, the 95% confidence intervals of the population coefficient were closely examined. A coefficient value less than 1, also supported by confidence interval ranges below 1, would indicate increasing water use efficiencies as population increases. This methodology therefore allows one to draw a more precise and reliable conclusion about the nature of the relationship between population growth and water use efficiencies.

4.5 Results

The modeling efforts to fit a parsimonious model to capture the variations in Total, Residential, and CII annual water consumption patterns, incorporating the random effects of climatic, socio-economic, and regional factors, reveal that there are efficiencies of scale being built into the municipal water systems across the 126 study cities and towns within CONUS.

4.5.1 Total Water Use

The final parsimonious mixed effects model for annual Total water consumption is presented in Equation 4.6. Tables 4.2 and 4.3. present the model estimates and their corresponding 95% confidence interval estimates, and the coefficients for the random effects, respectively. The R^2 of the CII water use model with and without the incorporation of random effects are 0.96 and 0.94, respectively. Figure SI-4.1 provides the test of Total water use model residual's normality, homoscedasticity, and randomness to illustrate its robustness and validity.

$$\begin{gathered}\log\left(Total\ Water\ Use_{Z_{j_k}}\right) \;=\; \log(\beta_0) + \;\beta_1 * \log(Population) + \beta_2 * \\ \log(Year) + \beta_3 * \log(Low\ developed\ acres) \;+\; \beta_4 \;* \\ \log(Singlefamily\ to\ Multifamily\ household\ ratio) \;+\; \log\left(Z_{j_k}\right)\end{gathered} \tag{4.6}$$

Table 4.2. Annual Total water consumption model estimates and their corresponding 95% CI.

Variable	**Coefficient**	**Estimate**	**2.5% CI**	**97.5% CI**
Intercept	β_0	10.610	10.225	10.996
log(Population)	β_1	0.887	0.867	0.927
log(Year)	β_2	-0.055	-0.082	-0.027
log(Low developed acres)	β_3	0.104	0.069	0.139
log(Single-family to Multi-family household ratio)	β_4	0.114	0.076	0.152

The model's results reveal significant insights into the scaling effects of population on total annual water use, particularly in the context of water use efficiency. The log-transformed population coefficient, at 0.90992, is notably less than 1. This value, along with its narrow confidence interval (2.5% = 0.883, 97.5% = 0.936), which also remains below 1, indicates that water use is increasing at a rate slower than population growth. This finding suggests that there are inherent efficiencies being realized in water use as the population expands.

Table 4.3. Random-effects intercepts of the annual Total water consumption model.

Z_j	Standard Deviation Explained	k	Classification	Intercept
Climate-Census Regions	0.169	1	Arid	0.206
		2	Continental-Central	-0.157
		3	Continental-East	-0.158
		4	Continental-West	0.065
		5	Temperate-South	-0.109
		6	Temperate-West	-0.062
		7	Tropical	0.214
Water Use Type	0.147	1	CII dominant	0.103
		2	Res dominant	-0.103

4.5.2 Residential Water Use

The final parsimonious mixed effects model for annual Residential water consumption is presented in Equation 4.7. Tables 4.4 and 4.5. present the model estimates and their corresponding 95% confidence interval estimates, and the coefficients for the random effects, respectively. The R^2 of the Residential water use model with and without the incorporation of random effects are 0.96 and 0.92, respectively. Figure SI-4.1 provides the test of Residential water use model residual's normality, homoscedasticity, and randomness to illustrate its robustness and validity.

$$\log\left(Residential\ Water\ Use_{Z_{j_k}}\right) = \log(\beta_0) + \beta_1 * \log(Population) + \beta_2 * \log(Year) + \beta_3 * \log(Singlefamily\ to\ Multifamily\ household\ ratio) + \beta_4 * \log(Low\ developed\ acres) + \log\left(Z_{j_k}\right) \tag{4.7}$$

The model for residential water use mirrors the total water use model in indicating efficiency gains as population increases, with a notable coefficient for log(population). The fact that both the estimate and the confidence intervals are below 1 allows for a more confident assertion that the observed efficiencies in water use are not merely incidental but a consistent trend across the studied locations. It implies that as communities grow, they are likely adopting more efficient water use practices or technologies, leading to a less-than-proportional increase in water consumption relative to population growth.

Table 4.4. Annual Residential water consumption model estimates and their corresponding 95% CI.

Variable	Coefficient	Estimate	2.5% CI	97.5% CI
Intercept	β_0	9.733	9.285	10.180
log(Population)	β_1	0.901	0.865	0.929
log(Year)	β_2	-0.065	-0.094	-0.035
log(Single-family to Multi-family household ratio)	β_3	0.162	0.119	0.205
log(Low developed acres)	β_4	0.100	0.063	0.137

Table 4.5. Random-effects intercepts of the annual Residential water consumption model

Z_j	Standard Deviation Explained	k	Classification	Intercept
Climate-Census Regions	0.216	1	Arid	0.236
		2	Continental-Central	-0.282
		3	Continental-East	-0.175
		4	Continental-West	0.216
		5	Temperate-South	-0.130
		6	Temperate-West	-0.043
		7	Tropical	0.178
Water Use Type	0.173	1	CII dominant	-0.122
		2	Res dominant	0.122

4.5.3 CII Water Use

The final parsimonious mixed effects model for annual CII water consumption is presented in Equation 4.8. Tables 4.6 and 4.7. present the model estimates and their corresponding 95% confidence interval estimates, and the coefficients for the random effects, respectively. The R^2 of the CII water use model with and without the incorporation of random effects are 0.92 and 0.79, respectively. Figure SI-4.3 provides the test of CII water use model residual's normality, homoscedasticity, and randomness to illustrate its robustness and validity.

$$\log\left(CII\ Water\ Use_{Z_{j_k}}\right) = \log(\beta_0) + \beta_1 * \log(Population) + \beta_2 * \log(Percent\ Multifamily\ households) + \beta_3 * \log(Income\ Gini - index) + \beta_4 * \log(Low\ developed\ acres) + \log\left(Z_{j_k}\right) \tag{4.8}$$

Table 4.6. Annual CII water consumption model estimates and their corresponding 95% CI

Variable	Coefficient	Estimate	2.5% CI	97.5% CI
Intercept	β_0	9.026	7.958	10.091
log(Population)	β_1	0.766	0.713	0.820
log(Percent Multifamily households)	β_2	0.330	0.202	0.458
log(Income Gini-index)	β_3	-0.212	-0.641	0.223
log(Low developed acres)	β_4	0.261	0.201	0.322

Table 4.7. Random-effects intercepts of the annual CII water consumption model.

Z_j	Standard Deviation Explained	k	Classification	Intercept
Climate-Census Regions	0.279	1	Arid	0.433
		2	Continental-Central	-0.072
		3	Continental-East	-0.366
		4	Continental-West	0.002
		5	Temperate-South	-0.189
		6	Temperate-West	-0.058
		7	Tropical	0.251
Water Use Type	0.530	1	CII dominant	0.374
		2	Res dominant	-0.374

The model for CII water use presents a slightly different scaling dynamic compared to Total and Residential water uses. The coefficient for 'population' is less than 1, suggesting increasing efficiency, but at a much lower rate. The significant coefficient for 'Percent Multi-Family Households' indicates an additional influence of housing type on CII water use. The coefficients of the random effects highlight variations in water use patterns across different climate and cluster categories, emphasizing the influence of regional and climatic factors on CII water consumption.

4.5.4 Robustness of Models

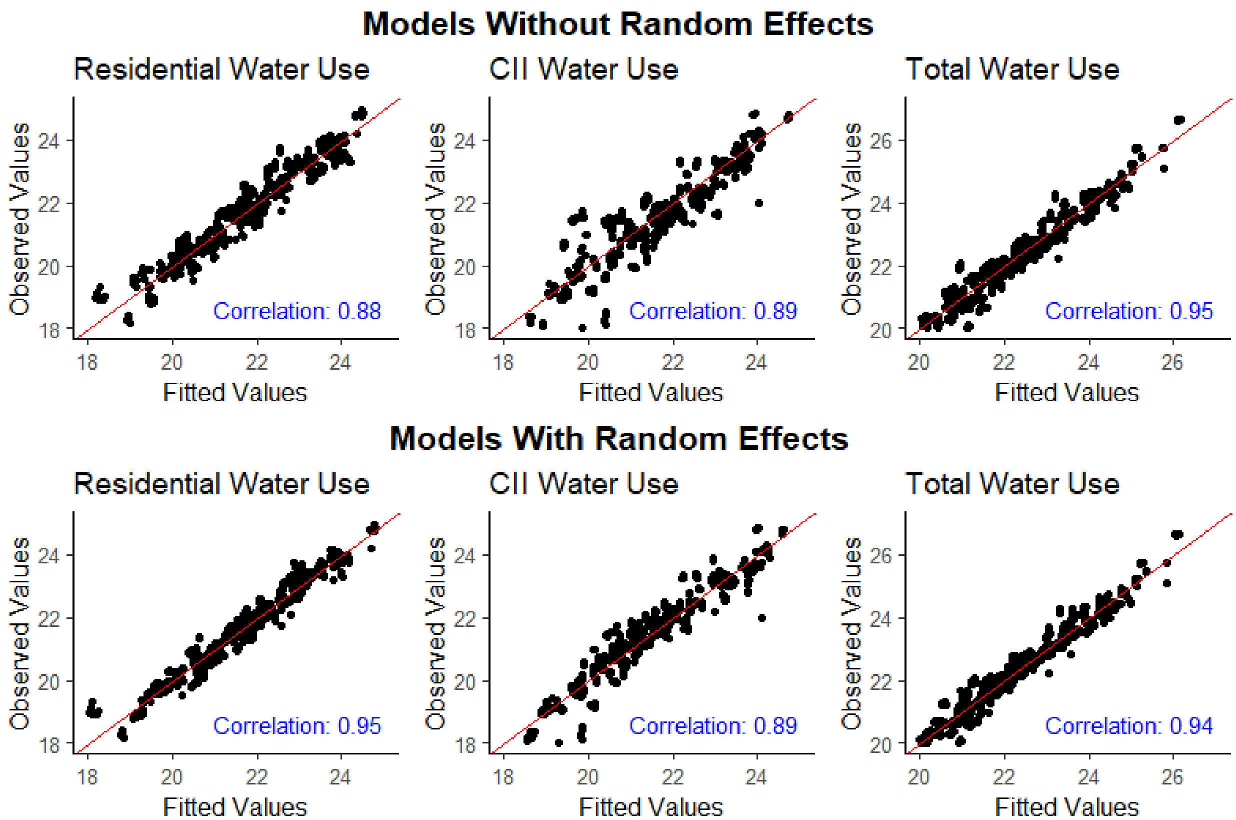


Figure 4.7. Difference in the goodness-of-fit of observed vs fitted values of Total, Residential, and CII annual water consumption estimated using the mixed-effects framework of the urban scaling's econometrics models, with and without the effects of random-effects factors incorporated.

Goodness-of-fit for observed versus fitted values of Total, Residential, and CII water consumption using log transformed mixed-effects models vary significantly based on the inclusion and exclusion of the random-effects variables (Figure 4.7). The top row plots show models without random effects incorporated and the bottom row plots includes random effects in the models. Overall, this figure highlights the improvement in correlation upon the inclusion of random effects, especially for Residential and CII water use. This implies that the inclusion of random effects in the models captures the variability in water use more effectively, thereby providing a better fit to the observed data. This enhancement in model fit supports the conclusions drawn from the model outputs, emphasizing the importance of accounting for both fixed and random effects in modeling complex water use patterns.

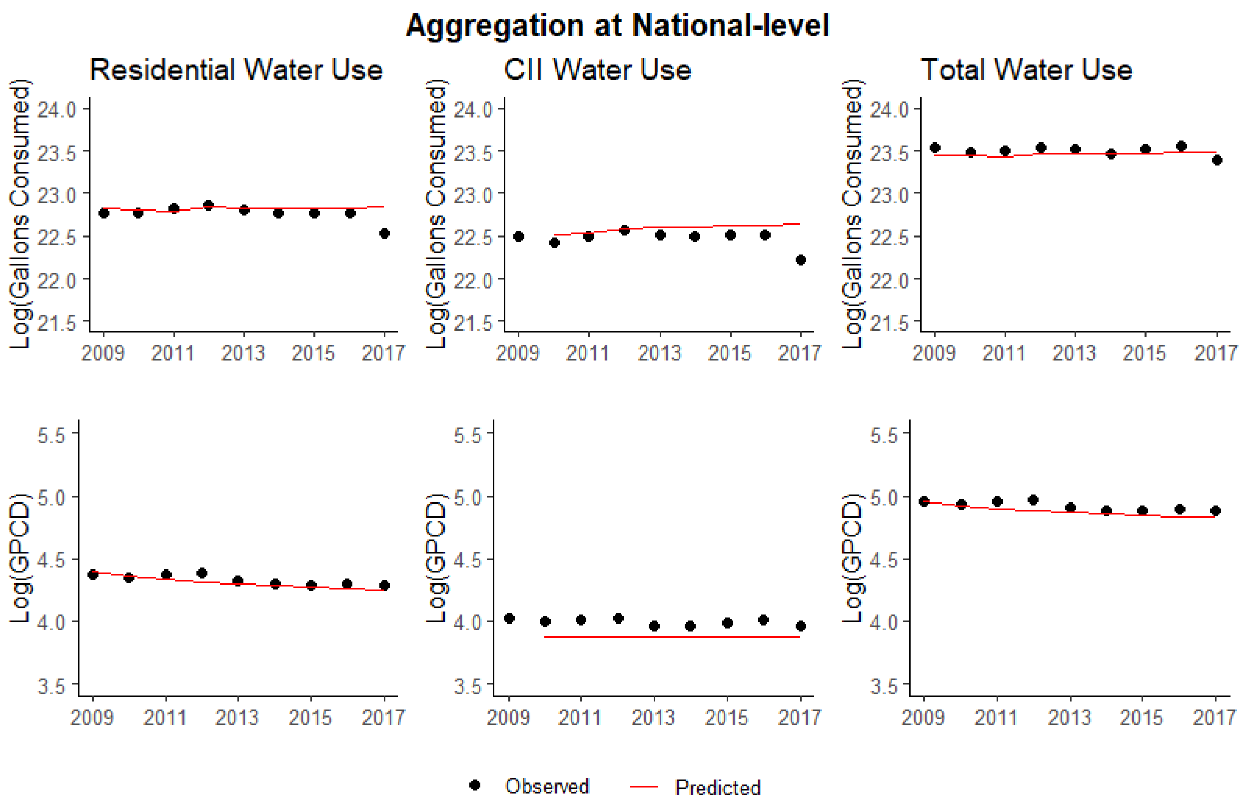


Figure 4.8. Subplots showing the time-series style goodness-of-fit of the mixed-effects models on the annual municipal water consumption patterns of cities aggregated at the national scale. The top row presents the water consumption patterns using the total gallons consumed metric and the bottom row using the GPCD metric.

When the log transformed mixed-effects model is fitted over the annual municipal water use data in a time-series, there is an overall ‘good’ fit; the modeled annual water consumption patterns pretty closely capture the observed trend in the consumption patterns across the water use categories (Figure 4.8).

Similar time-series fits of the mixed-effects model on the annual water consumption patterns for cities aggregated by the random regional effects of Climate-Census regions and Water Use type distinction are presented in Figures SI-4.3 and 4.4 of APPENDIX-C. With the incorporation of time and random effects, the mixed-effects models developed certainly capture the trend and range of water consumption patterns, but still underperform in the accuracy of their predictions. This leads to the hypothesis that the complexities of the municipal water use patterns cannot be fully captured by these linear models, and henceforth efforts must be undertaken to explore more sophisticated modeling approaches such as the neural network models and machine learning tools.

4.6 Conclusion

In the analysis of water use patterns across 126 cities and towns within CONUS, distinct dynamics were observed in Total, Residential, and CII water use. Each model incorporated log-transformed population, year, and other relevant variables as fixed effects, while accounting for the random effects of climate and regional clusters.

For Total water use, the model highlighted a significant relationship with population, year, low developed acres, and single-family/multifamily households' ratio (SF/MF ratio). The random effects of clustering cities by the Climate-Census regions and their Water Use Type revealed variability based on climatic and regional classifications. Residential water use followed a similar trend, with a notable impact of the SF/MF ratio on water use, indicating differences in consumption patterns based on housing types. CII water use, on the other hand, showed a slightly different pattern. While population still played a crucial role, the effect of the percent multifamily households and income Gini-index was more pronounced, suggesting the influence of socio-economic factors and housing structure on CII water consumption.

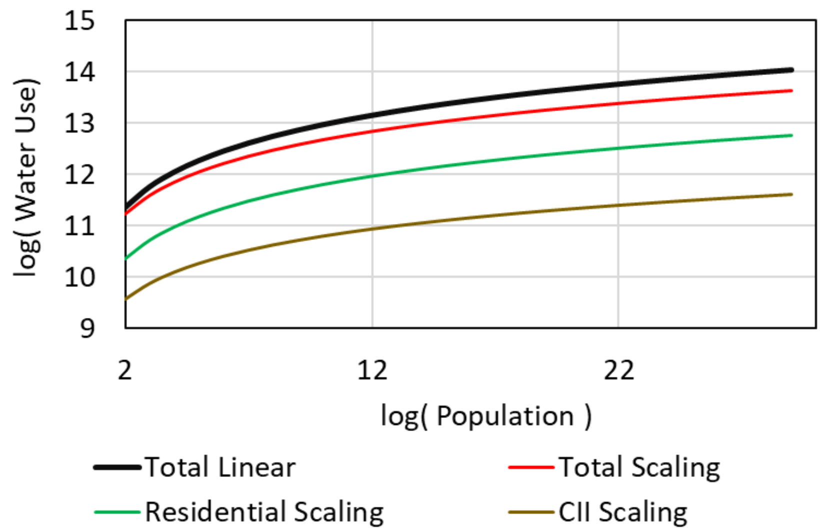


Figure 4.9. Comparison of the water use relationships explained between municipal water consumption and population by the linear model and the econometrics model with scaling relationships within CONUS for the period 2005 to 2017.

The scaling of water use with population was a key finding across all models. For Total and Residential water use, the coefficients for population were below 1 but close to it, indicating near-proportional increases in water use with population growth (Figure 4.9). This suggests that, although water consumption is rising with population, there are efficiencies being realized, possibly due to advancements in water-saving technologies or effective water management practices. The precision of these estimates, supported by the tight confidence intervals, adds a layer of reliability to this interpretation.

In contrast, the CII water use model presented a lower coefficient for population, indicating a less than proportional relationship between population growth and water use (Figure 4.9). This could imply that CII water use is more sensitive to factors other than population, such as economic activities, industrial practices, and regional policies. The presence of varying coefficients across different water use types underscores the importance of tailored strategies for managing different water use sectors.

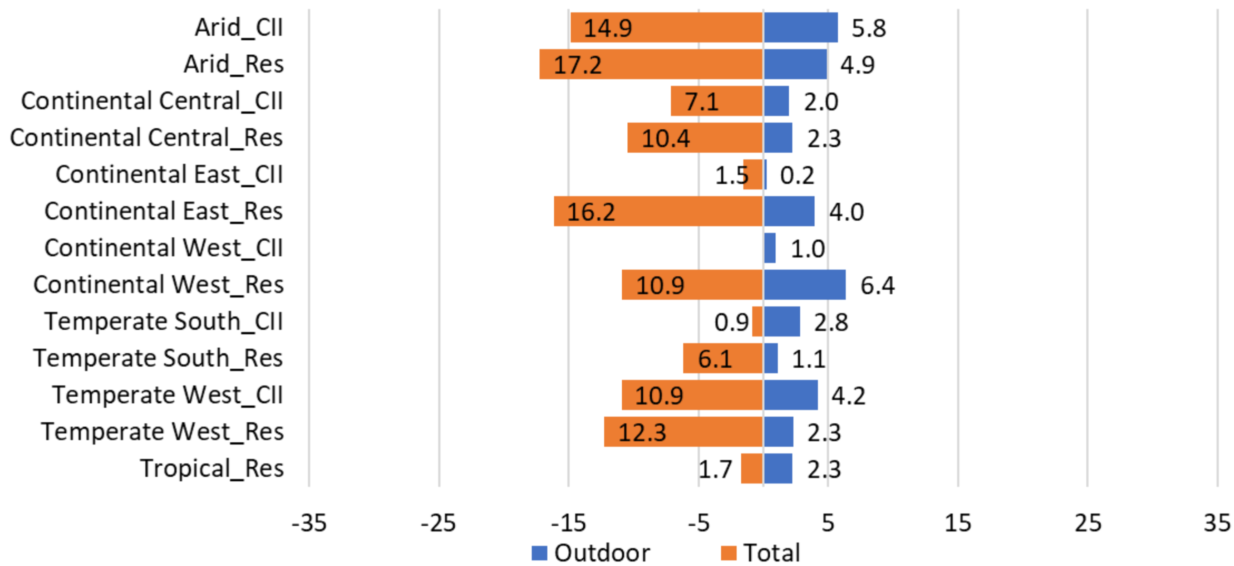


Figure 4.10. Magnitudes of difference in the annual estimation of Total and Outdoor water uses in GPCD by the linear water use model and the econometrics water use model with scaling relationships incorporated. The scaling models uniquely captures the water use variations by the climate and water use dominance type of a city.

Further investigation on the magnitude difference in the estimation of water use in GPCD by the linear and the econometrics scaling model yielded another way of estimating the unique variations in water use that is influenced by the climate region and the water use type dominance of a city. The econometrics scaling model well captures the water use patterns and trends across CONUS by uniquely estimating water use by the climate and water dominance type categorization of cities (Figure 4.10). Linear models comparatively overestimate the water needs as population growth continues and this can be detrimental to sustainable municipal water supply planning amidst water shortages and climate variation impacts. Therefore, effective identification of scaling relationships between water uses and population influenced by other climatic, socio-economic and built environment variables, as incorporated into the urban scaling theory equation, are crucial for long-term sustainable water resources management.

The application of urban scaling theory on municipal water consumption holds true as urban population grows. However, the nuances of the climate region type, and water use type dominance of a city further adds another dimension of influence on the scaling trends thereby proving that simple linear way of water use estimation is vague and must therefore incorporate the influence of factors like climate, socioeconomics, etc., for more reliable and accurate estimates of water consumption (Figure 4.10).

Overall, the analysis underscores the complexities in water use dynamics, influenced by a blend of demographic, socio-economic, and environmental factors. While there are clear indications of efficiencies of scale in water use, these efficiencies vary by type of water use. The results highlight the need for nuanced water management strategies that consider these different dynamics and the specific factors driving water consumption in various sectors. This understanding is crucial for developing effective policies and practices for sustainable water management and urban planning in rapidly evolving urban landscapes.

This insight is crucial for policymakers and urban planners who must consider not only the direct impact of population growth on water demand but also the mitigating effects of improved water use efficiencies. These findings underscore the importance of continued investment in water-saving technologies and practices, especially in urban areas experiencing rapid population growth.

CHAPTER 5 – SOPHISTICATED MODELING OF MUNICIPAL WATER CONSUMPTION PATTERNS USING MACHINE LEARNING TOOLKIT

5.1 Introduction

The management and projection of water supply and demand have been fundamental concerns in the United States for several decades. Historically, studies in this domain have evolved through various methodologies and scopes, reflecting the changing needs, and understanding of water use dynamics. Initial projections, dating back to the 1960s, were largely based on straightforward extrapolations of demographic trends. These early efforts laid the groundwork for more nuanced approaches that consider a range of drivers influencing water use, including socio-economic conditions, technological advancements, and environmental policies.

Since the pioneering work of Wollman (1960), a series of landmark studies have sought to refine the accuracy and relevance of water withdrawal projections. The Water Resources Council (1978), followed by research from Osborn (1986) and Waggoner (1990), progressively incorporated broader sets of variables and more sophisticated modeling techniques. However, it was not until the comprehensive analysis by Brown et al. (2013) that the projections began to adopt a more integrative approach, linking water use to a wider array of socio-economic factors for a variety of economic sectors that rely on water resources for their needs.

Building on this foundation, Warziniack et al. (2022) provided a set of projections that not only extended the temporal scope but also increased the spatial resolution to the county level. This shift towards finer granularity in data represented a significant step forward, allowing for more tailored and actionable insights for local water managers and policymakers.

Despite these advancements, there remains a gap between the methodologies used for projection and the emergent complexities of water use patterns. The mixed-effects models previously employed have revealed the influence of numerous variables beyond population and income, including climate, socio-economic, and urban-geologic factors. These factors have been shown to interact in intricate ways, influencing municipal water systems in both expected and unforeseen manners.

The limitations of the existing linear and non-linear models have become increasingly apparent. While they offer valuable perspectives, the complex and multi-faceted relationships that are inherent to hydrology and water consumption data are poorly captured by traditional models. The evolving realities of climate change, urbanization, and technological change necessitates a more robust and flexible modeling approach capable of adapting to and learning from the data (Zounemat-Kermani et al., 2021).

This research chapter, therefore, aims to bridge this methodological gap by introducing machine learning (ML) techniques into the modeling and projections of water use. ML's ability to learn from data and identify complex patterns makes it particularly suited for tackling the challenges posed by the dynamic nature of water use and its drivers (Jain et al., 2001; Jain and Ormsbee, 2002; Bougadis et al., 2005; Ghiassi et al., 2008; Adamowski et al., 2012; Vijai and Sivakumar 2018; Bata et al., 2020).

Machine learning, a subset of artificial intelligence, has revolutionized many fields of study by providing tools to decipher patterns from large datasets. In water resource management, ML techniques offer the promise of capturing the interactions between numerous variables that affect water use, from the macro-scale impacts of climate change to micro-scale behavioral changes in water consumption. Moreover, the emphasis on ML does not negate the value of the findings from mixed-effects modeling. Instead, it builds upon these insights, incorporating the identified fixed and random effects into a more sophisticated analytical framework. This integrative approach ensures that the strengths of traditional modeling are preserved while the advantages of modern ML techniques are fully harnessed.

Research using big data analytics that use machine learning and artificial intelligence tools are on the rise and are being employed in a variety fields ranging from weather forecasting (Camporeale, 2019; Scher & Messori, 2018), ecohydrological and water and energy demand modeling and forecasting (Quilty et al.,2019; Karandish & Šimůnek, 2016; Obringer & Nateghi, 2018; Obringer et al., 2020; Peters et al., 2007;Villarin & Rodriguez-Galiano, 2019; Wang & Srinivasan, 2015) to consumer level behavior modeling (Chaudhary et al., 2021;Valecha et al., 2018) for product marketing. Luukkonen et al. (2023) recently published their public supply water use reanalysis for the 2000-2020 period by HUC12, month, and year for the conterminous United States using a gradient boosting ML algorithm. These research efforts conducted using dynamic, sophisticated analytical tools have been well received and certainly point the direction forward for the application of such ML tools into a broader spectrum of economic and resource management sectors.

The goal of this work is not only to provide more accurate projections of water use but also to deepen our understanding of the factors that drive water demand. This, in turn, can inform the development of more sustainable water management practices and policies, essential for addressing the challenges of the 21st century. By leveraging the power of cutting-edge ML techniques and tools, this study aims to transform the way water use is modeled and projected - moving from traditional linear models to a more dynamic, responsive, and integrative approach.

5.2 Objectives:

The overarching aim of this research is to synthesize the myriad factors influencing municipal water uses from the previous chapters into a cohesive, predictive model that can serve as a tool for effective water resource management. The specific objectives of this chapter are:

1. To assimilate and integrate the insights gained from the mixed-effects models of previous chapters, to incorporated into the new modeling framework.

2. To develop and test an ensemble of Machine Learning models to capture the variations in the monthly municipal water consumption patterns influenced by the previously identified subset of climatic, socio-economic, built environment, and regional factors.

5.3 Data and Study Regions

Municipal water use data from Chinnasamy et al. (2021) at a monthly time-step, along with the compilation of all the climatic, socio-economic, built environment, and regional factors collected (Table 5.1) for study in Chapters 2 and 4 of this report will be used collectively for building a robust and a sophisticated model for municipal water consumption within CONUS.

Table 5.1. Explanatory factors used in this study.

Category	Groups of variables
Temporal	Month, season, year
Climate	Max temperate, max vapor pressure deficit, cumulative precipitation
Built environment	City area/ water service area, NLCD developed land cover area (open, low, medium, high, percent imperviousness), percentage of single-family and multi-family houses, ratio of single-family to multifamily households, median year household built
Socio-economic	Population, population density, number of employees and number of establishments in different NAICS sectors, median household income, income Gini-index, percent population with high/ some education
Regional	Latitude, Longitude, elevation, Climate-Census regions, Water Use Type

5.4 Selection of Machine Learning Tools

This research specifically explores various statistical regression models in the Python programming software. A variety of complex, machine learning based regression models were employed in this study. These machine learning based regressors included Random Forest Regressor (RF), Extra Trees Regressor (ET), Gradient Boosting Regressor (GBR), Light Gradient Boosting Machine (LightGBM), K Neighbors Regressor (KNN), and Adaptive Boosting Regressor (AdaBoost).

Zounemat-Kermani et al. (2021) provide an elaborate explanation of each of the above-mentioned machine learning regression tool, their previous applications in the field of hydrology and water resources planning and management, and present a compelling case of the adoption of these tools in future work. This study thereby has selected the above-mentioned machine learning tools based on their previous application in the field of water resources management and specifically applies those tools to model municipal water consumption patterns.

5.5 Modeling Framework

The modeling efforts are conducted on the monthly timestep of the aggregated municipal-level water use data from 2005 to 2017 for those 126 cities and towns within CONUS that are studied in this research work. An ensemble of machine learning tools was applied on this dataset, using the Python programming software, to create a robust modeling portfolio for the municipal water systems. Total monthly water consumption is attempted to be modeled in this chapter. Figure 5.1 shows the flowchart of the process of modeling the water consumption data using the ensemble ML algorithms.

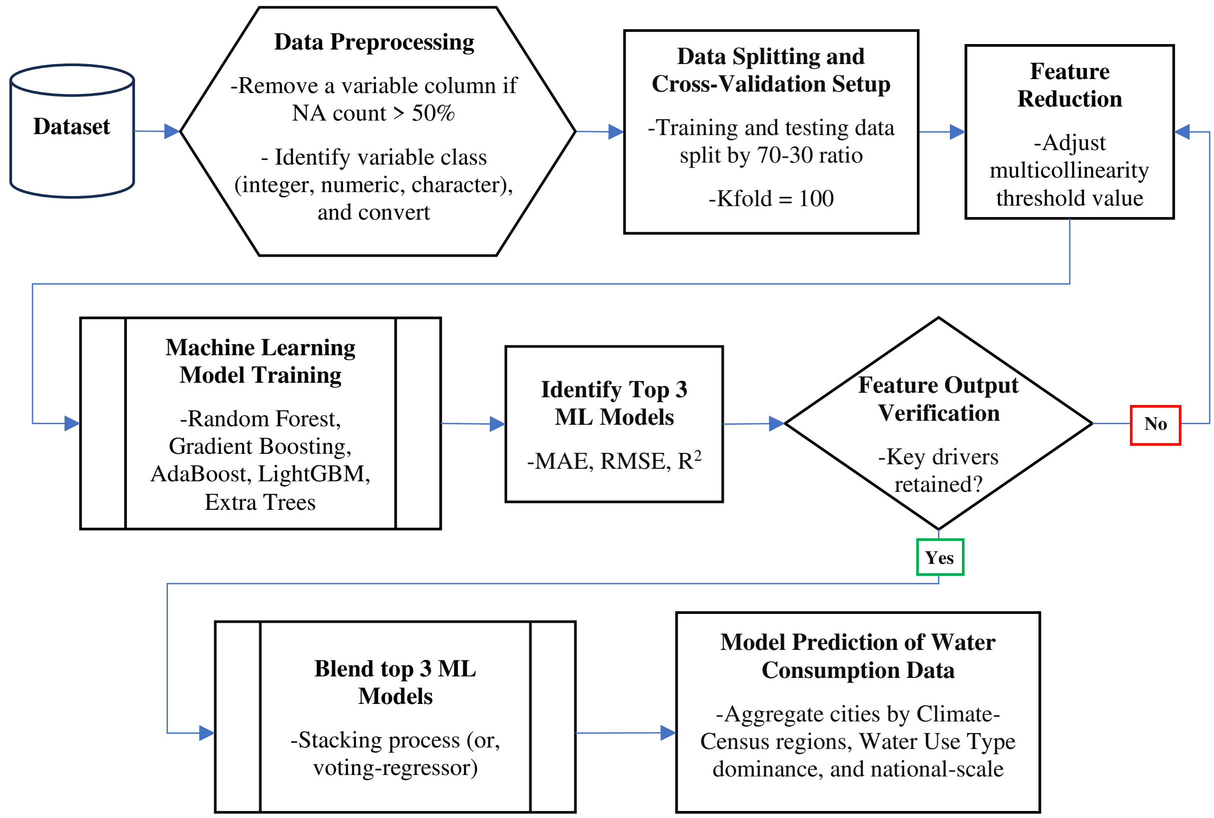


Figure 5.1. Flowchart of water consumption modeling using the ensemble ML algorithms.

5.5.1 Consolidation of Water Use Drivers

In conducting this machine learning approach to model municipal water consumption, the result from the municipal water uses characterization study (Chapter-2), along with the learnings from single-family residential water uses in Arizona (Chapter-3) and the results from the mixed-effects urban scaling (Chapter-4) were all consolidated for effective modeling of water consumption patterns. Table 5.1 presents all the features used in the modeling efforts conducted in this chapter. Special attention is given to the key drivers or important features identified in Chapter 2 for the CII/Res ratio classification tree described in Figure 2.9 and to the variables identified in the mixed-effects econometrics modeling of annual municipal water uses in

Section 4.5 of Chapter 4. These identified key drivers were ensured to be used, even after the feature reduction process, to build a parsimonious ensemble of ML models.

5.5.2 Data Preprocessing

The important step in any modeling pursuit is the preprocess stage where different strategies are undertaken to handle categorical variables, multicollinearity amongst the variables, and the handling of missing values. Variables with excessive missing values (>50% of entries) were removed from the analysis and those with some missing values are filled with the mean value of that variable. The final dataset provided to the machine learning tools consisted of 16,305 monthly total water use observations from 126 cities with 93 climatic, socio-economic, built environment, and regional variables or features.

5.5.3 Training - Testing Data Split and Cross-Validation Setup

Before the model building process, the entire dataset was split into training and testing datasets following a 70-30 split ratio. The training dataset is further split into '*K*' number of folds (or subsets). In each iteration of the cross-validation process, one of these folds is used as the validation set, and the remaining folds collectively serve as the training set for the model. This process is repeated '*K*' times, each time with a different fold serving as the validation set. This allows every observation in the original training dataset to be used both for training and validation at some point. In this study '*K*' folds parameter was set to 100.

5.5.4 Multicollinearity Detection and Feature Reduction

Feature reduction in the dataset was conducted by the PyCaret[3] package in Python using its inbuilt function. Multicollinearity threshold function is used to detect and remove highly correlated features during the model setup process. The default value for this parameter is 0.9. This means that if two features have a correlation greater than 0.9, one of those features will be removed to avoid multicollinearity issues. This helps prevent models from being affected by multicollinearity which can cause overfitting and instability. This value can be tuned as per the dataset requirements. Setting this threshold to a smaller value will be more aggressive in removing correlated features.

The process of identifying this multicollinearity threshold was iterative and manually conducted to ensure that variables identified in Figure 2.9 and Section 4.5 are retained. Strong subject knowledge is essential to discern the features that were important to be retained in the final model. The key drivers identified in Chapters 2-4 were crucial in deciding the appropriate multicollinearity threshold value so that less significant and multicollinear variables were removed in the final modeling process.

5.5.5 Model Building and Evaluation

The ML models mentioned in Section 4.5 are individually modeled using the training dataset comprising 70% of data entries from the original dataset. The training data is further split into *k* folds (here, 100), using stratified splitting to maintain equal distribution of target variable in each fold. For each ML model, *k* versions are training, each version using a different fold for validation and remaining *k-1* folds for training. This results in *k* performance scores per model

[3] https://pycaret.gitbook.io/docs/

based on out-of-fold predictions. For each ML model's k^{th} version, cross-validation performance metrics are evaluated. The iteration with the best average cross-validation score is selected as the best version for that particular ML model.

Model coefficient of determination (R^2), mean absolute error (MAE), mean standard error (MSE), and root mean squared error (RMSE) are the cross-validation performance metrics used in evaluating the ML model iterations. The best performing model is selected by minimized MAE and maximizing the R^2 metrics. The best model is retrained on full training data and tested on the unseen test set to get an unbiased estimate of its generalization performance on new data.

5.5.6 Model Blending or Stacking

In the final step, to prevent model biases and overfitting/ underfitting issues, the top '*n*' best performing models were blended or stacked statistical using the inbuilt *blend_models()* function of PyCaret package. The blended model combines predictions from the best 'n' models to improve the overall predictive performance of the stacked model. The model blending process was done with equal weight assigned to each of top *n* model, with the aim of leveraging the diverse strengths of individual models.

This streamlined and robust way of implementing machine learning tools on the monthly municipal water consumption dataset enables a rapid fitting and comparison of multiple ML algorithms to identify the top performing models for predicting the Total monthly water.

5.6 Results

The results of the ML algorithms or tools used to model the monthly Total water consumption across the 126 cities from 2005 to 2017 are presented in the order of their performance based on MAE, MSE, RMSE, and R^2 in Table 5.2. A multicollinearity threshold value of 0.4 was identified (for this particular dataset) through iterative process to distinctively subset the consolidated key drivers from Chapters 2-4 and retain them for the final model building process.

Table 5.2. Performance metrics of the ensemble ML models' fit on the Total monthly water consumption in the testing dataset.

Model	**MAE**	**MSE**	**RMSE**	**R^2**
Extra Trees Regressor (ET)	8.93E+07	4.74E+16	2.17E+08	0.996
Random Forest Regressor (RF)	9.95E+07	5.61E+16	2.36E+08	0.995
Light Gradient Boosting Machine (LightGBM)	1.12E+08	5.98E+16	2.41E+08	0.995
Gradient Boosting Regressor (GBR)	1.67E+08	1.02E+17	3.19E+08	0.991
K Neighbors Regressor (KNN)	1.97E+08	2.29E+17	4.78E+08	0.980
Adaptive Boosting Regressor (AdaBoost)	7.67E+08	7.58E+17	8.68E+08	0.934

The top three (*n*) ML models that emerged to be the best fits on the training and testing datasets were the Extra Trees Regressor (ET), Random Forest Regressor (RF), and Light Gradient Boosting Machine (LightGBM) models. Their model architectures are presented in Figures SI-5.1, 5.2 and 5.3 in APPENDIX-D. These models outperformed others significantly in key performance metrics (Table 5.2).

Among the top 3 ML models, ET Regressor achieved the highest R^2 score of 0.996, signifying an almost perfect prediction capability. It also maintained a low MAE and RMSE, indicating high precision and accuracy in the predictions with minimal overfitting. LightGBM model showcased a strong balance between speed and accuracy with the second highest R^2 score of 0.995. It had an appreciably low MAE and RMSE, indicating its robustness in dealing with various magnitudes of data. RF regressor also demonstrated high accuracy with an R^2 score of 0.995, closely following the ET Regressor.

To capture and employ all the unique characteristics of those three best performing ML models (ET, RF, and LightGBM), a “Blended” or “Stacked” or “Voting Regressor” model was built by assigning equal weights to each of them (Figure SI-5.4). The decision to blend these three models was based on their complementary strengths: the ET Regressor's precision, the LightGBM's efficiency and performance with errors, and the RF's balance between accuracy and complexity.

5.6.1 Ensemble ML Model Comparison and Validation

The cross-validation performance of the ensemble machine learning (ML) models – ET, RF, LightGBM, GBR, and AdaBoost regressors, during training and testing stages is presented separately in Figure 5.2. The model performance across the 100-fold cross-validation was measured using the coefficient of determination (R^2) metric. The training and testing cross-validation results are presented as violin plots showing the relative range of the model’s performance and also as individual dots representing the individual performance of k^{th} iteration of the ensemble ML model. The relative performance of the ensemble ML models are also visually compared with the “Blended” or “Stacked” or “Voting Regressor” model. The median value for each of those ML models’ performance is highlighted as a triangle (Figure 5.2).

During the training phase, the median cross-validation R^2 values for almost all the ensemble models were above 0.975 and with minimum variation, except for the AdaBoost model which had lower median performance metric with high variation (Table 5.2 and Figure 5.2). The median performance metric of the ET, RF, and LightGBM regressors are the highest compared to GBR, KNN and AdaBoost, and therefore were chosen to be stacked into a “Blended” model.

The performance of the Blended model during training is also better than the omitted ensemble ML models (Figure 5.2).

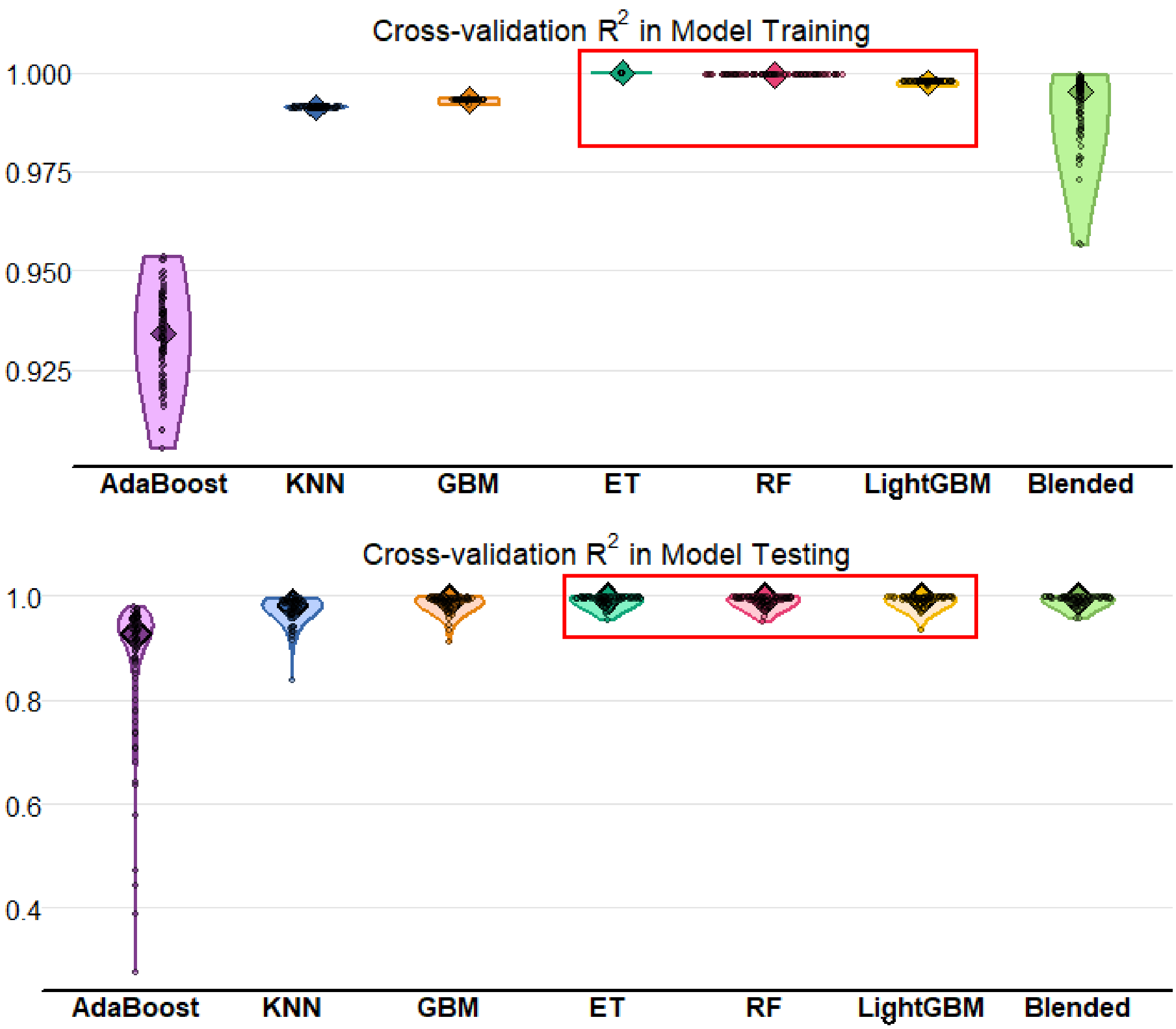


Figure 5.2. Cross-validation performance of the ensemble ML model's performance across the training (above) and testing (below) datasets.

Similarly, during the testing phase, the median and range of cross-validation R^2 values of the ET, RF, and LightGBM regressors, including the Blended model, outperform the other ML models. The distribution of plot of model residuals of the top 3 models during training and

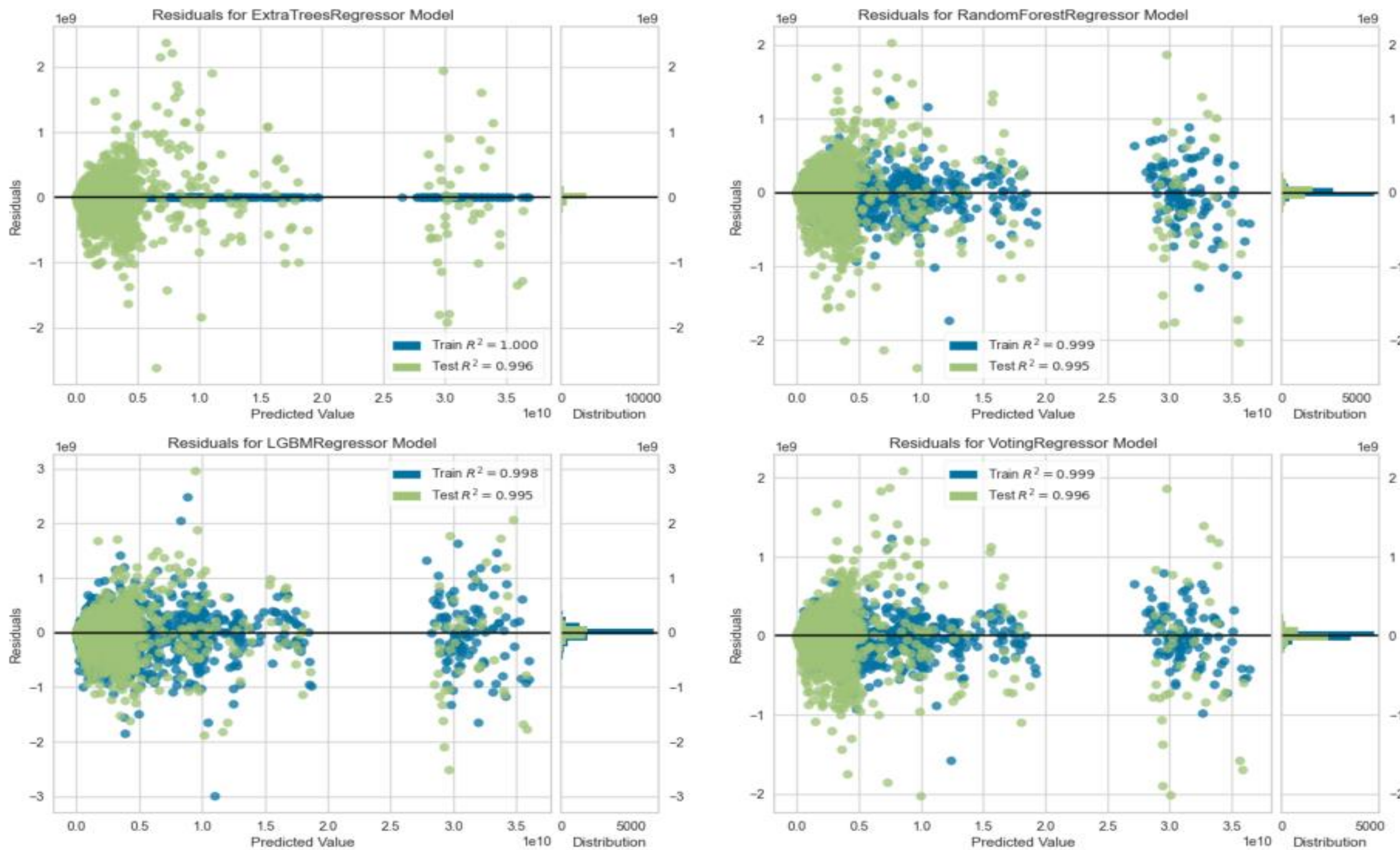


Figure 5.3. Model residual subplots of training and testing datasets from the 3 best performing ML algorithms with the blended model's.

testing are normally distributed and are randomly scattered around zero, indicating very accurate predictions and no clear signs of overfitting (Figure 5.3). Also, there is a very small difference between the R^2 value of training and testing datasets of these models, indicating that the models fit the unseen datasets well (Figures 5.2 and 5.3).

This model validation process emphasizes the importance of evaluating model performance through cross-validation techniques through visual comparison of the ensemble ML model's performance across the training and testing datasets. The results validate the importance of stacking or blending the top performing models to eliminate overfitting, bias and to capture the uniqueness of those individual ML models.

5.6.2 Important Model Features

The feature importance analysis (Figure 5.4, and Figures SI-5.5 and 5.6 in APPENDIX-D) of the top performing models revealed that 'population' was the most significant predictor across all models, highlighting the direct correlation between the size of a city's population and its water consumption. The 'SF.MF.ratio' followed in importance, suggesting that the residential composition of cities also plays a crucial role in determining water usage patterns. Climatic factor, represented by 'MaxVPD.hPa' or maximum vapor pressure deficit, is also consistently ranked as top predictors reflecting the importance of environmental conditions on water needs. Another housing related variable, 'Median year built' or median year the household built, offers further insight as the measure of efficiency in water consumption.

Differences in feature prioritization also was evident among these top 3 models, with 'Latitude' being a prime example. It held considerable sway in the RF model, indicating a sensitivity to geographical or climate variances that influenced water consumption patterns. The temporal variables 'Month' was notably important for the LightGBM and RF models, capturing the seasonal oscillations in water consumption. ‘Median_income' or median household income, was a salient feature for the ET and LightGBM models, highlighting the economic dimensions of water usage.

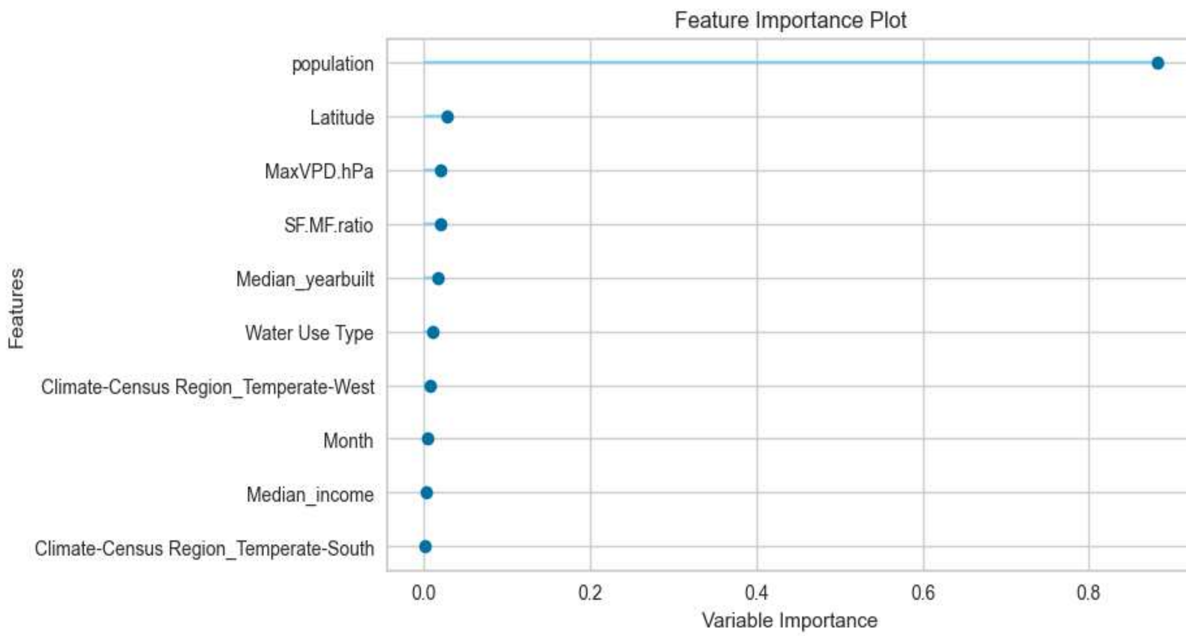


Figure 5. 4. Feature importance plot of the Extra Trees Regressor model.

The 'Climate-Census’ regions too exhibited a pronounced effect in the LightGBM model. 'Precipitation.in' variable stood out in the ET model, which may be due to its ability to consider nonlinear and complex interactions within its multitude of decision trees. Interestingly, the ‘Water Use Type’ categorical feature showed variable importance across models. In the ET and RF models, this feature was of moderate significance, implying that the water usage patterns might be clustered in ways that are not purely demographic or climatic but also behavioral or infrastructural. This feature was less prominent in the LightGBM model, which may suggest that the algorithm's approach to handling categorical features and its gradient-based optimization process gives it a different sensitivity to the nuances captured by Res and CII dominant classifications.

5.6.3 ML Model Fitting of Total Monthly Water Consumption

The following Figures 5.5, 5.6, and 5.7 provide the time series fit of Total water consumption aggregated by average for the whole nation, by Climate-Census regions and by the Water Use Type, respectively. Comparing these figures with the log transformed mixed effects model fit in Figure 4.8 and Figures SI-4.4 and 4.5 in APPENIX-C, it becomes very evident and conclusive that the complexities and variation in municipal water consumption patterns are best captured by the machine learning models.

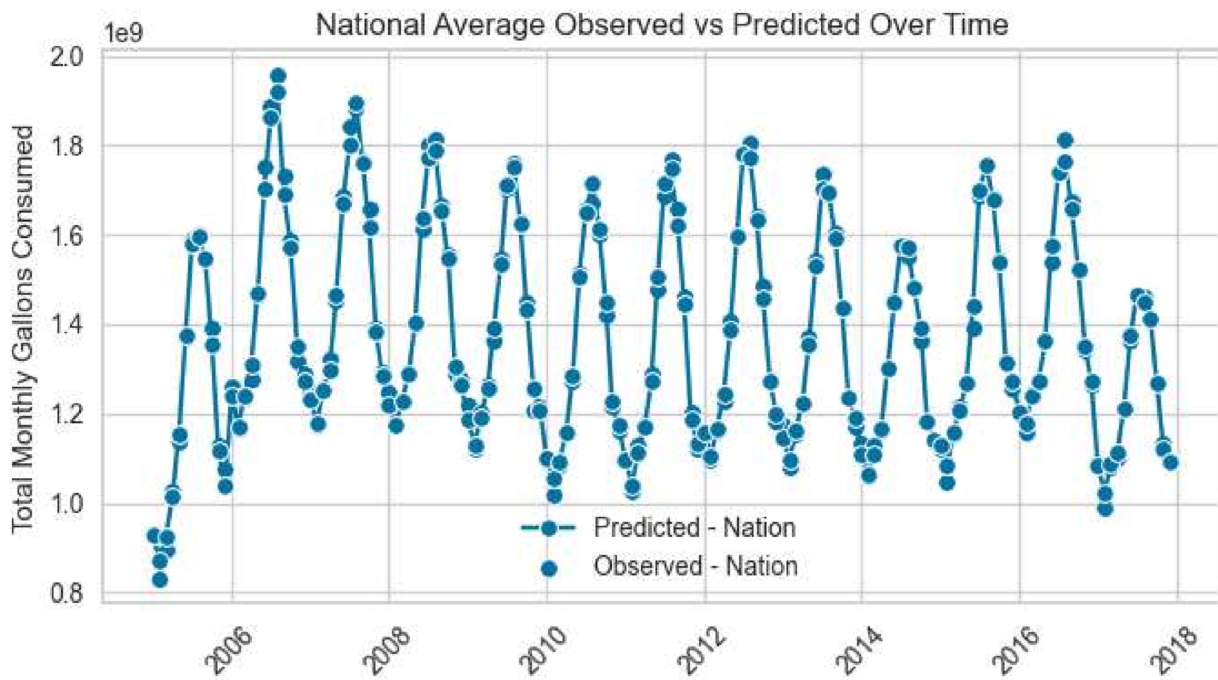


Figure 5.5. Blended model fit on Total monthly water consumption of 126 cities within CONUS aggregated into a national scale average.

Although the blended ML model fits the Total water consumption patterns very well, it must be approached with a lot of care and speculation. With ML models there are many cautionary tales about the overtraining of the model to the data set used to train the models leading to a bias in the model. Bias arises from the model overfitting the noise within the data rather than capturing the general pattern. This leads to very poor model performance when dataset outside the training data is fitted. So, in the case of forecasting future water consumption, the model may seem to make very accurate predictions within the training data's timespan but may completely miss longer-term trends or changes in usage.

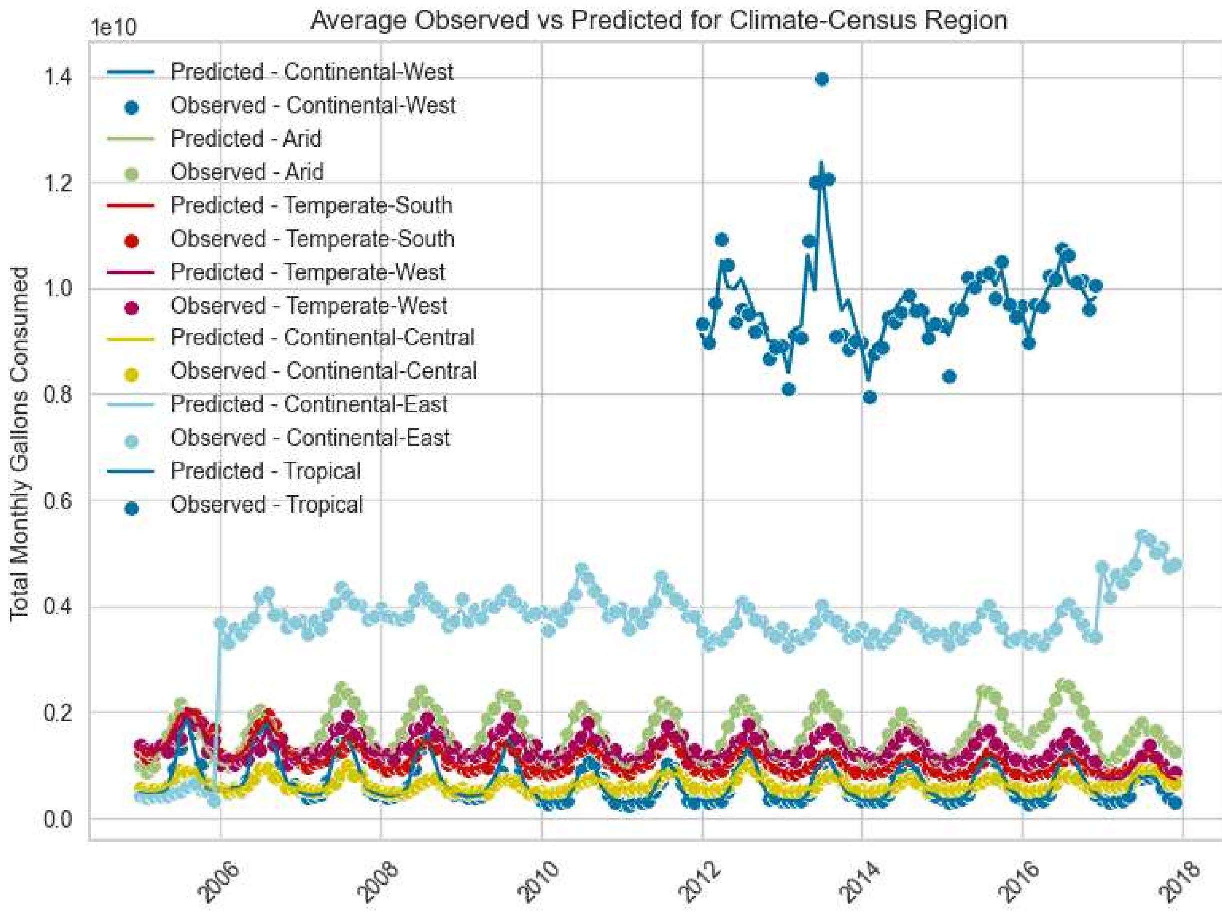


Figure 5.6. Blended model fit on Total monthly water consumption of 126 cities within CONUS aggregated into average based on Climate-Census classifications.

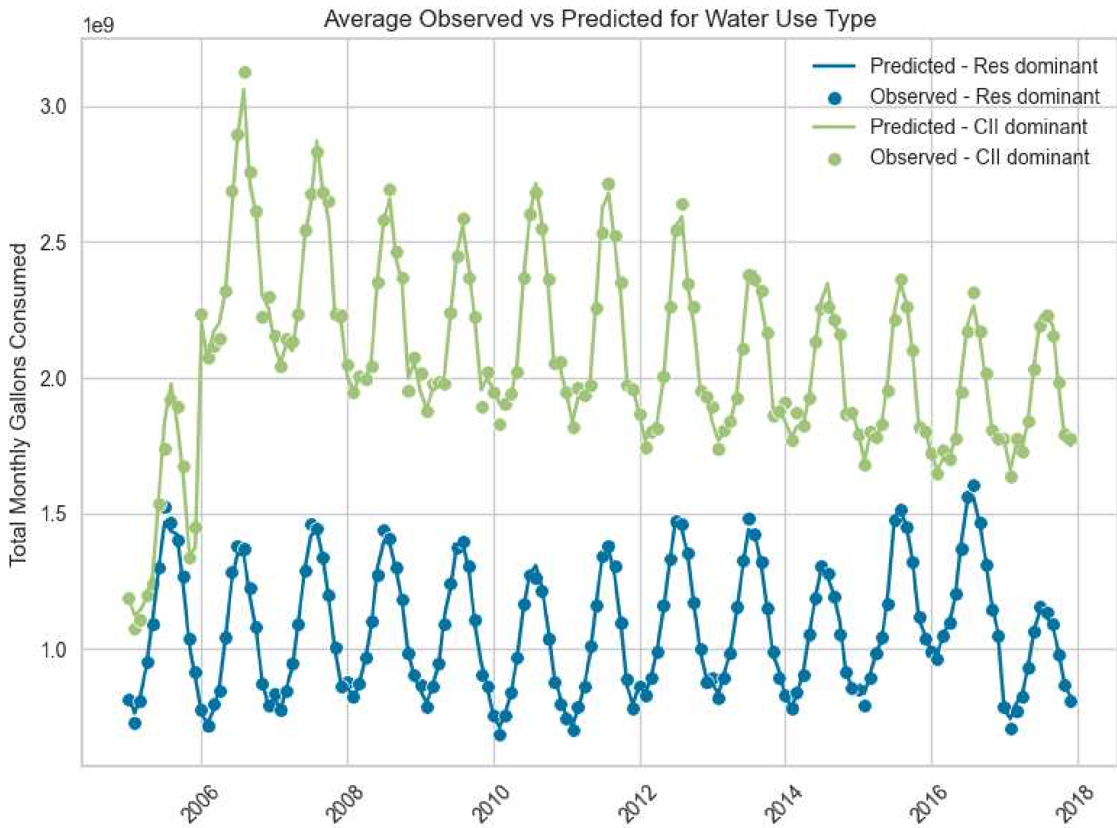


Figure 5.7. Blended model fit on Total monthly water consumption of 126 cities within CONUS aggregated into averages based on Water Use Type.

5.7 Conclusion

The synthesis of results from the feature importance analysis forms a compelling narrative about the multifaceted nature of urban water usage. The findings in this chapter are very important because the climatic, socio-economic, built environment and regional features identified by the ET, RF and LighGBM ML models confirmation the identification of key drivers of water consumption conducted in Chapters 2 and 4 of this report. It must also be noted that the subtleties in the models presented in this chapter and in chapters 2 and 4 are of different time scales and represent different aspects of municipal water consumption.

In spite of the bias and shortcomings, the ML models and the blended ML model validate the need for a nuanced approach to predicting and managing water resources—one that accounts for a diverse array of influences from population dynamics to housing structures, from climatic conditions to economic factors, and from geographical locations to the random effects exerted by the clustering methodologies. By blending the strengths of the top three ML models, a more holistic view on features that influence water consumption patterns is discovered. This comprehensive understanding of the contributing factors to water consumption not only aids in predicting usage patterns but also informs policy decisions, targeted conservation efforts, and infrastructure development, paving the way for more resilient and efficient urban environments and water management efforts.

CHAPTER 6 – PAVING THE WAY FOR FUTURE RESEARCH ON MUNICIPAL WATER SYSTEMS

This technical report presented an in-depth investigation into municipal water use patterns across 126 cities in the contiguous United States (CONUS) over the period 2005-2017. Leveraging extensive water use data compiled through surveys, this multi-faceted research yielded important insights into the trends, patterns, and the climatic, socio-economic, built environment, and regional factors influencing residential and commercial, industrial, institutional (CII) water consumption across CONUS. The study on high-resolution, smart metered water use data from sample set of single-family households in Arizona for the water year 2021 supplements this research with unique characterization results that are based on building level data, showcasing the strength and reliability of such data for more effective water resources management.

6.1 Key Findings

Several notable conclusions are drawn from this robust dataset and rigorous analytical approach spanning climatic, demographic, economic, and technological dimensions of municipal water systems.

6.1.1 Trends in Water Use and Conservation

There is clear evidence of declining per capita water use over time, indicating improving efficiency in municipal water systems despite a steady upward trend in population growth in cities across CONUS. Total water use has reduced at an annual rate of 2.6 GPCD between 2005 and 2017. Large cities and drier regions of the country achieved the highest reductions, underlying the presence of economies of scale in water conservation practices as urban areas expand, and also stressing the impact of climate variability and risks to water availability and allocation.

The learnings from the smart metered data in Ariona revealed that the adoption of water-efficient technologies and practices was found to be a key driver in reducing overall water consumption. Homes that implemented water-saving devices and systems, such as low-flow toilets and efficient irrigation systems, showed significantly lower water usage.

6.1.2 The Influence of Outdoor Water Use

Outdoor water use, primarily landscape irrigation, was found to be a major driver of variability in municipal water demand. Cities in hot, arid climates showed heavy reliance on outdoor water use, mainly in the residential sector. Reductions in outdoor water were particularly notable in large, arid cities, suggesting targeted water conservation policies and technological adoption to curb lawn and garden watering.

The analysis on outdoor water utilization points out that this component of municipal water use is highly variable and that municipal authorities and water managers through policy levers could manipulate to control demand spikes. With climate change projected to intensify risks to freshwater availability, outdoor water uses could be the place to try and enforce or manage consumption patterns better so more water could be squeezed for other essential municipal needs.

6.1.3 Urban Scaling Relationships in Water Use

Analysis of scaling relationships between water use and population pointed to the presence of sub-linear economies of scale for both total and residential municipal water consumption. For every 1% increase in population, total water use grew by 0.87% to 0.89%. This indicates the inherent efficiency gains in per capita terms as municipal areas expand in size. However, CII water use only grew at 0.76% for every 1% rise in population, implying a higher pace of efficiency improvements in non-residential, economic sector.

A deeper inference of these urban scaling relationships tends to favor the densification of people and expansion of cities that inherently favor water conservation effects. This is a rather an optimistic deduction indicating that the inevitability of urbanization can potentially make water use efforts more profound and impactful if infrastructure upgrades and administrative innovations maintain synchronicity with population influx and economic progression.

6.1.4 Drivers of Water Consumption

The report highlights the role of structural, infrastructural, and socio-economic features of urban centers that affect indoor water consumption as well as magnification of outdoor use. The type of housing – single-family versus multi-family households had demonstrable implications on the water use volumes and patterns, potentially due to ownership psyche versus rental mindset. Newer housing potentially built to contemporary building codes and households with high efficiency appliances showed to ingrain water efficiency relative. Higher income levels and presence of recreational amenities like swimming pools were associated with greater water use, however the behavioral associations were beyond the scope of this research. Economic and demographic aspects like manufacturing employment and multi-family housing were notable drivers of CII water consumption and showcase that these socioeconomic attributes induce variability in consumption across regions beyond just climate and demography.

The drivers of water consumption were more prominently identified with the smart metered data from Arizona. Factors such as household size, income level, and the of appliance efficiencies played a significant role in determining the patterns in household water use. The real-time nature of the smart metered data also showed greater sensitivity to local climate and are therefore more apt for studying the efficacy of water conservation policies and efforts.

6.1.5 Advances in Data Analytics

Machine learning techniques proved highly effective in modeling intricate municipal water use patterns, capturing subtleties that eluded traditional linear regression models. Blending the top ML algorithms yielded high acciracy predictive performance for monthly municipal water consumption across study cities based on climatic, socio-economic, and regional attributes. This underscores the merits of modern AI-based tools for gaining data-driven foresights to guide urban water management policy and practice.

6.2 Scope for Future Work

Given the dynamic nature of municipal water systems, continuous monitoring and regular updating of models are essential. This will ensure that the models remain relevant and accurate over time. This study highlights and attempts to the direct future studies on municipal water systems towards avenues for advancing data and methods to continuously improve characterization and forecasting of evolving municipal water demand.

Expanding the datasets used in this study further by inclusion of more cities over longer timeframes in the analysis can unravel more intricate spatial, socioeconomic, and technological shifts playing out at local and regional scales. In this regard, high-resolution, smart-metered data holds immense potential to illuminate water use variability at finer spatial and temporal scales.

Customized ML algorithms tailored to distinct factors influencing residential, CII and subsystem domains of municipal water use can aid precise targeting of conservation policies to varying customer groups. Distilled knowledge from behavioral science and urban planning research should inform feature selection and outcome metrics. Also, integrative modeling efforts that combines the empirical municipal water use patterns with auxiliary datasets on projected demographics, land use plans, climate patterns, and infrastructure investments can generate realistic simulations of future scenarios. This could foster sustainable planning for urban spaces and water resources.

Incorporating and analyzing water consumption patterns during disruptive events like the COVID pandemic can profoundly influence future preparedness of municipal water systems. Such exogenous shock events challenge conventional notions of municipal water demand stability and point to the need to integrate probability distributions of disruptive occurrences into projections.

6.3 Final Conclusion

Overall, as the complexity of urban water systems continue to grow in line with the climatic, socio-economic, technological, and environmental shifts occurring at an accelerated rate in the 21st century, reliance on data-driven approaches becomes critical for reliable foresight and for making informed strategies in managing water resources. This calls for continuous observation, knowledge discovery through cutting edge analytical methods, and integrative, systems-based practices to steward water resources planning and management sustainably. The methodologies and findings in this report contribute a useful foundation but will require constant revision to stay relevant to the challenges of securing water sustainability and resilience for communities of the future.

REFERENCES

Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B., & Sliusarieva, A. (2012). Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resources Research, 48(1), 1–14.

ADWR (Arizona Department of Water Resources). 2020. Water Your Facts. https://www.arizonawaterfacts.com/water-your-facts.

Ahams, I. C., Paterson, W., Garcia, S., Rushforth, R., Ruddell, B. L., & Mejia, A. (2017). Water Footprint of 65 Mid- to Large-Sized U.S. Cities and Their Metropolitan Areas. Journal of the American Water Resources Association, 53(5), 1147–1163.

Akaike, H. (1974), A new look at the statistical model identification. IEEE Transactions on Automatic Control, vol. 19, no. 6, pp. 716-723, December 1974.

Anele, A. O., Hamam, Y., Abu-Mahfouz, A. M., & Todini, E. (2017). Overview, comparative assessment and recommendations of forecasting models for short-term water demand prediction. Water (Switzerland), 9(11).

Angel, S., Sheppard, S.C, Civco, D.L., Buckley, R., Chabaeva, A., Gitlin, L., Kraley, A., Parent, J., and Perlin, M. (2005) The dynamics of global urban expansion. Washington, DC: World Bank, Transport and Urban Development Department; 2005 Sep.

Ault, T. R., Mankin, J. S., Cook, B. I., & Smerdon, J. E. (2016). Relative impacts of mitigation, temperature, and precipitation on 21st-century megadrought risk in the American Southwest. Science Advances, 2(10), 1–9

Baker, L. A., Brazel, A. T., & Westerhoff, P. (2004). Environmental consequences of rapid urbanization in warm, arid lands: Case study of Phoenix, Arizona (USA). Advances in Architecture Series, 18, 155–164.

Balling, R. C., Gober, P., & Jones, N. (2008). Sensitivity of residential water consumption to variations in climate: An intraurban analysis of Phoenix, Arizona. Water Resources Research, 44(10), 1–11.

Barenblatt, G.I., (2003). Scaling. New York: Cambridge University Press.

Basker, E., Becker, R. A., Foster, L., White, T. K., and Zawacki, A. (2019). Addressing Data Gaps: Four New Lines of Inquiry in the 2017 Economic Census. Center for Economic Studies, U.S. Census Bureau, CES 19-28.

Bata, M. H., Carriveau, R., & Ting, D. S.-K. (2020). Short-Term Water Demand Forecasting Using Nonlinear Autoregressive Artificial Neural Networks. Journal of Water Resources Planning and Management, 146(3), 1–9.

Becker, R. A. (2016). Water Use and Conservation in Manufacturing: Evidence from U.S. Microdata. Water Resources Management, 30(12), 4185–4200.

Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988). The New S Language. Wadsworth & Brooks/Cole.

Bentéjac, C., Csörgo, A., & Martínez-Muñoz, G. (2020). A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review, 54, 1937 - 1967.

Bettencourt, L. M. A. (2013). The origins of scaling in cities. Science, 340(6139), 1438–1441.

Billings, B., and Jones, C. (2008). Forecasting urban water demand, 2nd Ed., American Water Works Association, Denver, CO.

Bloom, D.E., Canning, D., and Fink, G. (2008). Urbanization and the wealth of nations. Science: (319) 772–775.

Bolin, B., Barreto, J.D., Hegmon, M., Meierotto, L., York, A. (2013). Double Exposure in the Sunbelt: The Sociospatial Distribution of Vulnerability in Phoenix, Arizona. In: Boone, C., Fragkias, M. (eds) Urbanization and Sustainability. Human-Environment Interactions, vol 3. Springer, Dordrecht.

Bougadis, J., Adamowski, K., & Diduch, R. (2005). Short-term municipal water demand forecasting. Hydrological Processes, 19(1), 137–148.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. Time Series Analysis: Forecasting and Control. 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1994.

Breiman, L. (2001), Random Forests. Machine Learning 45, 5–32.

Breyer, B., Zipper, S. C., & Qiu, J. (2018). Sociohydrological Impacts of Water Conservation Under Anthropogenic Drought in Austin, TX (USA). Water Resources Research, 54(4), 3062–3080.

Brown, T. C., Foti, R., & Ramirez, J. A. (2013). Projected freshwater withdrawals in the United States under a changing climate. Water Resources Research, 49(3), 1259–1276.

Brown, T.C., Foti, R., Ramirez, J. (2012). Chapter 12 - Water resources. In: Future of America's forest and rangelands: Forest Service 2010 Resources Planning Act Assessment. Gen. Tech. Rep. WO-87. Washington, DC: U.S. Department of Agriculture, Forest Service. Vol. 87(109-121).

Bureau of Reclamation. 2023. Colorado River Basin Drought Contingency Plans. https://www.usbr.gov/dcp/.

Bush, D. B., and William E. M. (1986). Potential costs and benefits to Arizona agriculture of the Central Arizona Project. College of Agriculture, University of Arizona (Tucson, AZ).

Butts, R., & Gasteyer, S. (2011). Environmental reviews & case studies: More cost per drop: Water rates, structural inequality, and race in the United States - The case of Michigan. Environmental Practice, 13(4), 386–395.

California State Water Resources Control Board (2019). Available from: https://data.ca.gov/organization/california-state-water-resources-control-board (Accessed: 10 November 2019).

Camporeale, E. (2019). The Challenge of Machine Learning in Space Weather: Nowcasting and Forecasting. Space Weather, 17, 1166 - 1207.

Cattell, R. B. (1966). The Scree Test for The Number of Factors. Multivariate Behavioral Research, 1(2), 245–276.

Cesanek, W., Elmer, V., and Graeff, J. (2017). Planners and Water, PAS Report 588. American Planning Association

Chambers, J. M., Cleveland, W. S., Kleiner, B. and Tukey, P. A. (1983). Graphical Methods for Data Analysis. Wadsworth & Brooks/Cole.

Chang, H., Bonnette, M. R., Stoker, P., Crow-Miller, B., & Wentz, E. (2017). Determinants of single family residential water use across scales in four western US cities. Science of the Total Environment, 596–597, 451–464.

Chatterjee, S., & Hadi, A. S. (2006). Regression analysis by example. John Wiley & Sons.

Chaudhary, K., Alam, M., Al-Rakhami, M.S., & Gumaei, A. (2021), Machine learning-based mathematical modelling for prediction of social media consumer behavior using big data analytics. J Big Data 8, 73 (2021).

Chave, J., and Levin, S. (2003). Scale and scaling in ecological and economic systems. Environmental and Resources Economics (26) 527–557.

Chinnasamy, C. V., Arabi, M., Sharvelle, S., Warziniack, T., Furth, C. D., & Dozier, A. (2021). Characterization of Municipal Water Uses in the Contiguous United States. Water Resources Research, 57(6), 1–18.

Chinnasamy, C. V., McIntyre, W. C., & Mays, D. C. (2018). Technical and administrative feasibility of alluvial aquifer storage and recovery on the South Platte River of northeastern Colorado. Water Policy, 20(4), 841–854.

Christensen, N. S., Wood, A. W., Voisin, N., Lettenmaier, D. P., & Palmer, R. N. (2004). The Effects of Climate Change on the Hydrolgy and Water Resources of the Colorado River Basin. Climatic Change, 62(1–3), 337–363.

Cominola, A., Preiss, L., Thyer, M., Maier, H. R., Prevos, P., Stewart, R. A., & Castelletti, A. (2023). The determinants of household water consumption: A review and assessment framework for research and practice. Npj Clean Water, 6(11), 1–14.

Commission for Environmental Cooperation Working Group, 1997, Ecological regions of North America – toward a common perspective: Montreal, Commission for Environmental Cooperation, 71 p.

Cosgrove, W. J., & Loucks, D. P. (2015). Water management: Current and future challenges and research directions. Water Resources Research, 51(6), 4823–4839.

Decker, E. H., Elliott, S., Smith, F. A., Blake, D. R., & Rowland, F. S. (2000). Energy and material flow through the urban ecosystem. Annual Review of Energy and the Environment, 25.

Decker, E. H., Elliott, S., Smith, F. A., Blake, D. R., & Rowland, F. S. (2000). Energy and Material Flow Through the Urban Ecosystem. Annu. Rev. Energy Environ (Vol. 25).

Denver Water. (2019). 2018 Comprehensive Annual Report. Retrieved from https://www.denverwater.org/about-us/investor-relations/financial-information/annual-reports

Deoreo, W. B., & Mayer, P. W. (2012). Insights into declining single-family residential water demands. Journal - American Water Works Association, 104(6), 383–394.

DeOreo, W. B., Dietemann, A., Skeel, T., Mayer, P. W., Lewis, D. M., & Smith, J. (2001). Retrofit realities. Journal / American Water Works Association, 93(3), 58–72.

DeOreo, W. B., Heaney, J. P., & Mayer, P. W. (1996). Flow trace analysis to access water use. American Water Works Association, 88(1), 79–90.

DeOreo, William B, Mayer, P. W., Dziegielewski, B., & Kiefer, J. (2016). Residential End Uses of Water, Version 2. Water Research Foundation.

Dettinger, M., Udall, B., & Georgakakos, A. (2015). Western water and climate change. Ecological Applications, 25(8), 2069–2093.

Dewitz, J., and U.S. Geological Survey (2021). National Land Cover Database (NLCD) 2019 Products (ver. 2.0, June 2021): U.S. Geological Survey data release.

Dieter, C. A., & Maupin, M. A. (2017). Public Supply and Domestic Water Use in the United States, 2015. U.S. Geological Survey Open-File Report 2017-1131.

Dieter, C.A., Maupin, M.A., Caldwell, R.R., Harris, M.A., Ivahnenko, T.I., Lovelace, J.K., Barber, N.L., and Linsey, K.S. (2018), Estimated use of water in the United States in 2015: U.S. Geological Survey Circular 1441, 65 p.

DOE (Department of Energy), 2012. Energy Conservation Program: Energy Conservation Standards for Residential Clothes Washers. Direct Final Rule. May 31, 2012.

DOE (Department of Energy), 2016, Energy Conservation Program: Energy Conservation Standards for Residential Dishwashers; Final Rule. December 13, 2016. www.regulations.gov/document?D=EERE-2014-BT-STD-0021-0033

Donkor, E. A., Mazzuchi, T. A., Soyer, R., & Alan Roberson, J. (2014). Urban Water Demand Forecasting: Review of Methods and Models. Journal of Water Resources Planning and Management, 140(2), 146–159.

Donnelly, K., & Cooley, H. (2015). Water Use Trends in the United States. Pacific Institute, (April). Retrieved from http://pacinst.org/wp-content/uploads/sites/21/2015/04/Water-Use-Trends-Report.pdf

Dornbierer J, Wika S, Robison C, Rouze G, Sohl T. (2021). Prototyping a Methodology for Long-Term (1680–2100) Historical-to-Future Landscape Modeling for the Conterminous United States. Land. 2021; 10(5):536.

Dozier, A. Q., Arabi, M., Wostoupal, B. C., Goemans, C. G., Zhang, Y., & Paustian, K. (2017). Declining agricultural production in rapidly urbanizing semi-arid regions: Policy tradeoffs and sustainability indicators. Environmental Research Letters, 12(8).

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

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

Eden, S., Scott, C. A., Lamberton, M. L., and Megdal, S. B., (2011). Water-energy interdependencies and the Central Arizona project. The Water–Energy Nexus in the American West 14271.

EPA (Environmental Protection Agency), 2021.WaterSense® Specification for Homes, Version 2.0 Supporting Statement. February 2021. https://www.epa.gov/sites/default/files/2021-02/documents/watersense_final_homes_v2.0_supporting_statement.pdf

Etale, A., Jobin, M., & Siegrist, M. (2018). Tap versus bottled water consumption: The influence of social norms, affect and image on consumer choice. Appetite, 121, 138–146.

Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in Regression Analysis: The Problem Revisited. The Review of Economic and Statistics, 49(1), 92–107.

Flume Inc., and Mayer, P. (2022), Flume Data Analysis Methodology. February 28, 2022. Flume Data Labs Index. https://ss-usa.s3.amazonaws.com/c/308481883/media/1622620d9c10cf05e28658655552106/Flume%20Data%20Analysis%20Methodology%20-%202022.pdf

Flyr, M., Burkhardt, J., Goemans, C., Hans, L., Neel, A., & Maas, A. (2019). Modeling commercial demand for water: Exploring alternative prices, instrumental variables, and heterogeneity. Land Economics, 95(2), 211–224.

Folger, P. (2017). Drought in the United States: Causes and Current Understanding. Retrieved from www.crs.gov

Folke, C. (2003). Freshwater for resilience: A shift in thinking. Philosophical Transactions of the Royal Society of London—B, 358(1440), 2027–2036.

Fortier, C., & Mailhot, A. (2015). Climate Change Impact on Combined Sewer Overflows. J. Water Resour. Plann. Manage., 141(5): 04014073.

Foti, Romano; Ramirez, Jorge A.; Brown, Thomas C. 2012. Vulnerability of U.S. water supply to shortage: a technical document supporting the Forest Service 2010 RPA Assessment. Gen. Tech. Rep. RMRS-GTR-295. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 147 p.

Gawryszewski, V. P., & Costa, L. S. (2005). Social inequality and homicide rates in Sao Paulo City, Brazil. Revista de saude publica, 39, 191-197.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. Machine Learning, 63(1), 3–42.

Ghiassi, M., Zimbra, D. K., & Saidane, H. (2008). Urban Water Demand Forecasting with a Dynamic Artificial Neural Network Model. Journal of Water Resources Planning and Management, 134(2), 138–146.

Glaeser, E. L., Kolko, J., Saiz, A., Journal, S., January, N., Glaeser, E. L., et al. (2001). Consumer City. Journal of Economic Geography, 1(1), 27–50.

Glaeser, E.L., Sacerdote, B.I., and Scheinkman, J.A. (2003) The social multiplier. Journal of the European Economic Association1: 345–353

Gleick, P. H. (2003). Water use. Annual Review of Environment and Resources, 28, 275–314.

Gleick, P. H. (2010). Roadmap for sustainable water resources in southwestern North America. Proceedings of the National Academy of Sciences of the United States of America, 107(50), 21300–21305.

Gleick, P. H., Haasz, D., Henges-Jeck, C., Srinivasan, V., Wolff, G., Kao Cushing, K., & Mann, A. (2003). Waste Not, Want Not: The Potential for Urban Water Conservation in California. Pacific Institute. Retrieved from http://www.pacinst.org/reports/urban_usage/

Gleick, P. H., Srinivasan, V., Henges-Jeck, C., & Wolff, G. (2004). Urban water conservation: A case study of commercial and industrial water use in California. In: Gleick, P (eds) The Worlds Water, pp. Island Press, Washington.

Gober, P., & Kirkwood, C. W. (2010). Vulnerability assessment of climate-induced water shortage in Phoenix. Proceedings of the National Academy of Sciences of the United States of America, 107(50), 21295–21299.

Gober, P., Quay, R., & Larson, K. L. (2016). Outdoor Water Use as an Adaptation Problem: Insights from North American Cities. Water Resources Management, 30(3), 899–912.

Gober, P., Wentz, E. A., Lant, T., Tschudi, M. K., & Kirkwood, C. W. (2011). WaterSim: A simulation model for urban water planning in Phoenix, Arizona, USA. Environment and Planning B: Planning and Design, 38(2), 197–215.

Goldstein, B., Birkved, M., Quitzau, M. B., & Hauschild, M. (2013). Quantification of urban metabolism through coupling with the life cycle assessment framework: Concept development and case study. Environmental Research Letters, 8, 035024.

Gordon, P., & Richardson, H. W. (1997). Are compact cities a desirable planning goal? Journal of the American Planning Association, 63(1), 95–106.

Grespan, A., Garcia, J., Brikalski, M. P., Henning, E., and Kalbusch, A. (2022). Assessment of water consumption in households using statistical analysis and regression trees. Sustainable Cities and Society, 87 (2022) 104186.

Grigg, N. S., Connor, T., & Maas, A. (2018). Financing Integration of Urban Water Systems: From Service Provision to Resource Management. Public Works Management and Policy, 23(2), 186–198.

Grimm, N. B., Faeth, S. H., Golubiewski, N. E., Redman, C. L., Wu, J., Bai, X., & Briggs, J. M. (2008). Global Change and the Ecology of Cities. Science, 319(February).

Grisham, A., & Fleming, W. M. (1989). Long-Term Options for Municipal Water Conservation. American Water Works Association, 81(3), 34–42.

Hagler, Y. (2009). Defining U.S. Megaregions. America 2050, November, 1–8. https://s3.us-east-1.amazonaws.com/rpa-org/pdfs/2050-Paper-Defining-US-Megaregions.pdf

Heidari, H., Arabi, M., Ghanbari, M., & Warziniack, T. (2020). A probabilistic approach for characterization of sub-annual socioeconomic drought intensity-duration-frequency (IDF) relationships in a changing environment. Water, 12(6), 1522.

Heidari, H., Arabi, M., Warziniack, T., & Sharvelle, S. (2021). Effects of Urban Development Patterns on Municipal Water Shortage. Frontiers in Water, 3(July), 1–11.

Hering, J. G., Waite, T. D., Luthy, R. G., Drewes, J. E., & Sedlak, D. L. (2013). A changing framework for urban water systems. Environmental Science and Technology, 47(19), 10721–10726.

Hipel, K. W. & McLeod, A.I. (1994), Time Series Modelling of Water Resources and Environmental Systems. New York: Elsevier Science. ISBN: 0444892702.

Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., et al. (2015). Completion of the 2011 national land cover database for the conterminous United States—Representing a decade of land cover change information. Photogrammetric Engineering & Remote Sensing, 81(5), 345–354.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 24(6), 417–441.

Husson, F., Le, S., & Pagès, J. (2017). Exploratory Multivariate Analysis by Example Using R (2nd Edition). CRC Press.

Hutson, S.S., Barber, N.L., Kenny, J.F., Linsey, K.S., Lumia, D.S., and Maupin, M.A. (2004). Estimated use of water in the United States in 2000: Reston, Va., U.S. Geological Survey Circular 1268, 46 p.

Jain, A., and Ormsbee, L. (2002). Short-term water demand forecast modeling techniques—conventional methods versus AI. J. Am. Water Works Assoc., 94(7), 64–72.

Jain, A., Varshney, A., and Joshi, U. (2001). Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. Water Resour. Manage., 15(5), 299–321.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: With applications in R. synthesis lectures on Mathematics and Statistics (Vol. 11). Springer.

Jolliffe, I. T. (1986). Principal Component Analysis. Springer, New York, NY. ISBN: 0387954422.

Joyce, L. A., & Coulson, D. (2020). Climate scenarios and projections. A Technical Document Supporting the USDA Forest Service 2020 RPA Assessment, General Technical Report RMRS-GTR-413. USDA Forest Service, Rocky Mountain Research Station.

Karandish, F., & Šimůnek, J. (2016). A comparison of numerical and machine-learning modeling of soil water content with limited input data. Journal of Hydrology, 543, 892–909.

Kaufman, L., & Rousseeuw, P.J. (1990). Finding groups in data: An introduction to cluster analysis. Wiley, New York.

Kaufmann, L.,& Rousseeuw, P.J. (1987). Clustering by Means of Medoids. Data Analysis based on the L1-Norm and Related Methods. 405-416.

Kennedy, C., Baker, L., Dhakal, S., & Ramaswami, A. (2012). Sustainable urban systems: An integrated approach. Journal of Industrial Ecology, 16(6), 775–779.

Kenney, D. S., Goemans, C., Klein, R., Lowrey, J., & Reidy, K. (2008). Residential water demand management: Lessons from Aurora, Colorado. Journal of the American Water Resources Association, 44(1), 192–207.

Kenny, J.F., & Juracek, K.E. (2012). Description of 2005–10 domestic water use for selected U.S. cities and guidance for estimating domestic water use: U.S. Geological Survey Scientific Investigations Report 2012–5163, 31 p.

Kenny, J.F., Barber, N.L., Hutson, S.S., Linsey, K.S., Lovelace, J.K., & Maupin, M.A. (2009). Estimated use of water in the United States in 2005: U.S. Geological Survey Circular 1344, 52 p.

Kim, H. K., & Swanson, N. R. (2014). Forecasting financial and macroeconomic variables using data reduction methods: new empirical evidence. Journal of Econometrics, Elsevier, vol. 178(P2), pages 352-367.

Koeller, J., & Gauley, W., (2004). Effectiveness of Data Logging Residential Water Meters to Identify and Quantify Toilet Flush Volumes: A Pilot Study. Los Angeles, California.

Kupel, D. E. (2003). Fuel for Growth: Water and Arizona’s Urban Environment. Tucson: University of Arizona Press.

Kyl Center (Kyl Center for Water Policy at Morrison Institute), (2023). Colorado River Shortage. Arizona Water Footprint. Arizona State University.

https://storymaps.arcgis.com/stories/a1a782ce054d4ad28a0d7d0845e6c03d. Accessed July, 2023.

Lafi, S. Q., & Kaneene, J. B. (1992). An explanation of the use of principal-components analysis to detect and correct for multicollinearity. Preventive Veterinary Medicine, 13(4), 261–275.

Langner, L. L., Joyce, L. A., Wear, D. N., Prestemon, J. P., Coulson, D., & O'Dea, C. B. (2020). Future scenarios. A technical document supporting the USDA Forest Service 2020 RPA Assessment. General Technical Report RMRS-GTR-412,

Larson, K. L., Polsky, C., Gober, P., Chang, H., & Shandas, V. (2013). Vulnerability of water systems to the effects of climate change and urbanization: A comparison of Phoenix, Arizona and Portland, Oregon (USA). Environmental Management, 52(1), 179–195.

Lindstrom, M. J., & Bates, D. M. (1990). Nonlinear Mixed Effects Models for Repeated Measures. Biometrics, 46(3), 673–687.

Linton, J., & Budds, J. (2014). The hydrosocial cycle: Defining and mobilizing a relational-dialectical approach to water. Geoforum, 57, 170–180.

Lobo, J., Bettencourt, L. M. A., Smith, M. E., & Ortman, S. (2020). Settlement scaling theory: Bridging the study of ancient and contemporary urban systems. Urban Studies, 57(4), 731–747.

Luukkonen, C.L., Alzraiee, A.H., Larsen, J.D., Martin, D., Herbert, D.M., Buchwald, C.A., Houston, N.A., Valseth, K.J., Paulinski, S., Miller, L.D., Niswonger, R., Stewart, J.S., and Dieter, C.A. (2023), Public supply water use reanalysis for the 2000-2020 period by HUC12, month, and year for the conterminous United States. U.S. Geological Survey data release.

Ma, X., Xue, X., González-Mejía, A., Garland, J., & Cashdollar, J. (2015). Sustainable water systems for the city of tomorrow-A conceptual framework. Sustainability, 7, 12071–12105.

Maas, A., Goemans, C., Manning, D., Kroll, S., Arabi, M., & Rodriguez-McGoffina, M. (2017). Evaluating the effect of conservation motivations on residential water demand. Journal of Environmental Management, 196, 394–401.

MacDonald, G. M. (2010). Water, climate change, and sustainability in the Southwest. Proceedings of the National Academy of Sciences of the United States of America, 107(50), 21256–21262.

Mack, E. A., & Wrase, S. (2017). A Burgeoning Crisis? A Nationwide Assessment of the Geography of Water Affordability in the United States. PLOS ONE, 12(1), e0169488.

Magionni, E. (2014). Water demand management in times of drought: What matters for water conservation. Water Resources Research, 51, 125–139.

Mahjabin, T., Garcia, S., Grady, C., & Mejia, A. (2018). Large cities get more for less: Water footprint efficiency across the US. PLoS ONE, 13(8), 1–17.

Maidment, D. R., & Miaou, S. P. (1986). Daily Water Use in Nine Cities. Water Resources Research, 22(6), 845–851.

Marston, L. (2022). Water use in a changing world. Nature Climate Change, Nature, 12, 317–319 (2022).

Marston, L. T., Abdallah, A. M., Bagstad, K. J., Dickson, K., Glynn, P., Larsen, S. G., Melton, F. S., Onda, K., Painter, J. A., Prairie, J., Ruddell, B. L., Rushforth, R. R., Senay, G. B., & Shaffer, K. (2022). Water-Use Data in the United States: Challenges and Future Directions. Journal of the American Water Resources Association, 58(4), 485–495.

Marston, L., Ao, Y., Konar, M., Mekonnen, M. M., & Hoekstra, A. Y. (2018). High-Resolution Water Footprints of Production of the United States. Water Resources Research, 54(3), 2288–2316.

Mayer, P. W. (2016). Water Research Foundation Study Documents Water Conservation Potential and More Efficiency in Households. Journal - American Water Works Association, 108(10), 31–40.

Mayer, P. W., Deoreo, W. B., Opitz, E. M., Kiefer, J. C., Davis, W. Y., & Dziegielewski, B. (1999). Residential End Uses of Water. American Water Works Association Research Foundation. Retrieved from https://www.waterdm.com/sites/default/files/WRF (1999) Residential End Uses of Water.pdf

Mayer, P.W. (2005). End Uses of Water: Practical Data Collection, Analysis, and Utility. Arab Water World. May/June 2005.

McGill, J. T. (1964). Growing importance of urban geology, circular 487. U.S. Geological Survey.

McMahon, G., Gregonis, S.M., Waltman, S.W., Omernik, J.M., Thorson, T.D., Freeouf, J.A., Rorick, A.H., and Keys, J.E. (2001). Developing a spatial framework of common ecological regions for the conterminous United States: Environmental Management, v. 28, no. 3, p. 293-316.

Mekonnen, M. M., & Hoekstra, A. Y. (2016). Four billion people facing severe water scarcity. Science Advances, 2(2), 1–6.

Mekonnen, M. M., & Hoekstra, A. Y. (2016). Sustainability: Four billion people facing severe water scarcity. Science Advances, 2(2), 1–7.

Merrett, S., Allan, J. A., & Lant, C. (2003). Virtual water - The water, food, and trade nexus useful concept or misleading metaphor? Water International, 28(1), 106–113.

Mills, G. (2007). Cities as agents of global change. In International Journal of Climatology (Vol. 27, pp. 1849–1857).

Mini, C., Hogue, T. S., & Pincetl, S. (2014). Estimation of residential outdoor water use in Los Angeles, California. Landscape and Urban Planning, 127, 124–135.

Mini, C., Hogue, T. S., & Pincetl, S. (2015). The effectiveness of water conservation measures on summer residential water use in Los Angeles, California. Resources, Conservation and Recycling, 94, 136–145.

Monks, I., Stewart, R. A., Sahin, O., & Keller, R. (2019). Revealing unreported benefits of digital water metering: Literature review and expert opinions. Water (Switzerland), 11(4), 1–32.

Mostafavi, N., Shojaei, H. R., Beheshtian, A., & Hoque, S. (2018). Residential water consumption modeling in the integrated urban metabolism analysis tool (IUMAT). Resources, Conservation and Recycling, 131, 64–74.

Murray, C. R. (1965). Estimated Use of Water in the United States. Issues 556-570.

Murrell, P. (2005). R Graphics. Chapman & Hall/CRC Press.

Nakicenovic, N., et al. (2000), Emissions Scenarios: A Special Report of Working Group III ofthe Intergovernmental Panel on Climate Change, 599 pp., Cambridge Univ. Press, Cambridge, U. K.

Niemczynowicz, J. (1999). Urban hydrology and water management—Present and future challenges. Urban Water, 1(1), 1– 14.

Nordbeck S (1971) Urban allometric growth. Geografiska Annaler. Series B, Human Geography (53) 54–67.

Obringer, R., & Nateghi, R. (2018). Predicting Urban Reservoir Levels Using Statistical Learning Techniques. Scientific Reports, 8(1), 1–9.

Obringer, R., Kumar, R., & Nateghi, R. (2020). Managing the water–electricity demand nexus in a warming climate. Climatic Change, 159(2), 233–252.

Odai, S. N. (2009). Preserving the hydrology of urban Ghana through implementing integrated water resources management. Water and Urban Development Paradigms: Towards an Integration of Engineering, Design and Management Approaches - Proceedings of the International Urban Water Conference, 1, 45–52.

Olmstead, S. M., & Stavins, R. N. (2009). Comparing price and nonprice approaches to urban water conservation. Water Resources Research, 45(4), 1–10.

Opalinski, N. F., Bhaskar, A. S., & Manning, D. T. (2020). Spatial and Seasonal Response of Municipal Water Use to Weather across the Contiguous U.S. Journal of the American Water Resources Association, 56(1), 68–81.

Osborn, T.C., Schefter, J. E., & Shabman, L. (1986). The Accuracy of Water Use Forecasts: Evaluation and Implications. Water Resources Bulletin 22:101-109.

Ouyang, Y., Wentz, E. A., Ruddell, B. L., & Harlan, S. L. (2014). A multi-scale analysis of single-family residential water use in the Phoenix metropolitan area. Journal of the American Water Resources Association, 50(2), 448–467.

Partlow, J. (2023). Arizona city cuts off a neighborhood's water supply amid drought. The Washington Post, January 16, 2023. https://www.washingtonpost.com/climate-environment/2023/01/16/rio-verde-foothills-water-scottsdale-arizona/

Paterson, W., Rushforth, R., Ruddell, B. L., Konar, M., Ahams, I. C., Gironás, J., Mijic, A., & Mejia, A. (2015). Water footprint of cities: A review and suggestions for future research. Sustainability (Switzerland), 7(7), 8461–8490. https://doi.org/10.3390/su7078461

Pesantez, J. E., Berglund, E. Z., & Kaza, N. (2020). Smart meters data for modeling and forecasting water demand at the user-level. Environmental Modelling and Software, 125(104633).

Peters, J., De Baets, B., Verhoest, N. E. C., Samson, R., Degroeve, S., Becker, P. De, & Huybrechts, W. (2007). Random forests as a tool for ecohydrological distribution modelling. Ecological Modelling, 207(2–4), 304–318.

Pimentel, D., Houser, J., Preiss, E., White, O., Fang, H., Mesnick, L., Barsky, T., Tariche, S., & Schreck, J. (1997). Water resources: Agriculture, the environment, and society: An assessment of the status of water resources. BioScience, 47(2), 97–106.

Pinheiro, J. C., & Bates, D. (1996). Unconstrained parametrizations for variance-covariance matrices. Statistics and Computing, 6, 289–296.

Pinheiro, J. C., Bates, D. M., & Lindstrom, M. J. (1995). Model Building for Nonlinear Mixed Effects Models. Technical Report 931, Department of Biostatistics, University of Wisconsin- Madison.

Polebitski, A., Palmer, R., & Waddell, P. (2011). Evaluating water demands under climate change and transitions in the urban environment. J. Water Resour. Plann. Manage., 10.1061/(ASCE)WR.1943-5452.0000112, 249–257.

Polebitski, A.S., & Palmer, R.N. (2010). Seasonal residential water demand forecasting for census tracts. J. Water Reour.Plann. Manage. 136 (1), 27e36.

PRISM Climate Group (2019). 30-year normal 4 km resolution. Oregon State University. Retrieved from http://prism.oregonstate.edu

Puri, R., & Maas, A. (2020). Evaluating the Sensitivity of Residential Water Demand Estimation to Model Specification and Instrument Choices. Water Resources Research, 56(1), 1–14.

Quilty, J., Adamowski, J., & Boucher, M. A. (2019). A Stochastic Data-Driven Ensemble Forecasting Framework for Water Resources: A Case Study Using Ensemble Members Derived from a Database of Deterministic Wavelet-Based Models. Water Resources Research, 55(1), 175–202.

Rahman, A., Srikumar, V., & Smith, A. D. (2018). Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Applied Energy, 212(October 2017), 372–385.

Raj, S. D. (2005). Bottled Water: How Safe Is It? Water Environment Research, 77(7), 3013–3018.

Rajagopalan, B., Nowak, K., Prairie, J., Hoerling, M., Harding, B., Barsugli, J., et al. (2009). Water supply risk on the Colorado River: Can management mitigate? Water Resources Research, 45(8), 1–7.

Rees, W., & Wackernaegel, M. (1996). Urban ecological footprints: why cities cannot be sustainable and why they are a key to sustainability, The Urban Sociology Reader, Eds. J. Lin and C. Mele. Environmental Impact Assessment Review, 16, 223–248.

Regional Plan Association (2006). America 2050: A Prospectus. New York, September 2006.

Rockaway, T. D., Coomes, P. A., Joshua, R., & Barry, K. (2011). Residential water use trends in North America. Journal of American Water Works Association, 103(2), 76–89.

Romano, M., & Kapelan, Z. (2014). Adaptive water demand forecasting for near real-time management of smart water distribution systems. Environmental Modelling and Software, 60, 265–276.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20(C), 53–65.

Salehabadi, H., Tarboton, D., Kuhn, E., Udall, B., Wheeler, K., Rosenberg, D., Goeking, S., and Schmidt, J.C. (2020). The Future Hydrology of the Colorado River Basin. Center for Colorado River Studies, Quinney College of Natural Resources, Utah State University White Paper No. 4.

Sampson, D. A., Quay, R., & Horrie, M. (2022). Building Type, Housing Density, and Water Use: Denver Water Data and Agent-Based Simulations. Journal of the American Water Resources Association, 58(3), 355–369.

Sanchez, G. M., Smith, J. W., Terando, A., Sun, G., & Meentemeyer, R. K. (2018). Spatial Patterns of Development Drive Water Use. Water Resources Research, 54(3), 1633–1649.

Saurí, D. (2013). Water Conservation: Theory and Evidence in Urban Areas of the Developed World. Annual Review of Environment and Resources, 38(1), 227–248.

Scher, S., & Messori, G. (2018). Predicting weather forecast uncertainty with machine learning. Quarterly Journal of the Royal Meteorological Society, 144, 2830 - 2841.

Schlapfer, M., Bettencourt, L.M.A., Grauwin, S., Raschke, M., Claxton, R., Smoreda, Z., West, G.B., & Ratti, C. (2014). The scaling of human interactions with city size. Journal of the Royal Society Interface 11 (98):20130789.20130789.

Schultz, W., Javey, S., & Sorokina, A. (2018). Smart Water Meters and Data Analytics Decrease Wasted Water Due to Leaks. Journal - American Water Works Association, 110(11), E24–E30.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's Tau. Journal of the American Statistical Association, 63(324),1379–1389.

Sharvelle, S., Dozier, A., Arabi, M., & Reichel, B. (2017). A geospatially-enabled web tool for urban water demand forecasting and assessment of alternative urban water management strategies. Environmental Modelling and Software, 97, 213–228.

Solley, W.B., Merk, C.F., & Pierce, P.R. (1988). Estimated use of water in the United States in 1985: U.S. Geological Survey Circular 1004, 82 p.

Solley, W.B., Pierce, R.R., & Perlman, H.A. (1993). Estimated use of water in the United States in 1990: U.S. Geological Survey Circular 1081, 76 p.

Solley, W.B., Pierce, R.R., & Perlman, H.A. (1998). Estimated use of water in the United States in 1995: U.S. Geological Survey Circular 1200, 71 p.

Sønderlund, A. L., Smith, J. R., Hutton, C. J., Kapelan, Z., & Savic, D. (2016). Effectiveness of Smart Meter-Based Consumption Feedback in Curbing Household Water Use: Knowns and Unknowns. Journal of Water Resources Planning and Management, 142(12).

Sprigg, W. A., & Hinkley, T. (2000). Preparing for a Changing Climate: The Potential Consequences of Climate Variability and Change. A Report of the Southwest Regional Assessment Group.

Stoker, P., & Rothfeder, R. (2014). Drivers of urban water use. Sustainable Cities and Society, 12, 1–8.

Stoker, P., Albrecht, T., Follingstad, G., & Carlson, E. (2022). Integrating Land Use Planning and Water Management in U.S. Cities: A Literature Review. Journal of the American Water Resources Association, 58(3), 321–335.

Stone, M. (1979). Comments on model selection criteria of Akaike and Schwarz. (Series B). Journal of the Royal Statistical Society. 41:276–278.

Stone, P. A. (1972). The Economics of the Form and Organisation of Cities. Urban Studies, 9(3), 329–346.

Sugar, L., & Kennedy, C. (2021). Urban Scaling and the Benefits of Living in Cities. Sustainable Cities and Society, 66(November 2020), 102617.

Sze, V., Chen, Y. H., Yang, T. J., & Emer, J. S. (2017). Efficient Processing of Deep Neural Networks: A Tutorial and Survey. Proceedings of the IEEE, 105(12), 2295–2329.

Tonkiss, F. (2014). Cities by design: the social life of urban form. John Wiley & Sons.

Turral, H., Burke, J., Faures, J. M., & Faures, J. M. (2011). Climate change, water and food security. Rome: Food and Agriculture Organization of the United Nations., 204.

U.S. Census Bureau (2019). American community survey 5-year estimates; economy-wide key statistics: 2012; 2010 census tract TIGER/line shapefiles. Generated by Cibi V. Chinnasamy, using American FactFinder. Retrieved from http://factfinder2.Census.gov

U.S. Census Bureau (2010). 2005–2009 American Community Survey 5-year estimates: accessed March 30, 2020, at http://factfinder2.census.gov/faces/nav/jsf/pages/index.xhtml.

U.S. Census Bureau (2019). 2013-2017 American Community Survey 5-Year Estimates; Economy-Wide Key Statistics: 2012; 2010 Census Tract TIGER/Line shapefiles.

Generated by Cibi V. Chinnasamy, using American FactFinder: http://factfinder2.Census.gov, (Accessed 20 January 2019).

U.S. Environmental Protection Agency (2009). 2006 Community Water System Survey, Volume I Overview: U.S. Environmental Protection Agency, Office of Water, 57 p.

Udall, B., & Overpeck, J. (2017). The twenty-first century Colorado River hot drought and implications for the future. Water Resources Research, 53(3), 2404–2418.

United Nations (2015). World urbanization prospects: The 2014 revision. United Nations Department of Eco- nomics and Social Affairs, Population Division: New York, NY, USA. 2015.

United Nations, Department of Economic and Social Affairs, Population Division (2019). World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420). New York: United Nations.

United States Geological Survey (1977). Chapter 11 - Water use, National handbook of recommended methods for water data acquisition. Reston, Va: Geological Survey, U.S. Dept. of the Interior. GOVDOC: I 19.15/3:W 45/2/CHAP.11

Uyanık, G. K., & Güler, N. (2013). A Study on Multiple Linear Regression Analysis. Procedia - Social and Behavioral Sciences, 106, 234–240.

Valecha, H., Varma, A., Khare, I., Sachdeva, A., & Goyal, M. (2018). Prediction of Consumer Behaviour using Random Forest Algorithm. 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), 1-6.

Vijai, P., & Sivakumar, B. P. (2018). Performance comparison of techniques for water demand forecasting. Procedia Computer Science, 143, 258–266.

Villarin, M. C., & Rodriguez-Galiano, V. F. (2019). Machine Learning for Modeling Water Demand. Journal of Water Resources Planning and Management, 145(5), 1–15. https://doi.org/10.1061/(asce)wr.1943-5452.0001067

Wackernagel, M., & Rees, W. (1998). Our ecological footprint: reducing human impact on the earth (Vol. 9). New society publishers.

Waggoner, P E. (1990). Climate change and US water resources. United States.

Walker, D., Creaco, E., Vamvakeridou-Lyroudia, L., Farmani, R., Kapelan, Z., & Savić, D. (2015). Forecasting domestic water consumption from smart meter readings using statistical methods and artificial neural networks. Procedia Engineering, 119(1), 1419–1428.

Wang, Y., Chen, P. C., Ma, H. W., Cheng, K. L., & Chang, C. Y. (2018). Socio-economic metabolism of urban construction materials: A case study of the Taipei metropolitan area. Resources, Conservation and Recycling, 128, 563–571.

Wang, Z., & Vivoni, E. R. (2022). Individualized and Combined Effects of Future Urban Growth and Climate Change on Irrigation Water Use in Central Arizona. Journal of the American Water Resources Association, 58(3), 370–387. https://doi.org/10.1111/1752-1688.13005

Warziniack, T., & Brown, T. C. (2019). The importance of municipal and agricultural demands in future water shortages in the United States. Environmental Research Letters, 14(8).

Warziniack, T., Arabi, M., Brown, T. C., Froemke, P., Ghosh, R., Rasmussen, S., & Swartzentruber, R. (2022). Projections of freshwater use in the United States under climate change. Earth's Future, 10, e2021EF002222.

Water Resources Council (1978). The Nation’s Water Resources 1975-2000. U.S. Government Printing Office, Washington D.C.

Wear, D. N., & Prestemon, J. P. (2019). Spatiotemporal downscaling of global population and income scenarios for the United States. PLoS One, 14(7), e0219242.

Wirth, L. (1938). Urbanism as a Way of Life. American Journal of Sociology. 44(1), 1–24.

Witcher, T. R. (2022). The storied history of the Central Arizona Project. Civil Engineering Source. American Society of Civil Engineers.

Wollman, N. (1960). A Preliminary Report on the Supply of and Demand for Water in the United States as of 1980 and 2000. Resources for the Future (RFT) Press, Washington, D.C.

Yang, L., Yang, S. H., Magiera, E., Froelich, W., Jach, T., & Laspidou, C. (2017). Domestic water consumption monitoring and behaviour intervention by employing the internet of things technologies. Procedia Computer Science, 111, 367–375.

Yigzaw, W., & Hossain, F. (2016). Water sustainability of large cities in the United States from the perspectives of population increase, anthropogenic activities, and climate change. Earth's Future, 4(12), 603–617.

Wang, Z., & Srinivasan, R. S. (2015). A review of artificial intelligence based building energy prediction with a focus on ensemble prediction models. 2015 Winter Simulation Conference (WSC), Huntington Beach, CA, USA, 2015, pp. 3438-3448.

Zhou, C., Gong, M., Xu, Z., & Qu, S. (2022). Urban scaling patterns for sustainable development goals related to water, energy, infrastructure, and society in China. Resources, Conservation and Recycling, 185(July), 106443.

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., & Hinkelmann, R. (2021). Ensemble machine learning paradigms in hydrology: A review. Journal of Hydrology, 598(April), 126266.

APPENDIX – A

SUPPORTING MATERIALS FOR CHAPTER 2 – CHARACTERIZATION OF MUNICIPAL WATER USES ACROSS 126 CITIES AND TOWNS USING AGGREGATED CITY- AND UTILITY-LEVEL WATER SUPPLY DATA

Water Use Data

Instructions: Fill in orange cells - delete example data... Estimate where needed and leave a note on method

Utility Name: What is the official name of your utility?

PWSID(s): What is/are the PWSID(s) of your treatment facilities?

Units of volume:

Service Area: Will you somehow send us your service area boundary (as a shapefile perhaps)?

Notes: Tell us about any nuances of the data you foresee being a problem... Or, what you might think could be an interesting outcome of this research that would benefit your organization...

Notes on data request:

"Residential use" includes multi-family and primarily **residential** mixed-use
"CII use" includes **commercial, industrial, and institutional (CII)**, and primarily commercial mixed-use
"Master meter" includes any large customers with their own sub-metering system
If there is no way to split use into residential and CII, just place numbers in Total Use category.

"Population" is the served population within the service boundary, and if there are large fluctuations of served populations throughout the year (tourists, students, etc.), provide the average number of people being served each day of the year.

Month	Year	Water Used (Units:)			
		Residential	CII Use	Master Meter	Total Use
Jan	2005				
Feb	2005				
Mar	2005				
Apr	2005				
May	2005				
Jun	2005				
Jul	2005				
Aug	2005				
Sep	2005				
Oct	2005				
Nov	2005				
Dec	2005				
Jan	2006				
Feb	2006				
Mar	2006				

Yearly Summary					
Year	Served Population	Water Used (Units:)			
		Residential Use	CII Use	Master Meter	Total Use
2005					
2006					
2007					
2008					
2009					
2010					
2011					
2012					
2013					
2014					
2015					
2016					
2017					
Average					

Figure SI-2.1. Template used to compile aggregate, monthly- municipal water uses data from each one of the study cities through survey.

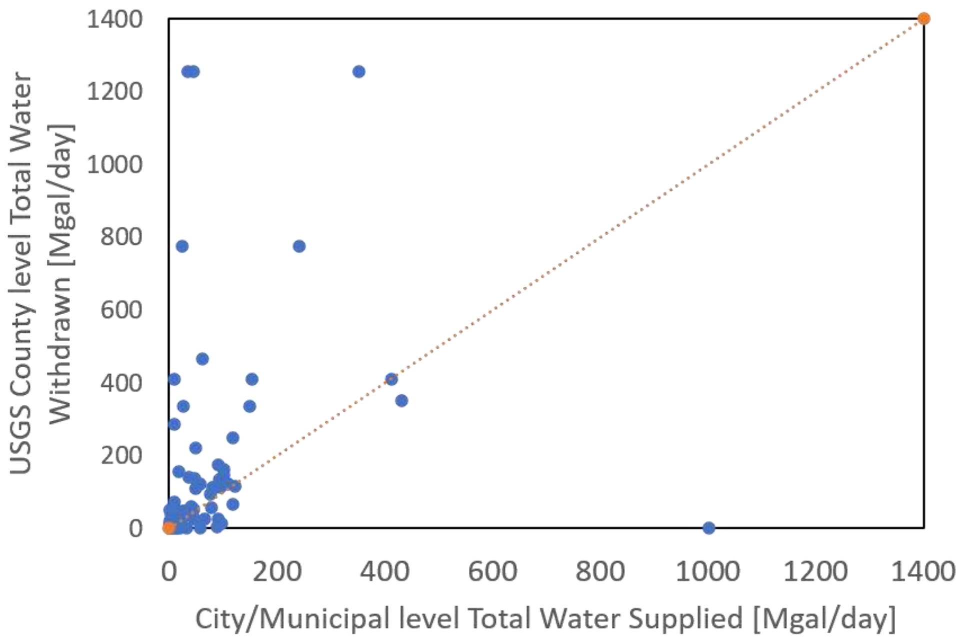


Figure SI-2.2. Comparison of 2015 USGS County level water withdrawal for Public Supply with the average of 2005 to 2017 municipal water use at the city/ town within that USGS county for which monthly water use data was obtained for this study.

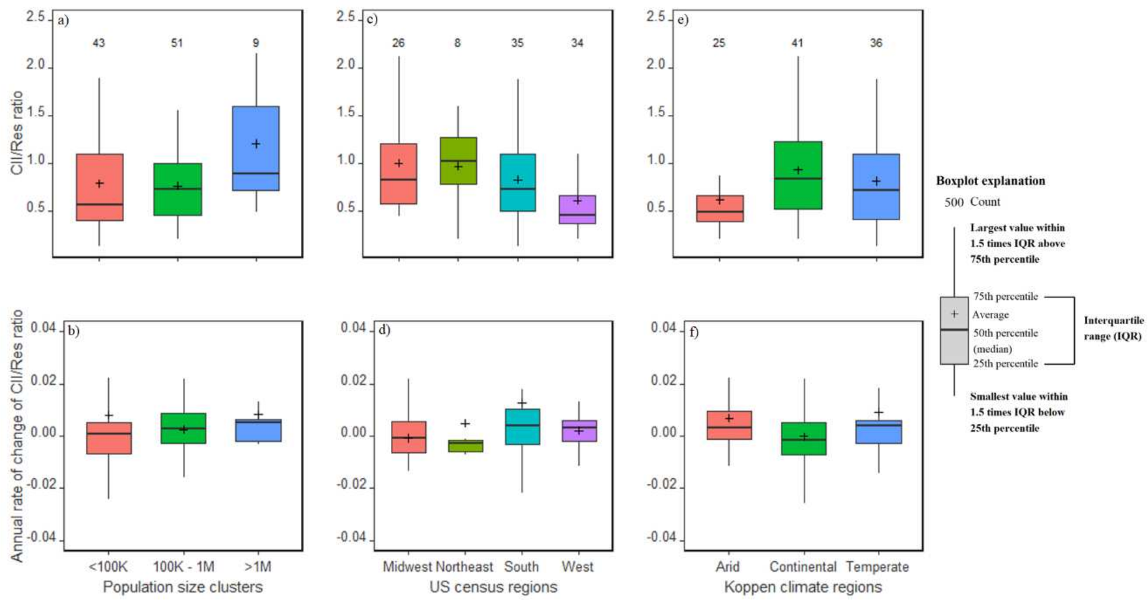


Figure SI-2.3. Variation in average and median rate of change of CII/Res ratio across the study regions during 2005 to 2017.

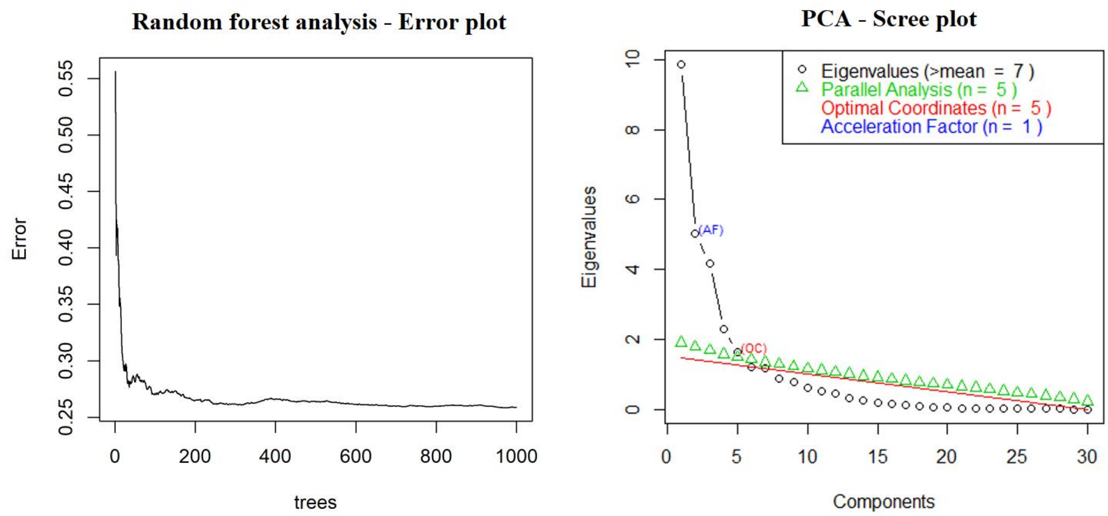


Figure SI-2.4. (Left panel) Error plot of the random forest analysis built using 1000 trees. (Right panel) Scree-test plot, based on the important variables from the random forest analysis, used to select optimal number of factors from a PCA biplot to build a classification tree for CII/Res ratio.

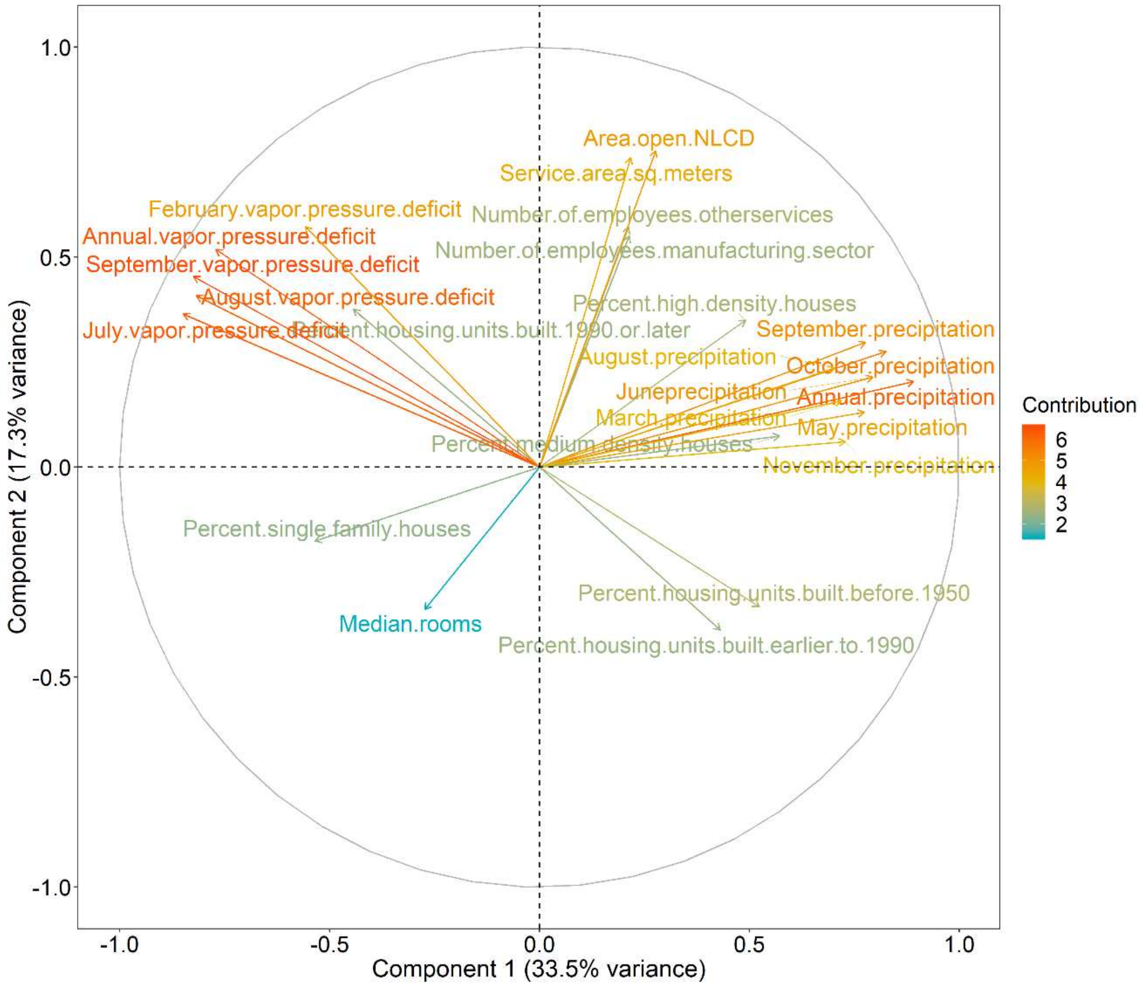


Figure SI-2.5. A PCA biplot of top 25 climatic, urban-ecologic and socio-economic variables from random forest analysis, exhibiting their uniqueness and multi-collinearity to explain the variation in CII/Res ratio.

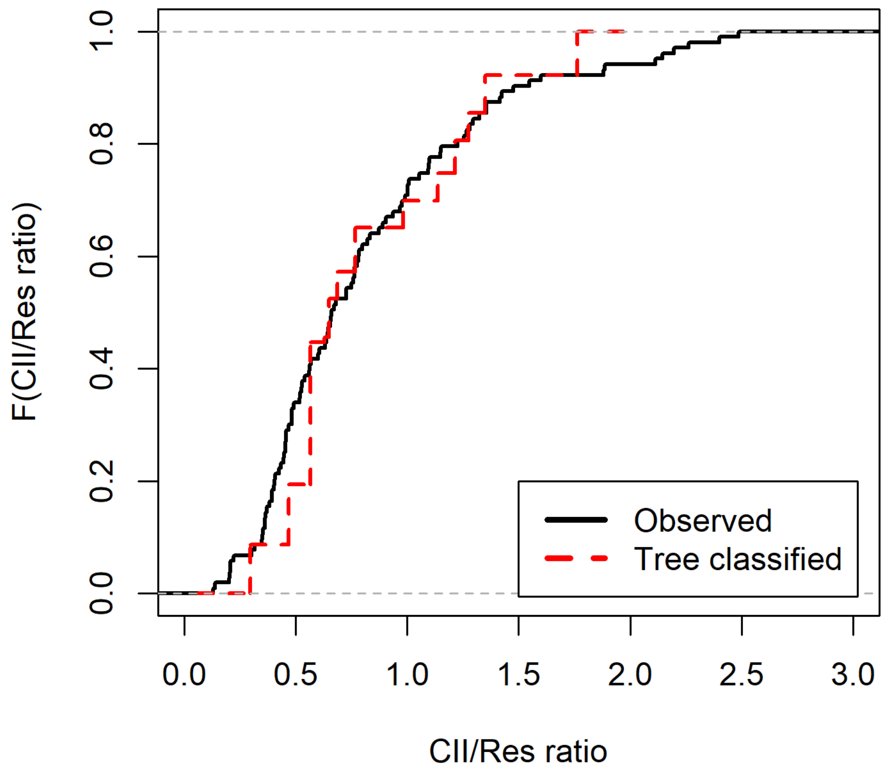


Figure SI-2.6. Observed and modeled cumulative distribution function of CII/Res water use ratio demonstrating a good fit between the empirical data and the classification tree model.

Table SI-2.1. Complete list of cities and towns with the corresponding census, climate and population size clusters used in this study.

City	U.S. Census Region	Koppen Climate Region	2010 Population	Population Cluster
Abilene, Texas	South	Temperate	123,729	Medium
Aiken, South Carolina	South	Temperate	29,406	Small
Alamogordo, New Mexico	West	Arid	32,986	Small
Albuquerque, New Mexico	West	Arid	614,980	Medium
Altoona, Pennsylvania	Northeast	Continental	54,369	Small
Amarillo, Texas	South	Arid	196,728	Medium
Annapolis, Maryland	South	Temperate	38,059	Small
Arlington, Texas	South	Temperate	366,503	Medium
Asheville, North Carolina	South	Temperate	125,000	Medium
Auburn, Maine	Northeast	Continental	16,996	Small
Augusta, Maine	Northeast	Continental	18,910	Small
Aurora, Colorado	West	Continental	330,815	Medium
Austin, Texas	South	Temperate	903,672	Medium
Bangor, Maine	Northeast	Continental	32,605	Small
Baytown, Texas	South	Temperate	72,488	Small
Belgrade, Montana	West	Continental	7,855	Small
Bend, Oregon	West	Arid	57,553	Small
Blacksburg, Virginia	South	Continental	33,804	Small
Bloomington, Indiana	Midwest	Continental	116,825	Small
Bloomington, Minnesota	Midwest	Continental	85,532	Small
Boston, Massachusetts	Northeast	Continental	1,047,462	Large
Boulder City, Nevada	West	Arid	20,694	Small
Bowie, Maryland	South	Temperate	21,703	Small
Bozeman, Montana	West	Continental	39,063	Small
Brownsville, Texas	South	Temperate	177,909	Medium
Casper, Wyoming	West	Continental	54,462	Small
Charlotte, North Carolina	South	Temperate	864,289	Medium
Chillicothe, Missouri	Midwest	Continental	9,241	Small
Cincinnati, Ohio	Midwest	Continental	1,047,462	Large
Clovis, California	West	Arid	101,950	Medium
Dallas, Texas	South	Temperate	1,236,107	Large
Dalton, Georgia	South	Temperate	100,182	Small
Denver, Colorado	West	Temperate	1,176,112	Medium
Des Moines, Iowa	Midwest	Continental	461,162	Medium
Detroit, Michigan	Midwest	Continental	660,000	Medium
Dickinson, North Dakota	Midwest	Continental	19,398	Small
Durham, North Carolina	South	Temperate	246,825	Medium
El Paso, Texas	South	Arid	758,484	Medium

Table SI-2.1. (Continued ...)

City	U.S. Census Region	Koppen Climate Region	2010 Population	Population Cluster
Enid, Oklahoma	South	Temperate	49,293	Small
Farmington, New Mexico	West	Arid	44,139	Small
Flagstaff, Arizona	West	Continental	67,288	Small
Fort Collins, Colorado	West	Arid	142,810	Medium
Fort Smith, Arkansas	South	Temperate	123,781	Medium
Fort Worth, Texas	South	Temperate	759,416	Medium
Gainesville, Florida	South	Temperate	120,392	Medium
Garland, Texas	South	Temperate	228,816	Medium
Grand Forks, North Dakota	Midwest	Continental	53,862	Small
Grand Island, Nebraska	Midwest	Continental	47,696	Small
Green River, Wyoming	West	Arid	12,576	Small
Greensboro, North Carolina	South	Temperate	272,192	Medium
Havre, Montana	West	Arid	9,516	Small
Hereford, Texas	South	Arid	15,225	Small
International Falls, Minnesota	Midwest	Continental	6,295	Small
Iowa City, Iowa	Midwest	Continental	68,884	Small
Jonesboro, Arkansas	South	Temperate	84,857	Small
Kansas City, Missouri	Midwest	Continental	464,122	Medium
Kearney, Nebraska	Midwest	Continental	31,189	Small
Kenner, Louisiana	South	Temperate	437,123	Medium
Kennewick, Washington	West	Arid	74,334	Small
La Crosse, Wisconsin	Midwest	Continental	52,000	Small
Lafayette, Louisiana	South	Temperate	122,005	Medium
Lancaster, California	West	Arid	168,223	Medium
Lansing, Michigan	Midwest	Continental	114,659	Medium
Laredo, Texas	South	Arid	241,118	Medium
Las Vegas, Nevada	West	Arid	1,970,912	Large
Long Beach, California	West	Arid	470,659	Medium
Los Angeles, California	West	Arid	3,852,565	Large
Louisville/Jefferson County metro government (balance), Kentucky	South	Continental	895,770	Medium
Lubbock, Texas	South	Arid	229,159	Medium
Madison, Wisconsin	Midwest	Continental	241,140	Medium
Manchester, New Hampshire	Northeast	Continental	158,116	Medium
Marietta, Georgia	South	Temperate	57,507	Small
Memphis, Tennessee	South	Temperate	779,487	Medium
Miami-Dade County, Florida	South	Tropical	2,215,016	Large
Milwaukee, Wisconsin	Midwest	Continental	863,800	Medium

Table SI-2.1. (Continued ...)

City	U.S. Census Region	Koppen Climate Region	2010 Population	Population Cluster
Minot, North Dakota	Midwest	Continental	64,110	Small
Monahans, Texas	South	Arid	7,104	Small
Montrose, Colorado	West	Arid	18,451	Small
New York, New York	Northeast	Continental	8,327,118	Large
Omaha, Nebraska	Midwest	Continental	442,818	Medium
Orlando, Florida	South	Temperate	425,100	Medium
Page, Arizona	West	Arid	10,436	Small
Peoria, Arizona	West	Arid	148,812	Medium
Philadelphia, Pennsylvania	Northeast	Continental	1,535,038	Large
Phoenix, Arizona	West	Arid	1,487,417	Large
Pierre, South Dakota	Midwest	Continental	13,800	Small
Pikeville, Kentucky	South	Temperate	50,077	Small
Plainview, Texas	South	Arid	21,663	Small
Pocatello, Idaho	West	Continental	54,529	Small
Ponca City, Oklahoma	South	Temperate	24,971	Small
Poplar Bluff, Missouri	Midwest	Temperate	17,153	Small
Portland, Maine	Northeast	Continental	158,478	Medium
Portland, Oregon	West	Temperate	570,555	Medium
Pueblo, Colorado	West	Arid	107,656	Medium
Rochester, Minnesota	Midwest	Continental	106,704	Medium
Rochester, New York	Northeast	Continental	697,308	Medium
Rock Springs, Wyoming	West	Arid	19,530	Small
Rockford, Illinois	Midwest	Continental	152,638	Small
Safford, Arizona	West	Arid	22,996	Small
Salina, Kansas	Midwest	Continental	47,700	Small
Salt Lake City, Utah	West	Continental	331,003	Medium
San Angelo, Texas	South	Temperate	95,137	Small
San Diego, California	West	Arid	1,304,648	Large
San Francisco, California	West	Temperate	2,330,770	Large
Santa Barbara, California	West	Temperate	91,922	Small
Santa Maria, California	West	Temperate	99,003	Small
Santa Rosa, California	West	Temperate	167,567	Medium
Sherman, Texas	South	Temperate	38,770	Small
Sioux Falls, South Dakota	Midwest	Continental	158,716	Medium
Spearfish, South Dakota	Midwest	Continental	10,607	Small
Spokane, Washington	West	Continental	206,307	Medium
Springfield, Ohio	Midwest	Continental	61,815	Small
St. Peters, Missouri	Midwest	Continental	45,321	Small
Steamboat Springs, Colorado	West	Continental	12,165	Small

Table SI-2.1. (Continued ...)

City	U.S. Census Region	Koppen Climate Region	2010 Population	Population Cluster
Tacoma, Washington	West	Temperate	318,177	Medium
Thomasville, Georgia	South	Temperate	18,423	Small
Tucson, Arizona	West	Arid	659,942	Medium
Tyler, Texas	South	Temperate	96,083	Small
Vacaville, California	West	Temperate	91,884	Small
Valentine, Nebraska	Midwest	Continental	2,779	Small
Vancouver, Washington	West	Temperate	234,369	Medium
Vernal, Utah	West	Arid	9,755	Small
Watertown, New York	Northeast	Continental	26,711	Small
Wichita Falls, Texas	South	Continental	140,000	Medium
Wichita, Kansas	Midwest	Temperate	453,621	Medium
Winston-Salem, North Carolina	South	Temperate	222,938	Medium

Table SI-2.2. List of cities and towns with the corresponding water providers.

City	Water Provider
Abilene, Texas	City of Abilene - Water Utilities
Aiken, South Carolina	City of Aiken Water, Sewer & Stormwater
Alamogordo, New Mexico	City of Alamogordo Domestic Water System
Albuquerque, New Mexico	Albuquerque Bernalillo County Water Utility Authority
Altoona, Pennsylvania	Altoona Water Authority
Amarillo, Texas	City of Amarillo - Water Utilities
Annapolis, Maryland	City of Annapolis
Arlington, Texas	City of Arlington Water Utilities
Asheville, North Carolina	City of Asheville
Auburn, Maine	Auburn Water District
Augusta, Maine	Greater August Utility District
Aurora, Colorado	Aurora Water
Austin, Texas	Austin Water
Bangor, Maine	Bangor Water District
Baytown, Texas	City of Baytown Water Department
Belgrade, Montana	City of Belgrade
Bend, Oregon	City of Bend, Utility Department
Blacksburg, Virginia	Town of Blacksburg
Bloomington, Indiana	City of Bloomington, IN Utilities
Bloomington, Minnesota	City of Bloomington, MN Utilities
Boston, Massachusetts	Boston Water and Sewer Commission
Boulder City, Nevada	Boulder City Water Utility
Bowie, Maryland	City of Bowie Utilities Division
Bozeman, Montana	City of Bozeman Water and Sewer Division
Brownsville, Texas	Brownsville Public Utilities Board
Casper, Wyoming	City of Casper, Wyoming
Charlotte, North Carolina	Charlotte Water
Chillicothe, Missouri	Chillicothe Municipal Utilities
Cincinnati, Ohio	Greater Cincinnati Water Works
Clovis, California	City of Clovis
Dallas, Texas	Dallas Water Utilities Department
Dalton, Georgia	Dalton Utilities
Denver, Colorado	Denver Water
Des Moines, Iowa	Des Moines Water Works
Detroit, Michigan	Detroit Water and Sewerage Department
Dickinson, North Dakota	City of Dickinson - Water Utilities Department
Durham, North Carolina	City of Durham
El Paso, Texas	El Paso Water

Table SI-2.2. (Continued ...)

City	Water Providers
Enid, Oklahoma	City of Enid Utilities
Farmington, New Mexico	City of Farmington
Flagstaff, Arizona	City of Flagstaff Water Services
Fort Collins, Colorado	City of Fort Collins Utilities
Fort Smith, Arkansas	City of Fort Smith Utilities
Fort Worth, Texas	Fort Worth Water Department
Gainesville, Florida	Gainesville Regional Utilities
Garland, Texas	City of Garland - Water Utilities
Grand Forks, North Dakota	Grand Forks Water Department
Grand Island, Nebraska	Grand Island Utilities Department Water Division
Green River, Wyoming	Green River Water and Sewer Division
Greensboro, North Carolina	Greensboro City - Division of Water Resources
Havre, Montana	City of Havre - Public Works
Hereford, Texas	City of Hereford Municipal Water System
International Falls, Minnesota	International Falls Water Department
Iowa City, Iowa	Iowa City Utilities
Jonesboro, Arkansas	Jonesboro City Water & Light
Kansas City, Missouri	KC Water
Kearney, Nebraska	City of Kearney Nebraska Utilities Department
Kenner, Louisiana	Kenner Public Works Department
Kennewick, Washington	City of Kennewick
La Crosse, Wisconsin	City of La Crosse Utilities Water Sewer Storm
Lafayette, Louisiana	Lafayette Utilities System
Lancaster, California	Los Angeles County Waterworks District 40
Lansing, Michigan	Lansing Board of Water and Light
Laredo, Texas	City of Laredo Utilities
Las Vegas, Nevada	Southern Nevada Water Authority/Las Vegas Valley Water District
Long Beach, California	Long Beach Water Department
Los Angeles, California	Los Angeles Department of Water and Power
Louisville/Jefferson County metro government (balance), Kentucky	Louisville Water Company
Lubbock, Texas	Lubbock Public Water System
Madison, Wisconsin	Madison Water Utility
Manchester, New Hampshire	City of Manchester Water Works
Marietta, Georgia	Marietta Water
Memphis, Tennessee	Memphis Light, Gas and Water Division
Miami-Dade County, Florida	Miami-Dade County, FL
Milwaukee, Wisconsin	Milwaukee Water Works

Table SI-2.2. (Continued ...)

City	Water Providers
Minot, North Dakota	Minot City Water Utility
Monahans, Texas	Water Department - City of Monahans
Montrose, Colorado	City of Montrose Utilities (Water & Sewer)
New York, New York	New York City Department of Environmental Protection - Bureau of Water and Sewer
Omaha, Nebraska	Metropolitan Utilities District of Omaha
Orlando, Florida	Orange County Water Division
Page, Arizona	Page Utility Enterprises
Peoria, Arizona	City of Peoria
Philadelphia, Pennsylvania	Philadelphia Water Department
Phoenix, Arizona	Phoenix Water
Pierre, South Dakota	City of Pierre Utilities
Pikeville, Kentucky	Mountain Water District
Plainview, Texas	Plainview Municipal Water
Pocatello, Idaho	Pocatello Water Department
Ponca City, Oklahoma	Ponca City Utilities
Poplar Bluff, Missouri	Poplar Bluff Municipal Utilities
Portland, Maine	Portland Water District
Portland, Oregon	Portland Water Bureau
Pueblo, Colorado	Board of Water Works of Pueblo, Colorado
Rochester, Minnesota	Rochester Public Utilities
Rochester, New York	Monroe County Water Authority
Rock Springs, Wyoming	Rock Springs Municipal Utility
Rockford, Illinois	City of Rockford Water Division
Safford, Arizona	City of Safford Utilities
Salina, Kansas	City of Salina Utilities
Salt Lake City, Utah	Salt Lake City Public Utilities
San Angelo, Texas	City of San Angelo Water Utilities
San Diego, California	City of San Diego Public Utilities
San Francisco, California	San Francisco Public Utilities Commission
Santa Barbara, California	City of Santa Barbara Utilities
Santa Maria, California	City of Santa Maria Utilities
Santa Rosa, California	Santa Rosa Water
Sherman, Texas	City of Sherman Customer & Utility Services
Sioux Falls, South Dakota	Sioux Falls Utilities
Spearfish, South Dakota	City of Spearfish Utilities
Spokane, Washington	City of Spokane Utilities
Springfield, Ohio	City of Springfield Utilities
St. Peters, Missouri	City of St. Peters Utilities
Steamboat Springs, Colorado	City of Steamboat Springs Utilities

Table SI-2.2. (Continued ...)

City	Water Providers
Tacoma, Washington	Tacoma Water
Thomasville, Georgia	City of Thomasville Utilities
Tucson, Arizona	Tucson Water
Tyler, Texas	Tyler Water Utilities
Vacaville, California	Vacaville Utilities
Valentine, Nebraska	Water & Sewer Department
Vancouver, Washington	Utilities Department-Vancouver
Vernal, Utah	Vernal City Utilities
Watertown, New York	Watertown City Water Department
Wichita Falls, Texas	City of Wichita, Kansas - Wichita Water Utility
Wichita, Kansas	City of Wichita Falls Utilities
Winston-Salem, North Carolina	City/County Utilities (Winston-Salem/Forsyth County)

APPENDIX – B

SUPPORTING MATERIALS FOR CHAPTER 3 – CHARACTERIZATION OF SINGLE-FAMILY RESIDENTIAL WATER USES EMPLOYING HIGH-RESOLUTION, SMART METERING DATA

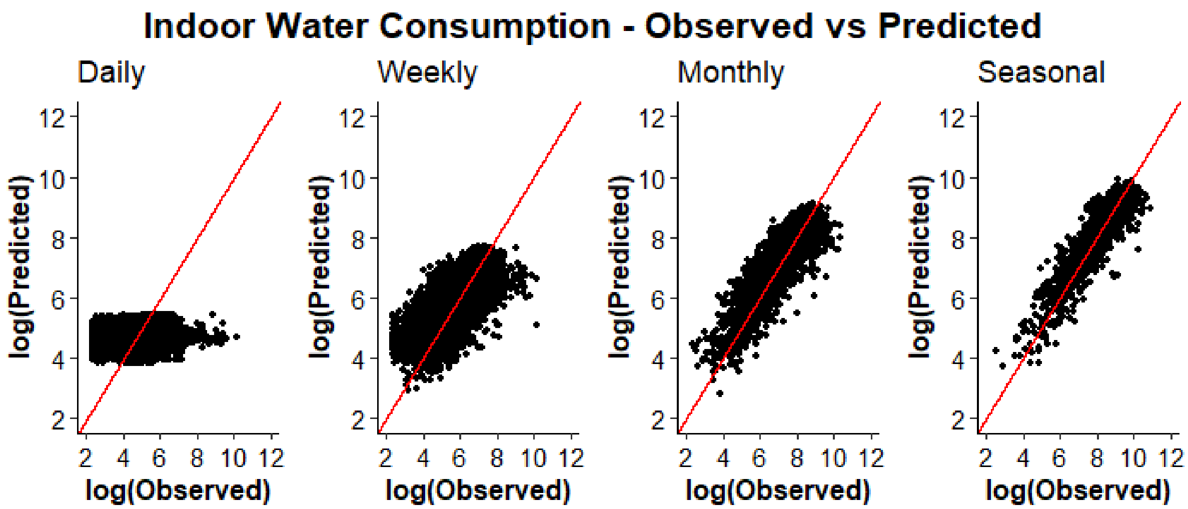


Figure SI-3.1. Goodness-of-fit plots of observed versus predicted household level indoor water consumption data at different time-steps using the log transformed mixed effects model. R^2 for these models are presented in Table 3.3.

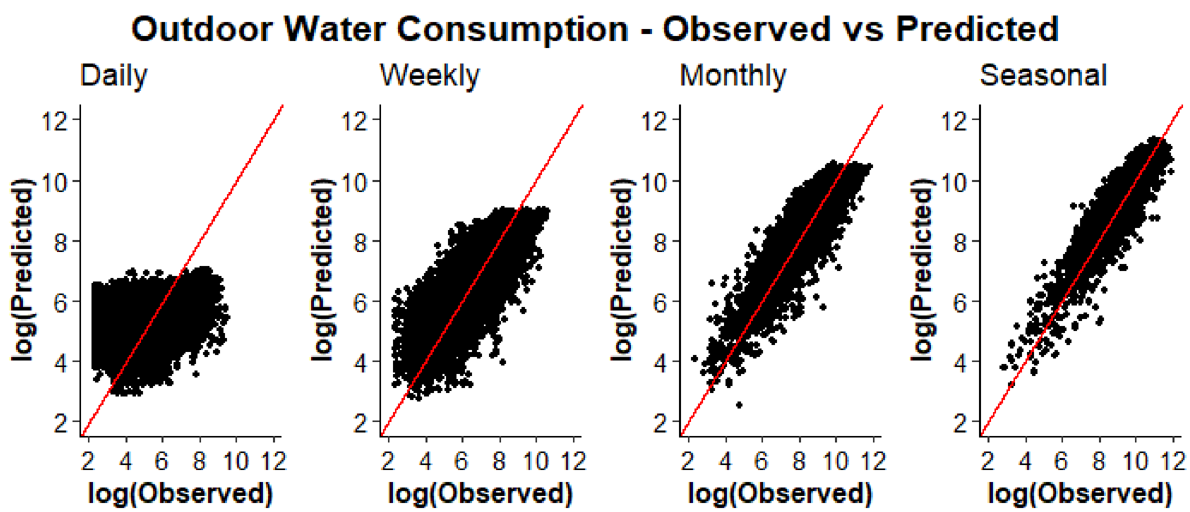


Figure SI-3.2. Goodness-of-fit plots of observed versus predicted household level outdoor water consumption data at different time-steps using the log transformed mixed effects model. R^2 for these models are presented in Table 3.3.

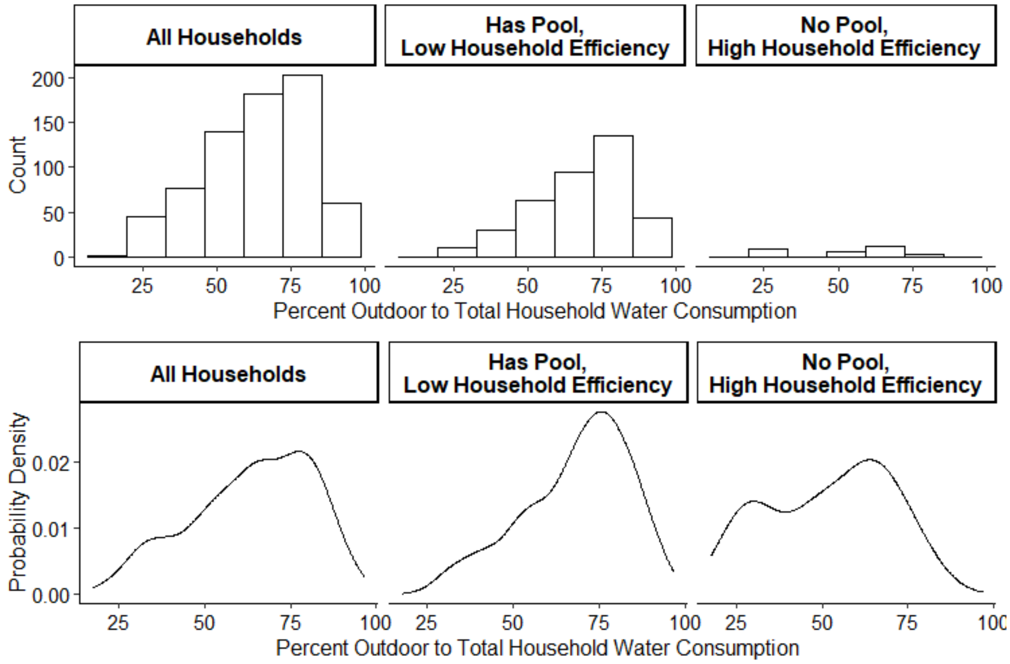


Figure SI-3.3. Histogram and density plots of household level outdoor water use ratio with total water use under different classifications of the households based on their water conservation mindset

Table SI-3.1. Characteristics of the normal distribution of percent outdoor to total water uses in households grouped based on their conservation mindset, compared to all households grouped together.

Category	N	Mean	Standard Deviation	Skewness	Standard Mean Error	95% Confidence interval of Mean
All households	706	63.7	17.5	-0.45	0.66	62.4 – 65.0
Has Pool, Low household Efficiency	377	68.2	15.6	-0.62	0.81	66.6 – 69.8
No Pool, High Household Efficiency	32	52.0	17.6	-.021	3.1	46.0 – 58.1

Table SI-3.2. Important factors of single-family household's indoor and outdoor water consumption in Arizona at the monthly and seasonal time steps, and their corresponding coefficient estimates in the log transformed mixed-effects models. "NA" denotes variable insignificance in that time-step. Astrix (*) is used under the random-effect column if that variable has a fixed-effect nature, and vice versa.

Water Use Category	Variables	Fixed effect coefficient (α)		Random effect coefficient (β)	
		Monthly	Seasonal	Monthly	Seasonal
Indoor GPH	*Intercept*	296.6	8153.93	*	*
	Air Temperature (Max)	-3.94	-2.89	*	
	Dishwasher Efficiency	*	*	0.17	0.17
	Evapotranspiration	1.05	1.07	*	*
	Faucet Efficiency	*	*	-0.12	NA
	Has Pool	*	*	0.19	0.19
	Month of Year	0.29	NA	*	NA
	Number of Bathrooms	*	*	0.14	0.17
	Number of Residents	*	*	0.43	0.41
	Postal Code	*	*	-22.11	-20.42
	Showerhead Efficiency	*	*	0.13	0.14
	Toilet Efficiency	*	*	0.12	0.14
	Year	NA	-1034.81	NA	*
	Year Built	*	*	-2.98	-3.52
Outdoor GPH	*Intercept*	578.77	9787.14	*	*
	Air Temperature (Max)	-3.36	-2.04	*	*
	Evapotranspiration	1.21	1.20	*	*
	Has Pool	*	*	0.15	0.14
	Home Value	*	*	0.25	0.30
	Irrigation Efficiency	*	*	1.28	1.30
	Irrigation Frequency	*	NA	0.10	NA
	Irrigation Type	*	*	0.61	0.67
	Lot Size	*	*	0.14	0.12
	Month of Year	0.36	NA	*	NA
	Number of Bathrooms	*	*	-0.14	-0.17
	Number of Residents	*	*	0.13	0.14
	Postal Code	*	*	-44.54	-36.09
	Year	NA	-1222.69	NA	*
	Year Built	*	*	-7.57	-8.02

APPENDIX – C

SUPPORTING MATERIALS FOR CHAPTER 4 - ROBUST URBAN SCALING METRICS OF MUNICIPAL WATER USE CONSUMPTION PATTERNS

Tables SI-4.1.A and 4.1.B below present total population and annual change in population in under metro, micro and rural classification of CONUS counties from 1985 to 2015. Data source: (Dieter et al., 2018; Maupin et al., 2014; Kenny et al., 2009; Hutson et al., 2004; Solley et al., 1988, 1993, 1998).

Table SI-4.1.A

	1985	1990	1995	2000	2005	2010	2015
Metro	198,297,670	208,832,910	221,744,470	240,343,040	254,903,967	266,074,321	278,661,079
Micro	24,373,260	24,303,830	25,435,150	26,503,990	27,110,271	27,746,147	27,835,808
Rural	19,680,230	19,199,520	19,888,160	18,498,850	18,650,367	18,757,571	18,521,977
Total	242,351,160	252,336,260	267,067,780	285,345,880	300,664,605	312,578,039	325,018,864

Table SI-4.1.B

	1990	1995	2000	2005	2010	2015	Median	Average
Percent change in Metro	5.0	5.8	7.7	5.7	4.2	4.5	5.4	5.5
Percent change in Micro	-0.3	4.4	4.0	2.2	2.3	0.3	2.3	2.2
Percent change in Rural	-2.5	3.5	-7.5	0.8	0.6	-1.3	-0.4	-1.1

Table SI-4.2. Classification of cities and towns by Climate-Census region and Water Use Type.

City	Climate-Census Region	Water Use Type	CII/Res ratio
Abilene, Texas	Temperate-South	Res dominant	0.727
Aiken, South Carolina	Temperate-South	Res dominant	0.295
Alamogordo, New Mexico	Arid	Res dominant	0.205
Albuquerque, New Mexico	Arid	Res dominant	0.38
Altoona, Pennsylvania	Continental-East	Res dominant	0.454
Amarillo, Texas	Arid	Res dominant	0.653
Annapolis, Maryland	Temperate-South	Res dominant	0.392
Arlington, Texas	Temperate-South	Res dominant	0.35
Asheville, North Carolina	Temperate-South	Res dominant	0.782
Auburn, Maine	Continental-East	Res dominant	0.686
Augusta, Maine	Continental-East	Res dominant	0.686
Aurora, Colorado	Continental-West	Res dominant	0.3
Austin, Texas	Temperate-South	Res dominant	0.569
Bangor, Maine	Continental-East	CII dominant	1.295
Baytown, Texas	Temperate-South	CII dominant	1.763
Belgrade, Montana	Continental-West	Res dominant	0.205
Bend, Oregon	Arid	Res dominant	0.558
Blacksburg, Virginia	Continental-Central	Res dominant	0.564
Bloomington, Indiana	Continental-Central	Res dominant	0.765
Bloomington, Minnesota	Continental-Central	Res dominant	0.604
Boston, Massachusetts	Continental-East	CII dominant	1.6
Boulder City, Nevada	Arid	CII dominant	1.885
Bowie, Maryland	Temperate-South	Res dominant	0.138
Bozeman, Montana	Continental-West	Res dominant	0.645
Brownsville, Texas	Temperate-South	Res dominant	0.295
Casper, Wyoming	Continental-West	Res dominant	0.467
Charlotte, North Carolina	Temperate-South	Res dominant	0.527
Chillicothe, Missouri	Continental-Central	CII dominant	1.254
Cincinnati, Ohio	Continental-Central	CII dominant	2.114
Clovis, California	Arid	Res dominant	0.204
Dallas, Texas	Temperate-South	CII dominant	1.322
Dalton, Georgia	Temperate-South	CII dominant	2.197
Denver, Colorado	Arid	Res dominant	0.71
Des Moines, Iowa	Continental-Central	Res dominant	0.489
Detroit, Michigan	Continental-Central	CII dominant	0.968
Dickinson, North Dakota	Continental-Central	Res dominant	0.564
Durham, North Carolina	Temperate-South	Res dominant	0.727
El Paso, Texas	Arid	Res dominant	0.466
Enid, Oklahoma	Temperate-South	Res dominant	0.765
Farmington, New Mexico	Arid	CII dominant	1.355
Flagstaff, Arizona	Continental-West	Res dominant	0.661

Fort Collins, Colorado	Arid	Res dominant	0.872
Fort Smith, Arkansas	Temperate-South	CII dominant	1.548
Fort Worth, Texas	Temperate-South	CII dominant	0.99
Gainesville, Florida	Temperate-South	Res dominant	0.643
Garland, Texas	Temperate-South	Res dominant	0.423
Grand Forks, North Dakota	Continental-Central	CII dominant	2.4
Grand Island, Nebraska	Continental-Central	CII dominant	1.147
Green River, Wyoming	Arid	Res dominant	0.402
Greensboro, North Carolina	Temperate-South	Res dominant	0.78
Havre, Montana	Arid	Res dominant	0.467
Hereford, Texas	Arid	CII dominant	1.475
International Falls, Minnesota	Continental-Central	Res dominant	0.454
Iowa City, Iowa	Continental-Central	Res dominant	0.467
Jonesboro, Arkansas	Temperate-South	CII dominant	1.763
Kansas City, Missouri	Continental-Central	CII dominant	1.423
Kearney, Nebraska	Continental-Central	Res dominant	0.467
Kenner, Louisiana	Temperate-South	Res dominant	0.433
Kennewick, Washington	Arid	Res dominant	0.564
La Crosse, Wisconsin	Continental-Central	CII dominant	2.486
Lafayette, Louisiana	Temperate-South	CII dominant	1.099
Lancaster, California	Arid	Res dominant	0.353
Lansing, Michigan	Continental-Central	CII dominant	1.226
Laredo, Texas	Arid	Res dominant	0.524
Las Vegas, Nevada	Arid	Res dominant	0.679
Long Beach, California	Arid	Res dominant	0.443
Los Angeles, California	Arid	Res dominant	0.647
Louisville/Jefferson County metro government (balance), Kentucky	Continental-Central	CII dominant	1.28
Lubbock, Texas	Arid	Res dominant	0.564
Madison, Wisconsin	Continental-Central	Res dominant	0.799
Manchester, New Hampshire	Continental-East	CII dominant	1.266
Marietta, Georgia	Temperate-South	Res dominant	0.63
Memphis, Tennessee	Temperate-South	CII dominant	1.152
Miami, Florida	Tropical	Res dominant	0.89
Milwaukee, Wisconsin	Continental-Central	CII dominant	0.937
Minot, North Dakota	Continental-Central	Res dominant	0.446
Monahans, Texas	Arid	Res dominant	0.221
Montrose, Colorado	Arid	Res dominant	0.393
New York, New York	Continental-East	CII dominant	1.137
Omaha, Nebraska	Continental-Central	Res dominant	0.822
Orlando, Florida	Temperate-South	Res dominant	0.637
Page, Arizona	Arid	Res dominant	0.564
Peoria, Arizona	Arid	Res dominant	0.361
Philadelphia, Pennsylvania	Continental-East	Res dominant	0.891

Phoenix, Arizona	Arid	Res dominant	0.483
Pierre, South Dakota	Continental-Central	Res dominant	0.538
Pikeville, Kentucky	Temperate-South	Res dominant	0.128
Plainview, Texas	Arid	Res dominant	0.762
Pocatello, Idaho	Continental-West	Res dominant	0.407
Ponca City, Oklahoma	Temperate-South	CII dominant	1.88
Poplar Bluff, Missouri	Temperate-South	CII dominant	0.904
Portland, Maine	Continental-East	CII dominant	0.977
Portland, Oregon	Temperate-West	Res dominant	0.647
Pueblo, Colorado	Arid	Res dominant	0.658
Rochester, Minnesota	Continental-Central	CII dominant	1.001
Rochester, New York	Continental-East	Res dominant	0.201
Rock Springs, Wyoming	Arid	Res dominant	0.518
Rockford, Illinois	Continental-Central	Res dominant	0.563
Safford, Arizona	Arid	Res dominant	0.481
Salina, Kansas	Continental-Central	Res dominant	0.467
Salt Lake City, Utah	Continental-West	CII dominant	1.007
San Angelo, Texas	Temperate-South	Res dominant	0.362
San Diego, California	Arid	Res dominant	0.647
San Francisco, California	Temperate-West	CII dominant	2.144
Santa Barbara, California	Temperate-West	Res dominant	0.316
Santa Maria, California	Temperate-West	Res dominant	0.457
Santa Rosa, California	Temperate-West	Res dominant	0.371
Sherman, Texas	Temperate-South	CII dominant	1.415
Sioux Falls, South Dakota	Continental-Central	Res dominant	0.758
Spearfish, South Dakota	Continental-Central	Res dominant	0.831
Spokane, Washington	Continental-West	Res dominant	0.456
Springfield, Ohio	Continental-Central	CII dominant	1.354
St. Peters, Missouri	Continental-Central	Res dominant	0.517
Steamboat Springs, Colorado	Continental-West	Res dominant	0.564
Tacoma, Washington	Temperate-West	CII dominant	1.094
Thomasville, Georgia	Temperate-South	Res dominant	0.765
Tucson, Arizona	Arid	Res dominant	0.405
Tyler, Texas	Temperate-South	Res dominant	0.763
Vacaville, California	Temperate-West	Res dominant	0.347
Valentine, Nebraska	Continental-Central	Res dominant	0.481
Vancouver, Washington	Temperate-West	Res dominant	0.365
Vernal, Utah	Arid	Res dominant	0.6
Watertown, New York	Continental-East	CII dominant	1.052
Wichita Falls, Texas	Continental-Central	Res dominant	0.774
Wichita, Kansas	Temperate-South	CII dominant	1.002
Winston-Salem, North Carolina	Temperate-South	CII dominant	1.097

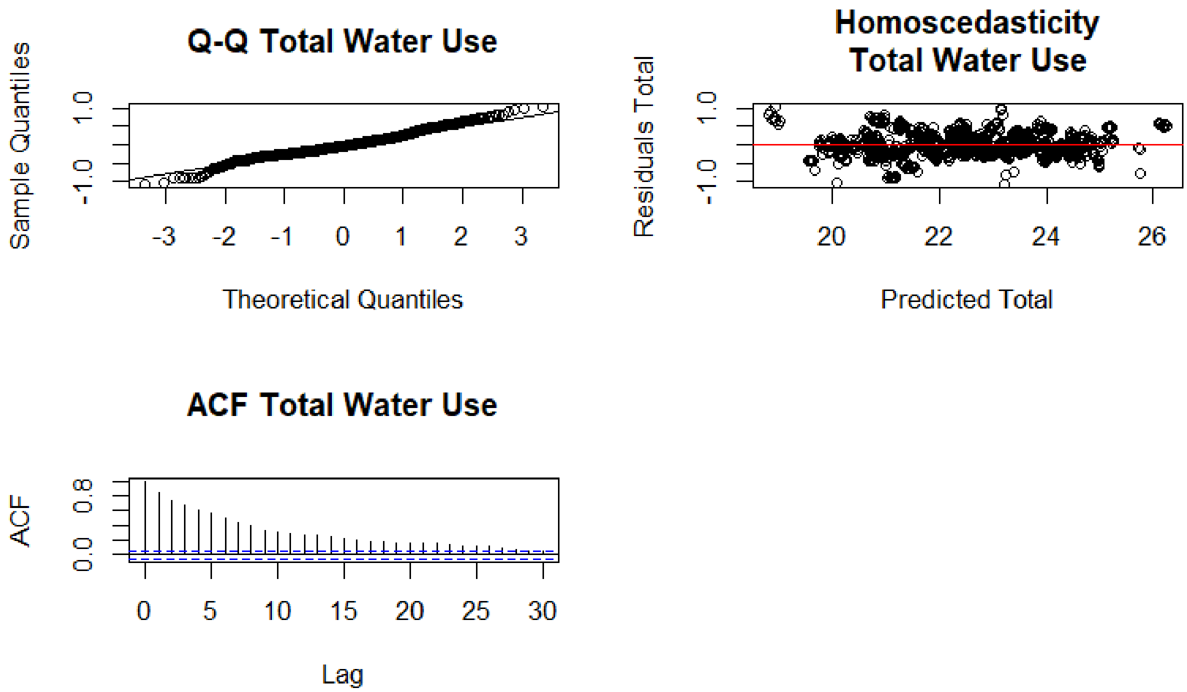


Figure SI-4.1. Test of normality, homoscedasticity, and randomness of the Total water use mixed-effects model's residuals.

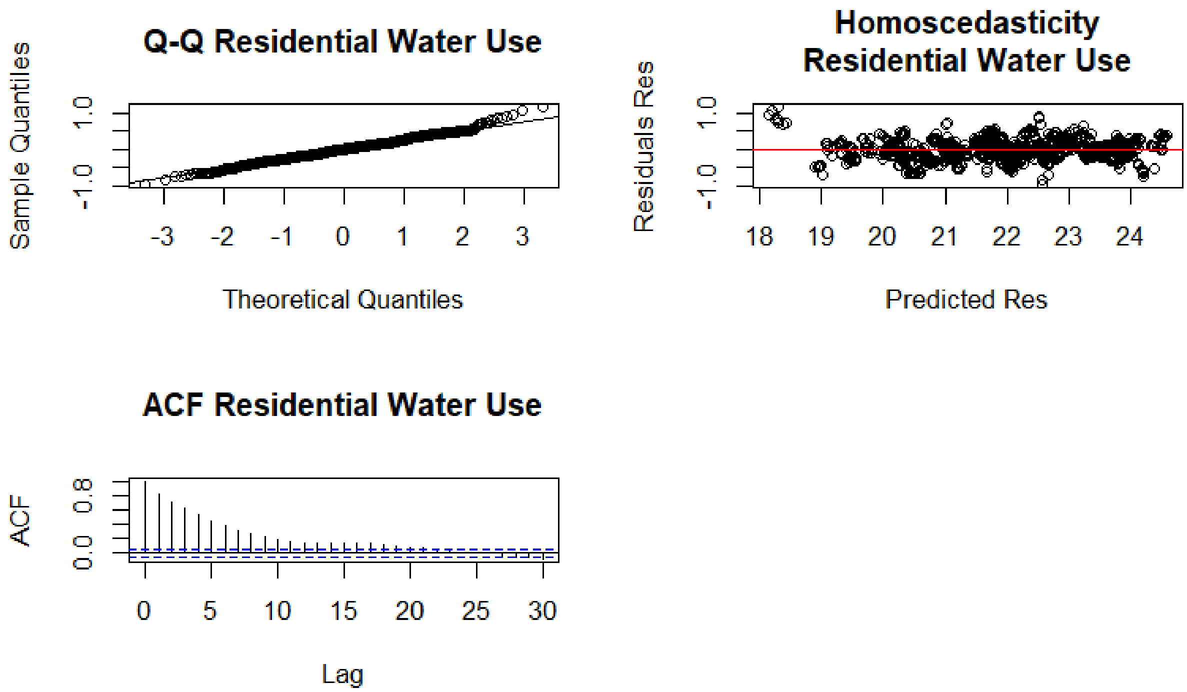


Figure SI-4.2. Test of normality, homoscedasticity, and randomness of the Residential water use mixed-effects model's residuals.

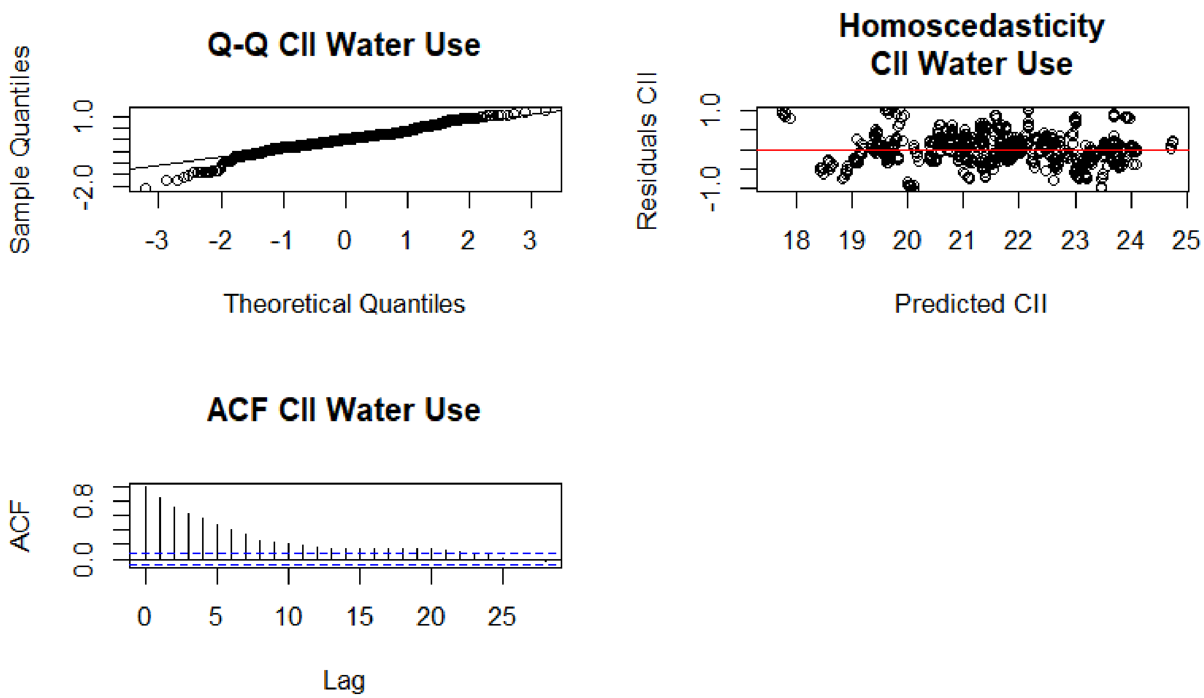


Figure SI-4.3. Test of normality, homoscedasticity, and randomness of the CII water use mixed-effects model's residuals.

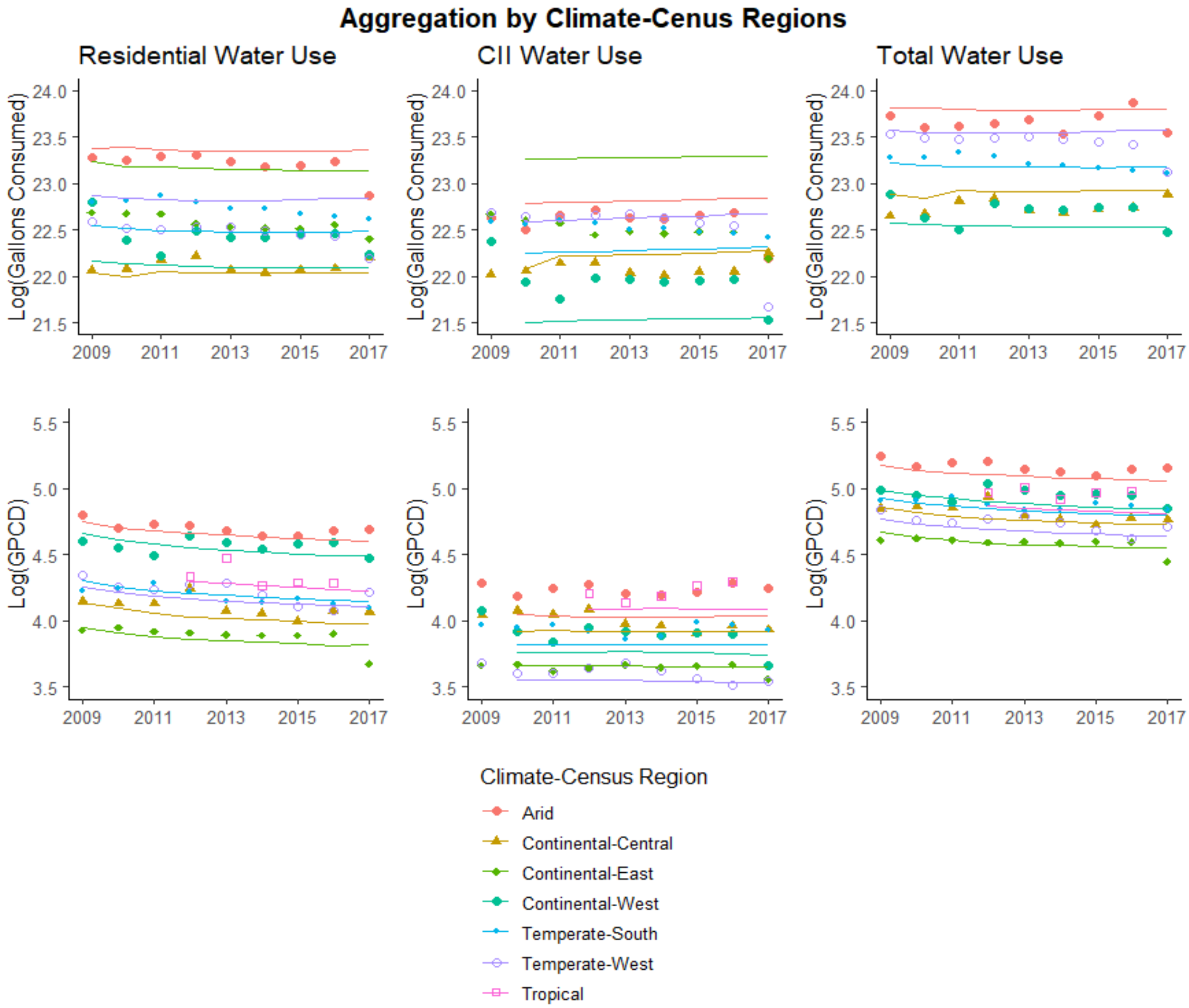


Figure SI-4.4. Subplots showing the time-series style goodness-of-fit of the mixed-effects models on the annual municipal water consumption patterns of cities aggregated using the Climate-Census regional classification. The top row presents the water consumption patterns using the annual gallons consumed metric and the bottom row using the GPCD metric.

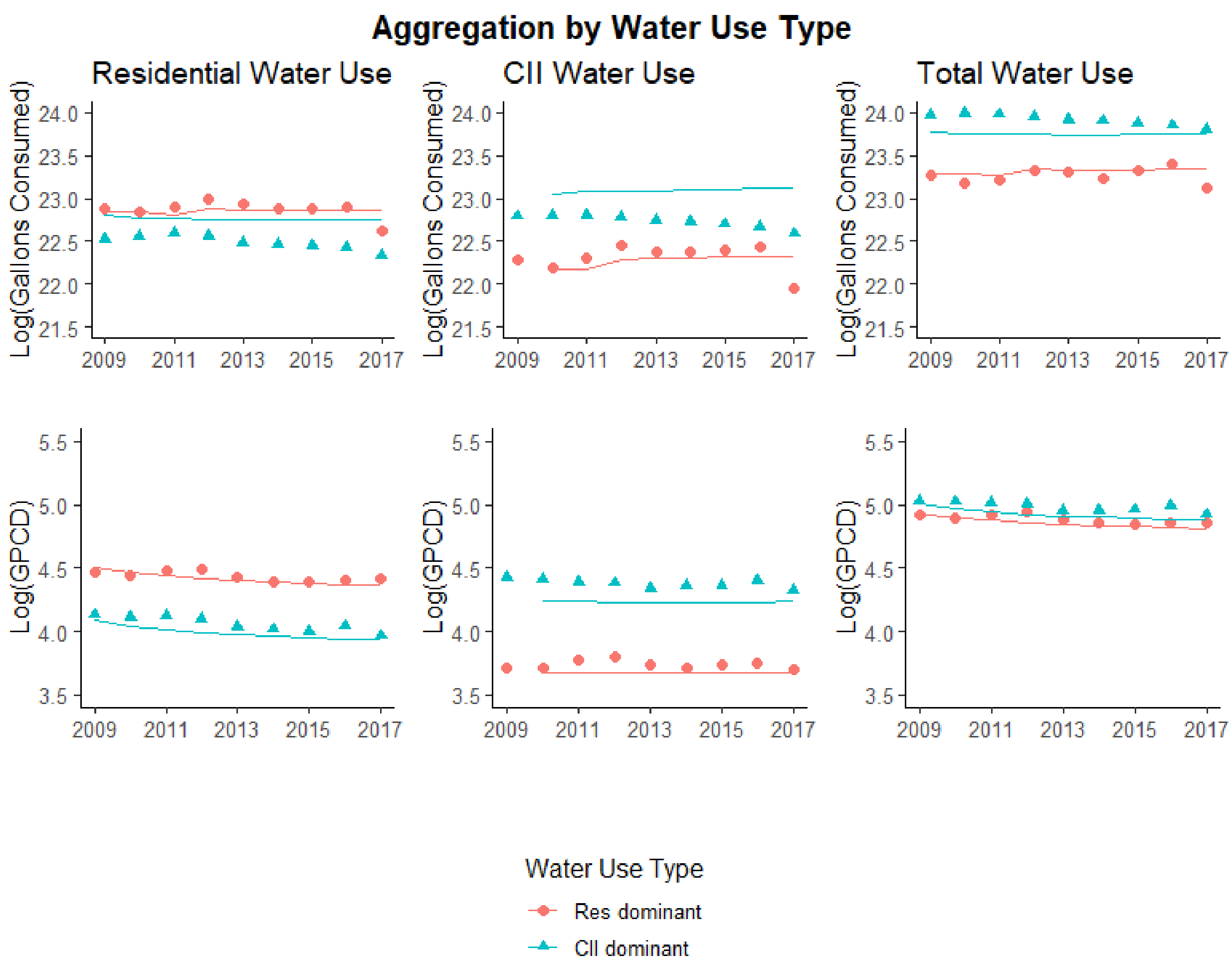


Figure SI-4.5. Subplots showing the time-series style goodness-of-fit of the mixed-effects models on the annual municipal water consumption patterns of cities aggregated based on their Water Use Type distinction. The top row presents the water consumption patterns using the total gallons consumed metric and the bottom row using the GPCD metric.

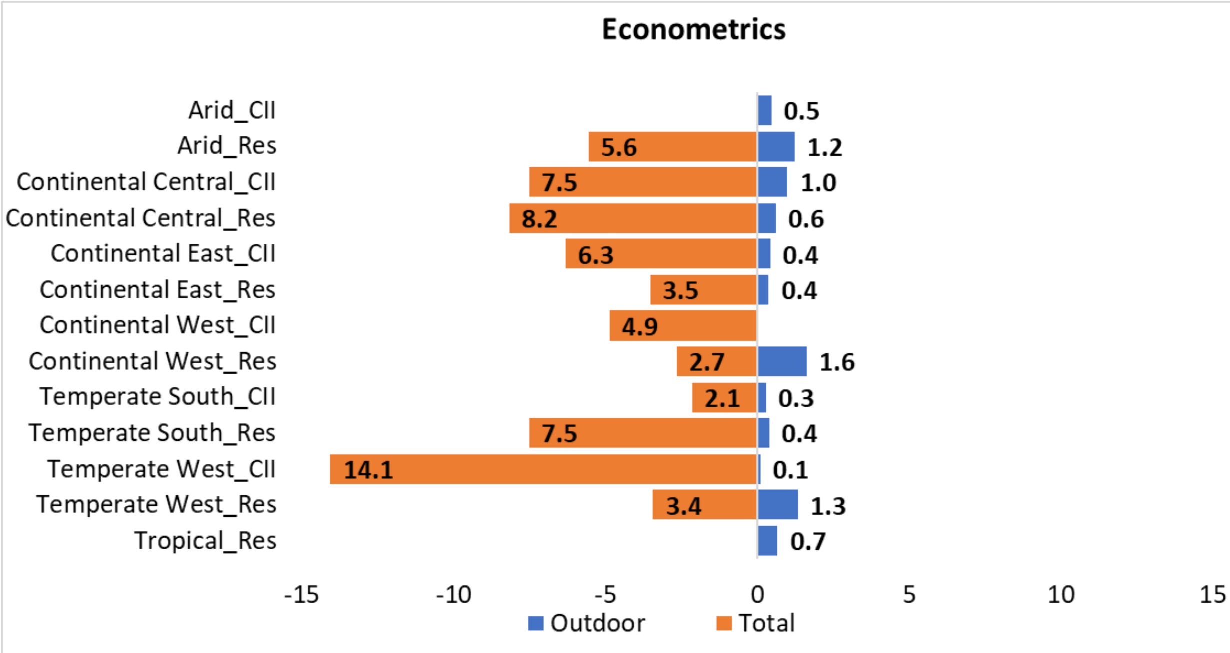


Figure SI-4.6. Magnitudes of difference in the annual estimation of Total and Outdoor water uses in GPCD by the linear water use model and the econometrics water use model with scaling relationships incorporated. Same variables used the econometrics scaling model are used for the linear model as well.

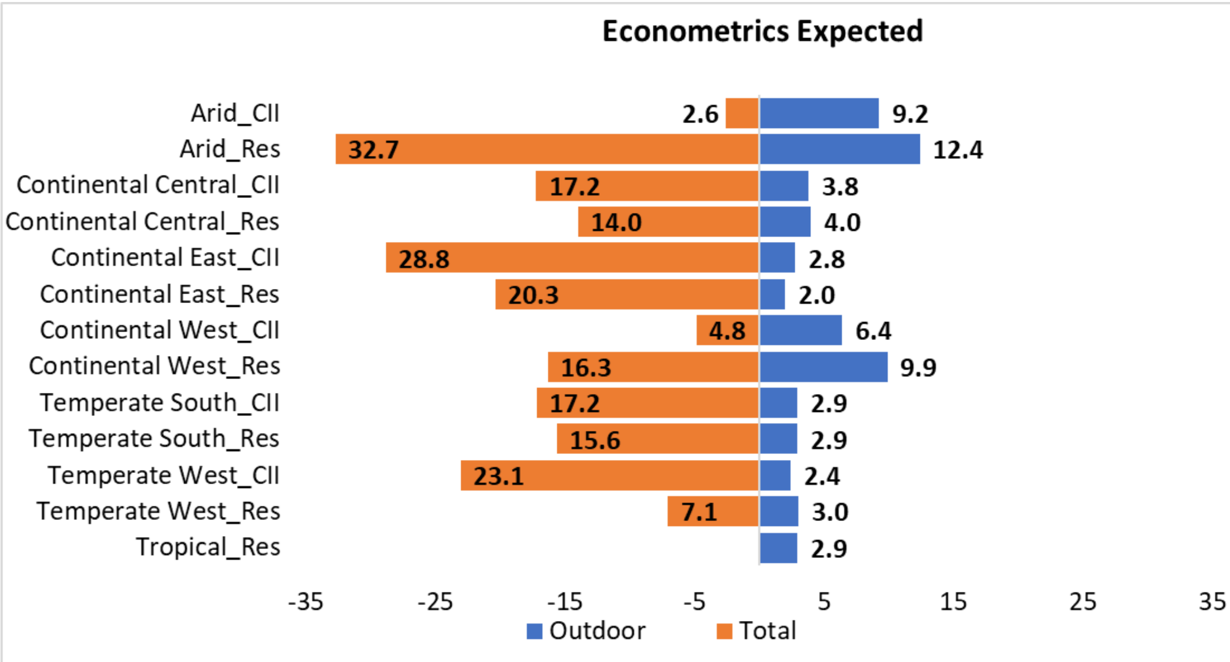


Figure SI-4.7. Magnitudes of difference in the annual estimation of Total and Outdoor water uses in GPCD by the linear water use model and the econometrics water use model with scaling relationships incorporated. Same variables used the econometrics scaling model are used for the linear model as well, and expected values of the model variables are used in the model estimation of water consumption.

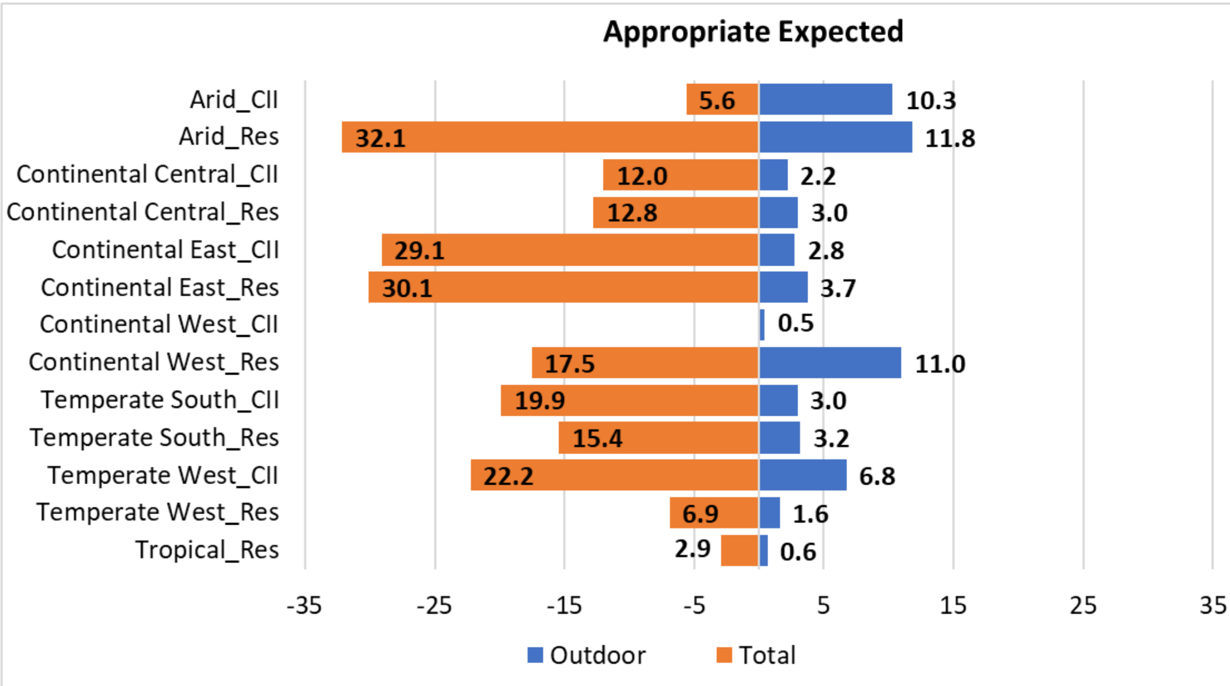


Figure SI-4.8. Magnitudes of difference in the annual estimation of Total and Outdoor water uses in GPCD by the linear water use model and the econometrics water use model with scaling relationships incorporated. Expected (average) values of the model variables are used in the model estimation of water consumption, and appropriate variables are used to build the linear model.

APPENDIX – D

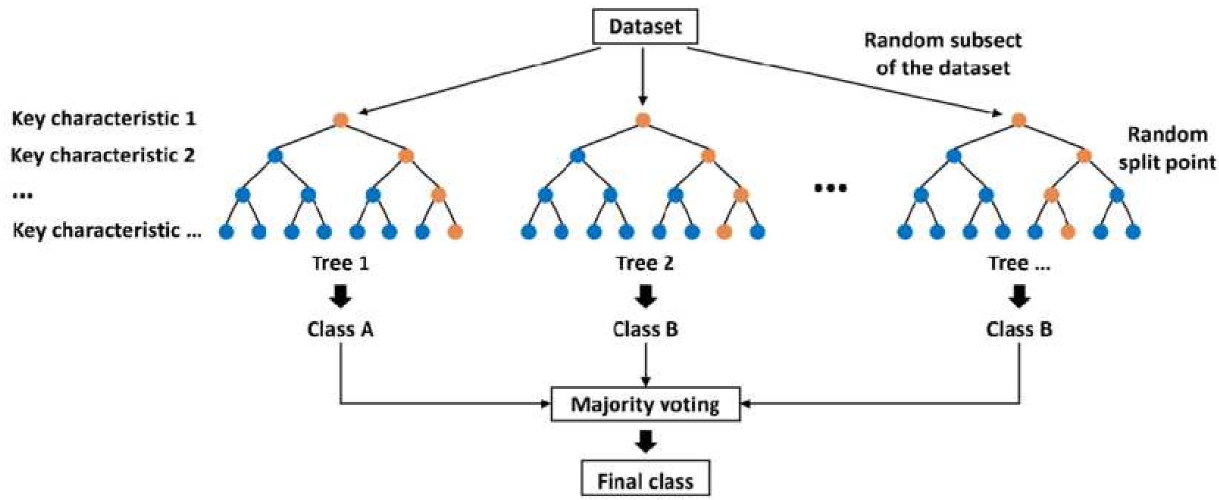


Figure SI-5.1. Structure of Extra Trees Regressor (Geurts et al., 2006). Extra Trees constructs the set of decision trees by randomly selecting a subset of the dataset. In the training of each decision tree, the split point to divide the tree at a particular node is randomly selected.

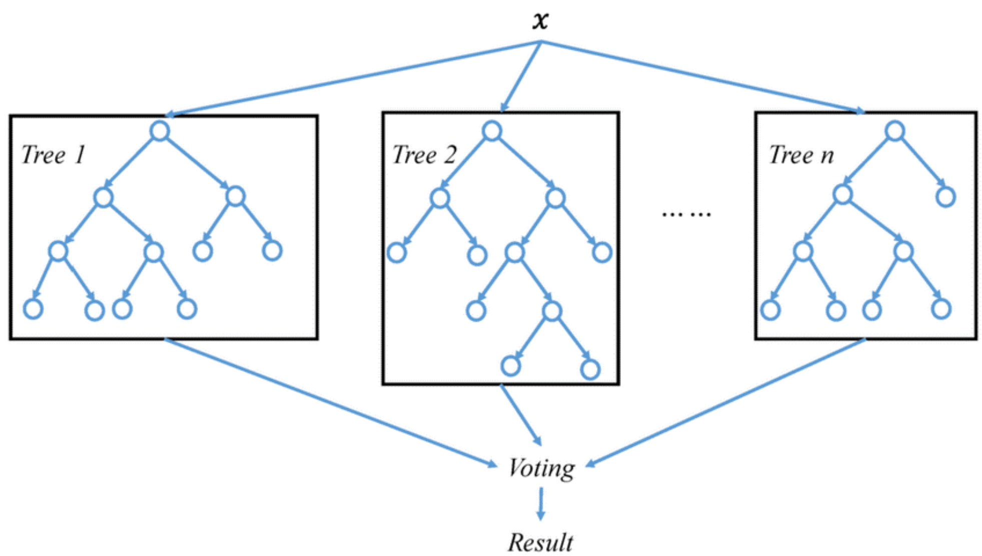


Figure SI-5.2. Structure of Random Forest Regressor. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest (Breiman, 2001).

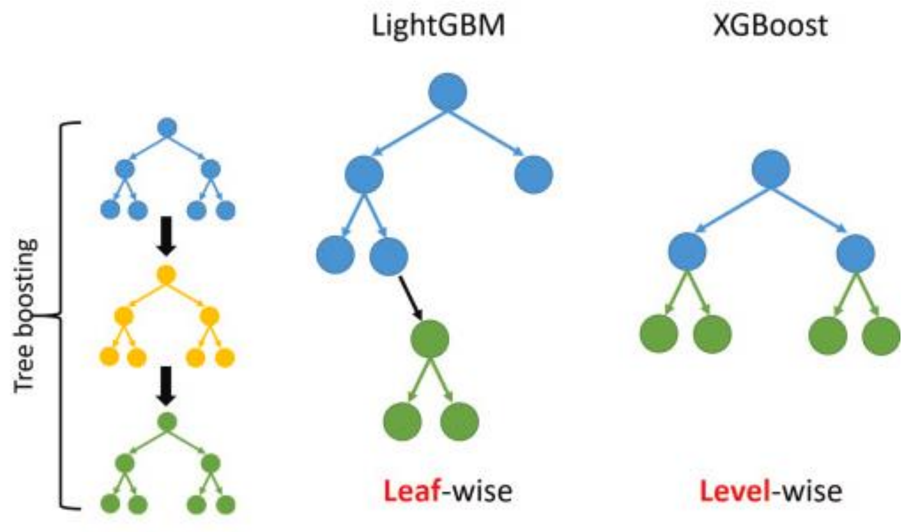


Figure SI-5.3. Structure of Light and Extreme Gradient Boosting Regressors (LightBGM and XGBoost). LightGBM and XGBoost follow the same principle except that LightGBM uses a leaf-wise policy to grow gradient boosting decision trees, which splits the tree only on the best nodes that can bring maximum reduction of the loss function, whereas XGBoost implements a level-wise policy that leads to a symmetrical tree structure (Bentéjac et al., 2020).

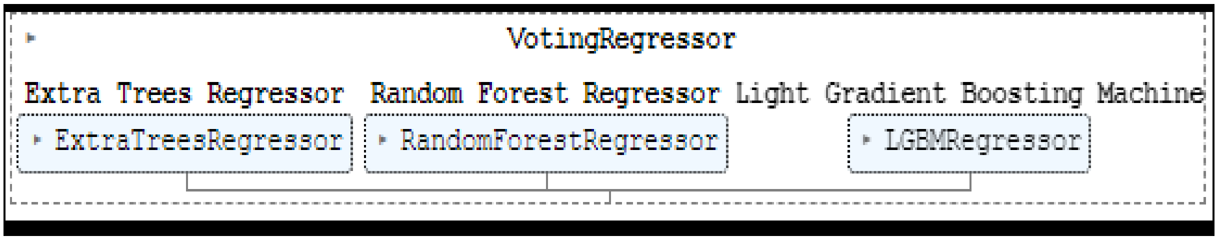


Figure SI-5.4. Pipeline of the Voting Regressor or the “blended” ML model that combines the predictive powers of the top 3 ML algorithms used to fit the Total monthly water use dataset.

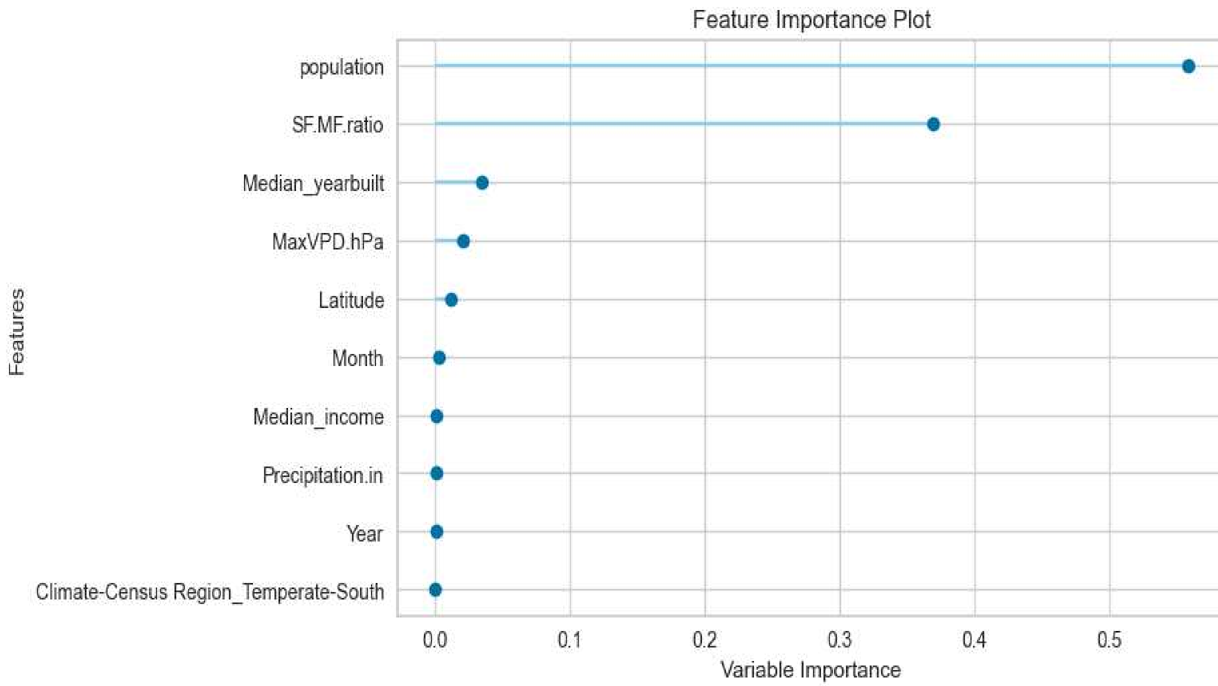


Figure SI-5.5. Feature importance plot of the Random Forest Regressor model.

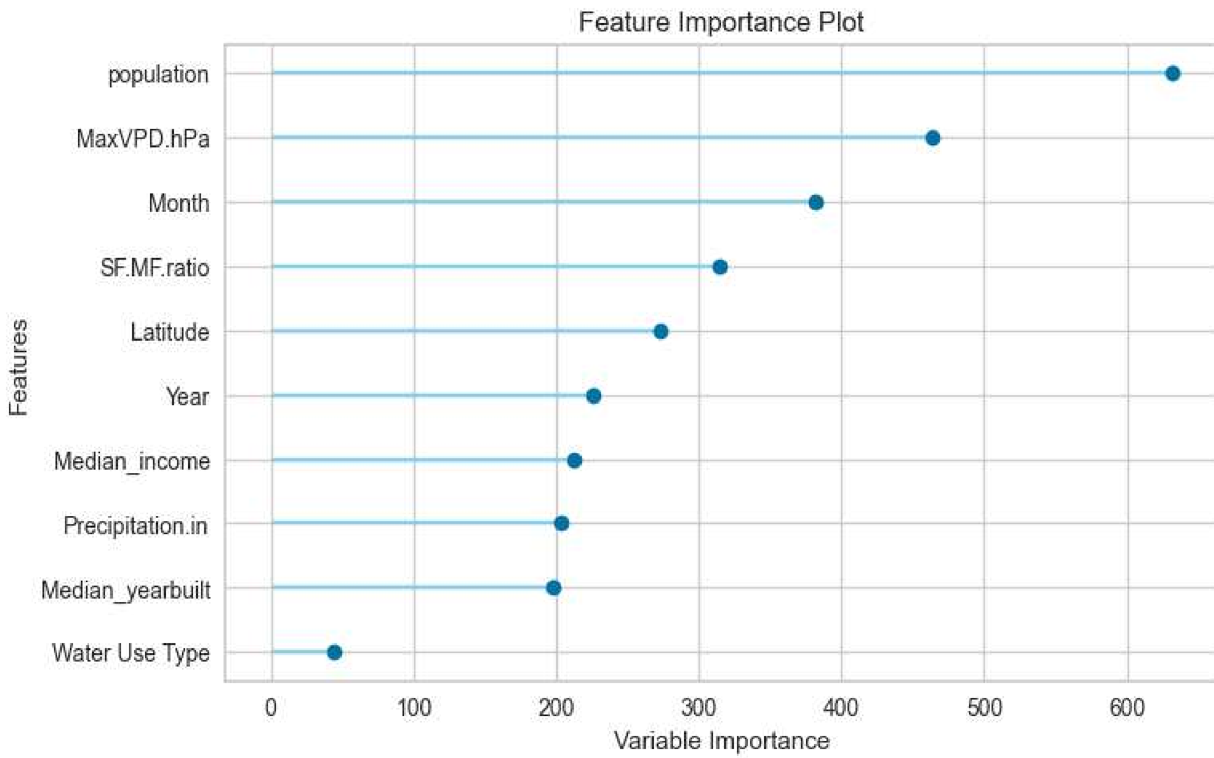


Figure SI-5.6. Feature importance plot of the Light Gradient Boosting Regressor model.