International Comparative Analysis of Urban Water Systems

by

Karen Marie Noiva

Bachelor of Science in Mechanical Engineering, Massachusetts Institute of Technology (2008)
Master of Science in Building Technology, Massachusetts Institute of Technology (2011)

Submitted to the Department of Architecture
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Architecture: Building Technology

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2018

Author . . . .
Karen Marie Noiva
Department of Architecture
September 15, 2017

Certified by . . . .
John E. Fernández
Professor of Architecture and Building Technology
Thesis Supervisor

Accepted by . . . .
Sheila Kennedy
Professor of Architecture
Chair, Department Committee on Graduate Students

Thesis Committee:

John E. Fernández
Professor of Architecture and Building Technology
Massachusetts Institute of Technology
Thesis Supervisor

Richard de Neufville
Professor of Data, Systems, and Society
Massachusetts Institute of Technology
Thesis Reader

James L. Wescoat
Aga Khan Professor of Urban Studies and Planning
Massachusetts Institute of Technology
Thesis Reader

International Comparative Analysis of Urban Water Systems

by

Karen Marie Noiva

Submitted to the Department of Architecture
on September 15, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Architecture: Building Technology

Abstract

This dissertation presents a new approach to structuring global diversity for a large number of urban water systems, so that trends observed in a small number of cases can contribute to a more general understanding of the spectrum of contemporary sustainability challenges faced by cities around the world. The two-part approach first uses a large number of cities (large-n, i.e. $n = 142$) to identify a typology which is used to guide the choice of two cases (small-n, i.e. $n = 2$) for further analysis.

In the first part of the approach, I compare a large number (large-n) of urban water systems. Simple profiles of key attributes of urban water supply and demand—population (N), water use intensity (w_N), and net annual water balance data (q_{Net})—are assembled from common global databases. Univariate and bivariate methods are used to identify global trends. I introduce two new indicators that benchmark urban water use intensity against climatic availability: the Water Use and Climate Index (WUCI, with units of m^2) and the Potential Self-Sufficiency Ratio (R_{SS}, unitless) and find that 65% of cities in the study have $R_{SS} \geq 1$. I then use exploratory statistical clustering algorithms to identify six type of urban water systems profiles, ranging from small, wet cities with low WUCI and high R_{SS} to large cities with high water use intensity, high WUCI, and lower R_{SS}.

In the second part of the approach, I demonstrate the use of that typology in framing case study choice for small-n international comparative analysis of urban water systems. I choose Los Angeles and Singapore from Type 4, which have large populations and high water use intensity but different climates. I apply univariate and bivariate methods to identify trends over time in water system profiles of LA and Singapore. Calculating WUCI and the Potential Self-Sufficiency Ratio for the two cases provides insight into historical behavior and future targets. Finally, I use these results to construct simple simulations to assess past behavior and future targets.

Thesis Supervisor: John E. Fernández
Title: Professor of Architecture and Building Technology

Acknowledgments

First, I would like to thank my advisor, Professor John E. Fernández for providing me with the opportunity to study sustainability through the lens of urban metabolism. Throughout the Master's and PhD, John encouraged my intellectual exploration while providing insight and steady guidance. He also provided unwavering support when I faced trying personal challenges.

Next, I would also like to thank the other two members of my committee — Professor James L. Wescoat and Professor Richard de Neufville — for pushing me intellectually, while offering encouragement and mentorship. This work would not have been completed without you.

I enjoyed working closely with many colleagues throughout the years — with Jonathan Krones, Melanie Wong, and Yuliya Kalmykova on various projects at MIT; my collaborators on the project with Rambøll—Alex Marks, Smita Rawoot, Matthias Wörlen; my IBM team— Shema Alain, Antony Kaguara, and Nathan Wangusi; my colleagues at the Sustainable Cities RCN—Elliot Cohen, Josh Sperling, Kate Wassel, and Janet Reyna.

I have also enjoyed learning from numerous other colleagues in the Building Technology program—David Q, David B, Tarek, Adam, Noel, Teri, Caitlin, Kate, Zahraa, Alejandro, Tim, Ale, Tea, and Sian— as well as numerous others, such as Daniel, Tamas, Juanjo, and Marianna.

I have had numerous mentors offer experience and advice along the years. Aside from those already mentioned, I would especially like to acknowledge including Uyi Stewart and Kamal Bhattacharya at IBM Research; Herbert Dreiseitl and Bettina Wanschura at Ramboll; Profs. Esfatih Eltahir, Dennis McLaughlin, Heather Lechtman, and Donald Sadoway at MIT; Felice Frankel and Rebecca Rosenberg at Harvard; and Prof. Prathiba Nerurkar at University of Hawai'i, Mānoa.

I was fortunate to have been supported financially throughout most of my graduate studies—first on a fellowship from MITEI, then as a research assistant with funding for various periods through the MIT-Singapore program and the Rambøll Foundation. I am also grateful to the Department of Architecture for funding during my two non-resident years at MIT.

Numerous staff members have supported me throughout the years. I would like to acknowledge the Architecture crew—Renée Caso; the BT support crew— Kathleen Ross, Alex Golledge, Alexandra Mulcahy; and the CRON crew—Duncan, Rickie, Phil, Tom, and Eduardo—all of whom have provided very welcome assistance throughout my time here.

I would like to acknowledge my immediate and extended family, for their financial, emotional, and moral support throughout the years, especially my grandparents—Linda Hsiao Ming Chiang Tse, Professor David Chen, Firmo Reis and Carol Noiva, and Ruth Cunningham Lyons; my parents—Susan Chen and Jeff and Connie Noiva; and my brothers—Eric, Trevon, and Kyle Noiva. My close friends, Kate and Darren, Smalltime, Ellen, Erik and Marty, Uche, Kat, Joana, Tarek and Nadja, Vivek, and Jeff. All of my lovely aunts, uncles, and cousins. And Tele, Cedar, Maple, Lance, Phoebe, Bear, and Tucker, for moral support and company on walks.

I would like to acknowledge and thank the University of Hawai'i's College of Tropical Agriculture and Human Resources for giving me the opportunity to study in Hawai'i and introduced me to the concepts of sustainability way back in 2003.

I dedicate this thesis to the memory of my beloved grandfather, Firmo Reis 'Pop' Noiva, who passed away suddenly on August 21, 2017. Pop was a man who knew what was important in life: family and friends, love, generosity, meaningful work—and a good sense of humor. Although there is never a good time to say goodbye, Pop lived a full life, and was loved by many people. Pop—I love you, I miss you, and this is for you.

Contents

List of Figures

List of Tables

Acronyms

CA	California. 147, 148
CASWP	California State Water Project. 147
CGIAR	Consortium of International Agricultural Research Centers. 160
CRS	Coordinate Reference System. 266
CWB	annual climatic water balance, i.e. or annual balance, net annual balance, or climatic water balance. 79, 282
GDP	Gross Domestic Product. 75, 81
HDI	Human Development Index. 130
IBNET	International Benchmarking Network for Water and Sanitation Utilities. 70, 76, 77, 133, 211
IQR	Inter-Quartile Range. 90, 92, 93, 106–109, 125–127, 164, 201, 213, 217
IWRM	Integrated Water Resources Management. 34, 49, 52, 61, 73
LA	Los Angeles. 140, 214–218, 220, 267
LA Aqueduct	Los Angeles Aqueduct. 145
LADWP	Los Angeles Department of Water and Power. 145–148, 150, 151
maf	million acre-feet. 147, 148
MAV	Moving Average Value analysis. 169, 187, 188
MDSD	Most Different System Design. 56, 58, 59
MIPS	Material Input required Per unit of Service (or good) produced. 156, 157
MIT	Massachusetts Institute of Technology. 260
MMA	Mill's Method of Agreement. 56, 58, 59
MMD	Mill's Method of Disagreement. 56, 58, 59
MSSD	Most Similar System Design. 56, 59
MWDSC	Metropolitan Water District of Southern California. 145, 147, 148, 150–152

OECD	Organisation for Economic Co-operation and Development. 28
PUB	Singapore's water agency, the Public Utilities Board. 145, 149, 151, 152, 159
qq	quantile-quantile. 70, 78, 82, 161
SgD	Singapore dollars. 148, 149
SUWM	Sustainable Urban Water Management. 27, 28, 36, 37, 42–44, 46, 49, 53, 55, 61, 63, 65–68, 73, 111, 181, 210, 217, 220
SWRM	Sustainable Water Resources Management. 32, 34, 36, 41
TARWR	Total Annual Renewable Freshwater Resources. 38
t-SNE	*t*-Distributed Stochastic Neighbor Embedding. 45, 70, 83, 85, 86, 88–90, 104, 112, 113, 130, 132, 211, 221
UrbMet	Urban Metabolism Group at MIT. 69, 70, 72, 73, 76, 85, 104, 105, 107–109, 111–113, 125, 127, 128, 131, 137, 211, 219, 255
US	United States of America. 112, 148, 215, 260
UWS	Urban Water System. 27, 28, 38, 41, 43–46, 49, 50, 52–55, 60–62, 66, 68, 73, 111, 112, 129, 209–212, 220–222
WebWIMP	Web-based, Water-Budget, Interactive, Modeling Program. 70, 73, 74, 76, 142, 143, 160, 211, 221, 256, 261, 262, 269
WHO	World Health Organization. 29, 30, 108, 109, 137
WRM	Water Resources Management. 27, 28, 32, 33, 39, 41, 44–46, 49, 50, 53, 55, 60, 61, 67, 70, 141, 188, 191, 207, 210, 212–215
WRS	Water Resources System. 41, 42
WSP	Water Service Provider. 35, 131
WUCI	Water Use and Climate Index. 70, 78, 124, 143, 157, 158, 189, 212, 214, 220
WUI	Water Use Intensity, i.e. average per capita water use. 282

Symbols and Notation

Sign	**Description**	**Unit**
A_N	City area	m^2 or km^2
q_{Net}	Net climatic water balance (height)	$m \cdot yr^{-1}$
N	Population	$capita$
k_N	Population proportional growth constant	$\% \cdot yr^{-1}$
q_P	Precipitation (height)	$m \cdot yr^{-1}$
R_{SS}	Self-sufficiency ratio	*dimensionless*
I_{UC}	Total Water Use and Climate Index	m^2 or km^2
w_N	Water use intensity	$m^3 \cdot yr^{-1}$
i_{UC}	Water Use and Climate Index	$m \cdot yr^{-1}$

Glossary

bivariate An analysis that focuses on two variables. Common bivariate methods include calculation of correlation coefficients or visualization as a scatterplot. 70, 78, 83

clade A branch in a dendrogram. 90

hydrologic unit The physical system, delineated by flow lines, from which water drains into a specified hydrologic feature, such as a groundwater table, river, or lake. 153

IPAT The IPAT equation is an identity equation, frequently used as an heuristic in sustainability studies, linking environmental impact (I) with drivers such as population (P), affluence (A), and (T: $I = P * A * T$. Affluence is often interpreted as GDP per person (e.g., $GDP/capita) and technology is often interpreted as the environmental impact generated in the production of a unit of GDP (e.g., environmental impact units/$GDP). 75, 219

large-*n* referring to a large number of cases in case study analysis. 41–45, 54, 60, 62–66, 69, 70, 72–76, 78, 85, 88, 101, 103, 129, 131–134, 211, 222

Most Different System Design A design for comparative research that compares cases with dissimilar system attributes and seeks to demonstrate that a relationship between dependent and independent variables holds across diverse conditions and is therefore robust. 57, 58

Most Similar System Design A design for comparative research that compares cases with similarities in dependent variables in which similarities in system attributes are viewed as controlled and dissimilarities are used to explain any differences observed in the dependent variables. 56

NEWater Singapore's recycled water system. 145

quantile-quantile plot A type of plot for visualizing distributions of data where actual observations of actual data are ordered by value and plotted against an equal number of observations, also ordered by value, that have been drawn from a theoretical distribution (such as the normal distribution). 45, 82, 219

`R` A functional language and environment for statistical computing (for more information, see `www.r-project.org`). 47, 61, 74, 80, 82, 83, 87, 89, 102, 131, 141, 159–162, 164, 194–196, 200, 201, 206, 214, 217, 218, 220, 221, 256, 283, 287, 293

small-*n* referring to a small number of cases in case study analysis. 42–44, 47, 54–56, 60, 63, 65–67, 69, 72, 76, 130, 134, 210, 212, 213, 215, 216, 219, 220, 222

univariate An analysis that focuses on a single variable. Common univariate methods include calculation of mean, standard deviation, or quantiles and visualization methods such as histograms. 70, 78, 83

Chapter 1

Introduction

1.1 The Global Fresh Water Challenge

In the next 35 years, the world is expected to acquire an additional 2.54 billion urban residents [309]. Within the next 15 years alone, water supplies must be expanded an additional 80% to meet the needs of growing cities [196]. Traditionally, most cities have relied on importing water from distant watersheds or pumping groundwater aquifers. However, the traditional approach will become increasingly difficult as 41% of global land area has already been tapped for municipal consumption and as pressures on freshwater resources—such as climate change, urbanization, and competition from food and energy sectors—mount [327, 154, 196, 293, 328]. This changing scene has ushered in a new era for Water Resources Management (WRM), one in which sustainability, environmental needs, and social equity are discussed alongside demand and supply projections [312, 196, 295, 29, 314].

These pressures have led to increased competition for freshwater sources, prompting cities to turn to less conventional approaches to WRM [100, 356]. Cities are increasingly looking to expand their supplies by treating lower quality sources—such as seawater, urban runoff, and wastewater—and to improve their use of their existing water supplies [278, 200, 255, 196, 120, 316]. However, related technologies and policies are often considered risky and relatively unproven in comparison with more customary approaches to WRM [222, 120]. Thus, the coming decades will be a time of experimentation for Urban Water System (UWS), as cities try to adapt innovative methods to their specific needs and to identify a portfolio of technologies and policies appropriate to their socioeconomic, cultural, political, environmental, and climatological contexts [300, 203, 41, 258, 188].

However, experimentation engenders uncertainty, delaying or deterring cities from transitioning to Sustainable Urban Water Management (SUWM) [26, 130, 188, 122, 185]. All cities require water for survival, and water security is paramount; managers and decision-makers cannot risk aberrations in supply [128, 119, 38, 278, 43, 113, 118, 136, 196, 222]. Water infrastructure is also costly, adding a financial dimension to the risk [183, 257, 188, 227, 100, 196, 316]. Further complicating matters, the new paradigm for WRM requires integration with dimensions unfamiliar to the conventional approach, such as land use, public outreach, landscape design, and ecology [130, 227, 196, 356]. Complicating matters, SUWM often requires a restructuring of local water governance, management, and finance [150, 53, 189, 312, 296, 81, 267, 227, 356].

Cities would benefit from methods that help them to identify portfolios of technologies and policies appropriate to their needs [245, 186, 196, 319, 131]. These methods could help decision makers to characterize the uncertainty and risk in adapting unfamiliar technologies and policies to new operating conditions [245]. The challenges that have delayed widespread adoption of SUWM have also fomented a remarkable diversity of models, methods, and approaches [215, 138, 75, 258, 285, 188, 16, 179, 320]. On the one hand, this spate of creativity will undoubtedly be essential to solving the emerging urban water challenges; on the other, the plenitude magnifies the uncertainties for decision-makers, since no single unifying or prescriptive approaches have yet emerged [258, 188, 100].

The underlying complexities of water resource systems, make it difficult to both account for the underlying variability in social, ecological, economic, and hydrological and climatic processes without resorting to panaceas [245]. Many of the more comprehensive methods and models are developed for a particular city or region and are difficult to adapt to new contexts. Due to the complexity of water systems and WRM, these approaches are typically not perfectly suited to new cases. Neither the diverse underlying factors nor our understanding of UWS and SUWM is stationary in time, so even a model that was developed for a particular city may no longer be used for the same application at a later date. Because of this, approaches have proliferated, with both similarities and differences.

Rigorous comparative analysis of case studies can help to break this impasse but remains relatively undeveloped in the study of UWS [174, 219, 116, 341, 218].

1.1.1 The Global Crisis

One of the most pressing challenges of our time is how to sustainably and equitably supply potable water to humanity in the coming years and decades [314].

In 2015, members of the World Economic Forum voted the risk of a global "water crisis" as the threat with the greatest potential impact—beating out "*spread of infectious disease*", *weapons of mass destruction*", "*interstate conflict*", and even "*failure of climate-change adaptation*" for the dubious distinction [337]. However, the threat of a global crisis in freshwater resources has been looming for some time [287, 110, 311, 259, 107, 111, 161]. Many people and places around the world already experience water stress, and experts are increasingly concerned that water scarcity in 2050 could be widespread [312, 314, 196, 314, 206]. Two-thirds of the global population already live in areas with severe physical water scarcity at least one month of the year, while 0.5 billion people experience such conditions throughout the year [206]. The Organisation for Economic Co-operation and Development (OECD) has projected that by 2050, with business-as-usual, 4 billion people will live in "severely water-stressed" river basins [100, p. 612].

Pressures on freshwater resources have become increasingly acute and continue to mount [287, 110, 259, 107, 111, 311, 196, 313]. Human appropriation of freshwater resources has already reached unprecedented levels, driven by population growth and economic development [287, 110, 259, 311, 312, 317, 313, 314, 314]. Socioeconomic development has not only driven increased water use; it has also engendered changes in land use and climate that have altered local and global hydrological processes, dire consequences for water management [259, 239, 202, 272, 313, 196, 314]. Many freshwater sources around the world have been depleted or contaminated by unsustainable water withdrawals, increased erosion, agricultural runoff, and industrial byproducts [92, 202, 302, 176, 330, 313, 196, 314]. Arid and semi-arid regions continue to withdraw large volumes of water for agriculture, depleting aquifers and leading to land subsidence [189, 147, 142, 302, 104, 89, 134, 114, 206]. Deforestation, habitat destruction, and surface impermeability have negatively impacted filtration and storage and increased the risk of pollution, flooding, and drought [253, 189, 312, 317, 313, 66, 272].

Water demand is expected to rise with population growth, increasing industrialization, and rising affluence [44, 310, 309]. While global population more than doubled between 1950 and 2014—from 2.52 billion to 7.24 billion—it was outpaced by growth in water use [110, 111, 337, 309]. This rise in per capita water use has been attributed to increasing affluence, and affluence has caused changes in diet as well as increased consumption of other goods and services, all of which entail some degree of water use [226, 111, 284, 133, 37, 313, 162].

Global population has been projected to reach 9.6 billion by 2050, with over 98% of this growth is expected to occur in developing countries [110, 111, 337, 309]. While in some affluent places in the world, total and per capita water use have leveled off or showed signs of decreasing, there is a large latent demand in developing countries, which are already home to over 80% of humanity [274, 139, 232, 77, 337, 309]. The majority of residents in developing countries are still impoverished, and many cannot afford to consume even the 20 liters per capita per day recommended by the World Health Organization (WHO) for basic health and sanitation [105, 108, 143, 345, 314, 282]. Exacerbating the development water challenge are the needs of the water-intensive food and energy sectors [104, 12, 273]. Therefore, even if water-efficient technologies become widespread, ensuring adequate water, food, and energy to all of humanity necessitates an increase in average per capita water [207, 134, 345, 243, 36].

1.1.2 The Urban Challenge

Of particular concern are cities, which are home to over half of humanity and which generate an estimated 47% of global GDP [196, 309].

In the coming decades, urban areas around the world face numerous water resource challenges, including rising water demand and the need for additional supplies, aging infrastructure, deteriorating water quality and environmental degradation, climate change, flooding, social equity, and competition for water from the food and energy sectors—to name a few [172, 312, 317, 302, 334, 200, 269, 313, 313, 14, 100, 333, 314, 315]. The reliability and integrity of urban freshwater resources are increasingly threatened by the general pressures on global freshwater resources [302, 334, 271, 4, 146, 198, 337, 117]. Compounding these issues, the infrastructure shortfall has grown since water infrastructure investments still lag growth in water demand [313, 346, 314, 315].

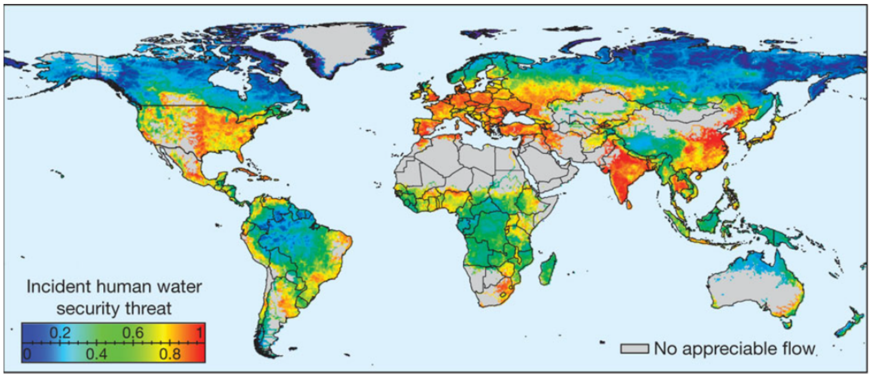
Urban areas tend to be the most water-scarce, due to high levels of water demand concentrated in a relatively small geographical area [239, 334, 196, 206]. An estimated 150 million urban residents already live in cities with perennial water shortage—defined by McDonald et al. as "having less than 100L per person per day of sustainable surface and groundwater flow within their urban extent"—and this number has been projected to rise to at least 1 billion people by 2050 [197].

At the same time, cities tend to be more affluent than more rural areas, and can therefore afford higher levels of water infrastructure, which helps mitigate conditions of water scarcity [334, 196, 198, 313]. The world's largest cities were found to move 183 km^3 yr^{-1} a distance of 27,000 $\pm$ 3800 km, with an upstream contributing area of 41% of global land area [198]! Figure 1.1 depicts the conventional view of urban water supply as a linear process: "raw" water is first obtained from surface water and aquifers; it then may be conveyed to storage in reservoirs or tanks before being treated and distributed to users. Wastewater from these users is then collected along with stormwater (often in a combined system) and is then treated before being released downstream [119]. Storage infrastructure helps reduce the variability of water supplies to precipitation, conveyance helps to transport distant water sources into the homes of many, and water treatment helps the safety and usability of lower-quality sources [107, 313, 346, 222, 315].

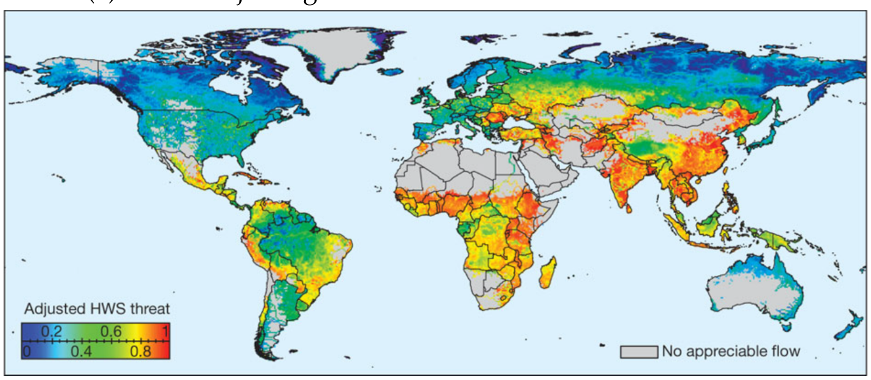
Levels of water infrastructure are not evenly distributed throughout the world, with cities in developing countries at especially high risk[1] [271, 346]. For instance, Vörösmarty et al. found that 80% of the global population in the year 2000 lived in areas with high threat levels as measured by a cumulative index of 23 stressors of water security—such as pollution, habitat destruction, and consumptive losses, and competition for water resources between sectors (as seen in Figure 1.0) [334]. After adjusting for estimated benefits from water infrastructure, it was found that areas with high levels of water infrastructure investment had reduced exposure to high incident threats by as much as 95%, benefiting 850 million people [334]. However, population growth in developing countries has tended to outpace investment in infrastructure, leaving more than 3.4 billion people at high risk [334].

The global shortfall in urban water infrastructure is large and growing [121, 196, 314, 315]. Throughout the 21st century, experts have projected growth in employment opportunities and population to be centered in cities [309]. Between 2014–2050, the urban population is expected

[1]Between 1990 and 2012, the number of urban residents without access to improved drinking water increased by 34%, from 111 to 149 million [314]. In sub-Saharan Africa, one of the most rapidly urbanizing areas of the world, the number with piped access decreased from 42% to 34% [314]. Many urban residents around the world consume less than the 20 L/cap/day recommended by the WHO for "basic access"—substantially less than the 100–200 L/cap/day recommended for "optimal access" [105, 344, 347, 313]. Unfortunately, the burden of infrastructure shortfall is largest in the developing world and generates enormous healthcare costs by increasing the prevalence of waterborne illnesses, which further impairs economic growth rates in a vicious cycle [2, 211, 347, 313, 271, 348, 337, 309, 346, 315].



(a) Before adjusting for the benefits of water infrastructure.



(b) After adjustment for water infrastructure.

Figure 1.0: Water infrastructure investment mitigates threats to water security. This map combines elements of Figures 1 and 4 from Vörösmarty et al. (2010) and depicts relative threats to water security in the year 2000 as measured by *Human Water Security* (HWS) (a cumulative index of 23 drivers) before (top) and after (bottom) adjusting for the benefits of water infrastructure [334]). Vörösmarty et al. found that, before accounting for the benefits of water infrastructure, 80% of the world's population lived in areas with high incident threat to water security, but that this risk was reduced by as much as 95% in affluent countries with high levels of water infrastructure investments [334].

to increase by 2.45 billion people—a number roughly as large as the total world population in 1950—with 95% of this growth concentrated in developing countries [309]. Urban water supplies must be expanded by an estimated 80% to meet projected water demand growth within the next 15 years alone [196].

While the coming decades require an unprecedented expansion of urban water supplies, increased water use could increasingly impair the ability of current and future generations and ecosystems to meet their needs [89, 313, 314, 315]. Experts are increasingly concerned that global renewable freshwater resources will be insufficient to sustainably meet future water demand [108, 47, 278, 335, 313]. Evidence from around has increasingly convinced experts that freshwater use is approaching—already surpassed—the limits of renewable sources worldwide [110, 112, 313, 196, 314, 315]. One-third of the world's population already lives in areas where water use exceeds sustainable yields, a number expected to rise to 50% by 2050 [314]. Moreover, many high quality, easily accessible, reliable natural freshwater sources have already been tapped for human use—whether for cities, energy-production, agriculture, or industry—heightening competition for water resources [259, 239, 156, 271, 196]. Additionally, freshwater sources—whether tapped or untapped for human use—are increasingly contaminated from byproducts of urbanization, deforestation, agriculture, requiring more energy- and cost-intensive treatment [253, 189, 312, 317, 313, 66, 272].

For all of these reasons, expanding water infrastructure is increasingly costly, with experts saying that the age of "peak water" has passed [112, 240]. The increasing competition for freshwater sources and decreasing water quality has generated uncertainty within WRM, which is further exacerbated by changing climate and an uncertain economy. These uncertainties—combined with the magnitude of the challenge—have triggered a paradigm shift in the conservative field of WRM.

1.2 A New Era for Water Resources Management

Conventional WRM developed during the 20th century and enabled both unprecedented socioeconomic development and water use [111, 47]. While some aspects of WRM will continue to be deployed in addressing the global water crisis, emerging issues will require adaptation and innovation in science, technology, engineering, management, and policy [180, 222].

A new paradigm for WRM, Sustainable Water Resources Management (SWRM), has emerged amidst contemporary challenges. The growing mandate for SWRM has propelled innovation in technology and policy and increased interest in less-conventional approaches, such as demand management, desalination, wastewater reuse, and so-called green infrastructure and are summarized in 1.2.2 and 1.2.2 [326]. While many of these innovative approaches are viable and effective, many barriers to their widespread adoption and effective use remain, especially the adaptation of solutions to specific use contexts, and best practices for doing so [222].

1.2.1 Conventional Water Resources Management and Its Legacy

The so-called "traditional" or "conventional" approach to WRM is to maintain a water supply that is much greater than water demand; i.e., to "build large and build soon" [119, 108, 111, 47]. The conventional approach tends to focus on the application of built technology to enhance water security, for which it has been nicknamed the "hard-path" solution [47, 106]. Following the conventional approach, water resource managers typically desire storage and conveyance systems that are much larger than normal operation requires, so as to accommodate growth and maintain performance levels even when supply is lower or demand higher than normal [119, 181, 291, 194, 313]. For instance, when projections of demand or supply suggest that the extra capacity is or will soon no longer provide a sufficient buffer, the conventional approach prescribes supply expansion—typically of surface waters and/or aquifers [181, 47, 106, 194].

In the 20th century, the conventional approach to WRM was pursued within a larger cultural context that generally sought to conquer nature and maximize socioeconomic development, and was characterized by a "bigger is better" and "if you build it they will come" mentality—culminating, for instance, in such projects as the Three Gorges Dam or exemplified, for instance, by the "first in time first in right" (i.e., prior appropriation) approach to water rights in the Western U.S. [70, 47, 289, 130, 140]. Within that paradigm, the conventional approach to WRM was considered a resounding success, undoubtedly enabling the unprecedented water use, population expansion, and economic development observed over the last century [102, 65, 297, 45].

Expansion of water supply and sewerage in the early to mid-20th century also led to widespread improvements in health and sanitation [102, 65, 39]. Cutler and Miller found that 40% of the reduction in mortality in the U.S. between 1900-1940 could be attributed tp such investments[2] [65].

However, even when properly maintained, built water infrastructure deteriorates over time, and technology may become obsolete. At some point, infrastructure reaches the point where it should be replaced or decommissioned [172, 180, 63, 194, 100]. Use of water infrastructure past its design service life risks severe service disruptions, contamination[3], and catastrophic system failure [172, 180, 63, 216, 18, 90, 269, 14, 100, 125].

Water infrastructure typically has a maximum design service life of 50–100 years[4] [269, 14]. This means that any urban water infrastructure that was constructed in the early to mid-20th century should be replaced within the next few decades or already replaced. In spite of the risks, many cities have deferred maintenance and replacement and frequently and continue to operate infrastructure pasts its design life [172, 199, 227, 269, 270, 14, 294, 15]. As a result, risks associated with aging infrastructure have become a major threat to water security for many cities around the world [334, 234, 288, 100, 332, 315, 338, 15].

Global climate change contributes to obsolescence. Historical weather and climate are used in infrastructure design and also in managing water resources throughout the year [137, 317, 194, 322, 185, 66]. Local temperature and precipitation affect not only water supply but also ecosystem and anthropogenic water demand [250, 104, 322, 66, 320, 129]. Climate change is The spatiotemporal variability of temperature precipitation is expected to change—exactly how is unknown, but global climate models suggest more and more frequent extreme climate events, which will further heighten the challenges of WRM [178, 7, 212]. The expected departure of future from historical climate would reduce the performance of extant infrastructure [212, 317, 194, 66, 337]. The departure of reality from managerial expectations could lead to inefficient infrastructure decisions that exacerbate urban water scarcity [277, 301, 197, 69, 22, 313, 113, 36, 117].

[2]With the total reduction in mortality in the U.S. over 1900–1940 estimated to be 30%, improvements in water supply and sewerage led to a 12% reduction in mortality over that period, with the return on investment estimated at 23:1 [65].

[3]For instance, a major a failure to replace obsolete lead plumbing fixtures led to elevated blood lead levels in children—and a major health crisis—in the city of Flint, Michigan [125].

[4]Different types of infrastructure components also differ in their typical design life.

In light of contemporary freshwater challenges, today the legacy of conventional WRM is less certain; while its impacts were not wholly negative, neither were they universally positive. Awareness has grown that the conventional approach to WRM was too limited in scope and contributed to the mismanagement that has led to contamination, depletion, and wasteful use of freshwater resources [180, 290, 47, 106, 69, 124].

1.2.2 Sustainable Water Resources Management

As the shortfalls of the conventional approach have been recognized, the WRM has gradually integrated additional concepts such as demand management, social equity, and regulatory frameworks [194]. *Integrated Water Resources Management (IWRM)* emphasizes the "systematic consideration of the various dimensions of water: surface water and aquifers, quality and quantity; (2) the implication that while water is a system it is also a component that interacts with other systems; and (3) the interrelationships between water and social and economic development" [194, p. 827]. More recently, the scope of IWRM has been extended to address additional factors such as sustainability and ecosystem services, in which case it is often referred to as SWRM[5] [180, 130, 188, 315].

The paradigm shift from conventional WRM to SWRM represents an expansion of scope, rather than a retraction. Mays defined water resources sustainability as "the ability to use water in sufficient quantities and quality from the local to the global scale to meet the needs of humans and ecosystems for the present and the future to sustain life, and to protect humans from the damages brought about by the natural and human-caused disasters that affect sustaining life" [194, p. 13]. This definition does not preclude socioeconomic development or economic considerations [313, 314, 315]. On the contrary, SWRM supports socioeconomic development as well as considerations of costs and profitability; after all, financial sustainability is an crucial dimension of SWRM [269, 270, 338].

SWRM expands the scope of the values and costs considered when assessing different technologies and policies in the context of water resource systems [313, 196, 314]. For instance, cost functions in SWRM ideally include some quantification of ecosystem damage, poverty, and resource depletion [89, 129, 293, 346]. Including these additional dimensions alters the relative attractiveness of potential solutions in comparison with the conventional approach. Moreover, the problems and pressures that propelled SWRM into the forefront have further altered the landscape of water resource systems.

This paradigm shift has helped technologies and policies that were previously considered too costly, risky, or unorthodox—such as wastewater reuse—to gain interest, attention, and investment throughout the water community [307, 313, 89, 316]. These strategies can be summarized within two categories:

[5]The terms SWRM and IWRM. Since SWRM can be considered to include concepts of IWRM, it is used preferentially in this dissertation.

- Treatment of low quality water sources, such as brackish water, seawater, or wastewater; and

- Improved use of existing resources (i.e. conservation) such as through demand management, wastewater reuse, reduction of unaccounted for water/leakage, improved water use efficiency, and improved resource management.

These options and barriers to their adoption are summarized below.

Treatment of Low-Quality Sources

Increasing investments have led to substantial breakthroughs in science and technology and brought about reductions in costs and energy use in treatment of low quality water sources—such as brackish water, wastewater, and even seawater [112, 240, 256, 255, 313, 314]. These breakthroughs help to improve the viability and tractability of using low quality water resources in water supply, which is increasingly important as rising water demand generates increased competition for natural freshwater resources, and as those resources become further contaminated and depleted [115, 112, 240, 313, 256, 255, 314, 120]. Beyond opening up new sources of potable water, many of these developments have additional attractive properties, such as modularity, which allows Water Service Provider (WSP)s greater flexibility and adaptability to changing demand, technology, or costs [229, 354].

While the technology supporting these treatments is rapidly advancing, perceived and actual risks continues to impede widespread adoption of less-conventional practices [59, 278, 322, 41, 130, 293, 222, 303, 36, 316]. By their nature, unconventional technologies and policies have typically not yet been proven under a wide range of conditions [59, 317, 222, 316]. Exacerbating the risk of failure, industry standards, policy structure, and financing options are less likely to be aligned with unconventional practices [257, 248, 19, 227, 356].

For instance, treating low quality water remains relatively costly and energy-intensive compared to conventional sources, and can be one or two orders of magnitude higher [313, 256, 255].

Negative perception can also generate real challenges to implementing of an unconventional technology or policy [312, 74, 360, 127, 356, 120, 316]. For instance, negative public opinion of wastewater reuse persists, even though treatment practices typically result in higher quality water than conventional treatment of surface or groundwater sources [56, 307, 317, 78, 316]. Public opinion of wastewater reuse has in some cases been so low that communities have staunchly resisted proposals for non-potable wastewater reuse in during severe droughts [51, 141].

However, public perception of less-conventional sources can be improved through education, public outreach, and time [307, 10, 305, 74, 127, 316]. Water scarcity has also played a role in advancing public acceptance; many cities that reuse water are water scarce with respect to climatic availability or other factors [305, 278, 188, 316].

Improved Use of Existing Sources

Conservation and demand management features prominently in discussions of SWRM [47, 106, 283, 206]. The costs of treating low quality water sources have decreased, but are still higher than using less water, or using water more efficiently[242, 241]. One option to improve the use of existing water sources is through policies that affect user behavior, e.g., by encouraging conservation or increasing water tariffs [242, 241, 283, 31]. Demand management may also encourage adoption of more water efficient technologies, such as low-flow toilets, which promise to allow the same quality of life to be achieved with less water use [107, 47, 46]. Water use can also be reduced through better leak detection and asset management, an area that has benefited from improved sensing and algorithms [3, 227].

Improving the use of existing urban water sources is also benefited by work in quantifying the benefits of conservation and water efficiency, which allow proposals for improved use to be compared more directly with options to expand water supply [240, 343, 196].

In addition to demand management and reduction of leakages, experts have also called for better monitoring of water use with respect to sustainable yields, and for "better sharing of limited freshwater resources" [206, p. 1]. Strategies may also include land management policies, which can help improve water quality and reduce the need for water treatment [305, 296, 297, 330, 297]. Some urban water supplies continue to serve agricultural needs at the urban, peri-urban, or rural fringe [76, 140]. In some cities, large volumes of water are used for landscaping, alone [213]. Due to the water-energy-food nexus, energy and agricultural policy play a role in SUWM [155, 200, 28, 175, 314, 36, 315].

There are numerous challenges in improving use of existing sources. Compared to the technologies for treating low quality water, demand management, asset management, setting up frameworks for "better sharing of freshwater resources", and land use policies are highly context-specific. Sharing freshwater resources and monitoring sustainable yields requires some alignment of political boundaries with hydrological units [61, 11, 29]. While many such *transboundary* agreements have been reached, SWRM requires the coordination of efforts across multiple dimensions [109, 2, 245, 61]. Examples from other industries have demonstrated that improving material efficiency often leads to *increased* total consumption [5, 308]. Similarly, since many associate conservation with reduced socioeconomic development, calls for frugality may be seen as politically untenable in both affluent and developing regions alike [283]. Efficient water use is only one strategy of SWRM; setting sustainability targets in other areas is necessary, but requires consideration of local culture, social equity, industries, and ecologies [347, 314].

1.3 Toward a Unifying Perspective

The promising technologies and policies for SWRM that have emerged do not yet have widespread adoption and implementation [203]. However, even as the threat of water scarcity and the need for SWRM undoubtedly looms large, there remains significant uncertainty about how best to meet the upcoming challenges, including what actions can or should be taken, and by whom; how to prioritize present and future needs, and how to as balance social, economic, and environmental uses [258, 100, 348, 314]. Since each city will require a different portfolio of both conventional and innovative technologies and policies for SWRM, the coming years and decades will to some extent be a period of experimentation [326, 130]. Portfolios must be chosen with consideration of additional dimensions of water resource management that are not typically included in the conventional approach [130, 131, 314]. Achieving sustainable urban water management requires looking beyond the physical sciences and engineering and into politics, law, economics, society, health, and religion [217, 205, 20, 116, 295]. Decision-makers are increasingly called on to not only be experts on water supply and demand, but also on the broader cultural, economic, political, and religious contexts in which technologies and policies are to be deployed [217, 313, 131].

This uncertainty has delayed action, increasing the imperative for concerted, integrated technology and policy responses in combating the growing water crisis [180, 258, 130, 314]. Any cooperative effort requires agreement on a tractable, unified, and integrated approach—or at the very least, compatibility between approaches, and a unified framework for sharing data and knowledge [245, 281, 258, 100, 29]. However, a unified perspective continues to elude SWRM with new approaches proliferating [100]. For instance, a systematic review of the literature by Plummer, Loë, and Armitage found over 710 different indicators used amongst only the top 50 water assessment tools surveyed [258].

SWRM would benefit from an approach that enables stakeholders to identify and adapt solutions to local needs; that supports sharing of data and knowledge to support adoption of unproven technologies and policies; and that helps integrate global action and assessment across multiple dimensions of water use. Experts generally agree that overcoming barriers to widespread implementation of SWRM requires concerted and integrated efforts within a unified framework [258, 100, 348, 314].

There is general agreement that it requires thoughtful and considerate balancing of environmental, social, ecological, and economic water needs [245, 30, 267, 194]. Determining what these needs are and how to balance them requires measurement and quantification of relevant system attributes and models of the dynamics and inter-relationships between water uses—not to mention models integrating these needs with models of climate and biogeophysical factors that affect local availability [94, 194, 126, 268]. While the concept has been around in some form for some time, our understanding of what constitutes "sustainability" (including SUWM) continues to evolve through active research [57, 267, 122].

However, identifying a tractable approach is easier said than done. Urban water systems are inherently complex[6], and cannot be fully understood by any single person or approach [299, 245, 96, 291]. For this reason, the processes and interactions underlying water resource systems must, by necessity, be distilled into more digestible units that can be understood by policy makers and stakeholders [299]. This step of simplifying the real world into useful models has been shown to be a crucial step in designing effective water management policies [326, 299, 205, 331, 298, 342].

Water infrastructure must be appropriately operated, maintained, and managed, which requires appropriate levels of political, regulatory, and financial support and coordination of related resources [299, 194, 336]. Effective management and decision-making for UWS also require the collection of numerous types of data, such as climate, hydrology, supply, demand, watershed characteristics, and wastewater production [119, 94, 52, 184]. These data are then used in conjunction with various types of models to assess system performance and to calibrate management of the urban water system to changing conditions [119]. Infrastructure and water resources are managed to maintain reliable performance levels, reduce variability, and minimize risk of water *stress* and *scarcity*[7] [128, 313]. For instance, since water infrastructure cannot be constructed instantaneously, the capacity of water infrastructure should be sufficient to meet current needs as well as those in the near future, which requires monitoring of factors affecting demand and supply [181, 291, 301, 194]. Finally, water resources and infrastructure should ideally be managed in a way that supports larger societal goals [180, 194].

Since processes affecting the outcomes of WRM policies and technologies vary over a wide range of spatial and temporal scales, characterizing the spatiotemporal context is important for effective policy design implementation [119, 180, 189, 247, 83, 58, 258, 313, 84, 131]. Towards this end, efforts have been made to characterize spatial variation in performance metrics [88, 201, 249, 250, 329, 84]. Calibrating policies to local needs also requires historical analysis to provide context for time-dependent factors that can affect outcomes [6, 2, 299, 9, 21].

For all of these reasons, in spite of the growing need for a unifying perspective, WRM must not resort to relying on panaceas, i.e., "the tendency, when confronted with pervasive uncertainty, to believe that scholars can generate simple predictive models of linked social–ecological systems and deduce general solutions to the overuse of resources" [245, p. 15176]. Many dangers lie in over-simplification of complex dynamics, such as those exhibited by water resource systems,

[6]The Oxford English Dictionary defines "complex" as "consisting of many and different parts" [235]. Water resource systems exhibit both *combinatorial complexity* and *dynamic complexity*. While both types of complexity have been found to tax the limits of human intuition, they require different tools for decision-making. Systems exhibiting combinatorial complexity have a large number of components, and require tools that help identify optimal solutions from an astronomical number of possible outcomes [299, 9]. Dynamic complexity arises in systems due to interactions between system components exhibit non-linearity, delays, and feedback, and may arise in systems with low combinatorial complexity [299, 9]. Systems exhibiting dynamic complexity benefit from simulation tools that allow users to explore how the system dynamics evolve under different conditions or policies over time [299, 9, 96].

[7]Water *stress* and *scarcity* sometimes have specific meanings in the water resources community, but can also be used more generally to refer to instances when there is less supply available than desired, i.e., when *availability* is low. When used more specifically, stress and scarcity are defined for a region in terms of Total Annual Renewable Freshwater Resources (TARWR): *stress* occurs TARWR is less than 1,700 $m^3 \cdot yr^{-1}$; *scarcity* when TARWR is less than 1,000 $m^3 \cdot yr^{-1}$, and *absolute/extreme scarcity* when TARWR is less than 500 $m^3 \cdot yr^{-1}$ [313, p. 124].

including the design and implementation of policies that lead to policy resistance or dreaded "unintended consequences" [6, 2, 299, 9, 21]. As an example, the over-exploitation of freshwater resources can be considered an unintended consequence of conventional WRM [290, 69]. Even experts in WRM often fall victim to the desire for a straightforward approach [332, 131].

Thus, the question remains: how can the need for a unified, global perspective be aligned with a meaningful consideration of local issues for sustainable water resources management?

1.3.1 The Urban Perspective

Cities exist in many places in the world and vary in size, climate, and numerous other factors [25, 279, 34, 264, 310, 32]. While no two cities are alike, viewing water resources management through the urban lens may provide a unifying perspective with which to facilitate the learning and innovation necessary to meet the world's myriad water challenges.

All cities require freshwater resources for survival [353, 119]. Throughout the millennia, successful cities all constructed vast water infrastructure and developed water resource policies for supply and sanitation [193].

Looking Outwards

Urban water supply in a growing city goes through several key phases. When an urban center first develops, centralized water infrastructure and resource management are typically lacking, and urban residents rely on local wells or surface waters for supply [228, 195]. As the city continues to grow, water demand surpasses reliable yields, and quality is compromised by pollution. Because of this dynamic, growing cities typically look to import freshwater resources from outside of the municipal boundary—i.e., watershed expansion—whether by canals or other means[8] [140, 151, 189, 297, 45, 68, 286, 333].

Figure 1.1 depicts typical components of conventional urban water systems. Water is collected from surface water and/or aquifer sources into a storage body, such as a reservoir, aquifer, or tank. Raw source water is then conveyed to treatment plant, where the water may undergo multiple types of processes until it meets standards for water quality and aesthetics. The treated, now-potable water is then pumped to local storage, from which it can be distributed to domestic, commercial, industrial, and municipal users in the urban service boundary. During use, water might be evaporated, transpired, or otherwise consumed—and therefore lost from the urban water balance. Alternatively, used water—i.e., wastewater—remains within the urban system and is collected within a sewerage system, which may also collect stormwater runoff generated from the service boundary. Water within the sewerage system is then conveyed to a wastewater treatment plant, where it is treated to environmental standards. The treated wastewater is then usually discharged into some receiving body that is typically far downstream from the source, but may even be in a completely different watershed [119, 189, 256].

[8] Demonstrating the importance of external water supplies to cities, McDonald and Shemie found that global totals of interbasin water transfers amounted to an estimated 43% of urban water supply [196].

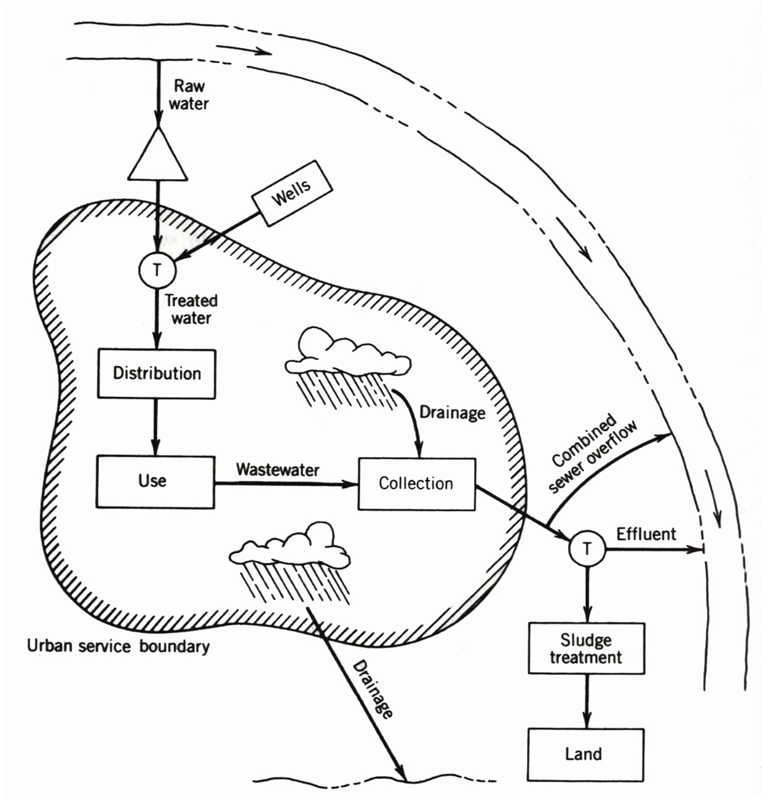


Figure 1.1: The conventional urban water system [119, Figure 2.1].

Looking Inwards

Because cities have limited financial resources, cost-related factors such as source proximity, quantity, quality, and reliability are important considerations in expanding urban water supplies [128, 119, 8, 181, 194, 329]. Ideal freshwater resources for municipal use are therefore reliable, abundant, nearby (but outside of the city), and relatively clean [189]. However, cities looking to expand their watersheds today must increasingly compete for water resources that are remote, low quality, and often unreliable—or even depleted or contaminated by other human activity [275, 196, 198]. As the costs of "looking outwards" have increased, cities are increasingly looking to improve their use of existing sources, and to expand their use of local, low-quality sources, such as brackish water, seawater, wastewater, and urban runoff [90].

Given the limitations of existing global freshwater resources, cities are increasingly turning to less conventional sources of potable water—seawater, stormwater runoff, and recycled wastewater [196]. These water sources have high concentrations of pollutants, dissolved chemicals, and organisms toxic to human health [189, 335]. Because of this, treatment of these less-conventional water sources is more energy- and cost-intensive [256, 255]. Historically, some cities have found purchasing land in distant watersheds to protect water quality to be more economical than the costs of treating polluted water [204, 148, 254, 297, 196]. However, decreasing costs for water treatment have begun to open these "taps" as potentially viable sources for potable water in cities.

1.3.2 The Role of Comparative Analysis

Instead of attempting to distill diverse complex systems *a priori*, Ostrom, Janssen, and Anderies instead recommends a more adaptive approach, with an emphasis on "diagnosis, monitoring, and learning in sustainability science" [245, p. 15177]. An adaptive approach based on monitoring and learning requires a shared knowledge base, e.g., such as one built from rigorous comparative analysis of case studies of Water Resources System (WRS), and would, therefore, benefit from a framework for integrating data, results, and knowledge. In other words, transition to SWRM would benefit from a unified approach that can provide both a global and historical perspective on water resource issues and context-specific insight across a multitude of spatiotemporal scales and other, less-conventional dimensions of WRM [88, 189, 313].

Conventional WRM has demonstrated that existing approaches to constructing portfolios of technologies and policies UWS can be adapted to different contexts. While urban contexts are all different, almost all have water systems to at least some degree, and numerous methods and best practices exist for adapting technologies and policies to new environments [119, 194, 144]. However, many of these methods tend to be prescriptive, and efforts to adapt WRM to a constantly changing mental model of water resources has proven more elusive. What is lacking in conventional WRM is a framework through which these methods and approaches can themselves be examined critically, tested, and revised according to new information.

As new technologies, policies, and approaches proliferate, researchers must critically assess the applicability of concepts, tools, and methods to the diversity in urban water systems [340]. Rigorous comparative analysis, when conducted within an appropriate research framework for inductive and deductive inference, can help to facilitate learning and transfer of knowledge during this period of experimentation [174, 340, 116, 218, 341]. However, while the need adaptation and learning has become apparent, there remains ongoing disagreement about "whether to use quantitative or qualitative methods, on whether to use a large or a small number of cases for comparison, on the unit of comparison and the indicators chosen to compare them, and whether to focus on similarities or differences" [116, p. 968]. Few studies conduct rigorous quantitative comparisons of different urban water resources management contexts. Comparison of international cases is even less frequent, and where it has been done the selection of cases and method for comparison is typically based on analyst familiarity rather than a rigorous, repeatable, and scalable approach [340, 193, 219, 116, 218, 320].

In summary, comparative analysis of UWS is not well-developed. Perhaps because of the inherent complexity of the underlying systems, in-depth case studies tend to focus on providing specific insight into a single cases *or* studies of large numbers of urban water systems tend to focus on providing summary statistics or descriptive trends for large numbers (large-*n*, i.e., with a sample size $n > 10$) of urban water systems [116]. While large-n studies may include an international sample of cities, comparison of cases is often limited to a single dimension or indicator [340, 341, 339]. These studies are less able to provide enough context to support rigorous choice of case studies for more in-depth analysis [218, 116]. Where in-depth case study analysis of WRS occurs, it is typically performed on a small-n sample (i.e. on 1–4 cases) that are often chosen for arbitrary reasons, such as researcher familiarity, which makes it difficult to gain more general insight. Because of this, in-depth case study analysis may focus on descriptive trends. In-depth comparative analysis is rarely performed on international cases, preventing insight into global trends and opportunities for transferring knowledge across country borders [341].

Meeting the growing threats to urban water security requires that WRM adapt and learn faster than in the 20th century. Since it is through systematic and rigorous comparison that deep learning and insight occurs,researchers must address the lack of systematic international comparative analysis [174, 340, 299, 116, 218]. To minimize the inherent risks associated with uncertain technologies, policies, and experimentation in general, cities would benefit from approaches that facilitate the identification of technologies and policies appropriate to their particular contexts [2, 244, 85]. One way to do this is by looking to solutions that have been successfully (or unsuccessfully) implemented in similar urban contexts.

1.4 Dissertation Overview

1.4.1 Research Statement

In the next 35 years, the world is expected to acquire an additional 2.54 billion urban residents [309]. Traditionally, most cities have relied on watershed expansion or water imports to meet demand growth [197, 196]. However, watershed expansion will become increasingly difficult as 41% of global land area has already been tapped for municipal consumption [196]. Amidst these pressures, cities are increasingly looking to improve their use of existing sources, as well as expand their use of low-quality sources—especially those that can be obtained within the urban boundary. This increased emphasis on self-sufficiency, R_{SS}—defined as the ratio of supply from local sources relative to total water use—has become a high priority for many municipal water systems. However, many of the technologies and policies associated with the emerging paradigm of SUWM are considered relatively risky and unproven. Amidst this time of experimentation, cities require an approach to experimentation and learning that provides a global perspective and encourages transfer of knowledge across international borders.

1.4.2 Research Overview

Research Questions

In this dissertation, I present a novel approach to rigorous, quantitative comparative international water research that aims to answer that question in the positive. To do this, I use large-*n* statistical analysis as a framework to guide the selection of cases for small-*n* comparison. This framework consists of two parts: a method for case study selection and a method for case study analysis. This work is motivated by the question:

Is it possible to gain a global perspective of urban water systems in a way that:

1. *does not require that the diversity in water supply and demand be simplified* a priori;
2. *that can adapt to changes in underlying data, as well as changes in society's understanding of water resources management (and related processes and challenges)*?

The motivating question is broken down into guiding research questions for each chapter in the dissertation as:

Chapter 2 How do we learn about UWS, how does comparative analysis contribute to learning, and how are UWS compared?

Chapter 3 Can simple profiles of supply and demand provide both global perspective *and* meaningful insight into regional challenges?

Chapter 4 To what extent could similar cities have been self-sufficient in the past with respect to local water resources, and how does this compare to their actual portfolio of water supply? How similar were their portfolios and to what extent could they be similar?

Chapter 5 Using data and results from Chapter 4, to what extent could these cities be self-sufficient in the future, and how does this compare to their projected portfolios of water supply? Does the comparative analysis in Chapters 4 and 5 offer case-specific or general insight?

Chapter 2

A primary goal for the dissertation is to approach the analysis of urban water metabolism in a methodical and systematic way to provide the basis for rigorous quantitative comparison. Ideally, we want to pursue small-*n* comparative analysis and research design that highlights local components and dynamics related to SUWM of UWS while providing insight into how the case studies compare on a larger, international scale. This requires a research design for large-*n* analysis that establishes a logical framework for choosing and situating case studies according to a clear inferential logic, which I discuss in Chapter 2. I also aim to develop an approach that facilitates scalability, replicability, and comparability to benefit future work.

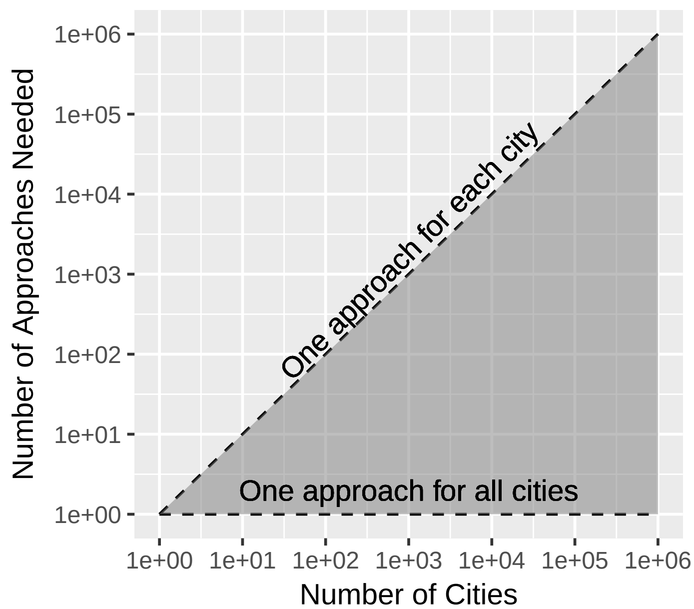


Figure 1.2: Diversity in cities and in technology and policy portfolios for urban water systems. There is a large diversity in urban attributes related to water supply and demand, but is each city uniquely different, or are there larger patterns in this diversity? Each city requires a portfolio of WRM technologies and policies appropriate for its particular profile. How can these portfolios be determined? If there is a large number of urban water contexts, do they each require a different approach? Is there a single approach that suffices for all cities and all water resource challenges, regardless of how different their profiles? Or does the number of approaches required fall somewhere in between?

In Chapter 2, A Framework for Comparative Analysis, I further discuss some of the ideas introduced in Sections 1.2 and 1.3 in Chapter 1. Section 2.1 explains a conceptual model of learning in complex systems and the role of comparative analysis. Section 2.2 summarizes the debate over whether it is "preferable" to perform *quantitative* or *qualitative* analysis on a *small* or *large* number of cases (small-*n* versus large-*n*). I conclude that the situation is not "either-or", and that both are important in advancing knowledge about UWS. In Section 2.3, I consider how emerging data-mining tools can contribute to identifying more complex patterns in large-*n*, multivariate data, and their use in WRM research. In the last section in Chapter 2, An Integrated Approach, I first synthesize these ideas into a research framework for adaptive learning for UWS, and then outline a research plan for the analysis pursued in Chapters 3, 4, and 5.

Chapter 3
"Large-*n*" Comparative Analysis

There are many cities in the world, with different shapes, densities, climates, and sizes—all of which affect local water demand, use patterns, and hydrology. Appropriate portfolios of technologies and policies for SUWM depend on these urban attributes. While one urban context may differ substantially from another, they may also share similarities. One can imagine a two-dimensional space with diversity of urban supply and demand profiles on one axis and approaches to WRM on the other, as illustrated in Figure 1.2. Are the profiles of cities so different that no over-arching patterns in or types of UWS or WRM challenges can be identified? Is there only one portfolio of appropriate technologies and policies that can be used to regardless of the diversity in urban profiles? Or is there only one type of profile but a large number of viable approaches?

Three simple metrics are chosen to represent urban water demand and supply profiles: population (N, in capita), water use intensity (w_N, in $m^3 \cdot yr^{-1} \cdot cap^{-1}$), and climatic water balance (q_P, in $m \cdot yr^{-1}$). Data for each of these metrics are curated from existing bases for a "large-n " (i.e., more than 100) international comparison of 142 cities, representing 6 continents and a diversity of sizes and climates. City population is chosen as a primary driver of total urban water consumption. Water use intensity, i.e., per capita water consumption, provides a relative measure of average urban affluence or infrastructure investment. Net climatic water balance provides a measure of regional climatic water availability.

In Chapter 3, I apply exploratory statistical methods to characterize the spectrum of simple water supply and demand profiles for 142 cities across six continents. In Section 3.2, I revisit concepts from Chapter 2 summarizing large-n analysis in WRM.

In Section 3.3, I discuss the methods that I use to construct a dataset for statistical analysis. In Section 3.3.1 I consider the choice of N, w_N, and q_{Net} to represent urban profiles of water supply and demand and in Section 3.3.2 I provide an overview of the common databases, scripts, and web-scraping methods that I use to assemble the data. In Section 3.4, I explain the methods used in the first phase of the large-n analysis, including the calculation of descriptive statistics and correlation and the plotting of histograms, order bar charts, quantile-quantile plots, and scatterplots to characterize the global spectrum of urban water profiles. I present the results from this first phase in Section 3.6.

In Section 3.5, I present the methods for the second phase of large-n analysis—statistical clustering. I use two statistical clustering algorithms, *hierarchical clustering* and *t*-Distributed Stochastic Neighbor Embedding (t-SNE) to test whether simple profiles of supply and demand can provide global perspective *and* meaningful insight into regional challenges. I provide background on these algorithms in Section 3.5.1 and describe the workflow in Section 3.5.2. In Section 3.5.3, I describe methods for visualizing the clustering results—*dendrograms*, the t-SNE scatterplot, and violin-/box-plots—which I present in Sections 3.7 and 3.8. In Section 3.8, I identify a typology of six types of urban water supply and demand profiles from the trivariate clustering results.

I then introduce a new index to bolster intuition into the clustering results. I define the Water Use and Climate Index (WUCI, or i_{UC}) as the ratio of water use intensity ($m^3 \cdot yr^{-1} \cdot cap^{-1}$) to annual precipitation ($m \cdot yr^{-1}$), giving WUCI units of area/capita. WUCI is a new type of water resources footprint, and provides insight into the local impact of urban water use relative to climatic water resources, as compared with other "footprint" measures in use [135, 206]. While the different types are not found to be completely distinct in terms of WUCI, WUCI is found to provide a lens to compare the groups on relative local environmental impact.

These results represent the first known application of clustering algorithms to international profiles of urban water supply and demand and demonstrate that even simple metrics—when chosen purposefully to represent important aspects of SUWM—are useful in constructing a global typology from multivariate data. Since the typology is constructed from a quantitative characterization of relative similarities, it is then used to guide the selection of two case studies for further analysis in Chapter 4.

Chapters 4 and 5
"Small-*n*" Comparative Analysis

In Chapters 4 and 5 I pursue the second phase of research: in-depth comparative analysis of two case studies over time. The goal of this analysis is to test whether insights into the paths to SUWM could be gained by examining simple urban water supply and demand profiles over time for two cases. The general aims of the analysis in Chapters 4 and 5 were to:

1. Demonstrate the use of the typology from Chapter 3 as a framework to guide the choice of two cases;
2. Provide additional insight into the results of Chapter 3 by characterizing the trajectories of urban water profiles for two cities *over time*;
3. Develop a method for analysis that can be applied to both cases;
4. Test whether this method provides deeper insight into WRM for both cases; and
5. Compare both cases, using the results from the analysis to identify similarities and differences between two cases.

Chapter 4 presents a three-part approach to historical analysis for comparative analysis of UWS. While the results from Chapter 3 provide insight into the relative differences and similarities of cities at an instant in time, but that analysis does not assess trajectories of urban water supply and demand, i.e., whether cities are becoming more or less sustainable.

The first part of analysis demonstrates one approach to using a typology, such as that developed in Chapter 3, to guide the selection of case studies in a rigorous, replicable, and scalable manner. Two primary cases are chosen from the same type to test the utility of the typology in identifying cities with similar WRM challenges. Type 4 cities are chosen as the type from which to choose two cases. Type 4 cities, characterized by large populations and high water use intensities, are more likely to have already experienced growing pressures on freshwater supplies, and therefore more likely to be concerned about SUWM. From within Type 4, Los Angeles and Singapore are then selected as the two primary cases, as these two cities represented the lower and upper ends, respectively, of the Type 4 spectrum of climatic water availability. The logic of "similar but different"—the first step of a recommended approach to small-*n* comparative research design that is discussed in Section 2.2.1—is used to frame this logic.

The second part of analysis in Chapter 4 assembles datasets of historical values for the urban water supply and demand profiles for the case studies chosen in part one, drawing from common datasets for climatic water availability and on commonly reported indicators—such as city area, population, and total or per capita water use. In the third part of analysis, this data is then analyzed with simple growth models and simple methods for statistical and time series analysis[9]. Summary statistics calculated for population growth, per capita water use, and precipitation were assessed, and results indicate that historical values can be reasonably represented by random distributions. However, when I examine historical values as time series, I find that the means and variations in these series were not stationary in time for either case study, especially with respect to population growth and water use intensity. Finally, I consider whether portfolios of water supply and demand are similar for cities within Type 4, even under different conditions of natural water availability.

In Chapter 5, I extend the analysis from Chapter 4 by creating a method to simulate simple scenarios of water supply and demand for each of the two cases. I then apply these methods to generate simulations for two periods: historical (1960–2016) and future (2017–2040).

Chapter 6

Chapter 6 summarizes the analysis, results, and conclusions from Chapters 3—5 and discusses the potential impact of the work within the broader research context. Section 6.2 summarizes the results from the previous chapters. In Section 6.3, I examine the results relative to the research questions and lay out a program for future work. Finally, Section 6.4 summarizes the conclusions for the work.

Appendices

Supporting methods, results, data, and code are provided in Appendices A—D:

Appendix A: additional information about methods used in earlier chapters

Appendix B: supplementary results

Appendix C: tables of data used in the analysis

Appendix D: `R` and Python scripts developed for the analysis

[9]These analyses were conducted using freely available platforms and packages.

Chapter 2

A Framework for Comparative Analysis Using Large-*n* Analysis to Situate Small-*n* Comparison

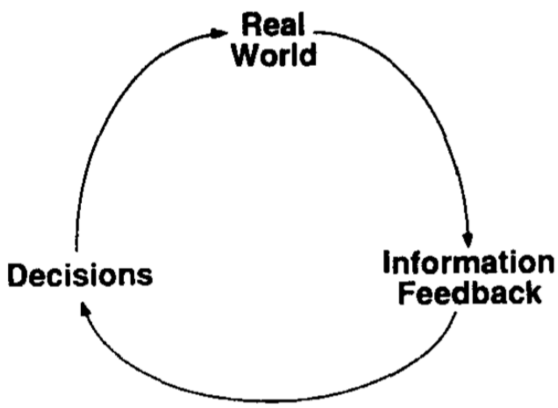
2.1 Learning About Complex Systems

As the call for a new paradigm for WRM has amplified, so too has have researchers from different disciplines come together to provide insight into the emerging challenges [27, 267, 188, 118, 295]. Water use and management have myriad dimensions ranging including cultural, religious, health, economic, environmental, and hydrological—to name only several—and there is an equally diverse spectrum of interdisciplinary approaches to WRM [119, 94, 317, 149, 194, 219, 295]. However, these efforts are all unified by the motivation to advance understanding of water resources on some level, and many are impelled by the desire to facilitate transitions to a more sustainable future [246, 48, 42, 41].

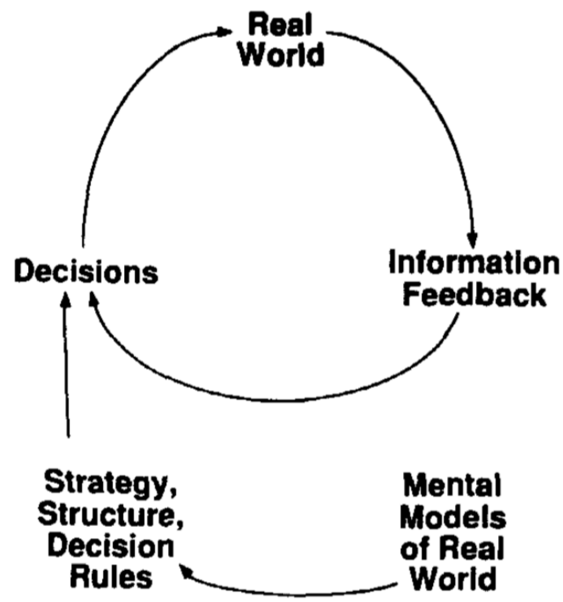
In the pursuit of such knowledge, it is important to be cognizant that UWS are complex human socio-technical systems, and so it is not possible for any single person to fully understand *all* aspects of IWRM, at all possible levels of detail, at any time [299, 94, 194]. Because of this, it UWS are difficult to manage, let alone to learn about.

Therefore, one motivating question for this dissertation is how can we improve the advancement of knowledge of UWS and its dissemination to facilitate transitions to SUWM? At the heart of this question lies the process of learning and methods for collecting, storing, managing, and sharing data. In this dissertation, learning—whether at the scale of an individual person or an organization—is viewed as an iterative feedback process. Figures 2.1a–2.1b show the learning process at its simplest—as a single loop feedback process. As seen in both Figure 2.1a and 2.1b, in the learning loop, information about the real world is used in decisions. Actions based on these decisions then cause changes in the real world (sometimes after a delay) and, after these

Figure 2.1: Sterman's model of learning as a feedback process [300, Fig. 1, p. 293]. Information about the real world is used in decision-making. These decisions then cause changes in the real world. After these effects have been observed, the data can then be used in new decisions. Figure 2.1a shows the simple model, while 2.1b emphasizes that these decisions are based on decision rules, which in turn are informed by mental models of the real world, all of which are embedded in a larger institutional context.



(a) The simple model



(b) The learning process, decision rules, and mental models.

effects have been observed, the data is then used in new decisions. Figure 2.1a highlights the main learning loop, while Figure 2.1b emphasizes that decisions are based on decision rules that are informed by mental models of the world.

The process depicted in Figure 2.1 highlights the learning about a single dynamic system, within the framework of a particular mental paradigm. However, as stated previously, the paradigm for WRM and UWS has itself begun to shift. How does this relate to the process depicted in Figure 2.1?

The diagram in Figure 2.2, adapted from Figure 1.14 in Sterman (2006), illustrates the learning loop from Figure 2.1 (Loop 1) along with two other learning loops [299]. Loop 1 might be characterized as *learning by experience*, since this type of learning requires actions to be propagated through the real system, observed, and then used the new information incorporated into new decisions. Loop 2 shows another mechanism for learning, *learning by simulation*. In Loop 2, mental models of the real world are formulated as computer simulation models [299]. These simulation models can be used to help test whether our mental models of the world lead to the outcomes we expect. If not, simulation provides a method by which theories can be tested much more quickly than an experiment in the real world [299].

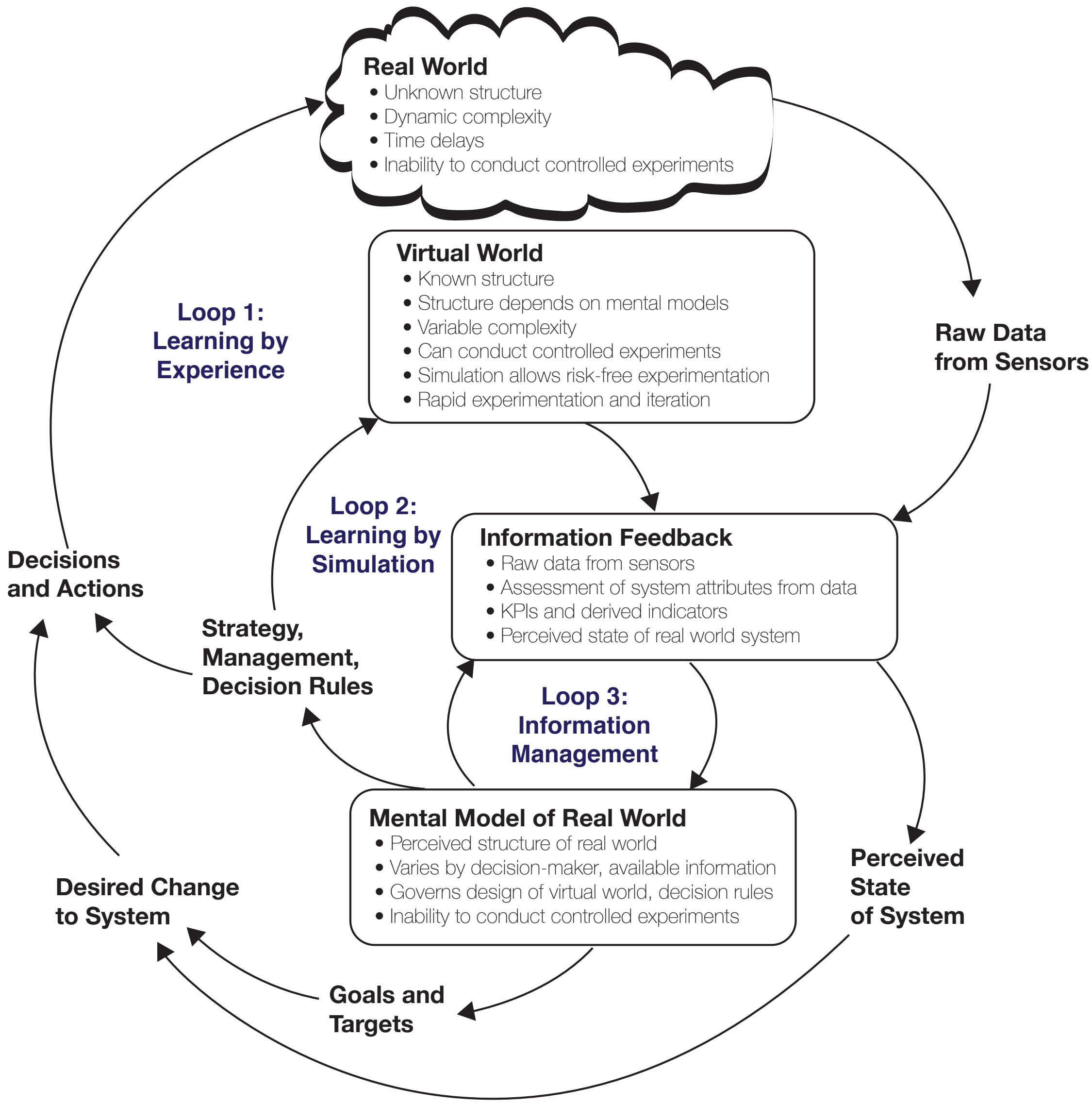


Figure 2.2: The role of learning in urban water systems. Adapted from Figure 1.14 in Sterman (2006, p. 34) [299].

In both processes, learning by experience and learning by simulation, data plays an important role. Researchers and experts on management and decision-making generally make a distinction between *data*, *information*, and *knowledge*. The word *data* may refer to any building blocks of knowledge, such as a series of precipitation data, while *information* "is data endowed with relevance and purpose"—for instance, the analysis of precipitation data to inform reservoir operation [94, p. 139]. *Knowledge* is only generated when information has been synthesized into actionable decision rules or into larger mental models of how the world works [94]. Ideally, for adaptive learning, the process—shown in Loop 3 in Figure 2.2—is iterative, and information is sought out and actively used to update mental models and our mental models [299, 94]. These updated mental models, in turn, are then being used to adapt data collection, monitoring, processing, and management to fit changing paradigms [299, 94].

Our mental models of the world may be updated through information *passively* or *actively*. *Actively* updating mental models requires that theories be explicitly stated in a testable way—i.e., that a hypothesis about the structure of the system be made, that an experiment be designed to generate data or that relevant data be acquired through other means, and that information acquired from such data be used to update the mental model. In *passive* learning, learning may be pursued in a less intentional or structured way, or the learning may happen implicitly, without the individual's (or organization's) full awareness.

However, it is difficult (if not impossible) for learning about UWS and IWRM to happen passively due to the combinatorial and dynamic complexity of the systems [299, 94, 219]. For this reason, it is necessary to pursue knowledge through a more explicit and systematic way—such as the scientific method [299, 94, 219].

The scientific method refers to any method or approach that follows the general philosophy of building knowledge about the world through rigorous observation, experimentation, and analysis of testable hypotheses. The first step of the scientific method might be considered to be observation or measurement of the world, whether the observation be of biological, physical, chemical, social, or cultural systems and phenomena. These observations are then used to formulate theories—i.e., models—about relationships between components of the system [174].

The relationships studied using the scientific method may be *causal* or *correlative*. A causal relationship is one in which an action by one part of the system causes a change in another; gaining insight into a causal relationship builds knowledge about the structure of the system itself. A *correlative* relationship is one in which a pattern between two components has been identified, but the relationship is not necessarily a structural one; gaining insight into correlative relationships can help provide context to construct testable hypotheses about causal relationships and system structure.

A traditional approach to identifying relationships between two variables is to design an experiment in which one of those two variables is varied as the independent variable and the response of the second variable is observed and measured. In this approach, the ideal is that any other system variables are held constant [174].

However, it is difficult to run controlled experiments on UWS. The large number variables that affect management of UWS, and feedbacks between them, make it difficult to ensure that a control variable is, in fact, controlled. Running controlled experiments is also difficult because UWS are in active use, and cannot simply be brought down so that researchers can run experiments. This is especially true since some dynamics of interest in UWS emerge over decades, rather than hours or days.

Even if it were common to run controlled experiments on select subsystems of UWS, the fashion for distilling complex systems into highly controlled—and increasingly isolated—subsystems has fallen out of favor. Empirical evidence for emergent behavior in complex systems—i.e., system dynamics that cannot be predicted from viewing subsystems in isolation—has begun to receive increased attention in the scientific community. In other words, running controlled experiments on complex systems like UWS, which exhibit both combinatorial and dynamic complexity, may not produce useful insight into SUWM.

One approach to overcoming these challenge, as mentioned earlier, is through the use of computer simulation of the real world through its representation in virtual models [299]. These models can test policies and ideas about the system more quickly than they can be tested in the real world [299]. This approach has become commonplace in some aspects of WRM, especially for physical systems or engineering applications [352, 69, 194, 214, 210, 16, 320].

Another approach is through analysis of historical dynamics of UWS. Historical data typically plays a crucial role in the aforementioned simulation studies; they may also be pursued as an end in themselves, as in comparative case study analysis [49]. Comparative analysis of case studies has been described as one of three approaches to *scientific explanation*, alongside experimental and statistical analysis [174]. The phrase, "comparative analysis," in common usage appears in the literature to refer to the comparison of a small- to medium-number of cases (small-*n* and medium-*n*), and typically on a qualitative basis. It is generally agreed that comparative research benefits from a design informed by a clear inferential logic, especially with regards to 1. the *number of cases* to be compared, and 2. whether to choose cases that are similar and/or different [174, 219, 218]. Essential components of this logic are essentially the same as those outlined earlier for the general scientific method, as well as for integrated urban water management [94]:

- a clear statement of the research motivation;
- a research question;
- well-defined unit of analysis, i.e., system definition; and
- the basis for comparison, i.e., the metrics or system attributes and dynamics being compared.

Once these essential ingredients have been assembled, the next step is to examine the research question and motivation and consider what type of inference is required to advance the research agenda.

One of the first steps in comparative analysis is deciding *pre facto* whether cases are similar or different. However, there are many ways of comparing urban water systems, and there remains disagreement about best practices. Comparative water research frequently lacks common indicators or standard methods across time or cases [217]. There remains ongoing disagreement about "whether to use quantitative or qualitative methods, on whether to use a large or a small number of cases for comparison, on the unit of comparison and the indicators chosen to compare them, and whether to focus on similarities or differences" [116, p. 968].

Data are required to formulate, validate, or test theories about the world, and may be *quantitative* or *qualitative* [299, 94]. Quantitative data are considered to be data about a phenomenon that has been systematically quantified and for which methods for standard instrumentation and measurement exist. Data about population, precipitation volume, and water use all fall into the category of quantitative data. *Qualitative* data are generally considered to be data about a phenomenon that is difficult to quantify or measure directly, or that may be subjective, i.e., variable with a subject's perspective; data about user satisfaction fits within this category.

This dissertation focuses on analyses of quantitative data. However, data from both categories is important to the management of urban water systems [94]. While analyses of quantitative data are often considered more scientifically acceptable, hypotheses focusing on qualitative data can also be rigorously formulated, tested, and modification—i.e., approached with the scientific method—with an appropriate design of experiments and analyses [174, 299, 266, 219]. For instance, even though social phenomena can be difficult to measure, even subjectively quantified data can be used to build testable models of human systems [299, 97]. Also, that data of a phenomenon are qualitative does not necessarily mean that the phenomenon is unquantifiable. Instead, the existence of qualitative data often precedes quantification through invention or implementation of appropriate instrumentation for measurement. For instance, the advent of infrastructure supporting global networks of social exchange has led to the collection of large streams of data about social and cultural phenomena that were previously considered unquantifiable [266, 94]. Additionally, the increased adoption of methods for rigorous physical instrumentation (such as brain imaging) by fields studying social phenomena (like psychology) has led to more advanced models relating subjective experience to measurable and quantifiable data. Finally, even unquantified qualitative data can provide useful context for comparative analysis, especially when assumptions are clearly stated [174].

Therefore, one of the challenges in comparative water research is in the choosing of a basis on which to compare water systems. Comparisons may be may explicit or implicit; for a small number of cases (small-n), a large number of cases large-n, or anywhere in between; quantitative or qualitative [219, 341]. Comparison also differs in perspective; UWS may be compared with respect to politics, institutions, health, and economics; or with a focus on engineering or physical systems, such as hydraulics, hydrology, climate, or infrastructure [219, 340, 341].

In 2.2, the pros and cons of comparing large or small numbers of cases is discussed. Following that, in 2.4, a new framework is presented for integrating "large-n" and "small-n" comparative analysis[1].

2.2 Comparative Analysis

There are at least two challenges in translating technologies, policies, or lessons learned from a precedent case to an entirely new urban context [219, 116, 218, 341, 196, 36]. The first lies in the selection of the cases themselves; the second lies in the basis for comparison.

[1]While there is no clear threshold between a "small", "medium", or "large" number of cases (n), it is generally agreed that "small" implies $n > 2$ and possibly $n < 10$, "large" implies $n > 100$, and "medium" implies any number lying somewhere in between.

The need for SUWM has become acute in many cities around the world, regardless of whether they are small, medium, or large; dense or sprawled; wet or arid; rich or poor; or new or old [94, 36]. While there are still more questions than concrete answers, the acuteness of the need has led to notable innovation in WRM methods, technologies, and policies, as well as in its implementation—in spite of the relative risks associated with adopted less conventional practices [51, 256, 255, 316]. However, because water security is paramount, it has typically been easier for wealthier cities to be innovative, since more affluent cities are more likely to be able to afford the integrated analysis that can assist in adapting the design of conventional technologies and policies to new contexts [175, 36, 314, 315, 316]. Yet SUWM is just as important for less affluent cities, if not more so. In those cities with an existing shortfall in water infrastructure, it would be preferable to implement innovative WRM solutions more appropriate to the coming water challenges rather than investing in outdated infrastructure.

In transitioning to SUWM, the identification of *appropriate* portfolios of technologies and policies is crucial for any city. To facilitate the transition, it can be helpful for a city to identify precedent cases, i.e., cities that have already begun to transition to more sustainable practices to look to in the design of their sustainable urban futures [340, 341, 218, 356]. However, while a city may look to others for best practices or other takeaways, it must construct its own portfolio to be most effective. In other words, one city may benefit from another's experience but is unlikely to be able to adopt those exact technologies and policies without adapting them to its own urban context. This requires comparing the two urban contexts in a way that can elicit specific lessons or recommendations—i.e., performing some degree of comparative analysis.

Since there is no major clearinghouse for case studies or comparative analyses, most cities looking to identify appropriate precedent cases for SUWM must start from the ground up [356]. Due to limited time and funding, it is generally not possible for one city to perform rigorous comparative analysis against all possible examples [219, 341, 356]. For this reason, the selection of cases tends to be afflicted with the same problems as most small-*n* comparative analysis—with cases chosen more for familiarity or availability rather than from a logical framework highlighting specific similarities or differences [340, 219, 116, 218, 341].

Even outside of urban management, it has apparently been difficult for researchers to design case studies for small-*n* comparative water research from which rigorous, generalizable conclusions can be drawn [340, 219, 116, 218]. Both quantitative and qualitative small-*n* comparative analysis of UWS alike have tended to suffer from a general lack of a systematic approach to case study choice and research design [194, 330, 29, 16, 100, 340, 219, 116, 218, 341]. Many studies examine single case studies—such as Garg (2007), and many others—some researchers select cases to reflect the authors' expertise, while others group case studies under predetermined headings, e.g. arid, tropical, low-income, megacities, etc. [174, 99, 318, 233, 116]. A special journal issue dedicated to the topic of comparative analysis of water research found the comparative logic to be "loose" at best [116, p. 971]:

> Looking at how comparisons were specifically conceived and carried out in the contributions to this issue, it is apparent that most of the comparisons are quite loose: there is often no common set of indicators or any other method defined to compare the cases systematically within each of the papers. Sometimes multiple sites are chosen within one case study, whilst other studies have taken on several cases. Further, some papers do not compare the cases at all but simply analyze them in parallel and aggregate the results. However, in view of the pressing issues facing the chosen case-study locations and the contributors awareness of this, all papers have chosen a certain critical mass of cases or sub-cases with the apparent aim of using these to derive results not specific to only one case but with wider potential applicability

In other words, loose research design and a general lack of standardization—both of indicators and methods—has hampered the validity and generalizability of the lessons learned from small-n comparative analysis [116]. When only a small number of cases are compared, concerns over the validity or generalizability of the results may be raised [340]. The issues at the heart of the matter are those of control, generalizability, and inference. As a result, the conclusions that can be drawn from small-n comparative analysis tend to be limited and qualitative.

Theoretically, however, a large number of cases is not required to corroborate or falsify generalizations [174]. This requires that the cases be chosen with reference to a clear, logical framework that reflects a well-posed research question and theory.

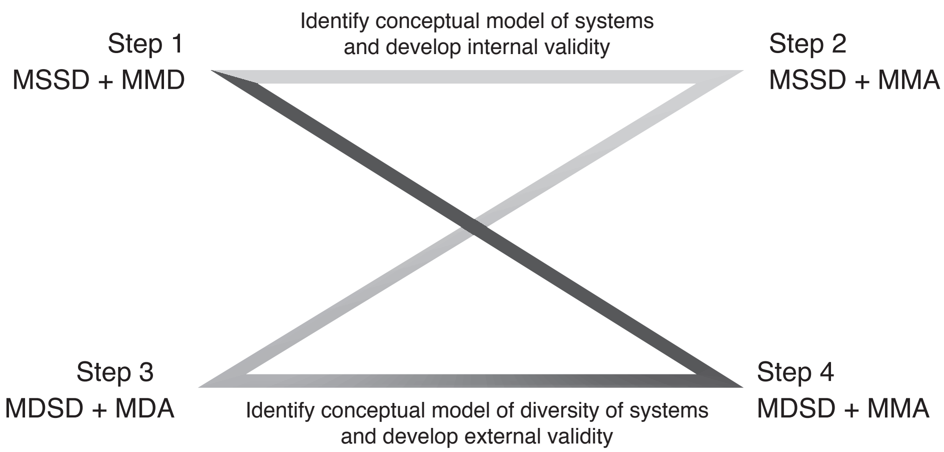
2.2.1 A Systematic Approach to Small-n Research Design

To address this shortfall, Gondhalekar, Mollinga, and Saravanan adapted a four-step, iterative approach for "small-n-and-medium-N"[2] comparative research from Levi-Faur (2006) for comparative water research [174, 116]. The iterative approach shown in Figure 2.3 is based around the idea that there are (at least) four types of inferential logic that can be used to advance knowledge through comparative analysis [174, 219, 116]. Przeworski and Teune (1970) provide two types of inference, Most Similar System Design (MSSD) and Most Different System Design (MDSD); and Mill (1843) provides the remaining two, Mill's Method of Agreement (MMA) and Mill's Method of Disagreement (MMD).

MSSD In Most Similar System Design, cases that are "as similar as possible with respect to as many features as possible" are compared Przeworski and Teune (1970, p. 32). The underlying assumption in MSSD for comparative analysis is that if the cases have similar outcomes, it will be easier to isolate system properties or other "intervening factors" that have led to the observed outcomes [174, p. 57]. Research that follows MSSD assumes that since systems are "as similar as possible", then it will be easier to identify the underlying attributes that have given rise to important differences in outcomes identified—if any are so identified.

[2] While there is no clear threshold between a "small", "medium", or "large" number of cases (n), it is generally agreed that "small" implies $n > 2$ and possibly $n < 10$, "large" implies $n > 100$, and "medium" implies any number lying somewhere in between.

Figure 2.3: An approach to structuring diversity through comparative analysis of small-*n* and medium-*n* cases, as introduced by Levi-Faur (2006) and adopted by Gondhalekar, Mollinga, and Saravanan (2013) for water research [174, Table 3.1, p. 59], [174, Figure 3.1, p. 61], [116, Figure 1, p. 973].



Przeworski and Teune's System Design		**Mill's Methods**	
		Method of Difference (MMD)	Method of Agreement (MMA)
	Most Similar (MSSD)	**Step 1** MSSD + MMD A set of "primary" cases are chosen for similarity in outcomes. These cases are compared, ideally in a consistent method, with the intent to formulate a conceptual model (i.e., hypothesis) of a system structure and its emergent properties. Ideally, primary cases differ in underlying properties to enhance the "robustness" of the initial theory.	**Step 2** MSSD + MMA The theory developed in Step 1 is applied to new cases that are similar to the original primary cases in terms of outcomes. The intent is to test the theory in situations where it would be expected to apply. Deviations from the theory are used to further develop the initial model.
	Most Different (MDSD)	**Step 3** MDSD + MMD Addition of "secondary" cases chosen for dissimilarity in both outcomes and underlying properties (compared to the primary cases). The intent in Step 3 is to test and develop the conceptual model on cases where it would not be expected to apply. Comparative analysis focuses on dissimilarity in cases to 1. characterize diversity, 2. expand the theory to encompass it, and 3. formulate a theory of "structured diversity" (i.e., a typology). Step 3 may be iterated by adding "tertiary" cases to further develop the typology.	**Step 4** MDSD + MMA While Step 3, focuses on differences in underlying properties and outcomes, Step 4 focuses on the similarities in the previous cases "in order to develop a theory encompassing the structured diversity mapped by the typology" Gondhalekar, Mollinga, and Saravanan (2013, p. 972) [116].

MDSD In Most Different System Design, cases are chosen based on "variation of the observed behavior at a level lower than that of systems" [260, p. 34]. In social inquiry, this means that the cases are chosen based on differences at the level of individuals, "groups, local communities, social classes, or occupations" [260, p. 34]. For MDSD, the assumption is made that the underlying "levels" of the system are "drawn from the same population", that both the original and sample populations are homogeneous in terms of underlying attributes [260]. As a corollary, it is assumed that systemic factors do not affect the outcomes; analysis first focuses on testing that assumption. If the initial assumption is not rejected, analysis focuses on the original "lower" system; otherwise, the unit of analysis must move "up"—i.e., the system boundary must be extended. Whether the initial assumption is rejected or not, the intent of Most Different System Design is to identify relationships between underlying system attributes [260, 174, 116].

MMA In MMA, researchers look for an agreement between a theory and actual and expected system attributes based on that theory. In other words, the goal of MMA is to identify a "description of an observation [that] affirms, beyond what is contained in the observation, an agreement among phenomena"—i.e., 'if A then B' [209, p. 647]. However, as Mill notes, that statement is general enough it has two interpretations, and which are distinct steps in scientific inquiry. First, "the comparison of phenomena to ascertain such agreements is a preliminary to induction"; in other words, identifying a relationship between observations is a way to identify a connecting theory; for instance, that "the proposition that the earth moves in an ellipse" connects observations made by Kepler [209, p. 647]. However, to satisfy the requirements for inductive logic, MMA requires that new observations be made. If they are then found to support the theory, this constitutes MMA [209]. In contrast, Levi-Faur (2006) uses MMA to describe both the initial step of preliminary theory formation and in theory validation (internal validation) [174].

MMD In MMD, according to Levi-Faur (2006), similarities in underlying attributes are considered with respect to differences in the underlying context, with the aim of uncovering a difference that can be used to explain the dissimilar outcome [174, 116]. The goal of Mill's original view of MMD is to establish 'if not A then not B'; Mill considers this to be an essential step in validating a proposed theory [209]. However, it is difficult to do in the case of complex systems. Mill provides an example of an "ideal" set of cases for MMD for social systems: "we require to find [at least] two instances, which tally in every particular except the one which is the subject of inquiry" [209, p. 882]. However, since that is very difficult to do for complex systems, Mill suggests the "Indirect Method of Difference": which, "instead of two instances differing in nothing but the presence or absence of a given circumstance, compares two classes of instances respectively agreeing in nothing but the presence of a circumstance on the one side and its absence on the other. To choose the most advantageous case conceivable, (a case far too advantageous to be ever obtained, *sic*) suppose that we compare one nation which has a restrictive policy, with two or more nations agreeing in nothing but in permitting free trade. We need not now suppose that either of these nations agrees with the first in all its circumstances; one may agree with it in some of its circumstances, and another in the remainder. And it may be argued, that if these nations remain poorer than the restrictive nation, it cannot be for want either of the first or of second set of circumstances, but it must be for want of the 'protective' system" [209, p. 882]. Since this is extremely unlikely, "there is thus a demonstrated impossibility of obtaining, in the investigations of the social science, the conditions required for the most conclusive form of inquiry by specific experience" [209, p. 882].

The MSSD and MDSD described in Przeworski and Teune (1970) *and* the MMA and MMD laid out by Mill (1843) are both seen as imperfect for systematic research of complex systems such as social systems [174, 219, 116, 218]. Mill (1843), viewed MMD as "inapplicable in the social science Mill (1843, p. 881)" and opined that there was a "danger of applying the methods of elementary chemistry to explore the sequences of the most complex order of phenomena in existence Mill (1843, p. 881)", a prejudice that continues to be held to this day [209, 97, 300, 299]. While Przeworski and Teune (1970) did not completely discount the application of scientific principles to social systems, they also left it unclear how to define systems and apply methods in a systematic way. This opaqueness has lingered throughout the decades; as quoted on Page 56, even in a special journal issue dedicated to the topic of comparative analysis of water research found the comparative logic to be "loose" at best, and that there was "no common set of indicators or any other method defined to compare the cases systematically within each of the papers" [116, p. 971].

The deficiencies in Przeworski and Teune (1970) and Mill (1843) in advancing systematic research of complex social systems and other complex systems (such as water resource systems) inspired Levi-Faur (2006) and Gondhalekar, Mollinga, and Saravanan (2013) to recast Przeworski and Teune's MSSD and MDSD and Mill's MMA and MMD for social systems [174, 116]. Levi-Faur (2006) and Gondhalekar, Mollinga, and Saravanan (2013) combined the two approaches into a new framework of:

Step 1 *MSSD + MMD*: dealing with differences in similar cases for new theory creation;

Step 2 *MSSD + MMA* dealing with similarities in similar cases, intended for theory development and establishing internal validity;

Step 3 *MDSD + MMD* dealing with differences in different cases, to promote external validity and/or to create a typology of systems; and

Step 4 *MDSD + MMA* dealing with similarities in different cases, to further advance knowledge about the systems [174, 116].

To expound upon this framework, the four-step process outlined by Levi-Faur and Gondhalekar, Mollinga, and Saravanan begins with a primary case and the development of a conceptual model of structure and emergent properties (MSSD + MDD) [174, 116]. The first analysis of the primary cases is then to be followed by an extension of the primary case, for instance by applying the conceptual model developed in the first step to a similar situation, in which the same theory would be expected to apply, (MSSD + MMA) [174, 116]. In the third step, MDSD + MMD, the conceptual model is tested through application to situations that were very different from the primary cases and in which one might not expect the same model to be valid [174, 116]. As the first and second steps built up or tested the internal validity of the conceptual model, the application of the model to secondary cases in the third step would be a test of its external validity [174, 116]. The fourth step, MDSD + MMA, revisits the original conceptual model and looks to further develop a theory that encompasses the similarities and differences uncovered by the previous three steps [174, 116]. In the final step, the research program requires returning to the first step in an iterative process of learning [174, 116]. As mentioned previously, this approach is outlined in Figure 2.3 (after Levi-Faur (2006, Table 3.1, p. 59), Levi-Faur (2006, Figure 3.1, p. 61), and Gondhalekar, Mollinga, and Saravanan (2013, Figure 1, p. 973) [174, 116]).

If small-*n* comparative analysis is pursued according to a logical approach such as that proposed by Levi-Faur and Gondhalekar, Mollinga, and Saravanan, then it *may* be possible to make valid generalizations using a small to medium number of cases [174, 116]. This has led some researchers to conclude that the "small-*n*-and-medium-*N*" approach seemed like "the most feasible way forward, and preferable to large-*n* methodologies, which may be viewed by some researchers as woefully "positivist"[3] and therefore "looked upon with suspicion" [219, p. 45].

The reality likely lies somewhere between. A small-*n* or medium-*n* city sample with additional resource consumption variables does not adequately fill the gap between cases and benchmarking. Small-*n* or medium-*n* comparative case study analysis, whether quantitative or qualitative, undoubtedly has, can, and does contribute to knowledge of WRM and UWS. However, Levi-Faur's and Gondhalekar, Mollinga, and Saravanan's step-wise approach is predicated upon being able to identify similarity and dissimilarity between cases, and this ideally requires an assessment of a large number of cases, which is benefited by quantitative methods.

Case study analysis that focuses on quantitative methods would also benefit from a more systematic approach, such as that outlined by Levi-Faur and Gondhalekar, Mollinga, and Saravanan, since emerging challenges in UWS require new theories and models. The idea of an iterative approach to building a theory from a small-*n* sample to a model, or set of models, supported by a large-*n* set of cases also fits within the framework of learning about complex systems that was outlined in Figure 2.2 on Page 51. However, it can still be difficult to identify patterns in complex systems for across many variables for a large-*n* set of cases. In the following subsections, Sections 2.2.2–2.3.2, an overview of emerging approaches to addressing this challenge is pursued. Finally, in Section 2.4, these small-*n* and large-*n* approaches are considered with respect to the larger goal—to build knowledge about UWS.

[3]Where positivism is defined as "a philosophical system that holds that every rationally justifiable assertion can be scientifically verified or is capable of logical or mathematical proof, and that therefore rejects metaphysics and theory" [236].

2.2.2 Using Large-*n* Analysis as a Framework for Small-*n* Case Study Selection

Adapting current WRM in cities to more sustainable practices requires distilling knowledge from data for a large number of variables and complicated dynamics [94].

Many studies have confirmed that humans have difficulty observing complex systems and deriving valid inferences about their structure and behavior [60, 299, 336, 229]. This is particularly true if:

- there are many variables;
- there are many observations (large-N);
- cause and effect are separated in space and/or in time;
- the system depends on processes operating over many orders of magnitude; and
- relationships between variables are non-linear.

Since all of the above are true for urban water systems, it is essentially impossible (or astronomically unlikely) for humans to be able to derive the knowledge required for a global transition to SUWM exclusively from small- to medium-*N* qualitative case study comparison.

Cities already collect data on many indicators for use in existing WRM practices, and it is likely that IWRM and SUWM will increase these data requirements [94]. There are large numbers of cities around the world, over 10,000, with new cities continually formed (for instance, see the database `world.cities` provided with the `maps` package in `R`). These cities are diverse in size, shape, climate, density, and numerous other factors [279, 310]. There is also substantial evidence of the large diversity in the existing structure of urban water systems, including built infrastructure and management practices; in cultural and socioeconomic factors; in urban form; and in climate, ecology, and biogeophysical factors—to name a few—all of which may affect the design of water technology and policy portfolios that will advance global water sustainability [94]. Therefore, data on many variables for a large number of cases will be necessary to fully characterize the diversity in urban water profiles at a single instant in time, let alone variations over time [94].

Fortunately, our increasing awareness of the complexity of human and natural systems has paralleled our ability to observe the world in which we live. Where in the past, limited data availability and comparability prevented researchers from comparing a large number of UWS in any depth, improved instrumentation and measurement have begun to generate huge amounts of data, often for metrics that were previously difficult to measure [94, 220, 123, 75, 43, 80, 84, 62, 315]. Data availability has advanced in many areas, from the physical to social phenomena [94, 220, 123, 75, 43, 80, 84, 62, 315]. Improvements have also increased the resolution of data in both time and space; this development has further multiplied the size of data streams [94, 84, 62].

In spite of this, few studies examine large urban datasets to conduct rigorous comparisons in a way that can help to uncover non-intuitive insight into urban water management on a global scale [340, 279, 192, 157, 87, 341, 218].

There is a lack of a large-*n* integrated multivariate international comparison of cities with respect water demand and supply. Where large-*n* integrated assessment is performed, it has been limited to univariate descriptive statistics, such as histogram plots, ordered bar charts, or the calculation of statistical moments and quantiles [317, 80, 219]. Sometimes the analysis has extended to simple bivariate methods, such as plotting two metrics on a scatterplot or calculating linear correlation coefficients [219]. These standard univariate and bivariate methods are an essential part of any comparative analysis, since it is useful to visualize the underlying data. However, when comparative analysis is confined to those methods, it limits the types of conclusions that can be drawn.

It is rare that a deeper exploration is performed, such as the construction qq-plots or transformation of the data with functions to improve visualization or test the (often implicit) assumption of normality [219, 116, 218]. And often, such comparison is not complete, in that the univariate or bivariate analysis has not been performed on all variables (or combinations of variables). While the results of such analyses can be suggestive, it is not possible to gain non-intuitive insight through incomplete analysis. Also, analysis is rarely done to identify rigorous distinctions. For instance, scatterplots may be used to group sets of observations that appear similar, which is intriguing, but such analysis would then need to be verified through clustering analysis, which is rarely performed.

Another popular approach to large-*n* analysis has been to examine urban processes through the lens of scaling laws [35, 34, 32, 23, 24]. The results of such studies appear to promote the idea that cities occupy a deterministic continuum rather than being fundamentally different; this has perhaps led other researchers to conclude that large-*n* quantitative analysis cannot lead to any legitimate conclusions about similarities and differences in UWS [32, 24, 219].

These simple approaches to large-*n* analysis can help to better characterize the spectrum of underlying variables. However, it can be difficult for human cognition to correctly and consistently identify significant patterns over time, more than two or three variables, or many orders of magnitude—as is the case for the data examined in Chapters 3—5 [299, 94, 75, 3, 336]. More complex methods are required to identify significant patterns in large-*n* multivariate datasets.

In summary, two trends—an increased demand for solving complex problems and increased availability of data—have spurred interest in the analysis of large numbers of observations or cases (i.e., large-*n*) or in large numbers of variables or large amounts of data [17, 94, 187, 361]. The number of variables and feedbacks between them poses a challenge for large-*n* comparative analysis, as the task of defining a system boundary in a way that is comparable between cases becomes much more difficult. Another challenge is that, even if such data were available for a large number of cases, it might be difficult to identify any general trends, patterns, or relationships, due to limitations in human cognition. This is especially true if analysis focuses exclusively on qualitative comparison, as is often done in comparative analysis focusing on social, cultural, or institutional aspects of UWS; however, the point is also relevant for quantitative analysis [174, 219, 116, 218]. In short, while increased data availability has generally been positive, its rapid expansion has also led to an acute need for improved methods and tools to assist humans—whether scientists, engineers, designers, managers, policy makers, or global citizens—in converting such vast data streams into information and insight about complex phenomena in a complex world [17, 299, 94, 187, 336, 229].

Fortunately, computers are better-suited to the task of exploring and identifying patterns in datasets with large numbers of observations across a multivariate space, and numerous tools have emerged for the task—a process generally known as data mining.

When looking to choose an appropriate set of cities in the design of small-*n* experimental design, a decision must be made as to the attributes on which the comparison will be made and the larger set of cities from which the small-*n* sample is chosen. Developing a typology can be a useful intermediate method for "chunking" this information into units that are more accessible to understanding by individuals, whether expert or non-expert. These patterns can be difficult to identify as the number of objects and attributes being compared increase. However, statistical clustering has emerged as one approach to data-mining that can assist in the identification of meaningful typologies from complex data.

In the following section, Identifying Patterns in Complex, Large-*n* Data, summarizes the application of algorithms from one category of data-mining methods, statistical clustering, to SUWM. The discussion then moves to an explanation of how clustering analysis of large-*n* quantitative data can be combined with small-*n* to provide a framework from which to guide case selection for small-*n* research.

2.3 Identifying Patterns in Complex, Large-*n* Data The Opportunity for Data-Mining

2.3.1 Clustering and Typologies

To compensate for the limitations of the human mind in processing complex streams of data, humans rely on a variety of internal and external aids to compensate for the limitations of memory and other cognitive faculties [70, 173, 221, 299]. Through iterative pattern-seeking—throughout each day and over our lifetimes—human cognition naturally attempts to simplify the complex data streams from the world into abstractions and generalizations [299, 94, 336]. These cognitive aids help reduce the complexity of the world into concepts that are more easily remembered and interpreted [299, 94, 336]. Generalizations may take the form of rules of thumb or mnemonics [299, 94, 336]. And perhaps one of the most widely adopted general approach to pattern seeking is that of the scientific method, discussed earlier in Section 2.1. And recently, computer-based tools have emerged that help researchers mine large quantities of data for patterns and trends. These data-mining tools can support—but do not completely replace—human-based pattern recognition [187].

Among the tools for data-mining are algorithms that identify groups of objects on the basis of similarities and differences in their attributes [17, 187, 182]. One approach to data-mining patterns from large-*n* datasets is that of statistical clustering, which generally refers to algorithms that seek to identify groups of observations within a larger dataset on the basis of quantifiable statistical attributes [187].

The algorithms and methods associated with statistical clustering, described in greater detail in Section 3.2, can seem complex and non-intuitive. However, the essential task of a clustering algorithm is simply a formalized, systematic method for a process that tends to come naturally to humans: that of creating a typology from diverse data.

The *Oxford English Dictionary* defines the word "typology" as the "study or analysis using a classification according to a general type", and "type" as "a category of people or things having common characteristics" ([238, 237]. This brings us back to the question: how does one go about determining whether two objects are similar or dissimilar?

Generally, the task requires 1. a listing of attributes of both objects; 2. some basis for comparison, i.e., assessing similarity and dissimilarity [17, 187]. As humans, this is a task that we have learned to do implicitly; for instance, in comparing an ax with a saw presumably elicits a different response than comparing an ax with a dessert or a kitten. However, each of these different objects has a number of attributes. The set of attributes possessed by each object may be the same as the other objects to which it is being compared, or different. The objects being compared may vary substantially over a particular attribute, or not very much. The extent to which we decide whether objects are similar or different also depends on the set of objects being compared.

As an example, consider the following thought experiment, starting with a set of five objects consisting of: an ax, a kitten, a lion cub, a tiger, and a bear cub.

It would be natural to divide this set of five objects into two groups: one group consisting of the ax (a weapon or tool), and the second containing the kitten, the tiger, and the two cubs (animals). Now remove the ax from consideration, so the set of four objects is a kitten, a lion cub, a tiger, and a bear cub. The natural new grouping is probably two groups: one with the kitten, lion cub, and tiger (all felines) and the second with the bear cub (of the family *Ursidae*).

However, the grouping depends on the perspective. For instance, say you were an expert on early animal development; the natural grouping might then be two groups, one with the kitten, lion cub, and bear cub (all young animals) and the second with the tiger (an adult animal). But what if you were instead an ecologist studying keystone predators in threatened habitats? Then you might divide the four into two groups: one consisting of the kitten and the lion cub, tiger, and bear comprising the second.

While the preceding thought exercise might seem somewhat foolish, it highlights several key aspects of the task of assessing similarity and dissimilarity: 1. it depends on the motivation behind the exercise, which affect 2. the attribute(s) being compared, and 3. it is affected by the range of properties.

2.3.2 Clustering Analysis in WRM

One can imagine how, as the number of objects and dimensionality of the attribute space increases, it becomes increasingly difficult for the human mind to assess the relative similarity of objects. A popular rule of thumb is that the number of distinct types of information a human can process simultaneously is typically limited to 7+/-2 [336]. Since statistical clustering can help reduce complex, multi-dimensional information, it can be used to help structure interdisciplinary research topics, such as resource management and sustainability [17, 280, 71, 103, 64]. Through the data-mining approach, cities that may appear incommensurable at first, such as Guangzhou, Singapore, Sydney, and Mexico City may, in fact, sit close together in the multi-dimensional space defined by their physical, economic, and technological traits.

The application of statistical clustering to SUWM is gradually being adopted across the various related disciplines [94, 265]. For instance, in water resources engineering, clustering has been widely adopted for forecasting of water supply and demand for a city at various temporal scales or in identifying leaks [73, 357, 252, 99, 1, 213, 359, 95]. Another application of clustering has been in the validation and updating of the Köppen Climate Classification System, a typology of global climate patterns [160, 276, 361].

Researchers have also begun to apply clustering techniques to inform general resource management and sustainability [153, 157, 86]. And a literature search identified several such studies focused specifically on water systems at a watershed scale and for American water utilities [192, 265]. However, to the author's knowledge, the work presented in this chapter and in Noiva, Fernández, and Wescoat (2016) is the first to perform clustering analysis of urban water supply and demand metrics integrated with climatic water balance data or to do this for a large number of international cases.

The next section, An Integrated Approach, presents a framework for linking large-*n* and small-*n* case research to facilitate adaptive learning for transitions to SUWM.

2.4 An Integrated Approach

Since statistical clustering assesses similarity and dissimilarity, it can be used to identify a typology with that is both accessible to decision-makers and is based on a systematic, multivariate analysis. Such a typology could then be used to structure the choice of cases for small-*n* comparative analysis.

Figure 2.4 shows a more detailed breakdown of the research that I pursue in this dissertation, considered within the framework depicted in Figure 2.3. One might consider as the conceptual model the notion encompassed by the phrase "looking inwards, looking outwards" that was proposed in Chapter 1.3.1, which underlies the choice of metrics and analyses in Chapters 3 and 4. From that perspective, the analysis pursued in Chapter 3 and the development of the typology might be considered the "primary case" in Gondhalekar, Mollinga, and Saravanan's framework; alternatively, one might take as the conceptual model that typology. Depending on which of

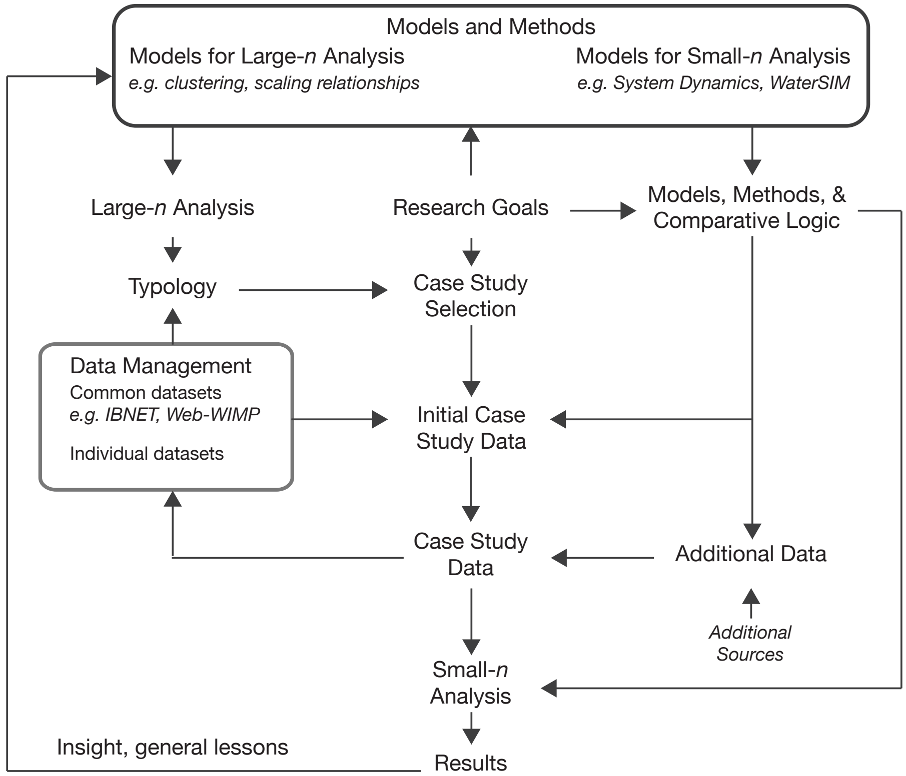


Figure 2.4: Both small-n and large-n comparative analysis can contribute to advancing general and case-specific knowledge about sustainability in urban water systems. large-n comparative analysis of relevant attributes and performance metrics can help to identify general trends in—or to structure—the diversity exhibited by UWS, while small-n comparison provides a useful scale for identifying regional variability and dynamics. While numerous viable approaches to SUWM may exist simultaneously, many share common metrics and/or methods. Therefore, it seems reasonable to suppose that small-n research could be designed with explicit reference to a systematic global perspective (and vice versa)—and that coordination of data and methods across the two scales of analysis could advance both perspectives more quickly than either could alone. This figure highlights the iterative nature of both scales of analysis, points for their connection, and the opportunity for integrating data and methods. large-n and small-n analysis could share some metrics and methods, the choice of which would be informed by a mental models of sustainability in urban water systems. The mental models are then tested through both large-n and small-n comparative analysis, and the results used to test and revise global and regional models of SUWM. The comparative analysis could lead to theory revision and to suggest alternative metrics and methods. At any time, multiple regional and global perspectives could exist simultaneously. However, the systematic approach would help to trace the evolution of knowledge.

those views is preferred, the two case studies in Chapters 4 and 5 would then be considered as primary cases or as extensions of the primary case. Figure 2.5 shows a conceptual diagram the framework I propose for linking large-n and small-n analysis—such as those that I pursue in Chapter 3 and Chapters 4 and 5, respectively—in an iterative feedback framework that can be

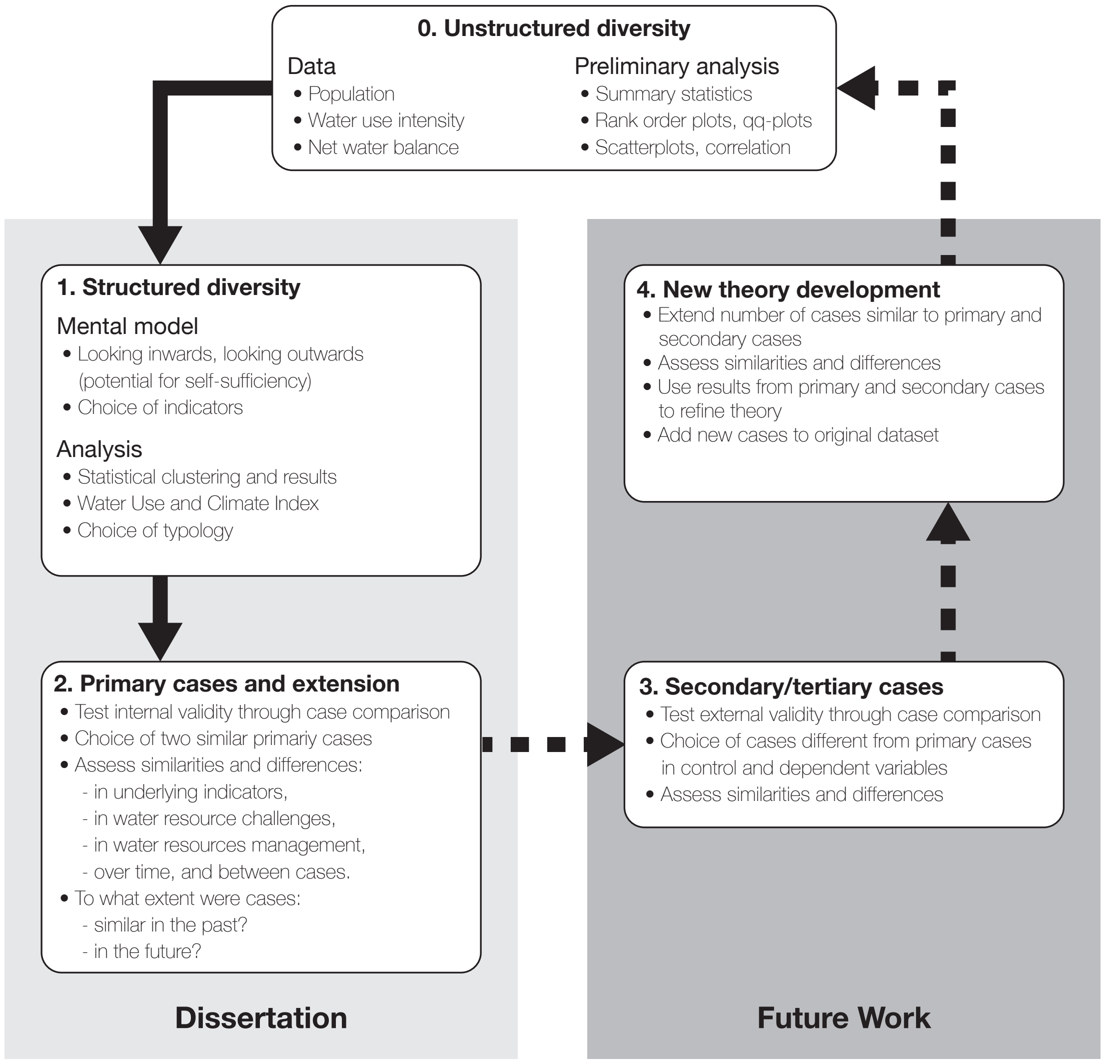


Figure 2.5: The structure of comparative analysis pursued in this dissertation.

used to support adaptive learning and WRM for SUWM. This framework combines ideas from Figure 2.2 (which depicted the feedback process of adaptive learning) and Figure 2.3 (which shows a step-wise approach for structuring diversity in small-*n* research).

A research program framed in this way would require sophisticated and considerate management not only of databases, but also of the mental models and tools used. Ideally, tools would be developed with an eye towards applicability; users would then test these tools to other urban contexts and highlight, in a methodical way, where the model failed or was valid. The model could then be adapted to the new context, and the adapted model would be included in the library of mental models for future use and testing.

While it might be difficult to imagine how to manage more complex models of UWS, it is perhaps easier to imagine such a research system if the research began with simple models—this will be attempted in this dissertation. Once such tools for collaboration were in place, more complex mental and simulation models—for which sophisticated data management exists to a certain extent—could more naturally fit within such a "system of systems" [94].

Of course, such a research program is not within the scope of a single dissertation or even a single person; to be truly effective, it would require coordination of efforts across many different scales. However, it is only a matter of time before such a system emerges, since there is a great need for it and since new tools for collaboration and data management are invented at an increasingly rapid pace. The work pursued in Chapters 3–5 develops methods and demonstrates one approach that would fit within this framework.

In summary, while learning can be advanced implicitly and passively, knowledge can be advanced more quickly by actively seeking to formulate, test, and update theories about SUWM in urban water systems. If it is possible to codify a system for adaptive learning, and if it enhances awareness of our existing knowledge base and its gaps, or can facilitate knowledge sharing, it could accelerate the transition of UWS to SUWM. In the following chapters, this dissertation attempts to demonstrate that structured comparative analysis can be one viable model for such a framework.

Chapter 3

Using a Large-*n* Analysis to Create a Global Typology of Urban Water Systems

3.1 Introduction

The motivating question for this work was: Can simple profiles of supply and demand provide both global perspective *and* meaningful insight into regional challenges?

While comparative water research is essential to advancing knowledge about global water resource issues, the sustainability of urban water systems is often compared informally or in small numbers of cases selected more for familiarity than from a rigorous approach to case study selection [340, 116, 218, 341]. The number of variables may be too few or too large to provide insight, or the structural relationships assumed may be too simple or too complex to effectively answer the research question posed. Few studies examine large urban datasets to conduct rigorous comparative analysis in a way that can provide unexpected insight into urban water challenges, especially among international cases [340, 116, 218, 341].

This chapter demonstrates one novel approach designed to fill this gap. In this study, I apply quantitative methods for data exploration and data-mining to analyze profiles of water supply and demand for 142 cities around the world. The aim of this research is to characterize urban water profiles for a large, international set of cities to 1. establish a baseline perspective of the global spectrum and 2. provide a rigorous, quantitative framework for identifying similar or dissimilar cities for small-*n* comparative analysis.

In Section 3.3, I choose population (N in cap$^{)}$, water use intensity (w_N in $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$), and climatic water balance (q_{Net} in $\mathrm{m} \cdot \mathrm{yr}^{-1}$) to represent simple profiles of urban water supply and demand, which I discuss in Section 3.3.1. I use the Urban Metabolism Group at MIT (UrbMet) dataset of 142 cities, as a useful starting point for constructing an appropriate dataset for the large-*n* comparative water research [86, 231]. These UrbMet cities are spread across six continents and four of the five major climate types and provide a (moderately) large-*n* sample of cities with reasonable international and climatic coverage. I then link population and water use intensity data from UrbMet to larger datasets, especially the `world.cities` database for population and the International Benchmarking Network for Water and Sanitation Utilities (IBNET) for water use intensity. I draw city location data from `world.cities` and compared these with data from the Google Maps API, and then use this data to scrape climatic data from the Web-based, Water-Budget, Interactive, Modeling Program (WebWIMP) for climatic water balance. These methods are described in Section 3.3.2.

Characterization of underlying data distributions is an essential part of any large-*n*, multivariate comparison. In the first part of the analysis, described in Section 3.4, I examine the distributions of population, water use intensity, and climatic water balance data using several basic exploratory statistical methods. For each univariate distribution, I calculate statistical moments (such as the mean and standard deviation) and quantiles and plot histograms, ordered bar charts, and quantile-quantile (qq)-plots. I also apply a $\log_{10}$ transformation to rescale the data for visualization and clustering, as suggested by the results of the plots of the untransformed data. To round off the initial data exploration, I perform a simple bivariate analysis by plotting bivariate combinations on scatterplots and calculating a correlation matrix. I present the results of this first phase of large-*n* analysis in Section 3.7.

In the second stage of data exploration, which I describe in Section 3.5, I employ two statistical data-mining methods, `hclust` and t-SNE, to the urban water profiles of the 142 cities. I explore the application of these methods to univariate, bivariate, and trivariate data to identify groups of cities with similar urban water profiles, as I explain in greater detail in Section 3.5.2. I find that the results of univariate and bivariate clustering, which I present in Section 3.7, do not results that are distinct across all three metrics that I use in the profiles. However, I do find that the results of trivariate clustering, which I present in Section 3.7, lead to six groups that are distinct and intuitive, and that these groups would have been difficult to identify *pre facto*. To provide further insight into the six groups that I identify through trivariate clustering, I develop a new indicator for environmental impact and water footprint, the Water Use and Climate Index, which I introduce in Section 3.5.4. I define Water Use and Climate Index (WUCI) as the ratio of precipitation height to water use intensity; it can be interpreted as the relative ability of the local climate to meet water demand. I present these results in Section 3.8.4.

Finally, in Sections 3.9–3.11, I summarize these results and conclude that they provide a step towards a global perspective and identify interesting opportunities for case study analysis.

In the next section, Comparative Analysis in Water Resources Research, I summarize other approaches to large-*n* analysis in WRM, before moving on to the choice of metrics and data methods in Section 3.3.

This chapter is joint work with James L. Wescoat and John E. Fernández.

3.2 Comparative Analysis in Water Resources Research

Since there are a large number of variables relevant to SUWM, and since relationships between them are often non-linear due to feedbacks or other natural phenomena, it is desirable to have a significant number of data points for many indicators and metrics over time. Having data for a larger number of points can help to provide a better characterization of the range of a phenomenon by descriptive statistics or other methods.

While larger surveys are beginning to benchmark various water management variables of water demand and water infrastructure performance, when statistical analysis of collected metrics is performed it is often limited to simple descriptive statistics such as averages and extremes for separate variables[1]. The full power of descriptive statistics is only rarely tapped: while the mean, median, and standard deviation are commonly reported, additional information such as skewness and kurtosis could also be included. The opportunity to analyze the relationships between and amongst variables (such as autocorrelation) through regression analysis and other inferential statistical techniques is only ever pursued for a very small number of variables, if at all. Not surprisingly, given the state of the field, benchmarking is sometimes discounted as insensitive to specific urban contexts. Given the large number of performance metrics often reported in such a benchmarking survey (e.g., the IBNET database, with over 200 variables) it is nearly impossible for a city decision maker (let alone a water infrastructure export) to get a sense for how comparable one city's water management context might be to that of another. Finally, these two types of analysis—case studies and benchmarking—are rarely linked with one another, which misses an important opportunity to provide a quantitative basis for situating important precedent projects within broader socioeconomic and biogeophysical patterns and trends in urban water management.

More recently, statistical clustering and machine learning have arisen that help to combine these approaches and to support the identification of patterns and trends that would be difficult for a human to identify without a computer.

In the field of water resources, early examples of comparison of larger sets of observations include the Köppen climate classification. First published by Wladimir Köppen in 1884, this early typology classified regions around the world using ranges and extreme values for climate variables, especially precipitation and temperature [280]. More recent versions of this system have grouped data using statistical data mining algorithms such as classification and regression trees [160, 249, 276, 361].

Large-N comparative studies relating social variables to resource use and population have been performed [33, 158]. Hall et al. examined storage per capita, institutional capacity, hydrologic variability, and the coefficient of monthly runoff for a large number of major watersheds around the world [121]. Hall et al. chose to visualize the data as a scatterplot of two composite

[1] e.g., employees or expenditures per 1,000 customers served

indices [121]. More recently, some large-n analyses have incorporated additional water utility metrics or watershed-focused analysis [265, 196]. However, these analyses have stopped short of using data mining techniques to further explore multivariate trends.

However, the use of data mining algorithms to classification problems for urban water issues is a more recent development, as is the classification of urban water systems within a particular country [358, 359, 265]. For example, clustering has been applied to the problem of forecasting short-term water demand within a single city or municipal water system, which falls under the category of a small-n analysis [99, 55, 357]. Other clustering studies fell into the medium-N category of comparative analysis, including one that includes a k-means clustering of cities based on water footprint, energy consumption, and municipal waste within the United Kingdom [157]. A study by Mayer, Winkler, and Fry used cluster analysis to classify watersheds in the Great Lakes basin according to social and environmental attributes [192]. And a large-n study by the Columbia University Water Center used hierarchical clustering to analyze utility rates in the United States with regards to financial sustainability [265].

Saldivar-Sali's 2010 master's thesis, "A Global Typology of Cities: Classification Tree Analysis of Urban Resource Consumption" generated a typology of the UrbMet database of cities (which was originally developed for the purpose) on the basis of material consumption profiles and identified 15 types of cities [279]. In contrast to the study pursued in this chapter, Saldivar-Sali used classification and regression tree analysis to identify threshold levels, using a training set of cities, that would then partition cities in the testing set of cities. These thresholds were found for the predictor (i.e., independent) variables of city GDP (estimated from national data), population, population density, and climate type (Köppen Climate Classification) [279]. There were eight dependent variables in the analysis, the resource consumption variables, which included average annual per capita water consumption in addition to energy use, biomass, construction materials, and others [279].

While Saldivar-Sali's was an important step in that it applied large-n data mining methods to an international set of cities, it was not clear from that the cities in resulting groups were more similar to each other than to cities in other groups [279]. Following that research, another UrbMet study examined the same dataset with hierarchical clustering to see if the same types of cities would be identified from a bottom-up versus a top-down type of approach [86]. In contrast to the previous results, the study by Fernández et al. identified only eight types of cities, and the groups were found to be different from those identified by Saldivar-Sali [86]. The study by Fernández et al. did not disprove the validity of Saldivar-Sali's results but did demonstrate that, especially in exploration of multivariate data, different data mining approaches can lead to very different results. This result is somewhat intuitive since complex multivariate data spaces do not necessarily lend themselves to conclusive results.

While the number of studies using data mining approaches to gain insight into large-*n* studies of cities, it is still relatively new in studies of SUWM [157, 218, 198, 231]. It therefore seemed worthwhile to apply clustering algorithms to profiles of urban water supply and demand for a large-*n* set of cities, even if that set of cities had been used in the past[2].

Another short-coming in both the application of data mining algorithms to and in more general large-*n* studies of IWRM or SUWM was that these studies had not been used to structure more in-depth case study research. This seemed like an oversight for two reasons: 1. large-*n* comparative analysis may not fully characterize the specific dynamics of complex systems like UWS and 2. in spite of this, even an imperfect large-*n* study could provide a useful perspective from which to pursue systematic case-study research.

In the next section, Data, the rationale for the choice of metrics used in this study has been described. Information is also provided on the data sources and methods.

3.3 Data

3.3.1 Choice of Metrics

The first step was to choose appropriate metrics and acquire data. Three metrics were chosen to represent urban water demand and supply profiles:

Population N,

Water use intensity w_N ($\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$), and

Climatic water balance net water balance[3] (q_{Net})

The starting point for the data was the UrbMet database of cities, which included data for population and water use intensity. The UrbMet database also included information on the main climate classification according to the Köppen climate classification scheme, but dataset lacked more detailed information about local natural water balance and climatic availability. Since it was desirable to look at urban water use within the context of more detailed information on local natural water resources availability, it was necessary to identify a source of data for climatic information. Average monthly and annual climatic water balance data were scraped from WebWIMP.

The first was an annual climatic water budget for each urban area, q_{Net}. q_{Net} (short for annual balance) is equal to precipitation (q_P, or **PREC**) minus snowmelt (q_M) and estimated adjusted potential evapotranspiration (q_{ET^0}), and was obtained from WebWIMP, an online interface that provided access to the results of a water balance analysis, which estimates average monthly components to 0.5 degree gridded results for a climatic water balance of global terrestrial locations.

[2]The UrbMet database of cities used in this study had been previously used in two other studies, as discussed previously [279, 86]

[3]Precipitation minus adjusted potential evapotranspiration and snowmelt. See Section 3.3.2, Section A.1.2, and Table A.2 in Appendix A for more information.

For instance, the WebWIMP interface provided estimates of climatic water supply (q_P) and local natural water demand due to evapotranspiration (q_{ET}), with a full listing of components shown in Table A.2 on Page 262. WebWIMP results were available as average monthly and annual values of historical climatic data or as estimates from said averages of water balance components. Derived components were estimated using the adjusted Thornthwaite approach, which has been described further in Appendix A.1.2 [351, 190]. Data from WebWIMP were scraped according to the methods described in Appendix A.1.2, with the scraped data provided in Appendix C.1.4.

Population and average annual water use intensity (i.e., per capita water use) were the two other metrics used in the large-n comparison.

According to many environmental and sustainability studies, population size is a primary driver in resource consumption and other environmental impacts. If water use intensity (w_N in $m^3 \cdot yr^{-1} \cdot cap^{-1}$) is average per capita water use, then total water use (W_N) for a population of size N can be written as:

$$W_N = w_N * N \tag{3.1}$$

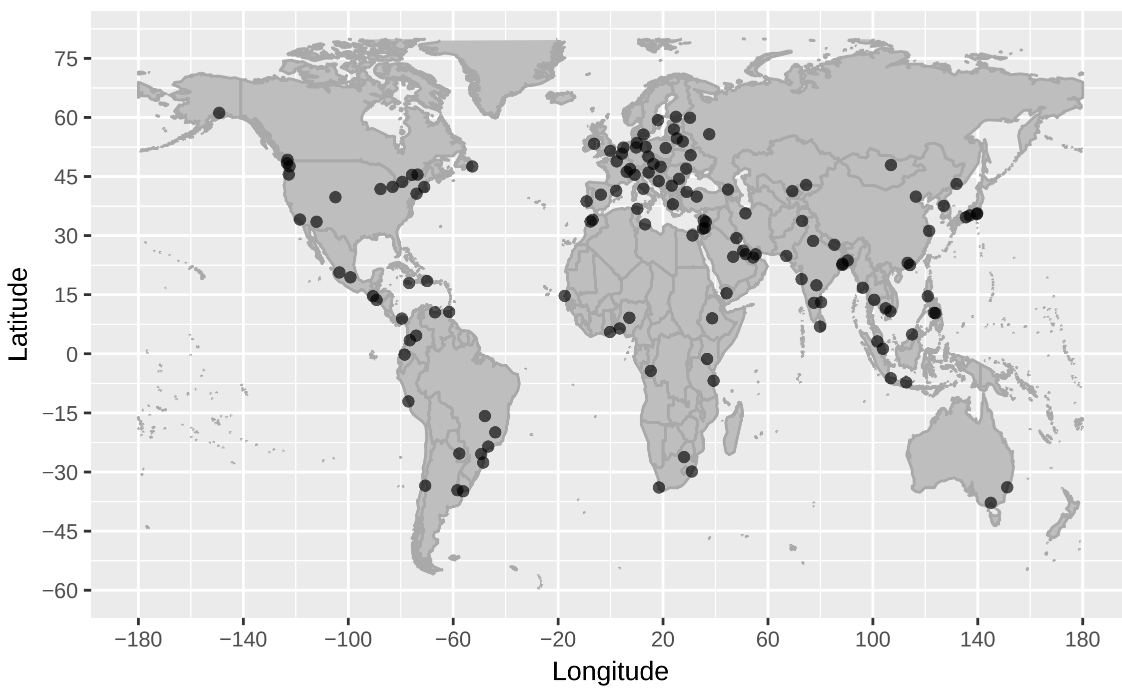


Figure 3.1: Map of the 142 cities used in the large-n comparative analysis, illustrating their distribution throughout the world. The cities represented 91 countries, six continents, and four of the five major climate types [279]. Country borders and outlines of continents were provided by the *maptools* package for R.

Equation 3.1 is one variation of the IPAT equation, a popular sustainability heuristic for understanding drivers of environmental impact. The IPAT equation, has multiple forms, but its titular form is that shown in Equation 3.2:

$$I = P \cdot A \cdot T \tag{3.2}$$

In Equation 3.2, a total environmental impact, I (e.g., resource use), is the product of population size, P (i.e., N); affluence, A (e.g., Gross Domestic Product (GDP)/capita); and technology, T. If T is interpreted as resource intensity, e.g., average resource input per unit of GDP output, then I has the same units as the resource input, e.g., mass or volume.

Equation 3.1 is similar to Equation 3.2, since W_N can be interpreted as an environmental impact (I), N in Equation 3.1 is equal to P in Equation 3.2, and w_N has the same units as the product of $A \cdot T$.

While IPAT is presented as an equation, it is more of a heuristic of driving and ameliorating factors [57].

$$Q_P = q_P * A_N \tag{3.3}$$

Equation 3.3 says that the total volume of water from precipitation (Q_P) generated within a city can be estimated from precipitation height (q_P) and city area (A_N). However, the volume of precipitation (Q_P) is generally not equal to the total surface runoff, due to losses from infiltration to the ground or evapotranspiration.

Since WebWIMP provided an estimate of the net climatic water balance (q_{Net}, called the "difference" or "DIFF" in WebWIMP), which was defined as the sum of rainfall (q_P) and estimated snowmelt (q_M) minus the estimated, adjusted potential evapotranspiration (q_{ET^0})[4]:

$$q_{Net} \; = \; q_P \; + \; q_M \; - \; q_{ET^0} \tag{3.4}$$

Net climatic water balance (q_{Net}) was chosen as the climatic variable of focus in this large-n comparative analysis instead of precipitation height (q_P), since it provided an estimate of natural water availability. It was reasoned that, all other things being equal, a city that is located in an area that has low natural water availability would be more likely to have different water issues from one with high natural availability. Taking q_{Net} as the variable of interest provided a way to distinguish between two (theoretical) cities with similar precipitation but different climates; one might be hotter than another, for instance. Since WebWIMP calculated q_{Net} using Thornthwaite's equation, which included an estimate of local evapotranspiration, which itself depended on (assumptions of) temperature and irradiance as well as local vegetation type, it was chosen as the data source, especially as data were available for all terrestrial locations.

WebWIMP precipitation (q_P) data were used in the calculation of a new type of environmental footprint, the Water Use and Climate Index, as described in Section 3.5.4.

[4]Descriptions of these variables have been listed in Table A.2 on Page A.2)

Table 3.1: Overview of variables used in clustering.

Abbr.	Metric	Unit	Minimum	Maximum
N	Population	capita	$2.728 \cdot 10^4$	$1.435 \cdot 10^7$
w_N	Water use intensity	$m^3 \cdot yr^{-1} \cdot cap^{-1}$	14	355
q_{Net}	Climatic water balance	$m \cdot yr^{-1}$	-1.446	3.833

3.3.2 Data Methods

Comparative analysis of urban water systems, whether small-n or large-n, is often hindered by questions about the comparability of data quality or system definitions and boundaries. While it was not the intention of this study to address these concerns directly, efforts were made to draw from standard datasets (e.g., from WebWIMP for climatic water balance or IBNET for water use intensity), as described in Section 3.3.2 and Appendix A.1.2.

The starting point for this study was a dataset developed by the UrbMet for use in typological exploration and analysis of resource consumption at the city level [279, 87, 231].

As described earlier, the UrbMet dataset was created by Saldivar-Sali in 2010 and consisted of 141 cities from 91 countries across six continents and four climate types[5] [279]. These cities have been shown plotted on a world map in Figure 3.1. Five variables from the UrbMet database were used in the large-n comparative analysis. Population (N) and water use intensity (w_N) were used in preliminary data exploration and in clustering analysis. City and country names were used to obtain location data for use in web-scraping climatic data from WebWIMP. City area (A_N) was used in the supplementary results.

To the initial data from the UrbMet database, data on net climatic water balance (q_{Net}) were added from WebWIMP as a measure of local climatic water availability, along with other components of water balance. The three variables, N, w_N, and q_{Net}, were the variables used in the statistical clustering analysis. These three metrics have been summarized in Table 3.1.

Population (N) City population (N) from the UrbMet database was included to provide a measure of city size and total environmental impact, in which population has been found to be a primary determining factor [57, 91]. Population in the estimates from the early 2000s were conservatively based on city boundaries vis-á-vis larger metropolitan regions [279].

Water Use Intensity (w_N) Per capita water consumption (w_N) provided a measure of the intensity of water use in the urban context. Data on w_N, in cubic meters per year ($m^3 \cdot yr^{-1} \cdot cap^{-1}$), were taken from the UrbMet dataset (Table A.1). The value for w_N in the UrbMet dataset was drawn in most cases from the World Bank-supported International Benchmarking Network (IBNET), supplemented by city-specific data when not available in IBNET [279, 144].

[5] There were no cities from Antarctica and no cities of the polar Köppen climate type.

IBNET sets common definitions for utility indicator and metrics and also provides guidelines to utilities for collecting these data [144]. For w_N, the metric chosen was IBNET Indicator 4.1, "Total Water Consumption" in units of liters/person/day (defined as the "Total annual water sold expressed by population served per day") [144, p. 3]. In other words, the "Total Water Consumption" provided by IBNET (v_U) was calculated by dividing the gross annual water sales of a water utility (V_U) by the number of people in its service area (N_U) over the number of days in a year ($d = 365$):

$$\text{Total Water Consumption} = \frac{V_U}{d * N_U} \tag{3.5}$$

Therefore, w_N was simply the value provided by IBNET for "Total Water Consumption" multiplied by 365 (the number of days in a year) and divided by the number of liters in a cubic meter (1 m^3 = 1000L).

$$w_N = v_U * \frac{d}{1000} = \frac{V_U}{1000 * N_U} \tag{3.6}$$

Later, when interpreting the results of the analysis, it was important to keep in mind limitations to using IBNET's Total Water Consumption as a measure of urban water demand. First, Total Water Consumption was calculated from the volume of water, V_U, that was *sold* by a utility over a year. Therefore, the value for w_N (which was derived from Total Water Consumption) is sensitive to the methods and accuracy with which water sales are monitored and accounted for by a utility. A second caveat of using IBNET's Total Water Consumption as a measure of urban water demand for a city is that of representation. Since data were reported by a utility for its service area and some cities may be serviced by more than one utility, this raises questions about the degree to which the reported value was representative of water use for the entire city. In addition, the data reported by the utility did not provide information about water use by those users that may use water from groundwater or other non-municipal sources either as primary or supplementary sources.

In other words, the value reported for IBNET on Total Water Consumption in some cases reflected water use by an unknown subset of a city's population. Another limitation of IBNET's Total Water Consumption data was that water demand was not divided into residential, commercial, or other components. Focusing on IBNET's Total Water Consumption data meant that the clustering analysis was unable to distinguish between water consumption due to industry and domestic water uses (though that would be a valuable extension of this research).

While these concerns limited the accuracy and specificity with which conclusions could be drawn or generalized, IBNET's Total Water Consumption data was considered to provide a reasonable approximate measure for the intent of the study in comparing the relative average water use intensity across cities.

Climatic water balance (q_{Net}) data were used as an estimate of local water availability and gross annual water supplies, using the University of Delaware's 0.5 degree grid WebWIMP tool [351]. WebWIMP (the Web-based, Water-Budget, Interactive Modeling Program) is an interface and dataset developed by Kenji Matsuura, C. Willmott, and D. Legates at the University of

Delaware in 2003, with updates made in 2009. The WebWIMP interface, available at `climate.geog.udel.edu/~wimp/`, provided access to data on climatically-averaged monthly water balances to 0.5 degree resolution using a modified Thornthwaite approach [351, 350, 191, 190][6].

Because of the relatively large number of cities for which the data was needed a Python script was written to scrape the requisite tables from WebWIMP. The Python script entered the latitude and longitude for each city into the WebWIMP interface and scraped the values for average annual and monthly water balance and its components as returned by WebWIMP (further detail and supporting code has been provided in Appendix A.1.2). The data for q_{Net} reflected the most recent values available from the online WebWIMP interface.

3.4 Preliminary Data Exploration

Statistical moments and quantiles were calculated for the univariate distributions of N, w_N, and q_{Net}. Univariate distributions were also examined using histogram, density, and qq-plots.

In the second part of the analysis, statistical exploratory and data mining techniques were used to characterize the urban water metabolism of the two case studies relative to a larger, international set of cities. The results were recast as an international typology of urban water sustainability situations. This typology is a first step in shedding light on the similarities and differences between the two case studies relative to other cities around the world with respect to their water supply and demand.

In the second part of the large-n comparative analysis, the data for population, water use intensity, and net water balance were characterized using basic, descriptive statistical methods. Univariate analysis, consisting of the calculation of summary statistics, visualization (ordered bar charts, histograms, and qq-plots), and a $\log_{10}$ transformation of the data. Basic Bivariate analysis consisted of the calculation of correlation coefficients and visualization as scatterplots.

Visualization methods provide visual context to support summary statistics. Visualization of the underlying data is a crucial step in exploratory data mining, as it provides clues as to outliers and the underlying distributions. Initially, population (N), water use intensity (w_N), water balance (q_{Net}), and WUCI (i_{UC}) were explored using histograms, qq-plots, box-/violin-plots, ordered bar charts, and scatterplots.

Each metric was plotted as an ordered bar chart—i.e., a bar chart with the cities ordered according to their value for that variable. The ordered bar charts for N, w_N, q_{Net}, and i_{UC} are shown in Figures 3.3–B.5. Figures 3.3 and 3.4 have a similar distribution of positive values, though Figure 3.3 has a longer tail of smaller cities than Figure 3.4. In contrast, Figure 3.5 ranged from negative to positive values for net annual water balance.

[6]These underlying data were generated from underlying spatial datasets on climate data (including precipitation and temperature), as well as information about local soil and vegetation and are available at Willmott, Matsuura and Collaborators' Global Climate Resource Pages from the University of Delaware.

Table 3.2: Overview of statistical moments.

Moment Number	Name	Symbol	Formula
1	Mean	μ	$\mu \equiv \mathrm{E}[X]$
2	Variance, or standard deviation	σ	$\sigma \equiv (\mathrm{E}[(x-\mu)^2])^{1/2}$
3	Skewness	γ	$\gamma = \frac{\mathrm{E}[(x-\mu)^3]}{(\mathrm{E}[(x-\mu)^2])^{3/2}}$
4	Kurtosis	κ	$\kappa[X] = \frac{\mathrm{E}[(x-\mu)^4]}{(\mathrm{E}[(x-\mu)^2])^2}$

3.4.1 Univariate

Summary statistics

Basic descriptive statistical methods can be a useful tool in characterizing the attributes by which cities or other cases are to be compared. The following descriptive statistics were calculated for the population (N), water use intensity (w_N), and annual climatic water balance (q_{Net} or annual climatic water balance, i.e. or annual balance, net annual balance, or climatic water balance (CWB)):

Quantiles: at 0%, 25%, 50%, 75%, and 100% probabilities;

Statistical moments: the mean μ, standard deviation σ, skewness γ, and kurtosis.

Moments In classical mechanics, objects are often described in terms of *total mass*, *center of mass*, and *rotational inertia*, which are known as the zeroth, first, and second moments. Since in statistics, data measurements are considered to exist as points in mathematical space, data can be considered to have aggregate properties analogous to matter in the physical world, such as mass, density, and shape. Statistical moments are a common basic method in descriptive statistics and are calculated for a data series from its *probability density*. While in principle there is a general formula for defining the nth-moment for a series, in practice, there are four moments which are commonly used to describe the shape of a probability density[7]: the *mean* (μ), the *standard deviation/variance* (σ), the *skewness*[8] (γ), and the *kurtosis* (κ). The formulas for these four moments are shown in Table 3.2.

The first and second moments, i.e., the mean and standard deviation, are by far the most common of the first four statistical moments. The mean is akin to the center of mass, while and the standard deviation is similar to the rotational inertia of an object in univariate physical space.

The third and fourth moments, skewness and kurtosis, are normalized (i.e., dimensionless moments). Skewness provides a measure of the symmetry of the distribution; a symmetric distribution has $\gamma = 0$. The magnitude of γ indicates the relative lopsidedness of the distribution,

[7]Or five, if the zeroth moment, mass (i.e., number), is included (considered to be the "zeroth" moment).
[8]Skewness is also known as Pearson's moment coefficient.

with the sign indicating whether it is skewed to the left ($\gamma < 0$) or right ($\gamma > 0$). Kurtosis is an indicator of the length and heaviness of the tail of a distribution relative to a normal distribution with the same variance. It is an indicator of the extent to which the mass of the object is concentrated towards the mean.

The mean, variance, skewness, and kurtosis were calculated using the `R` functions `mean`, `sd`, `skewness`, and `kurtosis`.

Quantiles Statistical quantiles also provide insight into the rough shape of a data distribution. Quantiles are calculated "cut" points that divide the range of values into partitions of equal probability, such that each partition contains an equal or approximately equal number of observations. For this study, the cut points for four equal quantiles (i.e., the *4-quantiles*, or *quartiles*) were found. The value x_k is defined as the k^{th} 4-quantile for a distribution X if $\Pr[X < x_k] \leq k/q$ where($\Pr[X < x_k]$ is the probability Pr that X is less than x_k). For example, the cutpoint x_2 second 4-quantile (also known as the *median*) for a set of observations is defined such that an equal number of observations fall above and below the median.

Four quantiles (i.e., at 0%, 25%, 50%, 75%, and 100% probabilities) were calculated using the `R` function `quantile`.

$\log_{10}$ transformation

In statistics, a data series can be transformed by application of a deterministic function to all points in a dataset to improve statistical analysis and/or visualization. Each data point, x_i, in a data series, $\mathbf{x}$, is transformed into $z_i = f(x_i)$, where f is a deterministic function. Typically, the function chosen is invertible and continuous.

Data transformation is performed in statistical analysis if a data series exhibits substantial skewness and the sample size is only moderate, the assumption of a normal distribution can lead to the wrong coverage probability (i.e., the probability that a confidence interval contains the true value). In this case, it is common to try to identify a transformation that improves the symmetry of the distribution. Applying a transformation to skewed data may also be done simply to improve visualization—for instance, to distribute tightly clustered data over a broader space.

Logarithms are one popular function for transforming data. For instance, a scatterplot of the population versus land area for countries around the world will tend to have a tight cluster of points at the lower one end of the graph; transforming both area and population by a logarithm can help to distribute the data more uniformly over the two-dimensional space.

There are also theoretical reasons to perform a logarithmic transformation, especially for population and water use intensity. Over the last several decades, improved collection and sharing of data has led to the emergence of a new science of cities, which is focused on uncovering patterns in urban form, socioeconomic activity, material flows, and other urban attributes [152, 24, 32]. In recent years, empirical evidence has accrued in support of the theory that cities grow randomly but proportionally and therefore the scaling of urban attributes with city size can be approximated by a log-normal distribution [98, 145, 72, 25]. In numerous studies, Bettencourt and others have found that many "urban properties, Y, vary continuously with population size [N] and are well described mathematically on average by power-law scaling relations of the form $Y = Y_0 N^{\beta}$, where Y_0 and β are constants in N [32, p. 1438]" [35, 25, 34, 32]. In two separate studies, Bettencourt et al. and Bettencourt and West analyzed data for numerous urban attributes, including patents, income, GDP, and housing and found scaling factors (β) with high R-squared values [35, 34].

In the histograms of the untransformed data, all three variables—N, w_N, and q_{Net}—demonstrated skewness to the left (see Figures 3.6a, 3.6c, and 3.6e). In the ordered bar charts of the untransformed data (see Figures 3.3–3.5), all three variables exhibited curvature suggestive of an exponential function. And substantial curvature was observed in the qq-plots of the untransformed data (see Figures 3.7a, 3.7c, and 3.7e).

For these reasons, a logarithmic transformation was applied to N, w_N, and q_{Net} of the form:

$$z_i = log_{10}(x_i + a) \tag{3.7}$$

where z_i is the $\log_{10}$ transform of x_i and a is a constant. A constant value of $a = 0$ was assumed in calculating the $\log_{10}$ transforms for N and w_N since both were greater than zero. For q_{Net}, the constant, a, was calculated as $a = \min(q_{Net})$ since some values of q_{Net} were less than zero.

The $\log_{10}$-transformed data were plotted in histograms (see Figures 3.6b, 3.6d, and 3.6f) and qq-plots (see Figures 3.7b, 3.7d, and 3.7f).

Ordered bar charts

For each variable used in clustering, i.e., population, water use intensity, and climatic water balance, cities were ranked from smallest to largest value. A bar chart for each variable was then plotted, with cities ordered by rank. Results are shown in Section 3.6.1.

Histograms

A **histogram** is a type of plot that approximates a density distribution for a finite data series. The range of values is divided into a series of even intervals. Then the number of observations that fall within each interval are counted and displayed as a bar, the height of which is proportional to the number of observations. Alternatively, the count may be normalized by the total number of observations to display a relative frequency.

Histograms were plotted using `ggplot` with the `geom_histogram` function from the `ggplot2` package for `R`[9] [349]. The command to generate one of these plots for variable x, with mean μ_{x} and variance σ_{x} took the general form:

$$\texttt{ggplot}(data = \text{dataset}, \ \texttt{aes}(x = \text{x})) \ + \ \texttt{geom\_histogram}(\texttt{aes}(y = ..\text{density})) \tag{3.8}$$

Two sets of histograms were generated for population, water use intensity, and climatic water balance, one for the untransformed data and the second of the data after transformation by $\log_{10}$. The results are shown in Figure 3.7 and discussed in Section 3.6.1.

Quantile-quantile plots

A quantile-quantile plot is a visualization method used to compare two statistical distributions. To generate a qq-plot for observed data requires an assumption of a theoretical distribution, such as the normal distribution. The theoretical distribution is used to generate a random sample equal in size to the number of length of observed data. The quantiles of the observed and theoretical data are then plotted against each other. If the two distributions are a good match, then the quantiles will be linearly correlated.

For the qq-plots, data were compared to the normal distribution. The shape of a normal distribution to which each series was compared was defined by the μ and σ of the series. Quantile-quantile plots were plotted using `ggplot` with the `stat_qq` function. The command to generate one of these plots for variable x, with mean μ_{var} and variance σ_{x} took the general form:

$$\texttt{ggplot}(data = \text{dataset}, \ \texttt{aes}(sample = \text{x})) \ + \ \texttt{stat\_qq}() \tag{3.9}$$

Two sets of qq-plots were generated for population, water use intensity, and climatic water balance. For the first set, the untransformed data for each of the three variables were plotted against the normal distribution; for the second set, the data were first transformed by $\log_{10}$. The results are shown in Figure 3.7 and discussed briefly in Section 3.6.1.

3.4.2 Bivariate

Scatterplots

For a scatterplot, data are plotted on Cartesian coordinates as defined by two variables. The following combinations were plotted as scatterplots:

- water use intensity and climatic water balance (w_N versus q_{Net}),
- population and climatic water balance (N versus q_{Net}), and
- water use intensity and population (w_N versus q_{Net})

[9] As with the descriptive statistics (above) values were removed.

Scatterplots were plotted for each bivariate combination using `ggplot` with the `geom_point` function. The command to generate one of these plots for the bivariate combination (var_1 and var_2) took the general form:

$$\texttt{ggplot}(data = \text{dataset}, \texttt{aes}(x = \text{var}_1, y = \text{var}_2)) + \texttt{geom\_point}() \tag{3.10}$$

Correlation Matrix

Correlations between N, w_N, and q_{Net} were calculated and plotted as a correlation matrix. The correlation matrix was plotted using the `ggcorr` function from the `ggcorr` library in `R`.

3.5 Statistical Clustering and t-SNE

The third part of the analysis explored the application of two exploratory data mining algorithms—hierarchical clustering (`hclust`) and t-SNE—to the 142 urban water demand and supply profiles. The hierarchical clustering algorithm was first applied to each variable (univariate) and then to two variables[10] (bivariate). Finally, the trivariate analysis (of all three variables, i.e., population, water use intensity, and net water balance) combined `hclust` with t-SNE. An overview of the workflow for clustering is shown in Figure 3.2.

An overview of clustering has been provided in the following subsection, Section 3.5.1, followed by the workflow and methods used in this study in Section 3.5.2.

3.5.1 Clustering Overview

Methods for descriptive statistical analysis may be considered essential to any analysis. However, when comparing a large number of observations on more than two variables, an analysis that is limited to descriptive statistics is unlikely to provide much general insight. Fortunately, the digital age has brought about advances in characterizing underlying patterns in multivariate data (i.e., data mining), even when the structural relationship between the variables is not known or defined *a priori*.

Clustering is one approach to data mining and is a creative, exploratory type of analysis. A researcher starts with a set of objects to cluster; the choice of attributes on which to cluster the data may be made before or after obtaining the appropriate dataset. Although statistical clustering does not require any specification of the structural relationship between variables, it does require some *a priori* knowledge for best results. At the very least, the choice of attributes reflects the researcher's understanding of which might lead to interesting results.

The main objective of any clustering method is to partition a set of observations into groups (i.e., clusters) such that members within a cluster are more similar to each other than to other groups. Numerous methods and algorithms for clustering exit for identifying trends in similarities and dissimilarities across members of large multivariate datasets. Common clustering

[10]The bivariate combinations used were the same as those used to plot scatterplots, as listed in Section 3.4.2.

Table 3.3: Comparison of common clustering algorithms

Algorithm	**Clustering model**	**Pros**	**Cons**
Hierarchical	Connectivity	Very visual results Can handle skewed clusters Number of clusters is flexible	No unique partitioning No notion of noise (i.e. outliers show up as distinct clusters or force merging) Data-hungry
k-means	Centroid	Computationally efficient	Number of clusters must be pre-specified Performs better with clusters of similar size Results dependent on seeding Assumes symmetric clusters
Expectation Maximization (EM)	Probability distribution (pdf)	Robust to missing values Results capture correlation and dependence of attributes	Outliers may be excluded from cluster results Tendency to overfit
DBSCAN	Density	Robust to outliers Pre-specification of cluster numbers not required	Soft membership (outliers may belong to more than one cluster) Results less intuitive

approaches include hierarchical clustering, *k*-means clustering, and model- or density-based clustering (summarized in Table 3.3). These algorithms differ in how the metric used to establish similarity, the formula used for calculating similarity, the method for iteration, and other factors.

Metrics for quantifying similarity are important in all of the clustering algorithms, which may also use as the basis of identifying clusters: areas of density within the data space, or statistical distributions. An example of a density-based algorithm is the so-called k-means algorithm. In contrast to hierarchical clustering, the k-means clustering approach is formulated as an optimization problem. Cluster centers are identified, and objects assigned to the nearest cluster, such that the squared distances of the objects from the cluster centers are minimized. However, this problem is considered to be NP-hard, and most algorithms find the local optimum rather than the global optimum.

Each of the clustering approaches has relative pros and cons, summarized in Table 3.3 on 84. For instance, the k-means approach works best when it can be assumed that cluster distributions follow a more or less normal distribution within the data space and that there is a relatively good separation between them. It does not do well with distributions that are skewed or have significant overlap. The expectation-maximization (EM) approach to clustering is an iterative method involving two main steps: first, the expectation step, which generates a value for the expectation of the log-likelihood, which is then evaluated using the current estimate of the parameters, and the second step, which computes parameters to maximize that expected log-likelihood. These parameters are then used in the expectation step, until a stable estimate has been reached. The hierarchical clustering approach uses a distance matrix and linkage criterion, described further in Section 3.5.1, to perform a step-wise grouping of each observation until all observations are linked in a single cluster. While hierarchical clustering can handle skewed and overlapping data, it is not robust to noise and does not produce a unique partition; groups are determined after the fact, as described in Section 3.5.1. However, the nested hierarchy that precludes a unique partitioning has the benefit of being easy to visualize.

The aim of the grouping exercise was to gain insight into relative sustainability of urban water metabolism around the world. The set of objects was the UrbMet database, consisting of 142 observations (the attributes for which are listed and described in Table A.1 in Appendix A.1.1). For this purpose, three attributes were chosen as indicators of water supply and demand:

1. population,
2. water use intensity, and
3. net climatic water balance;

this choice has been further described in Section 3.3.1. Since the main objective of the large-n comparison was exploratory, rather than predictive, it was decided to focus on the hierarchical clustering approach, which produces relatively intuitive and readable results. For the trivariate clustering, hierarchical clustering methods were combined with another clustering-like approach called t-distributed Stochastic Neighbor Embedding (t-SNE), which uses an iterative approach similar to the EM method for clustering. t-SNE is an approach for reducing high-dimensional data to a two-dimensional space that is easier to visualize. Instead of identifying specific groups, the t-SNE situated observations within a two-dimensional space that can then be plotted with a scatterplot.

3.5.2 Clustering Workflow

Table 3.4: Distance Formulas

Distance Type	**Formula**
Euclidean	$\lvert\lvert a-b \rvert\rvert_2 = \sqrt{\sum_i (a_i - b_i)^2}$
Maximum	$\lvert\lvert a-b \rvert\rvert_\infty = \max_i \lvert a_i - b_i \rvert$
Manhattan	$\lvert\lvert a-b \rvert\rvert_1 = \sum_i \lvert a_i - b_i \rvert$
Canberra	$d(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{n} \frac{\lvert p_i - q_i \rvert}{\lvert p_i + \lvert q_i \rvert}$
Minkowski	$\left(\sum_{i=1}^{n} \lvert a_i - b_i \lvert^p)^{1/p} \right)$

The hierarchical clustering, `hclust`, and t-SNE both used distances between observations as the main basis by which clusters are distinguished. Aside from the standard Euclidean distance, other common distance formulas include the Manhattan distance and the Mahalanobis distance. These distance formulas have been shown in Table 3.4, along with the Canberra and Maximum distances. The choice of the distance formula used may lead to significant differences in the results.

After the distance matrix was calculated, the hierarchical clustering algorithm uses the distance as a basis for growing a tree-like structure called a dendrogram. The growth of the dendrogram can either be agglomerative or divisive; the `hclust` algorithm used for hierarchical clustering in this study used an agglomerative approach. In agglomerative clustering, the algorithm starts with a number of groups equal to the number of observations. At each iterative step, each group is linked together with the group that is most similar to it until the final level is reached, which is one big cluster[11].

Table 3.5: Table of common linkage criterion used in hierarchical clustering, including mathematical formulations, based on the following notation: (Given n observations of vectors $\mathbf{x}$ and $\mathbf{y}$, where $d(\mathbf{x}, \mathbf{y})$ is any distance or dissimilarity measure between observations or vectors; and two clusters, C_K and C_L, merging to form a new cluster C_M, with centroids c_k, c_l, and c_m and chosen distance or dissimilarity metric, D_{KL} between two clusters; and mean vector $\bar{x}_K$ for cluster C_K, the distance between C_M and any other cluster C_J: of observations $A\mathbf{x}, \mathbf{y}$ and $B(\mathbf{x}, \mathbf{y})$.

Linkage Criterion	**Formulation**
Ward/minimum variance	$D_{KL} = \frac{\lvert\lvert \bar{x}_K - -\bar{x}_L \rvert\rvert^2}{\frac{1}{N_K} + \frac{1}{N_L}}$
Single/minimum	$D_{KL} = \min_i \in C_K \min_j \in C_L d(x_i, x_j)$
Complete/maximum	$D_{KL} = \max_i \in C_K \max_j \in C_L d(x_i, x_j)$
Average/mean	$D_{KL} = \frac{1}{N_K N_L} \sum_{i \in C_K} \sum_{j \in C_L} d(x_i, x_j)$
Mcquitty	$D_{JM} = [(D_{JK} + D_{JL})/2]$
Median	$D_{JM} = [(D_{JK} + D_{JL})/2] - [(D_{KL})/4]$
Centroid	$D_{KL} = \lvert\lvert \bar{x}_K - \bar{x}_L \rvert\rvert^2$

At each iterative step, the `hclust` identified similar groups on the basis of the linkage criterion, i.e., the method used for determining whether to merge two groups of observations. More explicitly, the linkage criterion was a function that calculates for, for two groups, a new attribute from the pair-wise distances between the observations that comprise them.

[11]For reference, the divisive approach is essentially the same but proceeds in the opposite direction. It begins with one large group consisting of all of the observations and progressively splits the group based on those that are most dissimilar.

Common linkage criteria include the *minimum* (complete) method, the *maximum* (single) method, the *mean/average* method, *centroid* method, and the *minimum energy* methods. These methods have been summarized in Table 3.5, along with the *Mcquitty*, *centroid*, and *Ward/minimum variance* methods. The choice of the linkage criteria affects the behavior of the algorithm towards outliers and also as to whether membership is discrete (i.e., whether an observation may belong to more than one cluster). For instance, in the *minimum/single* linkage method, the similarity of two clusters was the similarity of their most similar members. In other words, the *minimum/single* linkage criterion compared clusters locally; two clusters were considered similar on the basis of where the clusters were closest together, and overall cluster shape was not taken into account. In contrast to the *minimum/single* criterion, the *maximum/complete* method, considered two groups based on the similarity of their most dissimilar members; the *maximum/complete method* resulted in a preference for more compact clusters with small diameters and was more sensitive to outliers than the maximum single method. In the end, the *Ward/minimum variance method* was chosen for use in this study, which identified groupings that minimized variance for the clusters being merged. This was paired with the Euclidean distance formula, as described previously.

Since hierarchical clustering is a connectivity-based model, the results were not a unique partitioning of the dataset. Instead, the clusters were chosen *post facto* by 'cutting' the dendrogram to either the desired number of clusters or the desired height.

t-SNE, or t-distributed Stochastic Neighbor embedding, is an algorithm for reducing the dimensionality of high-dimensional datasets. Developed by Maaten and Hinton, the algorithm facilitates visualization of high-dimensional data by assigning each observation a location in two- or three-dimensions [182]. The observations can then be plotted on a scatterplot. The algorithm uses machine learning to assign locations to each observation such that observations that are most similar end up closer together than objects that are further away.

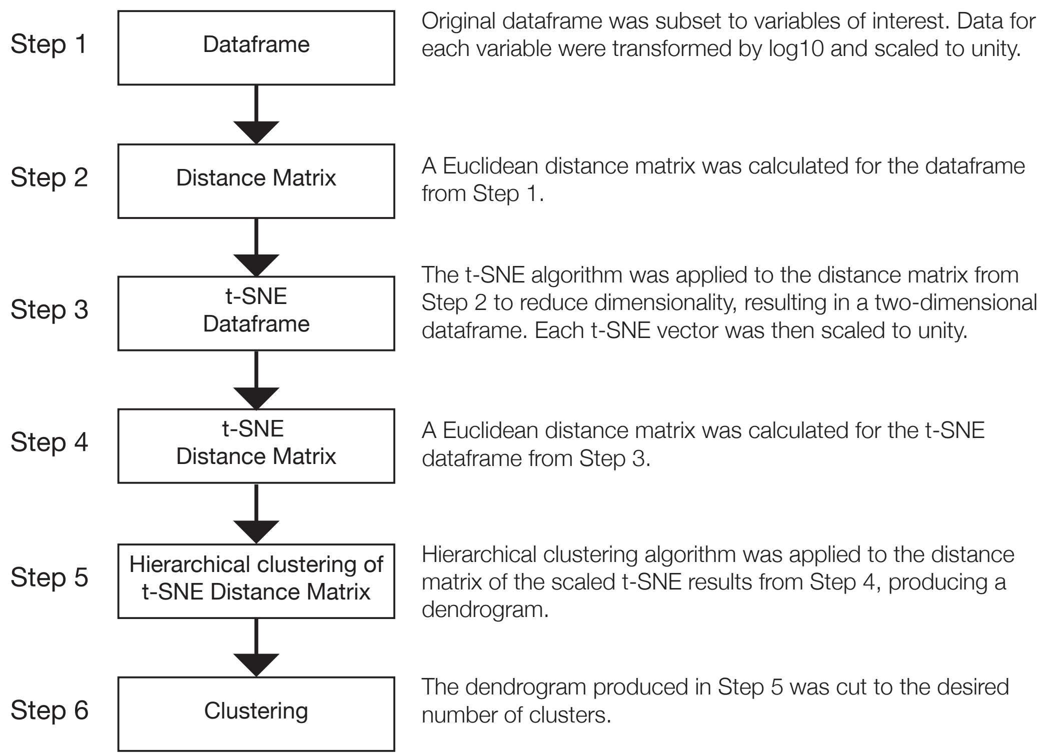
Step 1: Subset, transform, and scale the dataframe For Step 1, the original dataframe was subset to include only the variables used in clustering. For univariate clustering, the dataframe consisted of a single variable; for bivariate clustering, two variables; and the trivariate clustering used a dataframe of N, w_N, and q_{Net}. A $\log_{10}$ transformation was performed on each metric, as suggested by the results of the analysis from Section 3.6 and as described in Section 3.4.1. Each metric was then scaled to unity so that N, w_N, and q_{Net} would all be weighted equally.

Step 2: Calculate a Euclidean distance matrix In Step 2 of the clustering analysis, a distance matrix was calculated for the dataframe from Step 1. The distance matrix was calculated using the `dist` function from the `stats` library in `R`. There were six distance formulas available for `dist`[12], However, in the end, the basic Euclidean distance formula[13]. was chosen to produce the

[12]The distance formulas available for `dist` were the Euclidean, maximum, Manhattan, Canberra, binary, and Minkowski and are provided for completeness in Table 3.4 on Page 85.

[13]see Table 3.4 on Page 85.

Figure 3.2: Overview of the workflow for the clustering.



clustering results in this chapter, since there were no theoretical reasons to prefer a more complex formula and preliminary analysis did not find that other formulas produced substantially different or more interesting results.

Step 3: Reduce dimensionality using t-Distributed Stochastic Neighbor Embedding For the trivariate clustering analysis, the algorithm t-SNE, developed by Maaten and Hinton, was used to reduce the dimensionality of the dataframe from three to two variables [182]. The t-SNE algorithm reduces the dimensionality of datasets with many variables. The t-SNE algorithm iteratively calculates probabilities and statistics to assign observations to points in a two-dimensional space such that neighbors are more similar to each other than to distant observations. The algorithm returns with an embedding of all observations in a two-dimensional space after identifying a local maximum of probabilities.

There were two reasons for applying t-SNE to reduce dimensionality in the data.

First, the t-SNE algorithm was employed to facilitate visualization of the three-dimensional data by enabling the observations to be plotted on a single scatterplot. As previously discussed, the motivation behind the large-*n* comparative analysis pursued in this chapter was to characterize similarities and differences between urban water profiles, in which analysis visualization plays an important role. The embedding of the cities in a two-dimensional space provided another perspective for visualizing urban water profiles. In a two-dimensional variable space, the human eye can to some extent distinguish clustered groups of observations from one another. However, as dimensionality increases, the task becomes substantially more difficult and less intuitive. Therefore, in reducing the dimensionality of the trivariate urban water profiles, the t-SNE approach enabled observations and the resulting cluster results to be plotted in a way that was more intuitive to the human eye and provided visual verification and validation of the clustering results.

Second, the t-SNE algorithm was employed to improve the robustness of the hierarchical clustering results to outliers. Hierarchical clustering's step-wise approach to clustering did not ensure that the "best" clusters would be identified, especially with regards to the outliers, which can cause groups to merge or to appear alone in a cluster. The t-SNE algorithm improved the robustness by embedding each city in a two-dimensional space in a way that placed it nearest to other cities most similar to it.

For these reasons, it was decided to apply the t-SNE algorithm to the trivariate dataframe prior to hierarchical clustering to improve the validity of the clustering results.

For trivariate clustering, the distance matrix produced by Step 2 was passed the dissimilarity matrix given as the argument to the t-SNE algorithm. The product of t-SNE was a two-dimensional dataframe. Each vector was scaled to unity.

The t-SNE algorithm was not applied to the univariate or bivariate clustering exercises since t-SNE is an algorithm for dimensionality reduction. The univariate and bivariate clustering exercises skipped Step 2 and went directly from Step 2 to Step 5.

Step 4: Calculate distance matrix for t-SNE dataframe For trivariate clustering, a distance matrix was calculated for the dataframe produced from Step 3. This second distance matrix was then passed as an argument to the hierarchical clustering algorithm `hclust`, as described in the next paragraph.

Step 5: Hierarchical clustering of the distance matrix The distance matrix produced by Step 2 (for univariate and bivariate clustering) or Step 4 (for trivariate clustering) was passed to `hclust`, a hierarchical clustering algorithm available in `R`, which requires a dissimilarity (i.e., distance matrix). The `hclust` algorithm uses an agglomerative approach to clustering. In contrast to t-SNE, the `hclust` algorithm was only iterative only in the sense that it hierarchically grouped cities, in a step-wise fashion, into a nested hierarchy culminating in a single group. At each step, `hclust` merged groups produced in the previous step (beginning with each city in its own group) based on the linkage criterion, which determined at each step whether two clusters would be merged. The Ward's minimum variance method was chosen for the linkage criterion[14] (shown in Table 3.4 on Page 86), which merges pairs of clusters that minimize the error sum of squares compared to other possible combinations.

[14]The `hclust` function provided the option of seven linkage criteria: Ward, single, complete, average, Mcquitty, median, and centroid, which have been summarized for completeness on Page 86 in Table 3.5.

Step 6: Cutting the dendrogram The hierarchical clustering produced a connectivity model—i.e., a set of connections between observations—rather than a unique partitioning. Therefore, the final step in the clustering workflow was to choose the clusters *post facto*. The 'tree' (i.e., dendrogram) could be cut either to the desired number of clusters or a pre-determined height. Since there was no *a priori* reason to prefer a specific height, results cutting the dendrogram to $k = 2 – k = 8$ clusters were explored for the univariate, bivariate, and trivariate clustering results.

3.5.3 Visualization

Dendrograms

The most common visualization of the results of a hierarchical clustering analysis is that of the dendrogram, in which the nested hierarchy of relationships of similarity amongst the observations is represented as a tree-like structure. The dendrogram of the cluster results shows the step-wise pairing of existing subclusters, with each "branch" in the structure (known as a clade) and each terminal node (known as a *leaf*). In a dendrogram, the length of each branch between nodes indicated the relative dissimilarity between two clades. The dendrogram also helped to visualize the "cutting" of the clustering results to form clusters; the dendrogram was cut to the desired number of clusters[15] to determine the membership of each leaf (i.e., city).

t-SNE scatterplot

The 2-D embedding for the cities produced from the t-SNE method allowed the cities to be visualized in terms of their underlying attributes—N, w_N, and q_{Net}—on a single scatterplot. The cities, as embedded in this space, are adjacent to other similar cities. The scatterplot of the t-SNE space was produced to visually verify and validate the dendrogram results shown in Figure 3.9.

Box-/violin-plots

The cluster results were also plotted as violin plots with boxplots superimposed over them. The violin plot is essentially a box plot with a rotated kernel density plotted instead of a box. The boxplot (or box-and-whisker plot) depicts the median value of each cluster as a band within the box. The top and bottom of each box (the "hinges") represent the first and third quartiles. Boxplots may include "whiskers"—lines vertical lines extending from the tops and bottoms of the boxes. The length of the whiskers is determined by the Inter-Quartile Range (IQR), where $IQR = Q_3 - Q_1$: i.e., the difference between the third and first quartiles. The whiskers extend from each hinge to the value that is within 1.5*`IQR` of the hinge. Any value beyond the whiskers is an outlier and is plotted as a point.

[15]Dendrograms can also be cut to a desired height.

3.5.4 WUCI: the Water Use and Climate Index

To provide insight into the results of the trivariate clustering, Equations 3.2 and 3.1 were re-examined with respect to the variables used in clustering (i.e., N, w_N, and q_{Net}).

This led to the derivation of a novel indicator, the so-called **Water Use and Climate Index** (WUCI, or i_{UC}), calculated as:

$$\text{WUCI} = i_{UC} = \frac{w_N}{q_P} \tag{3.11}$$

In Equation 3.11, i_{UC} is calculated as the ratio of water use intensity (w_N, in $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$) to precipitation ($q_P$, in $\mathrm{m} \cdot \mathrm{yr}^{-1}$). Since w_N has units of $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$ and q_P has units of $\mathrm{m} \cdot \mathrm{yr}^{-1}$, i_{UC} has units of $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$, and can be interpreted as a type of water footprint.

WUCI can be interpreted as providing a rough indicator of the relative local area required to obtain enough water to provide a single average person in the city with enough water over the period of analysis (e.g., year). It is not intended as a close approximation of local water resources, since *actual* local natural water availability depends on numerous other factors, such as surface water and aquifer sources, and the calculation implicitly implies an assumption of a 100% precipitation collection efficiency.

Another issue with q_P is that focusing on precipitation in the calculation of WUCI is that it does not account for rainfall lost to interception, infiltration, or evapotranspiration. However, it was necessary to use precipitation height (q_P) instead of net climatic water balance (q_{Net}) or surplus (q_S), since both q_{Net} and q_S had values close to or equal to zero, which would have led to the calculation of an infinite area.

Multiplying per capita WUCI (i_{UC}) in Equation 3.11 by population (N), then provides a measure of total water footprint (total WUCI, or I_{UC}) for a city:

$$I_{UC} = N \cdot \frac{w_N}{q_P} \tag{3.12}$$

Since total WUCI (I_{UC}) is the product of WUCI and population, I_{UC} has units of area (e.g., m^2 or km^2).

3.5.5 The Potential Self-Sufficiency Ratio

Since total WUCI has units of area, it can be used to calculate an approximate self-sufficiency ratio by dividing total WUCI (I_{UC}) by the city area (A_N)[16]:

$$R_{SS} = \frac{q_P \cdot A_N}{N \cdot w_N} = \frac{A_N}{I_{UC}} \tag{3.13}$$

[16] Also see Equation 4.1 on 153.

As defined by Equation 4.1 on 153, R_{SS} is the ratio of local water supply relative to total local water use. When local water supply is taken to be the flux of precipitation through the city (i.e., the product of local precipitation (q_P, in *volume/time*) and the urban area (A_N, in *area*), and local water demand is taken to be the product of population size (N in *capita*) and per capita water demand (w_N, in *volume/time/capita*), the ratio for R_{SS} takes on the form of Equation 3.13[17].

Since the calculation of WUCI (i_{UC}) and total WUCI (I_{UC}) should be taken as indices rather than actual areas required for water collection, R_{SS} (as calculated in Equation 3.13) should not be taken as the *actual* self-sufficiency of a city. However, R_{SS} (Equation 3.13) does provide a rough indicator of the *potential* for self-sufficiency, and could be considered as a *potential self-sufficiency ratio*.

The Water Use and Climate Index may be compared with other sustainability heuristics such as water footprint analysis, which examines different types and amounts of water use in a system [133, 135]. The footprint idea was adapted into an aggregated metric, called the urban Water Use and Climate Index (i_{UC}, in $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$). This type of accounting is taken further in studies of virtual water trade and sustainable supply chain analysis (e.g., [67, 82, 135, 159, 304]).

As shown in Figures 3.8.4 (Page 124), 4.1 (Page 139), WUCI and total WUCI were found to provide some visual insight into the trivariate clustering results. The potential self-sufficiency ratio was explored more deeply in the case study analysis in Chapters 4 and 5.

3.6 Results of Preliminary Data Exploration

3.6.1 Univariate

Summary Statistics

The summary statistics for population (N in capita), water use intensity (w_N in $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$), and climatic water balance (q_{Net} in $\mathrm{m} \cdot \mathrm{yr}^{-1}$) are shown in Table 3.6.

For the cities in the dataset, population (N) ranged from a minimum of 2.729×10^4 cap$^($ Bandar Seri Begawan, Brunei) to a maximum of 14.35×10^6 capita (Shanghai, China). The median value occurred at 1.649×10^6, (between Guadalajara, Mexico and Damascus, Syria), while the IQR ranged from 6.467×10^5 capita (between Rabat, Morocco and Chisinau, Moldova) to 3.764×10^6 capita (between Hyderabad, India and Melbourne, Australia). These quantile breaks can be seen in the ordered bar chart of N, plotted in Figure 3.3.

Water use intensity (w_N) ranged from a minimum of 14 $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$ (Yangon, Burma) to 355 $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$ (Cairo, Egypt). The median value occurred at 86.5 $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$ (between Prague, Czech Republic and Ljubljana, Slovenia), while the IQR ranged from 57.25 $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$ (between Hamburg, Germany and San Salvador, El Salvador) to 148.5 $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$ (between Anchorage, USA and Geneva, Switzerland). These quantile breaks can be seen in the ordered bar chart of w_N, plotted in Figure 3.4.

[17] Equation 3.13 can also be rewritten in terms of population density, ρ_N, as $R_{SS} = i_{UC}^{-1} \cdot \rho_N^{-1}$.

Table 3.6: Summary statistics for population (N in capita), water use intensity (w_N in $m^3 \cdot yr^{-1} \cdot cap^{-1}$), and climatic water balance (q_{Net} in $m \cdot yr^{-1}$) for the database of cities.

Quantile	N capita	w_N $m^3 \cdot yr^{-1} \cdot cap^{-1}$	q_{Net} $m \cdot yr^{-1}$
μ	2883357	111	0.124
σ	3182849	75	0.752
γ	2	1	1.233
κ	2	1	4.781
Statistic	**N capita**	**w_N $m^3 \cdot yr^{-1} \cdot cap^{-1}$**	**q_{Net} $m \cdot yr^{-1}$**
0%	27285	14	-1.446
25%	646682	57	-0.240
50%	1649294	86	0.089
75%	3763940	148	0.447
100%	14348535	355	3.833

Net climatic water balance ranged from a minimum of -1.4460 $m \cdot yr^{-1}$(Abu Dhabi, United Arab Emirates) to 3.8330 $m \cdot yr^{-1}$(Anchorage, USA). The median value occurred at 0.0890 $m \cdot yr^{-1}$(between St. Petersburg, Russia and Riga, Latvia), with an IQR ranging from -0.2405 $m \cdot yr^{-1}$(between Bangkok, Thailand and Cali, Colombia) to 0.4465 $m \cdot yr^{-1}$(between Santo Domingo, Dominican Republic and Sydney, Australia). These quantile breaks can be seen in the ordered bar chart of q_{Net}, plotted in Figure 3.5.

Ordered Bar Charts

Histograms

A summary of descriptive statistics, including quartiles, are shown in Table 3.6. Histograms of population, water use intensity, and net water balance are shown with and without a $\log_{10}$ transformation in Figures 3.6a—3.6f.

Qq-plots

In comparison with the histogram of per capita water consumption (in Figure 3.6c), the histogram for climatic water balance was more symmetric about the median value (represented by the quantile break line at 50% probability).

As recommended by the qq-plots, each metric was transformed using a base-10 logarithmic transformation (log10) prior to clustering. As expected, this transformation was found to reduce skewness and improved the symmetry of their distribution (Figure 3.6)[18]. The distributions after transformation have been shown in Figure 3.6.

[18] A constant, a, was added to q_{Net} prior to the transform such that $a = 1 - \mathtt{min}(q_{net})$

3.6.2 Bivariate

Scatterplots and Correlation

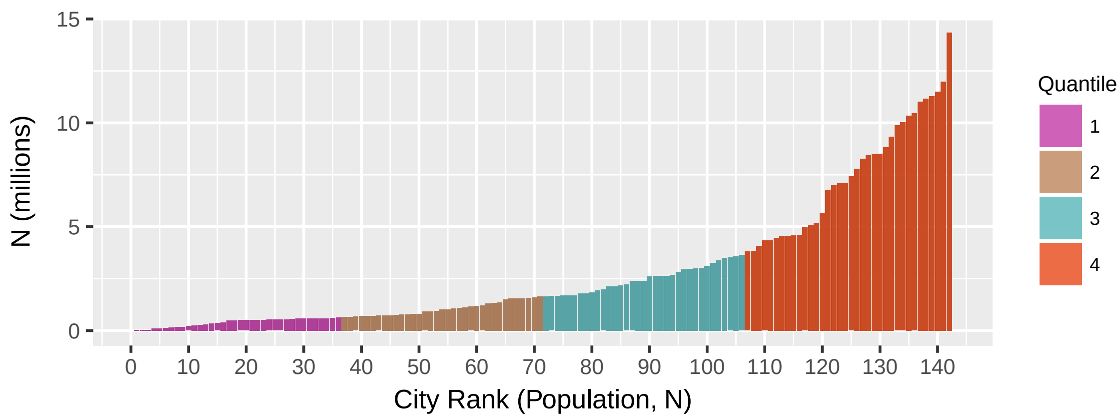
Scatterplots of the bivariate combinations of N, w_N, and q_P—along with quantile breaks—are shown in Figures 3.8a—3.8c on Pages 100–100, respectively.

Together, q_{Net} and w_N provided annual estimates of climatic water supply and per capita water demand. All other things being equal, it might be expected that per capita water consumption would be correlated with climatic water balance based on the assumptions that climatic water balance is a measure for available water resources, and per capita water consumption will be higher if available resources are higher. Therefore, the starting was that a correlation between per capita water consumption and climatic water balance would be observed. However, there was also no apparent correlation between w_N and q_{Net}, as seen in the correlation plot in Figure 3.8d; the correlation coefficient of these two variables was 0.03 (shown in Table 3.8e, which was not significant. Therefore, w_N and q_{Net} were treated as independent variables to cluster the sustainability of urban water patterns.

Since there was no strong correlation between these three variables, it was useful to move beyond quantiles to consider the relationship between case studies and to consider their context relative to all cities. To do this, clustering analysis was performed to provide more intuition into how to group the observations.

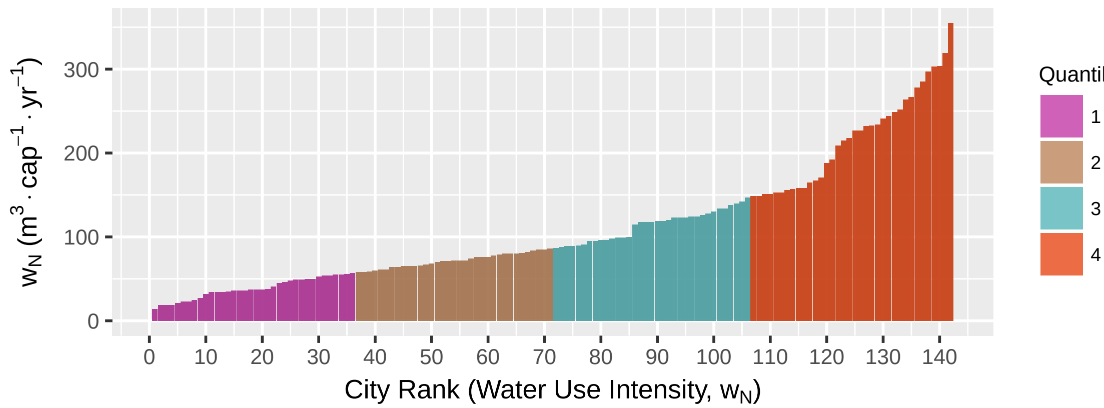
While it might be expected that to find high correlation between the size of the population and water consumption or local water availability (i.e., q_{Net}), the correlation was found to be low. As seen in Table 3.8e, r for $\log_{10}$ (w_N) vs. $\log_{10}$ (q_{Net} +a) was found to be 0.02 with a significance level of p = 0.8373, while the r for $\log_{10}$ (Population) vs. $\log_{10}$ (q_{Net} $+a$) was found to be 0.01 with a p = 0.9261. In other words, the $\log_{10}$ transforms of w_N and q_{Net} were found to be independent, as were the $\log_{10}$ transforms of q_{Net} and population. A small negative correlation of r = -0.12 was found between `log10`(w_N) vs. `log10`(Population), but the significance level was only 0.1684. In other words, these three variables seemed to be independent of each other, justifying an exploratory data mining approach to the data.

Figure 3.3: Ordered bar chart and table of population (N, in 10^6 cap) for UrbMet cities, colored by quantile (see 3.6). To identify the position for a particular city in this plot, first find its rank in the table.



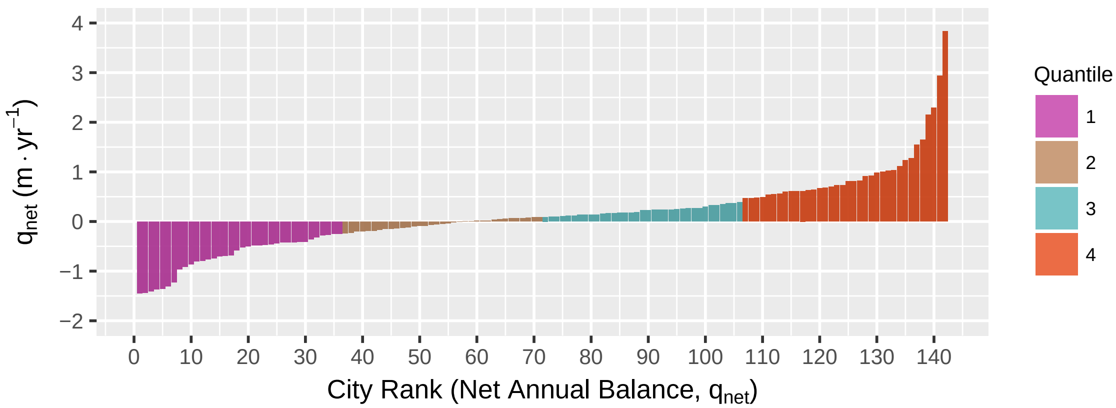
n	City	n	City	n	City	n	City
1	Bandar seri begawan	41	Phnom Penh	81	Bucharest	121	Cairo
2	Kuwait City	42	Cebu	82	Caracas	122	Shenzhen
3	Port-of-Spain	43	Riga	83	Paris	123	Tehran
4	St. John's	44	Jerusalem	84	Tashkent	124	Bogota
5	Abuja	45	Amsterdam	85	Quezon City	125	London
6	Bern	46	Johannesburg	86	Nagoya	126	Kinshasa
7	Brussels	47	Stockholm	87	Brasilia	127	New York, NY
8	Manama	48	Athens	88	Cali	128	Lima
9	Geneva	49	Bishkek	89	Belo Horizonte	129	Tokyo
10	Naihati	50	Ottawa	90	Surabaya	130	Guangzhou
11	Ljubljana	51	Santo Domingo	91	Rome	131	Jakarta
12	Anchorage, AK	52	Detroit, MI	92	Osaka	132	Karachi
13	Victoria, BC	53	Sana'a	93	Addis Ababa	133	Delhi
14	Doha	54	Ulaanbaatar	94	Kiev	134	Seoul
15	Beirut	55	Guatemala City	95	Chicago, IL	135	Dhaka
16	Florianopolis	56	Dakar	96	Nairobi	136	Moscow
17	Panama City	57	Dubai	97	Buenos Aires	137	Sao Paulo
18	Dublin	58	Tbilisi	98	Casablanca	138	Istanbul
19	Lisbon	59	Sofia	99	Ho Chi Minh City	139	Mexico City
20	Copenhagen	60	Prague	100	Madrid	140	Beijing
21	San Salvador	61	Amman	101	Montreal	141	Mumbai
22	Asuncion	62	Milan	102	Berlin	142	Shanghai
23	Hanover	63	Montevideo	103	Cape Town		
24	Abu Dhabi	64	Dar es Salaam	104	Ankara		
25	Sarajevo	65	Tripoli	105	Yokohama		
26	Islamabad	66	Kuala Lumpur	106	Hyderabad		
27	Vilnius	67	Phoenix, AZ	107	Melbourne		
28	Helsinki	68	Quito	108	Los Angeles, CA		
29	Kingston	69	Manila	109	Riyadh		
30	Vladivostok	70	Barcelona	110	Sydney		
31	Portland, OR	71	Guadalajara	111	Chennai		
32	Denver, CO	72	Damascus	112	Yangon		
33	Seattle, WA	73	Accra	113	St. Petersburg		
34	Boston, MA	74	Vienna	114	Kolkata		
35	Colombo	75	Budapest	115	Singapore		
36	Rabat	76	Warsaw	116	Toronto		
37	Chisinau	77	Hamburg	117	Santiago		
38	Durban	78	Curitiba	118	Bangalore		
39	Kathmandu	79	Minsk	119	Lagos		
40	Tunis	80	Vancouver, BC	120	Bangkok		

Figure 3.4: Ordered bar chart and table of water use intensity (w_N, in $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$) for UrbMet cities, colored by quantile. To identify the position for a particular city in this plot, first find its rank in the table.



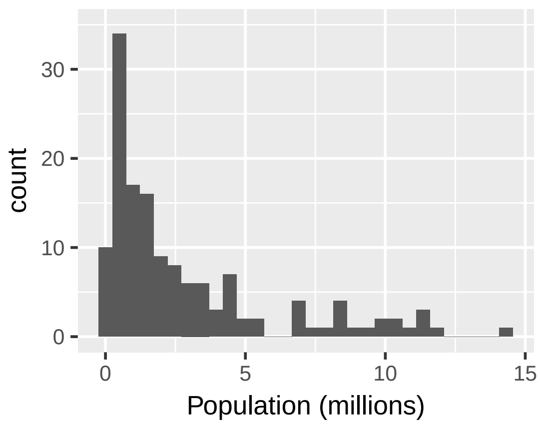
n	City	n	City	n	City	n	City
1	Yangon	41	Brussels	81	London	121	Manama
2	Accra	42	Vilnius	82	Paris	122	Phoenix, AZ
3	Addis Ababa	43	Florianopolis	83	Tehran	123	Abu Dhabi
4	Sana'a	44	Copenhagen	84	Singapore	124	St. John's
5	Colombo	45	Brasilia	85	Riga	125	Dubai
6	Dakar	46	Quito	86	Boston, MA	126	Port-of-Spain
7	Dar es Salaam	47	Cebu	87	Minsk	127	Denver, CO
8	Kinshasa	48	Nairobi	88	Detroit, MI	128	Vladivostok
9	Naihati	49	Karachi	89	Seattle, WA	129	Victoria, BC
10	Tunis	50	Phnom Penh	90	Sofia	130	Nagoya
11	Dhaka	51	Warsaw	91	Bandar seri begawan	131	Osaka
12	Bangalore	52	Bangkok	92	Rome	132	Yokohama
13	Kolkata	53	Caracas	93	Melbourne	133	Ottawa
14	Chennai	54	Guadalajara	94	Manila	134	Tokyo
15	Delhi	55	Cape Town	95	Lisbon	135	Vancouver, BC
16	Guatemala City	56	Barcelona	96	Sydney	136	Montreal
17	Mumbai	57	Santo Domingo	97	Chicago, IL	137	Toronto
18	Bogota	58	Helsinki	98	Amsterdam	138	Dublin
19	Hyderabad	59	Chisinau	99	Jerusalem	139	Stockholm
20	Abuja	60	Sarajevo	100	Bern	140	Tbilisi
21	Tripoli	61	Sao Paulo	101	Beirut	141	Kathmandu
22	Surabaya	62	Vienna	102	Doha	142	Cairo
23	Jakarta	63	Santiago	103	Kuala Lumpur		
24	Hanover	64	Bucharest	104	Buenos Aires		
25	Montevideo	65	Madrid	105	Panama City		
26	Lagos	66	Kingston	106	Anchorage, AK		
27	Ho Chi Minh City	67	Mexico City	107	Geneva		
28	Rabat	68	Ulaanbaatar	108	New York, NY		
29	Durban	69	Istanbul	109	Shenzhen		
30	Quezon City	70	Seoul	110	Portland, OR		
31	Curitiba	71	Prague	111	Guangzhou		
32	Lima	72	Ljubljana	112	Kiev		
33	Belo Horizonte	73	Ankara	113	Beijing		
34	Islamabad	74	Riyadh	114	Los Angeles, CA		
35	Casablanca	75	Johannesburg	115	Shanghai		
36	Hamburg	76	Asuncion	116	Cali		
37	San Salvador	77	Budapest	117	Tashkent		
38	Athens	78	Amman	118	Milan		
39	Damascus	79	Kuwait City	119	St. Petersburg		
40	Berlin	80	Bishkek	120	Moscow		

Figure 3.5: Ordered bar chart and table of net annual balance (q_{Net} in $m \cdot yr^{-1}$) and table for UrbMet cities, colored by quantile (see Table 3.6). To identify the position for a particular city in this plot, first find its rank in the table.

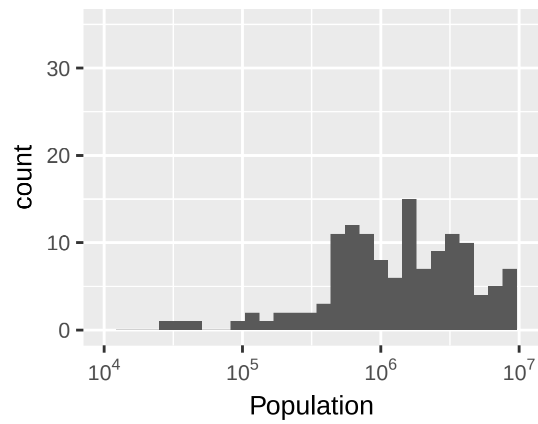


n	City	*n*	City	*n*	City	*n*	City
1	Abu Dhabi	41	Ulaanbaatar	81	Helsinki	121	Brasilia
2	Doha	42	Lisbon	82	Vladivostok	122	Dhaka
3	Manama	43	Bucharest	83	Detroit, MI	123	Curitiba
4	Riyadh	44	Bishkek	84	Toronto	124	Yangon
5	Dubai	45	Budapest	85	San Salvador	125	Shenzhen
6	Kuwait City	46	Chisinau	86	Amsterdam	126	Cebu
7	Karachi	47	Surabaya	87	Buenos Aires	127	Vancouver, BC
8	Phoenix, AZ	48	Nairobi	88	Barcelona	128	Panama City
9	Cairo	49	Warsaw	89	Quezon City	129	Bern
10	Tehran	50	Tashkent	90	Kathmandu	130	Portland, OR
11	Hyderabad	51	Berlin	91	Dublin	131	St. John’s
12	Dakar	52	Johannesburg	92	Port-of-Spain	132	Ljubljana
13	Tripoli	53	Guadalajara	93	Chicago, IL	133	Singapore
14	Delhi	54	Vienna	94	Shanghai	134	Nagoya
15	Amman	55	Istanbul	95	Kinshasa	135	Mumbai
16	Damascus	56	Kiev	96	Sofia	136	Tokyo
17	Lima	57	Milan	97	Victoria, BC	137	Colombo
18	Los Angeles, CA	58	Prague	98	Montevideo	138	Guatemala City
19	Chennai	59	Kingston	99	Rome	139	Bandar seri begawan
20	Caracas	60	Paris	100	Ottawa	140	Bogota
21	Jerusalem	61	Stockholm	101	Sarajevo	141	Sao Paulo
22	Dar es Salaam	62	Santiago	102	New York, NY	142	Anchorage, AK
23	Tunis	63	Kolkata	103	Manila		
24	Casablanca	64	Hanover	104	Asuncion		
25	Cape Town	65	Naihati	105	Montreal		
26	Rabat	66	Copenhagen	106	Santo Domingo		
27	Accra	67	London	107	Sydney		
28	Phnom Penh	68	Lagos	108	Seattle, WA		
29	Bangalore	69	Hamburg	109	Mexico City		
30	Islamabad	70	Minsk	110	Kuala Lumpur		
31	Tbilisi	71	St. Petersburg	111	Boston, MA		
32	Abuja	72	Riga	112	Beirut		
33	Ankara	73	Durban	113	Seoul		
34	Athens	74	Vilnius	114	Quito		
35	Madrid	75	Melbourne	115	Osaka		
36	Bangkok	76	Moscow	116	Guangzhou		
37	Cali	77	Ho Chi Minh City	117	Geneva		
38	Denver, CO	78	Addis Ababa	118	Florianopolis		
39	Sana’a	79	Brussels	119	Belo Horizonte		
40	Beijing	80	Jakarta	120	Yokohama		

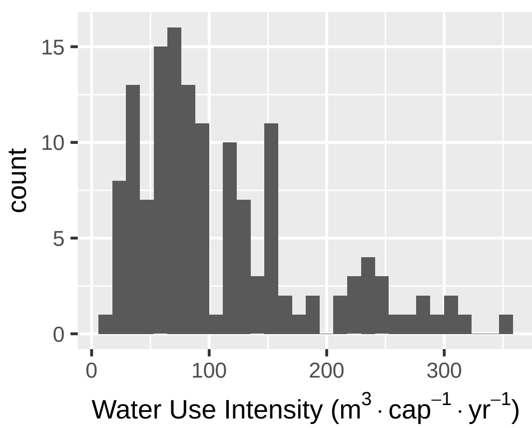
Figure 3.6: Histograms of population (N), water use intensity (w_N, in $m^3 \cdot yr^{-1} \cdot cap^{-1}$), and climatic water balance (q_{Net}, in $m \cdot yr^{-1}$), for cities in the UrbMet database, before and after a `log-10` transformation (left and right, respectively). The `log-10` transformation was found to improve symmetry of the data distributions.



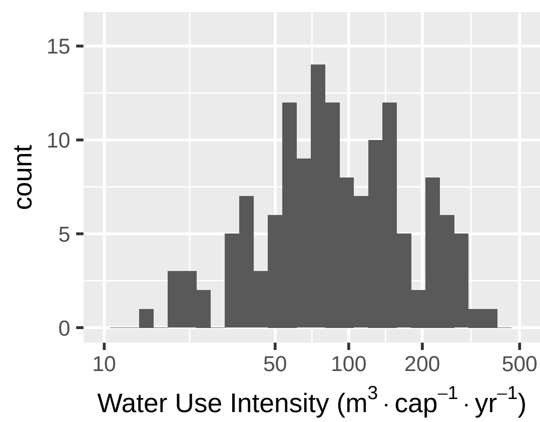
(a) N (without log10 transformation).



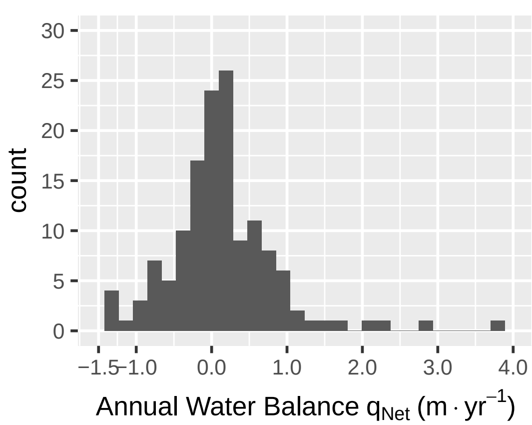
(b) N (transformed by log10).



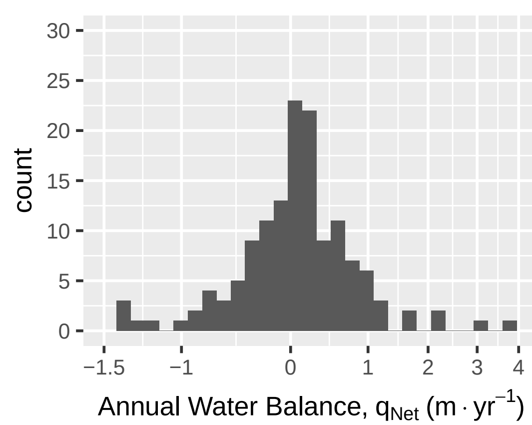
(c) w_N (without log10 transformation).



(d) w_N (transformed by log10).

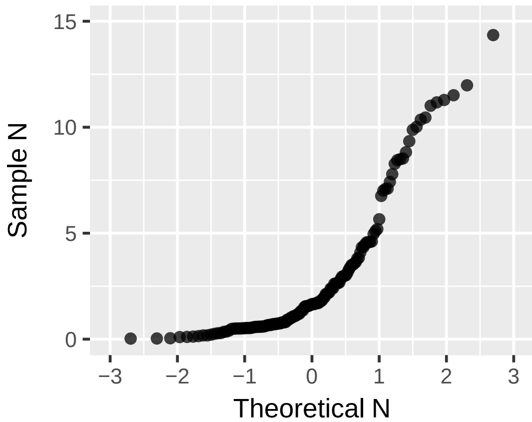


(e) q_{net} (without log10 transformation).

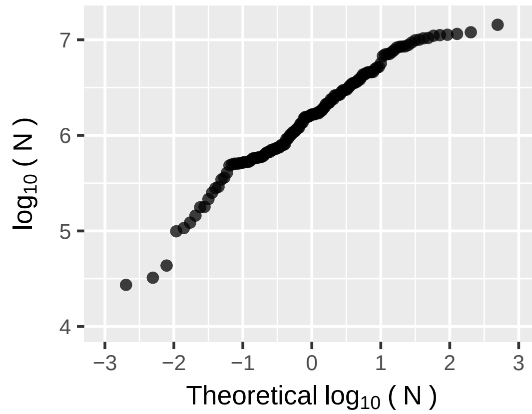


(f) q_{net} (transformed by log10).

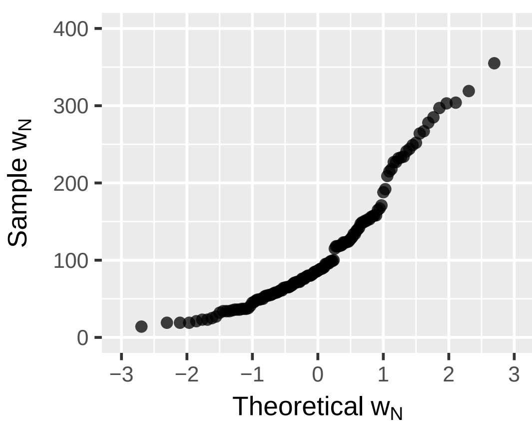
Figure 3.7: Quantile-quantile plots of population (N), water use intensity (w_N, in $m^3 \cdot yr^{-1} \cdot cap^{-1}$), and climatic water balance (q_{Net}, in $m \cdot yr^{-1}$) for UrbMet cities, before and after application of a log-10 transformation (left and right, respectively). The data for all three metrics — N, w_N, and q_{Net}— was more closely approximated by the normal distribution after application of the `log-10` transformation as indicated by the reduced curvature in the qq-plot.



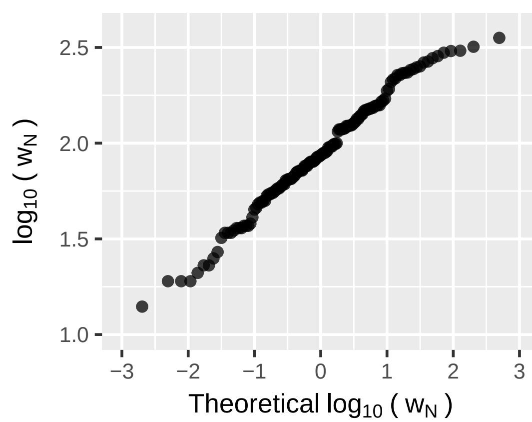
(a) N



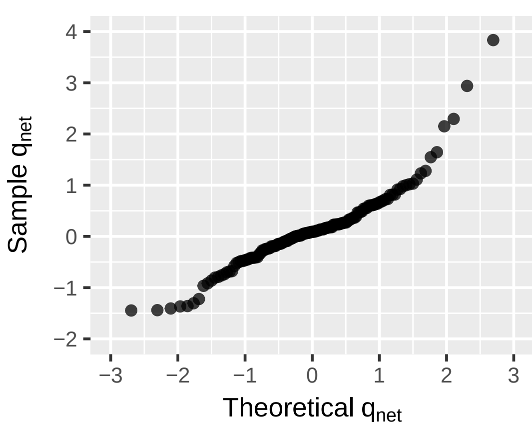
(b) $log_{10}(N)$



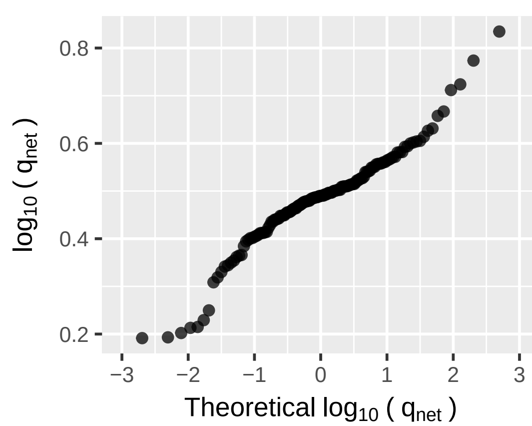
(c) w_N



(d) $log_{10}(w_N)$



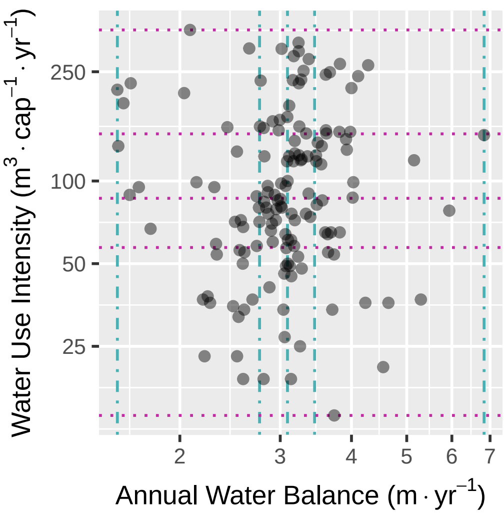
(e) q_{Net}



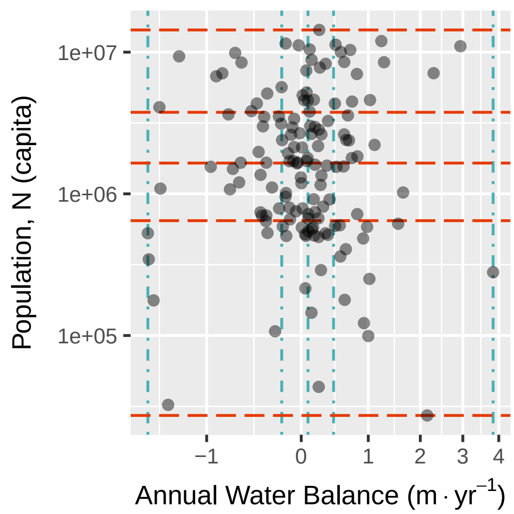
(f) $log_{10}(q_{net})$

Figure 3.8: Correlation plot and scatterplots of bivariate combinations of population (N, in capita), water use intensity (w_N, in $m^3 \cdot yr^{-1} \cdot cap^{-1}$), and net annual balance (q_{Net}, in $m \cdot yr^{-1}$) for cities in the Urban Metabolism database. The correlation plot is shown in Figure 3.8d at the bottom right. Scatterplots are shown with quantiles plotted as dotted lines.

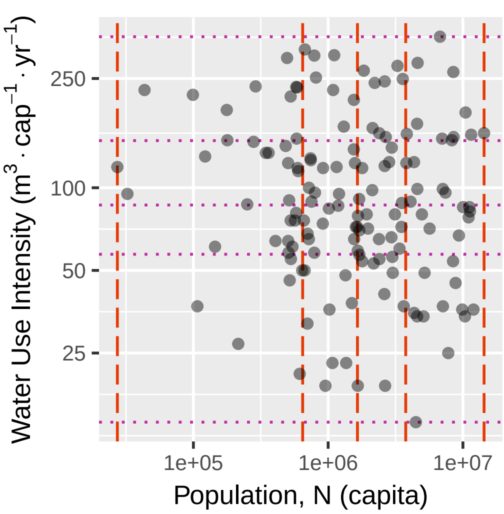
(a) Scatterplot of w_N versus q_{Net}.



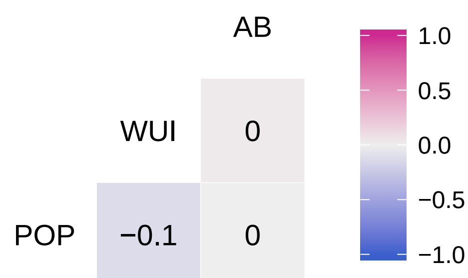
(b)



(c) Scatterplot of w_N versus N.



(d) Correlation plot.



(e) Correlation table.

Metric	$\log_{10}(w_N)$	$\log_{10}(q_{Net})$	$\log_{10}(N)$
$\log_{10}(w_N)$	1.00	0.02, p = 0.84	-0.12, p = 0.17
$\log_{10}(q_{Net})$		1.00	0.01, p = 0.93
$\log_{10}(N)$			1.00

3.7 Univariate and Bivariate Clustering Results

The results for univariate clustering are presented in Appendix B.1.4 and the results for bivariate clustering are presented in Appendix B.1.4.

The univariate and bivariate clustering results are shown with the number of clusters chosen for presentation and interpretability. The univariate clustering results are shown with cutting to dendrogram to $k = 4$ clusters so that they could be compared with statistical quantiles from Section 3.6.1. The bivariate clustering results have been shown with cutting the dendrogram to $k = 6$ clusters so that they could be visually compared with the trivariate clustering results (which are described in Section 3.8).

3.7.1 Univariate Clustering Results

The results of the univariate clustering of population, water use intensity (w_N), and annual balance are shown in Figures B.10–B.12. Each of the Figures B.10–B.12, has three subfigures:

- an ordered bar chart of the variable used in clustering, colored by quantile (top),
- an ordered bar chart of the variable used in clustering, colored by univariate clustering results with $n = 4$ clusters (middle), and
- a scatterplot of the two variables not used in clustering, colored by the univariate clustering results (bottom).

There were several takeaways from these results of univariate clustering.

For each of the univariate clusterings, a visual comparison of the ordered bar chart colored by quantiles (top) with that colored by univariate clustering (bottom) highlighted that the groups identified by clustering algorithms differ significantly from those delineated by quantiles. Recall that the results of statistical clustering identify groups of cities that are the most similar, quantitatively speaking, *when compared to all other cities*[19]. In contrast, statistical quantiles only delineate the large-n sample of cities *into groups of equal size*.

These results of univariate clustering suggest that the results of descriptive statistics may not provide meaningful groupings of cities and thresholds for comparison, *even on a univariate basis*.

This conclusion was also supported by the third subfigure in the univariate clustering results. As Figures B.10b–B.12b illustrate, the univariate clustering results led (unsurprisingly) to clear distinctions when viewed in the same univariate space as the variable on which clustering occurred. In contrast, the univariate clustering did not lead to meaningful distinctions when viewed on the two-dimensional space formed by the variables *not* used in clustering, as seen in Figures B.10c–B.12c.

That univariate clustering did not produce groups that were meaningfully distinct across all three variables was intuitive when considered in light of the correlation coefficients from Section 3.6.2, which found very low correlation across all three bivariate combinations.

[19] As measured by a distance matrix and as determined by a linkage criterion as discussed previously.

The results of univariate clustering were interesting to a certain degree, in that they provided visual intuition into the hierarchical clustering, especially for those less familiar with the methods (but were also able to provide context for the bivariate and trivariate clustering results, regardless of experience).

However, the conclusion was that the univariate results did not sufficiently characterize the full spectrum of urban water profiles.

3.7.2 Bivariate Clustering Results

Hierarchical clustering was performed on the three bivariate combinations of:

1. water use intensity and climatic water balance (w_N versus q_{Net}),
2. population and climatic water balance (N versus q_{Net}), and
3. water use intensity and population (w_N versus q_{Net})

A matrix of bivariate results for each of the combinations above are shown in Figures B.13–B.15. These plot matrices were produced using the `ggpairs` function from the `GGally` package in `R`. These matrices show the $n = 6$ clusters across all three variables (N, w_N, and q_{Net}), colored by bivariate clustering, as:

- histograms of the clusters on the diagonal,
- scatterplots (below the diagonal), and
- correlation coefficients (above the diagonal).

As with the results of univariate clustering presented in Section 3.7.1, there were several general observations from the results of bivariate clustering. This led to distinctions that were not necessarily the most intuitive in the bivariate spaces formed by the two variables used in clustering[20] can be seen in Figure B.13 (bottom row, middle) for w_N versus q_{Net}; in Figure B.14 (bottom row, left) for N versus q_{Net}; and in Figure B.15 (middle row, left) for w_N versus N. The corresponding correlations can be seen, respectively, in Figure B.13 (top row, middle) for w_N versus q_{Net}; in Figure B.14 (top row, right) for N versus q_{Net}; and in Figure B.15 (middle row, right) for w_N versus N. The distinctions between clusters in these bivariate spaces were not necessarily those that would have been drawn by hand. However, it was interesting to see the groups that were identified statistically through the application of clustering algorithms. The results illustrated that hierarchical clustering identifies groups that were not necessarily normal in distribution.

[20]E.g., the bivariate clustering of w_N and q_{Net} viewed in the scatterplot of w_N versus q_{Net} (see Figure B.13, bottom row, middle plot).

The results of bivariate clustering appeared more distinct in the two-dimensional spaces formed by the variables used in clustering than in the two dimensional space formed by the variable *not* used in clustering, and a variable used in clustering (as seen in the other two scatterplots in the plot matrices of bivariate clustering results[21]).

It was interesting to see the bivariate results across all three scatterplots of bivariate combinations. However, all three metrics—population, water use intensity, and water balance—were all chosen to represent important urban attributes that affect water supply and demand. Therefore, as with the univariate clustering results, while the bivariate clustering results were interesting to a certain degree, the conclusion was that the bivariate results were insufficient as a characterization of urban water profiles.

3.7.3 Conclusions

The results of univariate and bivariate clustering provided intuition into hierarchical clustering. Hierarchical clustering highlighted delineations between groups in the univariate or bivariate space formed by the variable(s) used in clustering. These results were interesting, as they differed from statistical quantiles and from visual intuition. At the same time, these clusters were supportive by rigorous quantitative statistical methods.

However, the results of the univariate and bivariate clustering analysis did not identify groups that were distinct across all three metrics. Therefore, it was concluded that univariate and bivariate analyses are insufficient to the task of meaningfully characterize profiles of urban water supply and demand around the world. These results suggest that the typical univariate or bivariate comparison of urban water systems are unlikely to produce meaningful insight for policy or decision-making.

The next step in the large-*n* comparative analysis was assessing whether urban water profiles could be meaningfully characterized through statistical clustering across all three variables.

3.8 Trivariate Clustering Results

3.8.1 Main Results

Dendrogram

The results of the hierarchical clustering of the trivariate combination of population (N, in capita), water use intensity (w_N, in $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$), and net annual water balance (q_{Net}, in $\mathrm{m} \cdot \mathrm{yr}^{-1}$) are shown as a dendrogram in Figure 3.9 on Page 105. Cities in the dendrogram are colored by cluster membership after cutting the dendrogram to a number of clusters, $k = 6$. I also explored the results of cutting the dendrogram to a number of clusters ranging from $k = 2$ to $k = 8$, and

[21] E.g., the bivariate clustering of w_N and q_{Net} viewed in the scatterplot of w_N versus q_{Net} (see Figure B.13, bottom row, middle plot) in contrast to those bivariate results viewed in the scatterplots of N versus q_{Net} (bottom row, left) or w_N versus N (middle row, left).

decided to cut the dendrogram to $k = 6$ clusters[22]. Cutting the dendrogram to $k < 62$ clusters resulted in some groupings that were too broad to be useful; cutting the dendrogram to $k > 6$ clusters resulted in groups that were too small to be useful (multiple clusters with a number of members, $n < 10$). Cutting the trivariate dendrogram to $k = 6$ clusters resulted in "Goldilocks" groups that were neither too larger nor too small—four groups with greater than 20 member cities (Types 1–4), one group with 7 members (Type 5), and one group with 15 members (Type 6).

I performed the remaining analysis on the clusters resulting from cutting the dendrogram to six clusters, which I recast as a typology. Since there was a choice in how the dendrogram is cut, it would be appropriate to view the combined steps of clustering and cutting the dendrogram as "creating" this typology.

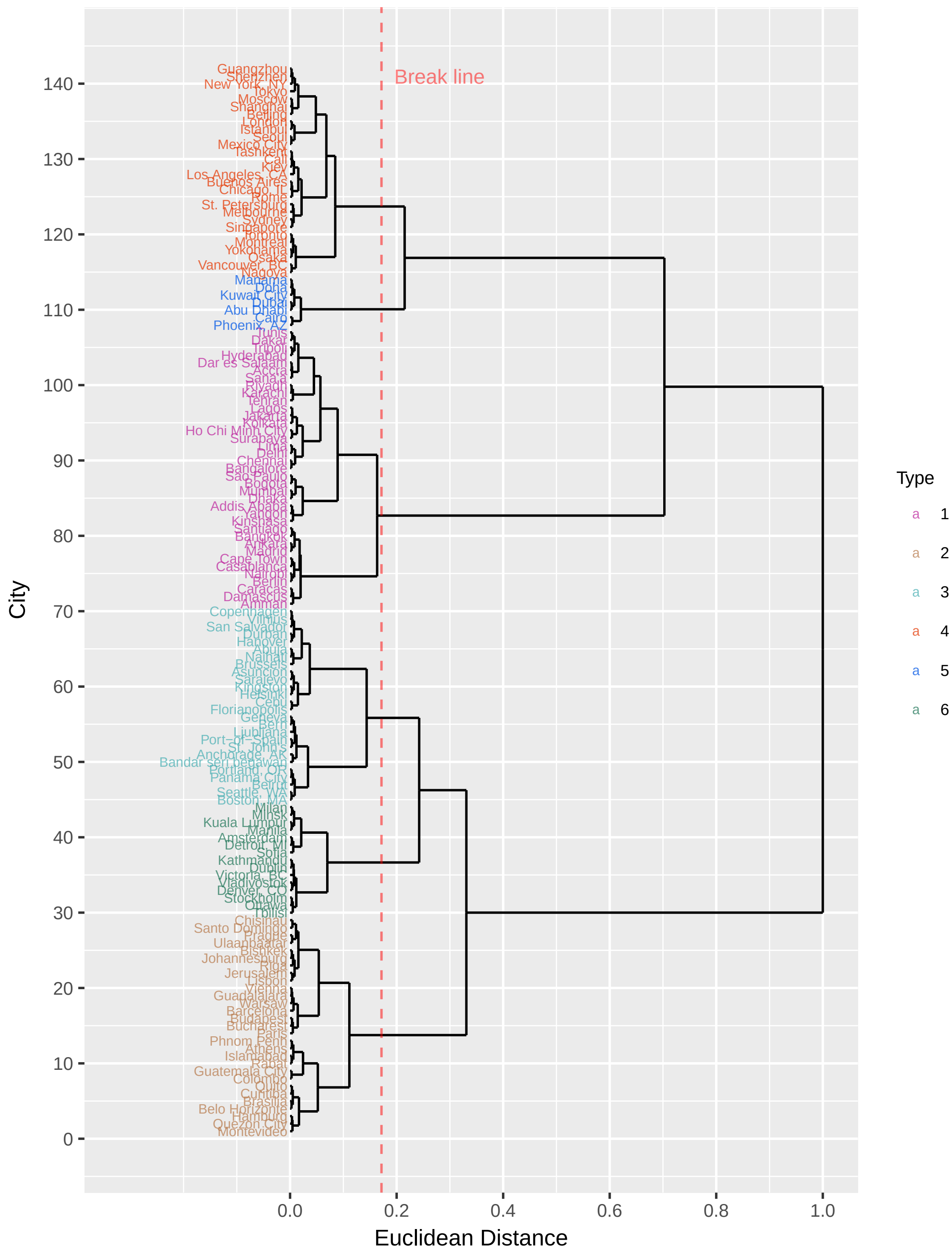
Given the large number of cities in the analysis, it was difficult to visualize all of the cities on a dendrogram on a single page in a legible way. Instead, close-ups of the dendrogram for each cluster type are provided in Figures 3.12a–3.17a, which are presented in Section 3.8.2 (starting on Page 115).

While not all of the UrbMet cities on the dendrogram are legible, Figure 3.9 still provides a global view of the trivariate results, as well as relative distances between cities and clusters at each level of the clustering. However, an important point to keep in mind when viewing a dendrogram is that each clade (i.e., branch) is free to rotate about its node, such that the dendrogram could be imagined as a kind of mobile (sculpture). As an example, consider Types 4 (in orange, at the top of Figure 3.9) and Type 2 (at the bottom). At the uppermost level of aggregation, each of the two "superclusters"—Types 1, 4, and 5 comprising the first and Types 2, 3, and 6 comprising the second—are free rotate (in the z-dimension, i.e., out of the page) about the node at the normalized Euclidean distance $d = 1.0$. When these two superclusters rotate, Type 4 can become closer to Type 2.

Therefore, while the dendrogram shown in Figure 3.9 helps to highlight connections, intuition into the relative similarities and differences between types of the typology is benefited by the t-SNE results, which are presented in Figure 3.11 on Page 113.

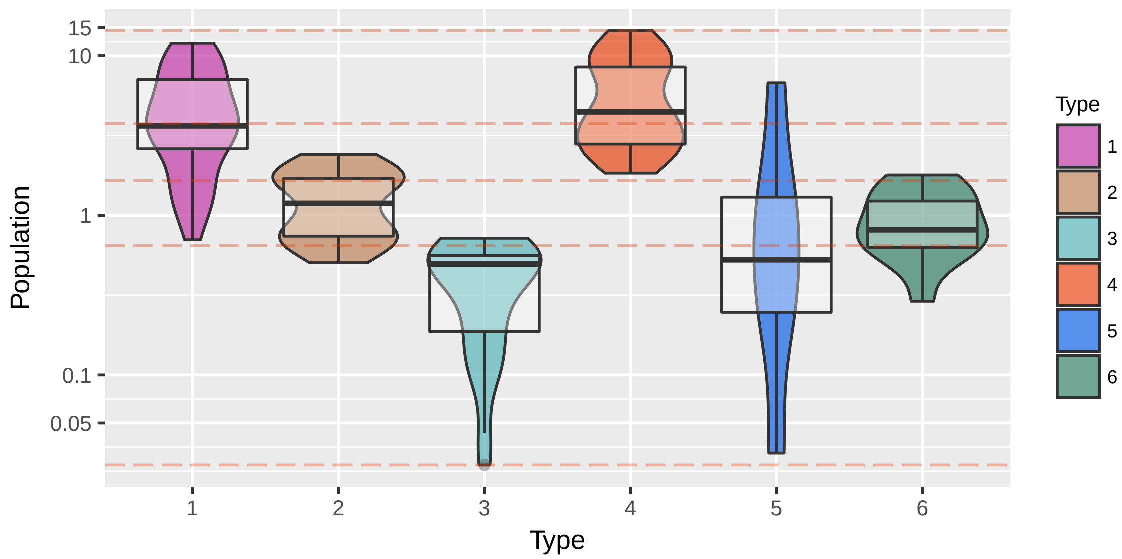
[22]The results are shown only for cutting the dendrogram to $k = 6$ clusters.

Figure 3.9: The results of hierarchical clustering of the UrbMet cities shown as a dendrogram, with cities colored by cluster type (cutting the dendrogram to $k = 6$ clusters). While the city labels in this figure are not clear, this figure provides a global view of the trivariate results, as well as relative distances between cities and clusters at each level of the clustering. Individual city labels are more legible in the close-ups of the dendrogram for each type, which are provided in Figures 3.12a–3.17a in Section 3.8.2 (starting on Page 115). This global view also shows, as a vertical dotted line, the "Break Line"—i.e., the quantitative threshold for cutting the dendrogram to $k = 6$ clusters.



Violin- and box-plots

(a) Violin- and box-plot of population N, by type.

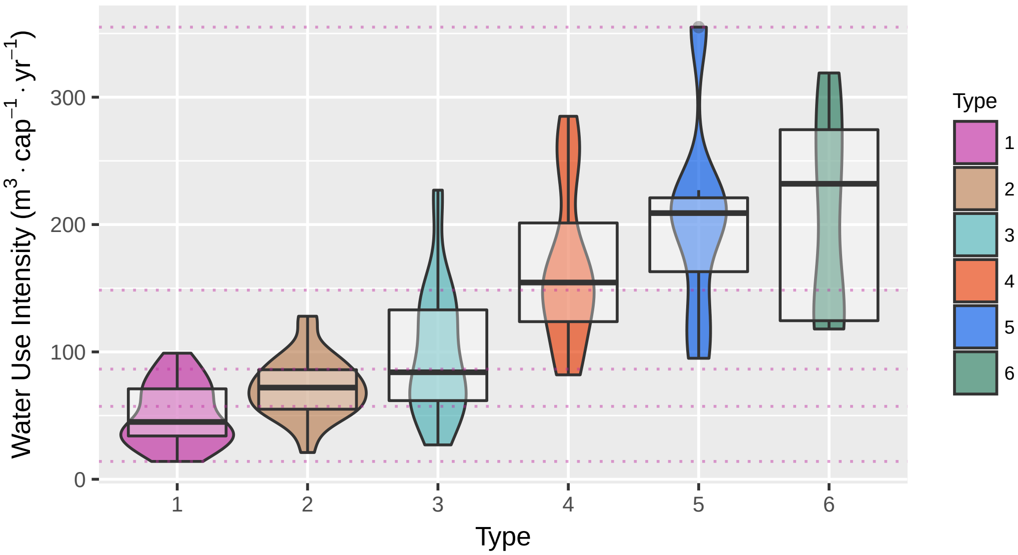


(b) Summary statistics for population, by type. Quantiles, μ_N, and σ_N are in units of 10^6 cap.

Statistic	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
μ	4.69	1.29	0.39	5.84	1.50	0.96
σ	3.14	0.60	0.21	3.63	2.38	0.44
γ	0.74	0.32	-0.32	0.70	1.46	0.41
κ	-0.62	-1.32	-1.44	-0.89	0.45	-1.14
Quantile	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**
0%	0.70	0.50	0.03	1.84	0.03	0.29
25%	2.61	0.74	0.19	2.80	0.26	0.63
50%	3.64	1.19	0.49	4.45	0.53	0.81
75%	7.09	1.70	0.56	8.50	1.32	1.23
100%	11.98	2.40	0.72	14.35	6.76	1.79

Relative univariate differences between types of the typology can be seen in violin-plots and boxplots of N, w_N, and q_{Net} in Figures 3.9a–3.9d. A table of summary statistics for each type is also provided with each violin-/box-plot for reference. These plots help highlight thresholds between different types and provide information about the symmetry and spread of the underlying distributions beyond summary statistics. The shape of each of the "violins" conveys information about the shape of the underlying distribution. Overlaying the violin for each type is a box-and-whisker plot (box-plot). The main box covers the inter-quartile range (IQR), i.e., the observations between the 25th and 75th percentiles, with the median value plotted as a horizontal bar. The "whiskers" of each box-plot extend to values that fall within ± 1.5 IQR of the lower and upper quantiles.

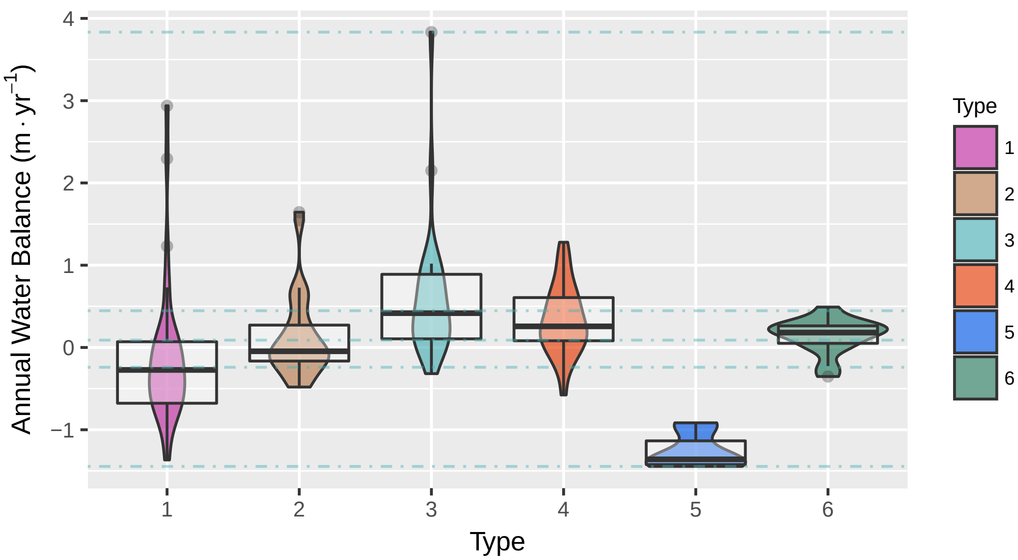
(c) Violin- and box-plot of water use intensity (w_N), by type.



Statistic	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
μ	50.38	72.52	99.73	166.61	203.86	205.53
σ	23.96	23.64	51.81	62.21	82.06	78.77
γ	0.37	0.32	0.82	0.55	0.47	0.12
κ	-1.08	0.04	-0.05	-0.96	-0.84	-1.77
Quantile	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**
0%	14.00	21.00	27.00	82.00	95.00	118.00
25%	34.00	55.00	61.75	123.75	163.00	124.50
50%	45.00	72.00	84.00	154.50	209.00	232.00
75%	71.00	86.00	133.00	201.25	221.00	274.50
100%	99.00	128.00	227.00	285.00	355.00	319.00

Population, N I plotted the violin-/box-plot of population (N, in capita) on a $\log_{10}$ scale, as seen in Figure 3.9a on Page 106. The quantiles for population, as calculated over all of the UrbMet cities, are plotted as horizontal, orange dashed lines. Figure 3.9a highlights that the IQRs for Types 1 and 4 fall above the 50th percentile (as calculated over all of the UrbMet cities); Types 1 and 4 could, therefore, be considered to have "large" populations. The 25th percentile of Type 1 occurs at 2.61 10^6 cap, while that for Type 4 occurs at 2.80 10^6 cap; in other words, cities in Types 1 and 4 tend to be larger than those values. Types 2 and 6 have "medium-size" populations, with IQRs that fall, approximately, between the 25th and 75th percentiles: the IQR of Type 2 is $[0.74, 1.70]$ 10^6 cap, while that for Type 6 is $[0.63, 1.23]$ 10^6 cap. In contrast to Types 1, 2, 4, and 6, Type 3 cities are "small", with a population $0.72 \geq N \geq 0.03$ 10^6 cap—the range of population for Type 3 cities lies almost entirely below the 25th percentile. Type 5 is the only type whose population range spans all four quantiles, with $6.75 \geq N \geq 0.03$ 10^6 cap. Figure 3.9a also shows that the population thresholds between types tend to match the global quantiles, with three exceptions: the IQR for Type 5 spans the first and second quantiles, while the IQRs for Types 1 and 4 lie between the third and fourth quantiles.

(d) Violin- and box-plot of net annual balance q_{Net}, by type.



(e) Summary statistics for net annual balance (q_{Net}), by type.

Statistic	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
μ	-0.13	0.13	0.61	0.34	-1.26	0.14
σ	0.85	0.52	0.83	0.43	0.23	0.22
γ	1.86	1.43	2.40	0.27	0.65	-0.78
κ	4.01	1.62	6.49	-0.39	-1.63	-0.05
Quantile	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**
0%	-1.37	-0.48	-0.32	-0.58	-1.45	-0.35
25%	-0.68	-0.17	0.11	0.08	-1.42	0.05
50%	-0.27	-0.05	0.42	0.26	-1.36	0.18
75%	0.07	0.27	0.89	0.61	-1.14	0.26
100%	2.94	1.64	3.83	1.28	-0.92	0.49

Water use intensity, w_N The distributions of water use intensity (w_N, in $m^3 \cdot yr^{-1} \cdot cap^{-1}$) for the six types are shown in Figure 3.9c on Page 107. The quantiles for water use intensity, as calculated over all of the UrbMet cities, are plotted as magenta dotted lines. Figure 3.9c shows that the IQRs of the six types tend to overlap, in contrast to the IQRs of population. While Type 1 cities have large populations they have "low" water use intensity, with an IQR of $[34, 71]$ $m^3 \cdot yr^{-1} \cdot cap^{-1}$; the lower end of the IQR for Type 5 is lower than 36.5 $m^3 \cdot yr^{-1} \cdot cap^{-1}$, i.e., the lower limit on per capita water access as recommended by the WHO. Type 2 cities have "low" to "medium" water use intensity, with an IQR of $[55, 86]$ $m^3 \cdot yr^{-1} \cdot cap^{-1}$. This is followed by Type 3 cities, which have "medium-low" to "medium-high" levels of water use intensity, with an IQR of $[62, 133]$ $m^3 \cdot yr^{-1} \cdot cap^{-1}$. Types 4–6 all have "medium-high" to "very high" levels of water use intensity—cities in Types 4–6 are nearly all in the upper 50th-percentile of water use intensity when compared with all of the UrbMet cities, and the median values for these three types all fall above the 75th-percentile. The IQR for water use intensity for Type 4 ranges over $[124, 201]$ $m^3 \cdot yr^{-1} \cdot cap^{-1}$, with a median value of 155 $m^3 \cdot yr^{-1} \cdot cap^{-1}$—over four times the

WHO minimum. The IQR for Type 5 cities lies entirely in the upper quantile (when compared to w_N for all of the UrbMet cities), and the median value—at 221 $m^3 \cdot yr^{-1} \cdot cap^{-1}$—is the second highest of all of the types; however, there are only 7 members in Type 5. Type 6 cities have an IQR of $[125, 275]$ $m^3 \cdot yr^{-1} \cdot cap^{-1}$ and, at 232 $m^3 \cdot yr^{-1} \cdot cap^{-1}$, have the highest median value of any of the types. Generally, the IQRs of water use intensity for the types do not align with the global quantiles, in contrast to those for population. Exceptions are Type 2 cities—with an IQR for w_N that approximately aligns with the second quantile—and Type 5, whose IQR lies above the 75th percentile.

Climatic water balance, q_{Net} The violin-/box-plot of net annual climatic water balance (q_{Net}, in $m \cdot yr^{-1}$) for each type is shown in Figure Figure 3.9d on Page 108. The quantiles for climatic water balance, as calculated over all of the UrbMet cities, are plotted as light blue dotted-dashed lines. As was the case for water use intensity, the IQRs for the six types tended to span multiple quantiles, although some alignment did occur. The lower value of the IQRs for Types 3, 4, and 6 were approximately aligned with the 50th percentile, which occurred at 0.089 $m \cdot yr^{-1}$; the IQRs for these types were $[0.105, 3.833]$, $[0.082, 0.607]$, and $[0.051, 0.263]$ (in $m \cdot yr^{-1}$), respectively. Since the IQRs for Types 3, 4, and 6 were greater than zero, the climatic water balance for these types can be considered "medium-high" to "high". Type 3 cities were the "wettest" cities, with almost half of Type 3 cities above the 75th percentile (which occurred at 0.447 $m \cdot yr^{-1}$). Type 4 cities (which also had some of the highest populations), also had "high" values for climatic water balance, with a median value of 0.26 $m \cdot yr^{-1}$. Type 2 cities had "medium-low"–"medium" levels of water balance, with an IQR ranging over $[-0.17, 0.275]$ $m \cdot yr^{-1}$and a median value centered close to zero, at -0.05 $m \cdot yr^{-1}$. Type 1 cities, which (along with Type 4) had some of the largest populations of the UrbMet cities, also had some of the lowest water balances, with an IQR generally less than zero and ranging over $[-0.68, 0.07]$ $m \cdot yr^{-1}$. The IQR (and range) for net climatic balance for Type 5 lies entirely below the 25th percentile, and its IQR does not overlap with the IQR of any other type. The IQR of q_{Net} for Type 5 is $[-1.45, -0.92]$—in other words, Type 5 cities are located in regions with a large climatic water deficit. Due to this climatic shortfall, it is not possible to supply Type 5 cities from local surface water sources. An example is Cairo, which has $q_{Net} = 0.916$; Cairo's primary water source is the Nile River, which has a drainage basin that spans eleven countries in Africa and occupies over $3 \cdot 10^6 km^2$ [208, 175].

The information gathered from inspection of the violin-/box-plots of N, w_N, and q_{Net} are summarized in Table 3.7 and inform the overview of types provided in Section 3.8.2 on Page 115.

Additional violin-plots and boxplots of the composite indicators described in Section 3.5.4 and 3.5.5 are provided in Section 3.8.4. These indicators are: the Water Use and Climate Index (WUCI or i_{UC}, in $m^2 \cdot cap^{-1}$), shown in Figure 3.17c on Page 124; the Total Water Use and Climate index (total WUCI or I_{UC}, in km^2), shown in Figure 3.17e on Page 125; and the Potential Self-Sufficiency Ratio, R_{SS} (unitless), shown in Figure 3.17g. Tables of summary statistics are also provided with each violin-/box-plot.

World Map of Typology

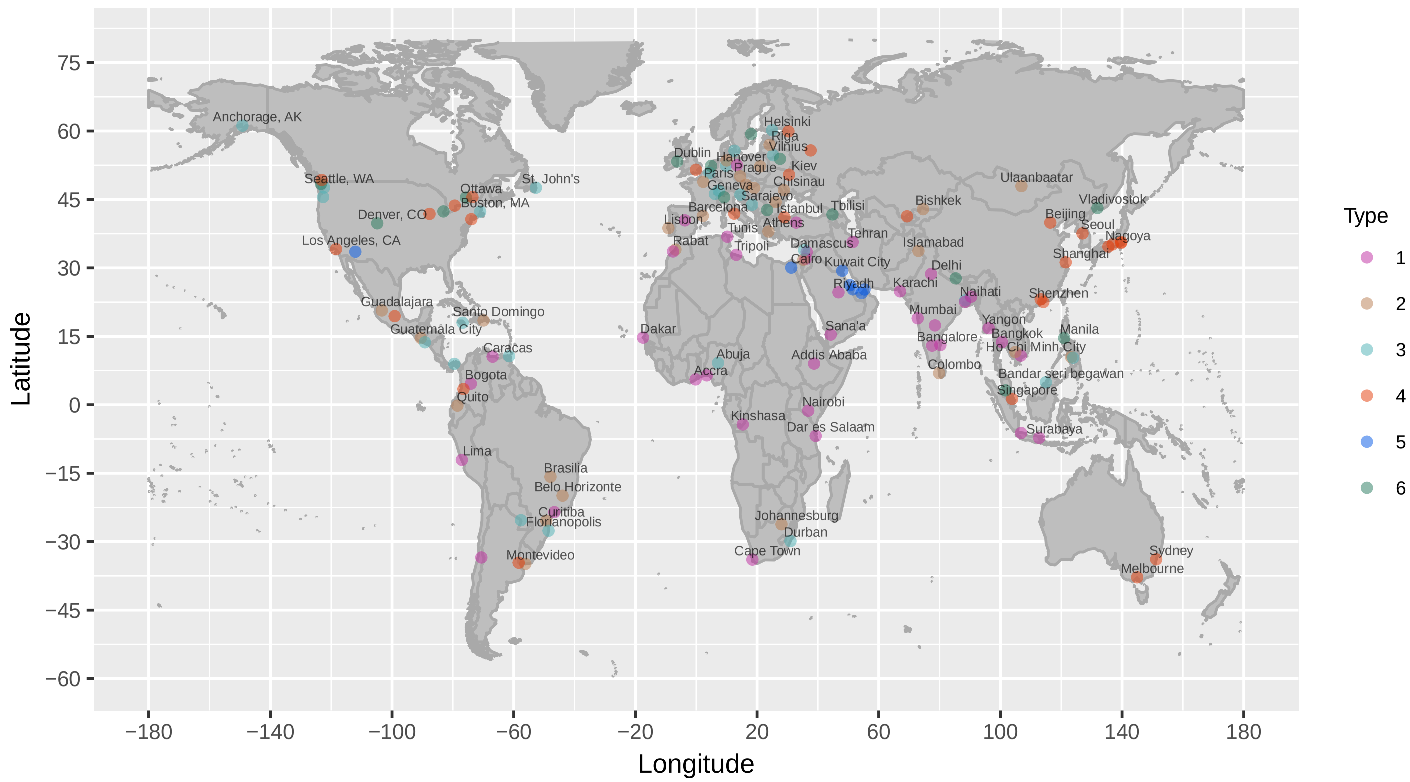


Figure 3.10: A world map of UrbMet cities, colored by type (as identified with the trivariate clustering). Viewing the typology in this way illustrates a few points. First, cities within the same country or continent are not necessarily of the same type, which suggests that political or geographical distinctions may not be the most useful basis for comparing UWS for SUWM. While political or geographical distinctions may still be useful for some studies, these results demonstrate another viable approach to identifying groups.

Figure 3.10 on Page 111 shows the UrbMet cities plotted on a world map and colored by type. Viewing the typology on a world map illustrates a few points. First, cities within the same country or continent are not necessarily of the same type, which suggests that political or geographical distinction may not be the most useful basis for comparing UWS. For instance, the cities from the United States of America (US) that were included in the study represent four of the six types. By highlighting similar cities that are not from the same country, the typology helps to identify opportunities for international comparison and knowledge exchange. Similarly, while it can sometimes be attractive to group cities by geography, e.g., "African cities", the typology highlights the limitations of that approach. For instance, while most of the African cities included in the study are members of Type 1, there are also cities that are members of Type 2 or 3. There are many Type 1 cities in Asia, Oceania, South America, and even Europe, and Types 2 and 3 are also distributed across multiple continents. At the same time, the approach was sensitive to dominant climatological factors—for instance, the extreme aridness of the Middle Eastern climate led to these cities being grouped together in Type 5 (which also included Arizona in the US). In short, while political or geographical distinctions may still be useful for some studies, these results demonstrate another viable approach to identifying groups of "like with like".

It is also useful to use this world map to highlight gaps in data coverage. For instance, some areas of the world (e.g., Europe) are better represented by cities in the UrbMet database than others. For instance, coverage in Africa is low, even though that continent is the location of some of the world's newest cities [310]. My coverage in China is also low, with only three coastal cities. Numerous other opportunities for expanding coverage also exist. The implications of this are discussed further in Section 3.8.5.

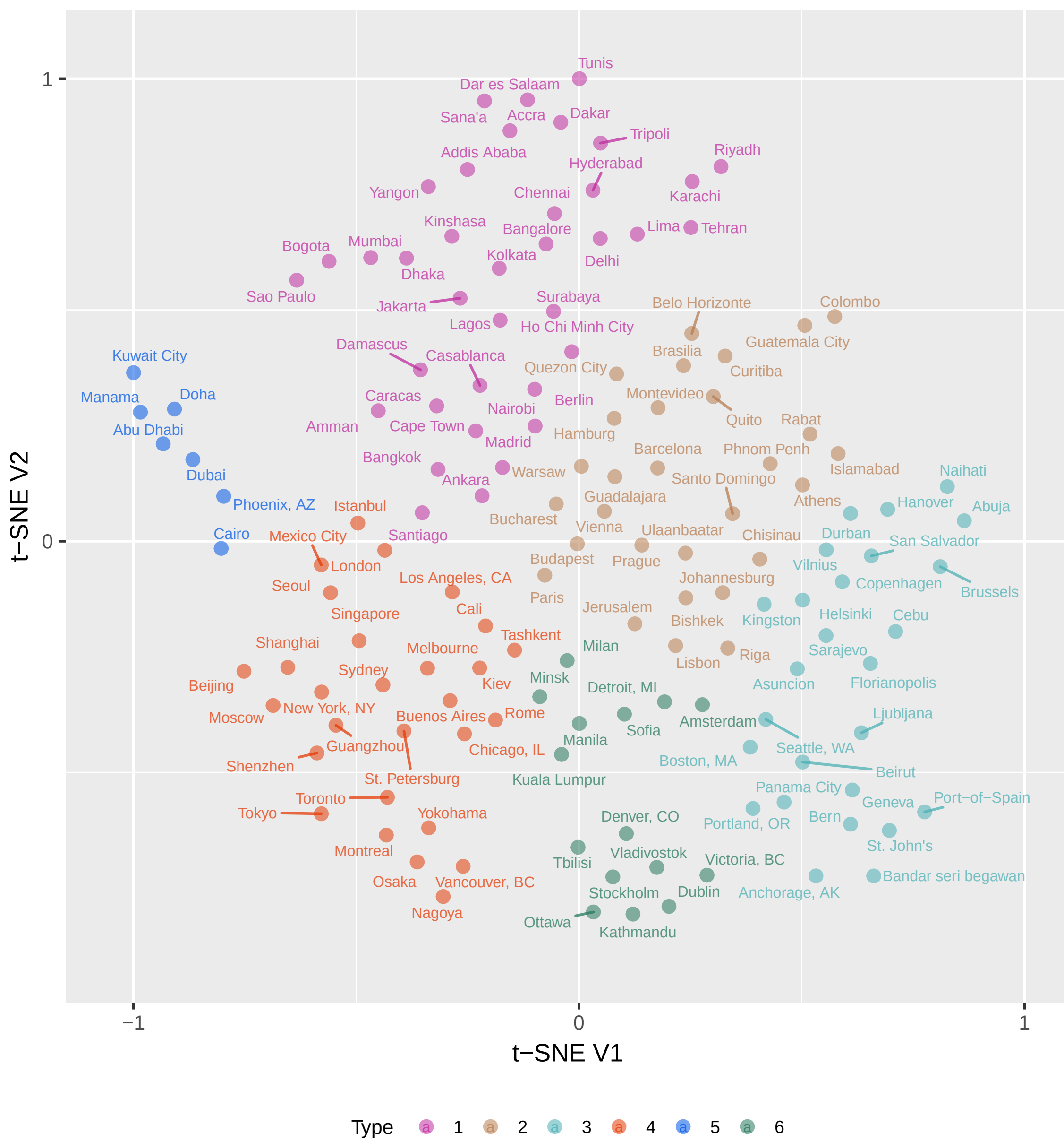
The dendrogram in Figure 3.8.1, the violin-/box-plots in Figures 3.9a–3.9d, and the world map in Figure 3.10 provide broad insight into the typology, but it is difficult to identify broader trends in similarity and difference from these plots alone. Next, a scatterplot of the t-SNE results is presented that helps to overcome this gap.

t-SNE scatterplot

One step in the trivariate clustering was the reduction of the trivariate space to a two-dimensional space using the t-SNE algorithm. Figure 3.8.1 on Page 112 shows the UrbMet cities plotted on that two-dimensional t-SNE space, with cities colored by the trivariate hierarchical clustering results. In other words, the scatterplot in Figure 3.8.1 shows the results of the t-SNE "projection" of the cities from the three-dimensional space formed by N, w_N, and q_{Net} onto a two-dimensional space determined by t-SNE. In the two-dimensional t-SNE space, each city's location was determined by the t-SNE algorithm with respect to its most likely nearest neighbors. In other words, the axes in Figure 3.11, 't-SNE V1' and 't-SNE V2', are the two vectors embedding each UrbMet city near its most similar neighbors. The scatterplot of t-SNE results helps to visualize the results of the hierarchical clustering and dendrogram in a way that is easier to gain intuition into the relative relationships between types *and* between specific cities.

Viewing the typology on the scatterplot of t-SNE results in Figure 3.8.1 helps to supplement the dendrogram in Figure 3.9. For instance, it is easier to see the relative similarity between Types 1, 4, and 2 on the t-SNE scatterplot. The t-SNE scatterplot also provides a view of the

Figure 3.11: A scatterplot of UrbMet cities, colored by type, plotted on 't-SNE V1' and 't-SNE V2', the two-dimensional vector embedding identified through t-SNE dimensionality reduction (see Section 3.5) of the trivariate space of population (N), water use intensity (w_N), and net climatic balance (q_{Net}). The types are clearly distinct in this diagram of the t-SNE results, and provide intuition into the relative relationships between the cities. This intuition is difficult to acquire from the dendrogram on Page 105, the combined violin- and box-plots on Pages 106–108, or the scatterplots of the three bivariate combinations of N, w_N, and q_{Net} (see Appendix B.1.4).



typology that would be difficult to infer from either the violin- and box-plots (Figures 3.9a–3.9d) or scatterplots of the bivariate combinations of N, w_N, and q_{Net} (see Figure B.16 on Page 288 in Appendix B.1.4).

3.8.2 Overview of Types

Table 3.7: Representative cities for each city type with descriptive attributes for population N, use intensity w_N, annual water balance q_{Net}, and i_{UC}. Designations 'high', 'medium', and 'low' were made with reference to the statistical quartiles (shown on the violin plots in dashed lines) and to the thresholds between types.

Type	Representative Cities	N	w_N	q_{Net}	i_{UC}
1	Lagos, Mumbai, São Paulo	Medium-low – very large	Very low	Low–high (generally ≤ 0)	Low–medium (generally low)
2	Hamburg, Montevideo, Ulaanbaatar	Medium-sized	Medium-low	Medium-low–Medium	Low–medium (generally low)
3	Boston, Copenhagen, Durban	Generally smaller	Medium	Medium–high (positive)	Low–medium (generally low)
4	London, New York, Tokyo	Medium-high – high (generally high)	Medium-high–high	Low–high (generally medium and ≥ 0)	Low–medium (generally medium)
5	Abu Dhabi, Cairo, Phoenix	Very small–very large	High–very high	Low–very low	Low–medium (generally medium)
6	Amsterdam, Denver, Kuala Lumpur	Low–medium-low	Very high	Medium (generally ≥ 0)	High

The cluster analysis yielded interesting results at several levels of division and grouping. Results from clustering were examined for two to eight clusters. It was found that six clusters yielded the most varied differences between clusters that were still legible through each of the four visualization methods. The typology generated by the six clusters is presented and discussed here. The cities were plotted on a world map, colored by type, as shown in Figure 3.10. A visual inspection suggested some suggestive patterns, such as East Asian, South Asian, and Central African cities, and other mixed/gradient patterns in Europe and North America. However, the plot had too much visual complexity to characterize any patterns from visual inspection alone.

The pairings of cities across the full range of clusters (i.e., from 1 to 142), were plotted as a dendrogram, shown in Figure 3.9 (with the six clusters selected shown in color). The scatterplot in Figure 3.11, produced using Ward's Linkage Criterion, depicts the six clusters more clearly in the two-dimensional space produced by t-SNE.

However, the six clusters are most readily interpreted using the violin-/box- plots shown in Figures 3.9a–3.17c. These graphic visualizations made it possible to discern the six urban clusters in qualitative and quantitative terms. Table 3.7 used these figures to define the general characteristics of each cluster and list several representative cities. I now recast these results as a working typology of urban water sustainability situations. The description of each "type" uses the terms "low", "medium", and "high" in quantitative as well as qualitative terms, as they are based on the quartile breaks and thresholds between clusters.

3.8.3 Initial Comparison

The Large Cities: Very Low w_N vs. High w_N (Types 1 and 4) The largest cities, with few exceptions, appear in Type 1 and Type 4. The main distinction between Type 1 and Type 4 is that Type 1 cities have very low w_N while Type 4 cities have high w_N. As would be expected, Type 1 cities have a i_{UC} that is much lower than that of Type 4 cities. However, Type 1 cities tend to have a low q_{Net}, and this means that the i_{UC} of Type 1 cities is more similar to that of Type 2 and Type 3 cities than it would be otherwise.

The high per capita water consumption exhibited by Type 4 cities suggest that these cities currently have sufficient water supply and urban water infrastructure. Type 4 cities also tend to be located in areas with higher natural water availability than several other city types, such as Type 1. However, because the population of Type 4 cities tends to be so large and consumption is so high, these cities likely face many challenges in securing sufficient water supply now and will continue to in the future. A cursory survey of the members of Type 4 appears to support this finding. For instance, even Type 4 cities with abundant natural resources, such as Singapore and New York, have recently needed to make substantial investments in promoting water conservation and in innovative water infrastructure (e.g., desalination and reclamation).

Both Type 1 and Type 4 cities are likely to have challenges in obtaining sufficient water resources for large urban populations. However, differences exist between the two types. Many Type 1 cities are located in developing countries and undergoing rapid urbanization. Their water resources challenges may be heightened by low natural water availability, rapid urbanization, and low access to financial resources. The low water consumption in these cities may be associated with relatively low levels of infrastructure. Even though Type 1 cities currently consume less water than Type 4 cities, it would not necessarily be appropriate to say they are "more sustainable" than Type 4 cities. If Type 4 cities can manage the resources within their watershed areas properly, Type 4 cities may be as sustainable relative to their water supply as Type 1 cities. As Type 1 cities continue to grow and develop, so too will the size of their catchment areas; Type 1 cities may eventually become Type 4 cities.

Type 1 cities also have the opportunity to implement new types of technologies or pioneering water management practices, such as more decentralized, neighborhood level water treatment and reuse, rather than trying to copy the development of water supplies of Type 4 cities exactly. This would be an excellent opportunity for Type 1 cities to collaborate with Type 4 cities for knowledge exchange.

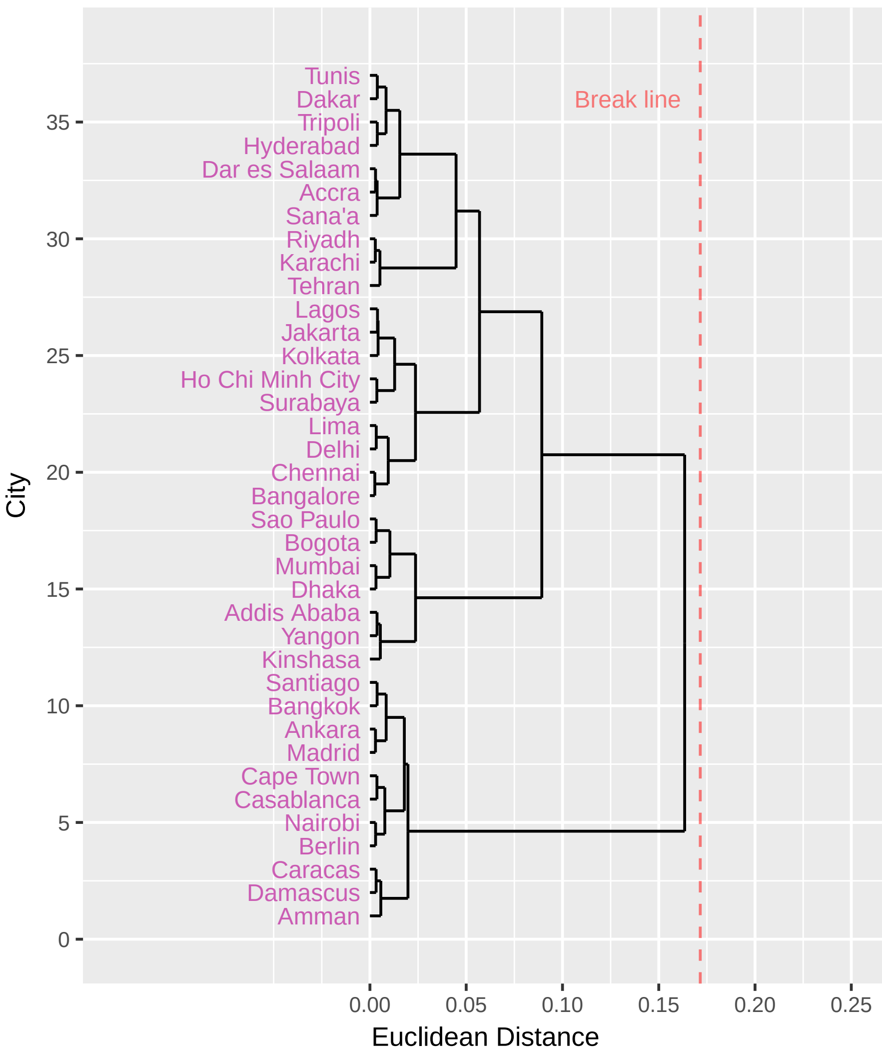
The Small, Wet Cities (Type 3) Type 3 cities are characterized by low populations that have low- to medium- levels of per capita water consumption and medium- to high- natural water supply. These cities are likely to have the fewest issues with sustainability compared to the other types; viewed from a slightly different perspective, Type 3 cities will likely find the transition to sustainability to be easier. Water supply pressures may be less acute for Type 3 cities. However, in spite of relatively low w_N and abundant water resources, many of the cities in Type 3 still face challenges with sustainable urban water management. These may include changes in water quality due to urbanization, aging infrastructure, and flooding.

The Medium-Sized Cities: Medium w_N vs. High w_N (Types 2 and 6) With a few exceptions, medium-sized cities (those in the 2nd and 3rd quartiles of population) tend to fall into Types 2 and 6. Type 6 cities tend to fall into the 3rd quartile of q_{Net}, while those of Type 2 tend to fall into the 2nd quartile. Type 6 cities have higher naturally available water resources. Another distinction between the two types is that Type 6 cities have very high w_N, which raises concerns about their sustainability, while those of Type 2 have medium-low w_N.

The Arid Cities: Low q_{Net} (Type 5) Type 5 cities were located in highly arid environments: these cities had negative values of q_{Net}, which means that there is much less rainfall than potential evapotranspiration. Surprisingly, these cities were also distinguished by their high levels of water consumption (w_N). These cities are therefore likely to rely extensively on water transfers or water imports for water supply, either as rivers or canals (e.g., Cairo and Phoenix), conversion of seawater to freshwater (e.g., Dubai and Abu Dhabi), and mining of fossil aquifers. Transporting water over large distances, desalination, and reclamation are all associated with relatively large costs [255]. Since these cities likely have high costs of supplying water. The water resources for cities in Type 5 are thus particularly vulnerable to economic shocks, energy availability, and political changes (since they may be obtaining water from outside their jurisdiction). Cost recovery and conservation may help Type 5 cities to increase the resilience of their urban water systems to climate change and to external pressures on their water supplies.

Figure 3.12: Overview of Type 1.

(a) Close-up of dendrogram of Type 1 from Figure 3.9.



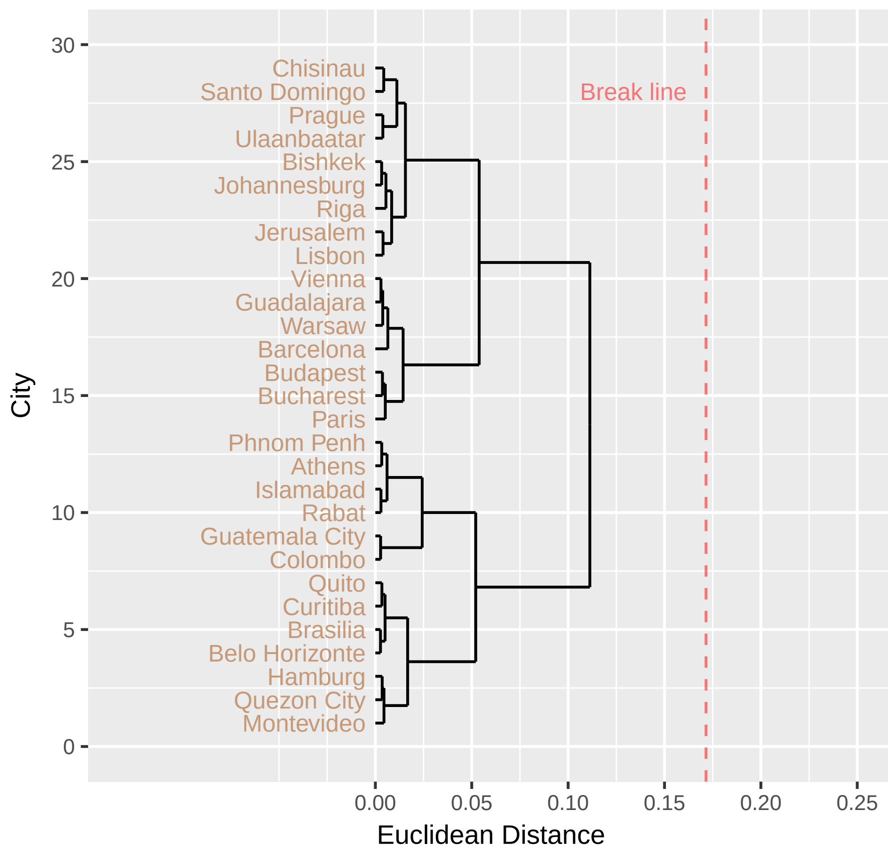
(b) Overview of Type 1.

N	w_N	q_{Net}	i_{UC}
capita	$m^3 \cdot yr^{-1} \cdot cap^{-1}$	$m \cdot yr^{-1}$	$m^2 \cdot cap^{-1}$
Medium-low—very large	Very low	Low—high (predominantly ≤ 0)	Low—medium (predominantly low)

Large cities with very low per capita water consumption. These cities represent a range of climate and supply conditions, but the net water balances are predominantly negative. The median value of population for Type 1 cities lies in the 3rd quartile. Type 1 cities fall predominantly below the median value for w_N, with a median value of 45.0 $m^3 \cdot yr^{-1} \cdot cap^{-1}$, that is within the first quartile. The median value for q_{Net} for Type 1 also lies within the 1st quartile, and the box for Type 1 on q_{Net} lies within the 1st and 2nd quartiles. In other words, Type 1 cities tend to have negative values for q_{Net}. To summarize, Type 1 cities tend to have large populations, low natural water availability on the low side, and very low per capita water consumption, which raises concerns about their sustainability in climatic and socio-economic terms Type 1 cities are most dissimilar from cities of Type 3 and 6, which are smaller and wetter. They show some overlap with Type 2 and Type 4 cities.

Figure 3.13: Overview of Type 2.

(a) Close-up of dendrogram of Type 2 from Figure 3.9.



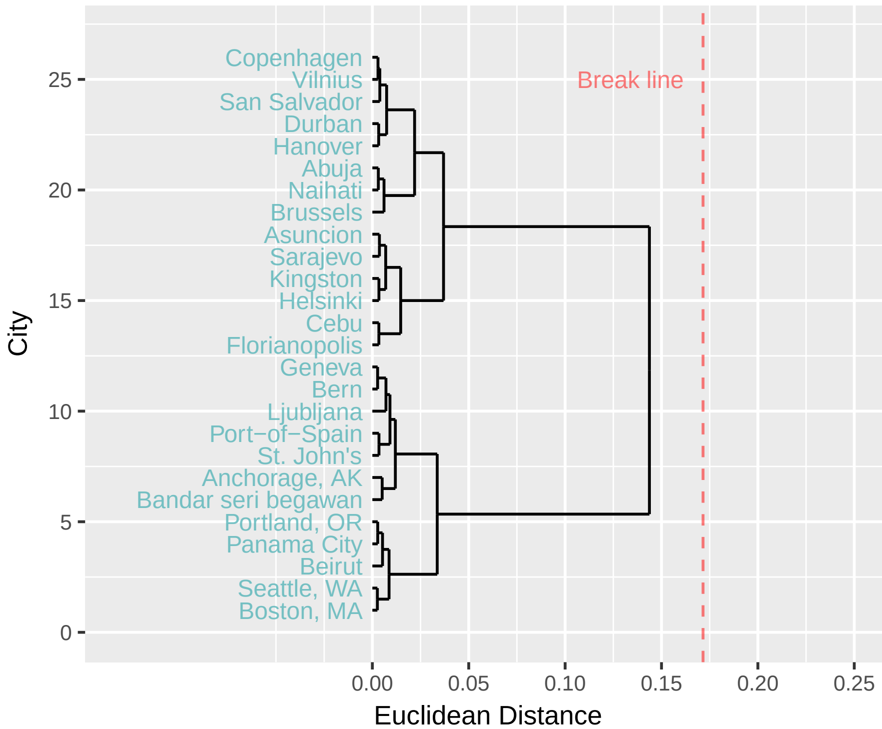
(b) Overview of Type 2.

N	w_N	q_{Net}	i_{UC}
capita	$m^3 \cdot yr^{-1} \cdot cap^{-1}$	$m \cdot yr^{-1}$	$m^2 \cdot cap^{-1}$
Medium-sized	Medium-low	Medium-low—Medium	Low—medium (predominantly low)

Medium-sized cities with medium-low water consumption and a net water balance of around zero. Type 2 cities have values for population that fall predominantly in the 2nd quartile, with a median value of 1.188 million; Type 2 cities tend to be medium-low to medium-sized. These cities have the second-lowest median value for w_N, after those of Type 1. Type 2 cities also have low to moderate water balances that (q_{Net}). In summary, Type 2 cities tend to be mid-size cities with q_{Net} close to zero (i.e., moderate water resources) and medium-low w_N, which are characteristics that may be sustainable.

Figure 3.14: Overview of Type 3.

(a) Close-up of dendrogram of Type 3 from Figure 3.9.



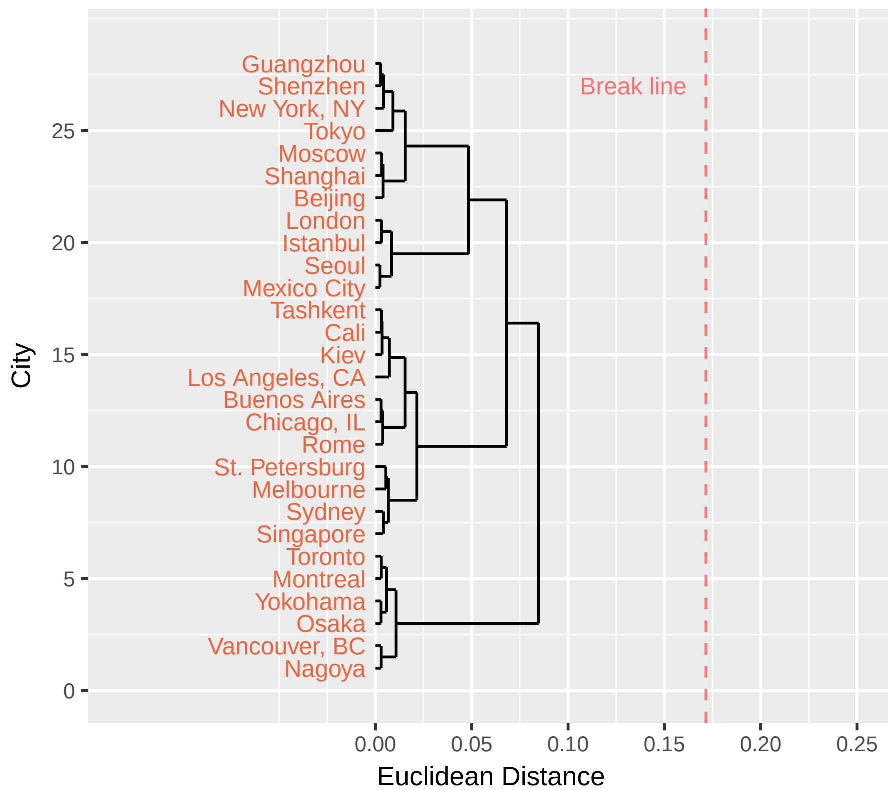
(b) Overview of Type 3.

N	w_N	q_{Net}	i_{UC}
capita	$m^3 \cdot yr^{-1} \cdot cap^{-1}$	$m \cdot yr^{-1}$	$m^2 \cdot cap^{-1}$
Generally smaller	Medium	Medium—high (positive)	Low—medium (predominantly low)

Small cities with medium levels of water consumption and positive net water balances. These cities tend to be small relative to the other cities in the dataset; their median value is the lowest of any of the six city types at 494,700 and nearly all Type 3 cities fall within the 1st quartile for population. They have a median value for w_N with a value of 84 $m^3 \cdot yr^{-1} \cdot cap^{-1}$. Type 3 cities have the highest median value for q_{Net} of any of the other types, at 0.417 $m \cdot yr^{-1}$, and most Type 3 cities have values for q_{Net} in the 3rd and 4th quartiles. In other words, Type 3 cities tend to be small, with relatively high naturally available water resources and medium levels of water consumption. These cities have the lowest Water Use and Conservation Index (i_{UC}), and might therefore be deemed the most sustainable in terms of use (though not in terms of risk which is not considered in this analysis). Type 3 cities show similarities with Type 2 and Type 6; and they are most dissimilar from Types 1, 4, and 5.

Figure 3.15: Overview of Type 4.

(a) Close-up of dendrogram of Type 4 from Figure 3.9.

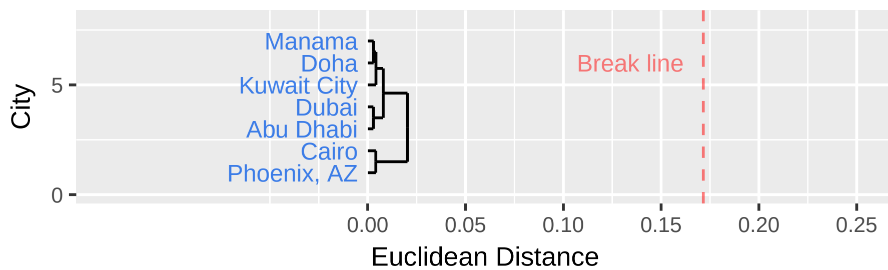


(b) Overview of Type 4.

N	w_N	q_{Net}	i_{UC}
capita	$m^3 \cdot yr^{-1} \cdot cap^{-1}$	$m \cdot yr^{-1}$	$m^2 \cdot cap^{-1}$
Medium-high—high (predominantly high)	Medium-high—high	Low–high (predominantly medium and ≥ 0)	Low—medium (predominantly medium)
Very large cities with high per capita water consumption and a positive net water balance. Type 4 cities have large populations with a median value of 4.45 million. Type 4 cities also have high water use intensity (w_N, with a median value of 154.5 $m^3 \cdot yr^{-1} \cdot cap^{-1}$, which falls in the 4th quartile). Almost all Type 4 cities are in the 3rd and 4th quartiles for w_N and q_{Net}. To summarize, they are very large cities with large natural water supplies and demand, and thus have greater potential to be sustainable.			

Figure 3.16: Overview of Type 5.

(a) Close-up of dendrogram of Type 5 from Figure 3.9.



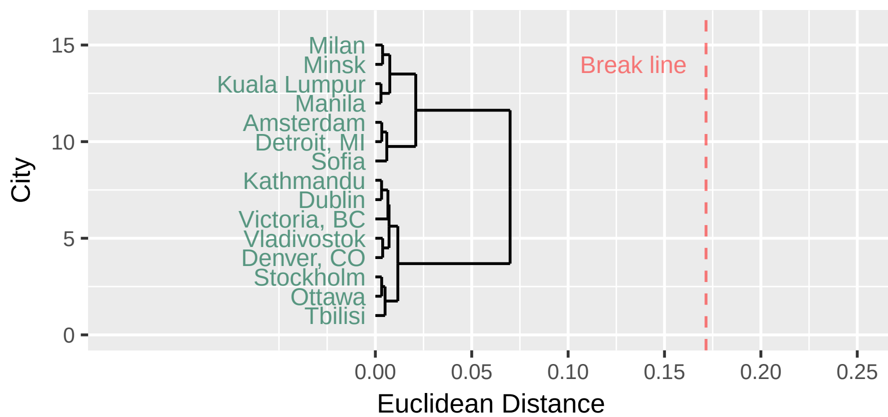
(b) Overview of Type 5.

N	w_N	q_{Net}	i_{UC}
capita	$m^3 \cdot yr^{-1} \cdot cap^{-1}$	$m \cdot yr^{-1}$	$m^2 \cdot cap^{-1}$
Very small—very large	High—very high	Low—very low	Low—medium (predominantly medium)

Cities of varied size with very high per capita water consumption located in highly arid environments. These cities have the second-highest median value for w_N, at 209.0 $m^3 \cdot yr^{-1} \cdot cap^{-1}$; almost all of the cities have w_N that falls within the 4th quartile. However, Type 5 has the lowest q_{Net} of any of the types, with a median value of -1.36 $m \cdot yr^{-1}$and a maximum value of -0.916 $m \cdot yr^{-1}$. The population of the Type 5 cities ranges from the smallest to the largest: from 32,400 to 6.759 million. To summarize, Type 5 cities have a range of populations; they are characterized by high w_N and very low q_{Net}. This pattern raises serious concerns about sustainability.

Figure 3.17: Overview of Type 6.

(a) Close-up of dendrogram of Type 6 from Figure 3.9.



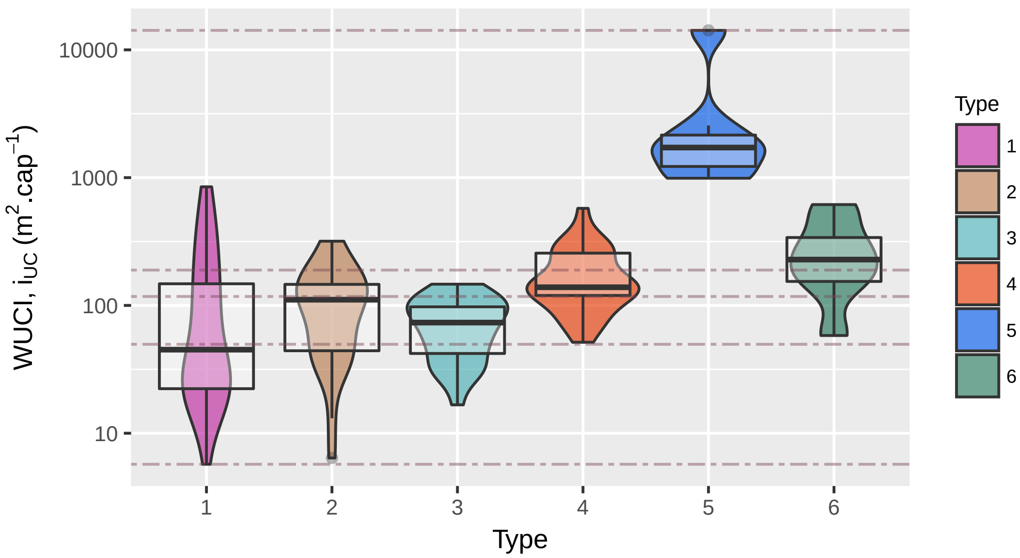
(b) Overview of Type 6.

N	w_N	q_{Net}	i_{UC}
capita	$m^3 \cdot yr^{-1} \cdot cap^{-1}$	$m \cdot yr^{-1}$	$m^2 \cdot cap^{-1}$
Low—medium-low	Very high	Medium (generally ≥ 0)	High

Medium-sized cities with very high per capita water consumption and low net positive water balance. These cities tend to be medium-sized cities, with high per capita water consumption and low, positive net water balances. They are one of the most rapidly growing forms of urbanization in developing countries, and this pattern raises concerns about sustainability. Type 6 cities are most similar to Type 1, Type 2, and Type 3 cities. They are most dissimilar from Type 1 and Type 5 cities, especially in terms of q_{Net}.

3.8.4 WUCI, Total WUCI, and Potential Self-Sufficiency

(c) Violin- and box-plot of Water Use and Climate Index i_{UC}, by type.



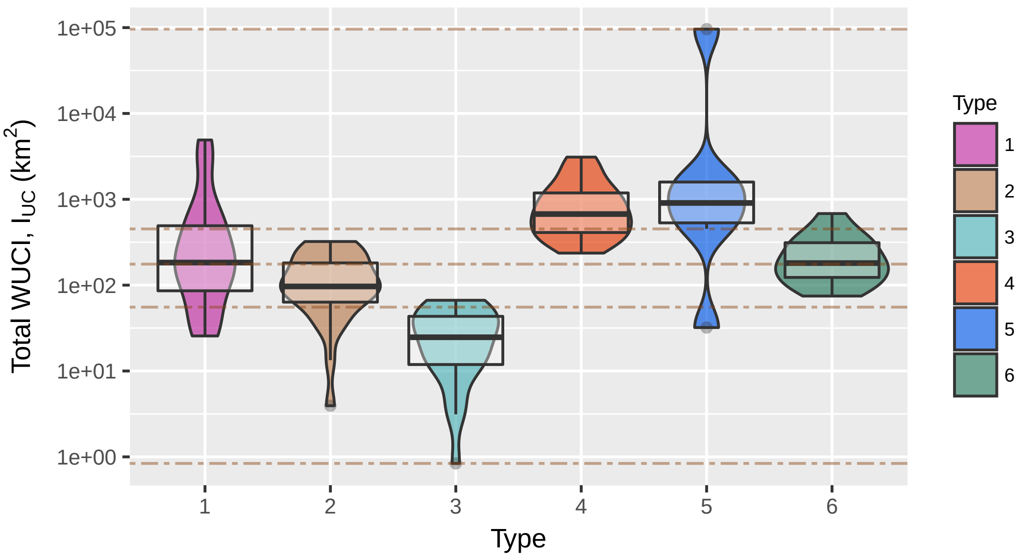
(d) Summary statistics for Water Use and Climate Index, by type.

Statistic	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
μ	135.79	111.42	73.89	177.77	3395.57	276.59
σ	195.62	79.43	36.13	110.90	4794.15	178.01
γ	2.12	0.94	0.18	1.70	1.58	0.69
κ	4.05	0.43	-1.22	3.41	0.71	-0.90
Quantile	**Type 1**	**Type 2**	**Type 3**	**Type 4**	**Type 5**	**Type 6**
0%	5.72	6.40	16.68	51.51	989.58	58.16
25%	22.31	44.20	42.24	119.74	1244.29	154.09
50%	45.08	110.86	73.58	138.73	1720.00	228.45
75%	147.76	146.15	97.54	256.71	2185.41	340.72
100%	847.62	318.18	146.70	575.09	14200.00	615.38

The WUCI, total WUCI, and potential self-sufficiency ratio (R_{SS}) were calculated for each of the six types (as described in Section 3.5.4) and plotted as violin-/box-plots. These plots are shown in Figures 3.17c–3.17g on Pages 124–126, respectively. Summary statistics and ordered bar charts, colored by quantile, for WUCI, total WUCI, and R_{SS} are provided in Appendix B.1.3. The composite indicators help to provide additional insight into the different types.

The Water Use and Climate Index: WUCI, or i_{UC} The Water Use and Climate Index (WUCI or i_{UC}, in $m^2 \cdot cap^{-1}$) for the six types was calculated and is shown as a violin-/box-plot in Figure 3.17c. Recall from Section 3.5.4 that WUCI was defined as $i_{UC} = w_N / q_P$; i.e., as the ratio of water use intensity (w_N, in $m^3 \cdot yr^{-1} \cdot cap^{-1}$) to precipitation ($q_P$, in $m \cdot yr^{-1}$). WUCI was defined in a way such that its value indicates an area such that collecting 100% of the precipitation falling on

(e) Violin- and box-plot of total Water Use and Climate Index (I_{UC}), by type.



(f) Summary statistics for total Water Use and Climate Index (I_{UC}), by type.

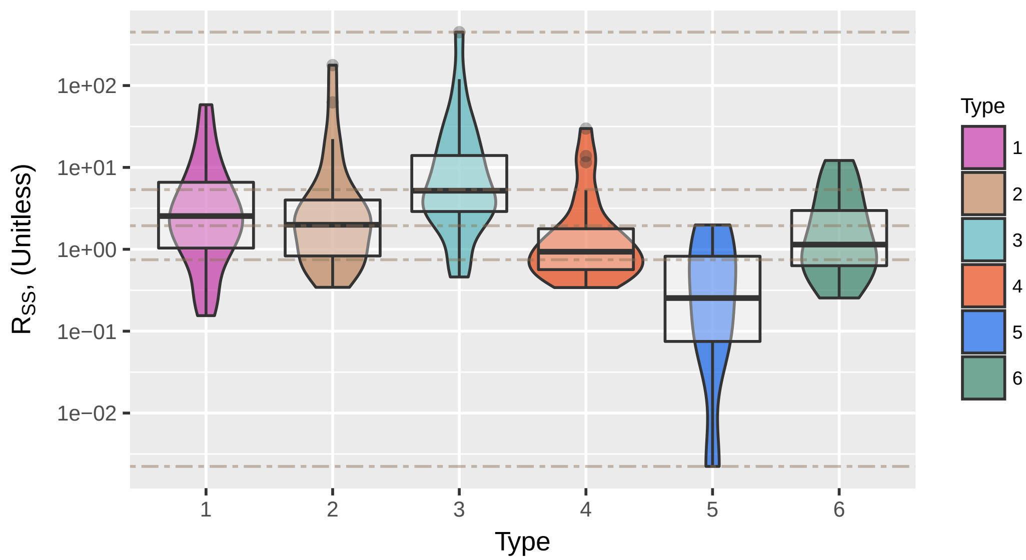
Statistic	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
μ	625.68	130.05	28.25	958.70	14452.56	233.27
σ	1130.26	90.73	19.87	770.44	35951.20	163.21
γ	2.47	0.69	0.37	1.44	1.62	1.35
κ	5.10	-0.73	-1.21	1.19	0.79	1.18
Quantile	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
0%	25.60	3.94	0.84	236.35	32.07	74.64
25%	85.97	63.42	11.93	410.35	538.75	123.43
50%	183.28	96.55	24.87	673.49	906.44	180.89
75%	491.72	181.60	43.32	1187.27	1590.04	312.34
100%	4907.28	322.23	66.65	3101.03	95971.85	682.22

that surface would meet the direct water use of a single average person in that city. Since that is not possible, WUCI should be interpreted as an index, indicating the relative difficulty in supply a city through climatic resources alone.

Water use intensity of UrbMet cities ranged over two orders of magnitude (from 14 $m^3 \cdot yr^{-1} \cdot cap^{-1}$ to 355 $m^3 \cdot yr^{-1} \cdot cap^{-1}$), while precipitation ranged over three (from 0.025 $m \cdot yr^{-1}$to 4.223 $m \cdot yr^{-1}$). WUCI ranged over four orders of magnitude, from 5.72 $m^2 \cdot cap^{-1}$ to 14,200 $m^2 \cdot cap^{-1}$. Therefore, like population, the violin-/box-plot for WUCI is plotted on a $\log_{10}$ scale, with purple dashed lines indicating the global quantiles.

Types 4, 6, and 5 had the highest WUCI of the six types, in that order. Cities in Type 5 had the highest WUCI, with i_{UC} ranging from 980 $m^2 \cdot cap^{-1}$ to 14,2000 $m^2 \cdot cap^{-1}$ (Cairo). In other words, the WUCI of Type 5 cities is on the same order magnitude as the total area of some cities. Type 6 cities have the next highest WUCI, with an IQR of $[154, 341]$ $m^2 \cdot cap^{-1}$ and a median value of 228 $m^2 \cdot cap^{-1}$. The IQR of Type 6 overlapped with that of Type 4, which was $[120, 257]$ $m^2 \cdot cap^{-1}$.

(g) Violin- and box-plot of the potential Self-Sufficiency Ratio (R_{SS}), by type.



(h) Summary statistics for R_{SS}.

Statistic	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
μ	6.90	11.60	31.59	3.09	0.58	2.77
σ	12.85	34.06	88.72	6.15	0.71	3.43
γ	2.96	4.08	4.02	3.19	0.93	1.50
κ	8.27	16.56	15.87	10.30	-0.67	1.17
Quantile	Type 1	Type 2	Type 3	Type 4	Type 5	Type 6
0%	0.15	0.34	0.46	0.34	0.00	0.25
25%	1.04	0.83	2.90	0.57	0.08	0.63
50%	2.55	1.99	5.22	0.94	0.25	1.14
75%	6.59	4.00	13.98	1.78	0.82	2.99
100%	58.30	177.43	448.87	29.84	1.98	12.15

Types 1, 3, and 2 had the lowest WUCI of the six types (ordered by increasing WUCI). Type 1 had the lowest WUCI of any of the other types, with an IQR of $[22.3, 148]$ $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$, and a median value of 45.1 $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$. The IQR of Type 1 overlapped with that of Type 2 and that of Type 3. Cities in Type 3 had the next lowest WUCI, with an IQR of $[42.2, 97.54]$ $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$ and a median value of 73.6 $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$. Type 2, with an IQR of $[44.2, 146]$ $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$ and a median value of 111 $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$ overlapped with that of Type 3.

The Total Water Use and Climate Index: total WUCI, or I_{UC} The Total Water Use and Climate Index (total WUCI or I_{UC}, in units of m^2 or km^2) was defined as WUCI multiplied by population: $I_{UC} = N \cdot i_{UC}$. As with WUCI, total WUCI provides a rough indicator of the difficulty in supplying water to an urban population with respect to average water use intensity and local climatic water availability. Another way to view WUCI and total WUCI is a type of per capita or total water footprint metric, respectively.

Figure 3.17c shows a violin-/box-plot of total WUCI for the six types, with global quantiles plotted as brown dotted lines. The relative position of the violins and boxes in the plot roughly follows those of population that were seen in Figure 3.9a, with some exceptions. As was the case with WUCI, Type 5 had the highest total WUCI of any type, with an IQR of $[539, 1590]$ km^2 and a median value of 906 km^2; this is in spite of the fact that the populations of Type 5 cities tended to be smaller[23]. Type 4 had the next highest total WUCI, with an IQR of $[410, 1187]$ km^2 and a median value of 673 km^2. Type 4 cities also had the largest populations and the second highest WUCI values. The IQRs of Types 1, 2, and 6 generally fell between the 25th and 75th global percentiles for this metric. The median values for Types 1 and 6 were approximately the same, at 183 $\mathrm{m} \cdot \mathrm{yr}^{-1}$and 181 $\mathrm{m} \cdot \mathrm{yr}^{-1}$, respectively; however, Type 1 had the larger IQR of $[86.0, 492]$ km^2 compared to Type 6's IQR of $[123, 312]$ km^2. Although Type 1 had some of the largest populations in the study, it also had some of the lowest water use intensities, so total WUCI for Type 1 cities was not as high as either Type 1 or Type 5. Type 2 cities had the second smallest total WUCI, with an IQR of $[63.4, 182]$ km^2 and a median value of 96.6 km^2. Type 3 cities, which tended to have small populations and wetter climates, had the lowest total WUCI of all of the six types, with an IQR of $[11.9, 43.3]$ km^2 and a median value of 24.9 km^2.

The Potential Self-Sufficiency Ratio, R_{SS} The Potential Self-Sufficiency Ratio (i.e., the Ratio of Potential Self-Sufficiency) was defined in Section 3.5.5 as the ratio of city area (A_N) divided by total WUCI: $R_{SS} = A_N / I_{UC}$—since both area and total WUCI have units of area, R_{SS} is dimensionless. As is the case with WUCI and total WUCI, R_{SS} is a theoretical construct relating the water footprint—total WUCI—representing the area required to supply the city with water using only climatic water resources—to the physical boundary of the city. If $R_{SS} \geq 1$, this suggests that *in theory*, 100% of a city's water use could be met with local climatic resources. In practice, this is not likely, since it would require all of my broad assumptions—such as uniform precipitation, 100% collection of precipitation with zero losses, etc.—to be true. Still, the calculation of R_{SS} helps to highlight, all other things being equal, whether a city is likely to be dependent on external water sources.

After calculating R_{SS} for all of the cities in the study, over 65% were found to have $R_{SS} \geq 1$.

A violin-/box-plot of R_{SS} for the six types is shown in Figure 3.17g. I found that R_{SS} for the UrbMet cities ranged from $449 \geq R_{SS} \geq 2.23 \cdot 10^{-3}$. Type 3 cities had the highest R_{SS}, with R_{SS} generally greater than one, an IQR of $[2.90, 449]$, and a median of $R_{SS} = 5.22$. Type 5 cities had the lowest R_{SS} of all of the types, with an IQR of $[0.0756, 0.823]$ and a median value of 0.254. The IQRs for Types 1, 2, 4, and 6 generally fell between the 25th and 75th global percentiles, which fell at 0.741 and 5.30, respectively. The median values for Types 1 and 2 fell close to the 50th global percentile (at 1.93), with values of $R_{SS} = 2.55$ and $R_{SS} = 1.99$, respectively. The median values for Types 4 and 6 were lower, occurring at $R_{SS} = 0.94$ and $R_{SS} = 1.14$, respectively.

[23] Although, remember that Type 5 has only seven members.

Surprisingly, R_{SS} for most of the cities in UrbMet were found to be larger than 1, suggesting that many cities could benefit from local stormwater collection and treatment if their water imports become threatened. Of course, the actual potential for self-sufficiency would likely be lower, since the precipitation (q_P) used in calculating R_{SS} does not account for interception, infiltration, or other losses and it is difficult to collect all of the runoff from a city, let alone treat it.

3.8.5 Discussion

Thresholds

The thresholds for q_{Net} suggest that cities in regions with naturally abundant water resources may be fundamentally different from those in more arid contexts. Types 3, 4, and 6 predominantly have values for q_{Net} greater than zero, which suggests that, all other things being equal, these cities may have relatively more ease at obtaining water resources than the other types.

The thresholds for population raise intriguing questions. Are there significant differences in providing urban water resources for cities larger than 1 million? Types 1 and 4 tend to be the largest cities in the database, but these two types differ in w_N and q_{Net}. Cities of Type 1 have very low w_N and low q_{Net}, while those of Type 4 have high w_N and a positive q_{Net}. This suggests that city size is not deterministically related to climate or consumption.

Outliers and Overlaps

It is important to remember that these types are based on the statistical clustering of continuous variables. Cities within a type are most likely to be more similar to each other than to other cities, and distinct from other types in some statistically significant way. The different types of cities identified in the clustering were relatively distinct on one or two variables but may have otherwise overlapped with another type on another. Because of this, cities with different membership may have similar metrics and therefore some common sustainability challenges. For instance, Cairo and Paris are members of Type 5 and Type 2, respectively, but due to their size are likely to share similar challenges with those of Type 1 and Type 4 respectively.

These overlaps and distinctions can be understood from Figure 3.11, the scatterplot of the t-SNE results, in which the cities that fall close to one another are most similar, while those that are further apart are most dissimilar. Two cities that lie in a transition zone between adjacent types may have different typological membership but be statistically similar to each other. An example of this is Los Angeles which is in Type 4 and Paris which is in Type 2. The relative adjacency of Types in Figure 3.11 also reflects the relative similarity between the types.

Within the variable space shown in Figure 3.11, it is also possible to visually identify subgroupings. For instance, the subgrouping of Denver, Tbilisi, Vladivostok, Victoria, Stockholm, Dublin, Ottawa, and Kathmandu can also be identified in the dendrogram Within this subgrouping there are non-intuitive pairings in the terminal leaf nodes, such as Kathmandu with Dublin and Vladivostok with Denver.

Intuitive vs. Unexpected Results

Some of the results are surprising while others are more expected. For instance, it is not a great surprise to find that cities with very arid climates were grouped together (in Type 5). However, prior to application of the clustering algorithm, this was not a foregone conclusion. This highlights the ability of the approach to yield meaningful results. It is perhaps also not much of a surprise that the largest cities were grouped together, as they were in Type 1 and in Type 4. At the same time, some may find it surprising to see London, Los Angeles, Singapore, Sydney, and New York together. However, examination of contemporary water issues within each of these cities supports this grouping. All of these cities have reached a size where desalination has been at least considered if not implemented; all of these cities have issues with water quality and stormwater runoff. Water plans for these cities may be expected to show some similarities. It might also have been expected that large, rapidly urbanizing cities in developing countries might be grouped together, as they were in Type 1. For instance, even prior to this clustering analysis it might have been possible to identify São Paulo and Mumbai as having similar challenges in meeting the demands of their burgeoning populations. However, it is notable that this meaningful pairing, and others, were identified through a relatively simple combination of metrics, with relatively widely available data, using a very scalable method.

The results also identified more surprising international groupings. Consider the subgrouping of Hanover, Durban, Copenhagen, Vilnius, and San Salvador in Type 3. While it is perhaps less surprising that Copenhagen, Hanover, and Vilnius are similar, it is surprising to include Durban and even more so San Salvador with those three. Yet, upon examining the scatterplots of these mid-range values on most variables, it makes sense why these cities were clustered. The identification of such sub-groups demonstrates the utility of applying quantitative data mining algorithms to uncover meaningful and intriguing relationships between cities around the world.

Limitations

Comparative water research, in general, is necessary to infer general lessons from specific cases, such as in the adaption of technology and policy portfolios from one city to a new urban context. The type of large-*n* comparative water research pursued in this chapter can help to advance knowledge about sustainability in UWS.

These results primarily demonstrate the utility of statistical clustering of a small number of metrics to support meaningful international comparison of cities. The relevance of this typology to agenda-setting and policy-making for sustainability remains an open question. One hypothesis for future work is that those cities are more likely to share similar sustainability challenges with other cities that have similar supply and demand metrics, as defined by their type and position in the t-SNE scatterplot.

To expand on this idea, it appears that some cities within a type, or adjacent to each other in Figure 3.11, may share similar issues in urban water resource management. For instance, Lagos, Jakarta, and Kolkata are all closely linked in Figure 3.12 on Page 118. These cities are all world mega-cities with issues of infrastructure shortfalls, sanitation, and flooding [175]. It is likely from this cursory inspection that these cities may be facing similar challenges in the coming years [175]. However, also within Type 1, Berlin is grouped with Cape Town, Casablanca, and Nairobi. These cities all have slightly drier climates and slightly smaller populations than other cities in Type 1. It is intriguing and surprising that Berlin would be grouped with these other cities, and would provide an interesting place for case study research, which could investigate in greater depth the extent to which these cities are similar or not. One difference is that the Berlin has a colder winter than the other three cities. Another difference is that Berlin is located in Germany, which has a high Human Development Index (HDI), while the other cities are located in South Africa, Morocco, and Kenya (respectively), which have lower HDIs. These dissimilarities do not necessarily mean that the cities do not share similar water challenges, but it also does not require that they do.

In other words, the relative similarity of two cities, as assessed in this chapter, does not require that those cities share pressing water resource and management challenges these cities face are, in fact, similar. However, as the previous discussion illustrates, the results presented in this chapter could be used directly to suggest groups of cities for small-*n* comparative research, an idea that is explored further in Chapter 4.

The results of statistical clustering highlighted some intuitive and non-intuitive results. Even if the results are intuitive, as in the grouping of Lagos, Jakarta, and Kolkata, this result is also interesting and a solid contribution to the field. While some may have suspected that these cities were similar, this is the first time that these cities have been quantified as similar when compared to many other cities in the world.

Some of the results are surprising, as is the grouping of Berlin, Cape Town, Casablanca, and Nairobi. Having spent a summer working in Nairobi on water resource issues, I may have considered Cape Town as a relevant city for comparison, but I probably would not have thought of Berlin. This type of result can also be useful. For instance, while Nairobi may have different short-term priorities—such as issues with access, sanitation, and unsustainable aquifer withdrawals and addressing a severe infrastructure shortfall—than Berlin, the pair of cities may share similar long-term challenges. If nothing else, that result suggests a study looking into the opportunity for knowledge-sharing.

However, it is important to keep in mind that these results are limited by the quality and comparability of the data. Were city boundaries used to calculate population defined in a similar way? How well does population data account for the population living in informal settlements, which in some cities represent a significant water demand?

For instance, in discussing the typology, "thresholds" in values for N, w_N, and q_{Net} were highlighted. Are these thresholds meaningful? One way this can be tested is through a classification and regression study, which would help establish internal validity. However, that study would not address the separate question of whether those thresholds would be stable if the number of cities in the analysis was expanded or if the analysis is repeated on the same set of cities but on data from a different instant in time. In other words, it would be useful for future work to examine the sensitivity of the clustering results to the underlying dataset.

3.9 Summary of Work

In this chapter, a large-*n* analysis of 142 cities was pursued to test whether a simple three metric (N, w_N, and q_{Net}) profile of urban water supply and demand could be used to provide insight into *both* global trends *and* more regional insight.

One of the main aims of this study was to use key metrics for water supply and demand to identify groups of similar and different cities. Statistical clustering identified six meaningful clusters of urban water sustainability conditions. These clusters were recast as a typology, and it was then demonstrated how these results could be used for stratified sampling of smaller numbers of cases for more fine-grained contextual comparative international water research.

Toward this end, the first part of the large-*n* comparative research focused on choosing appropriate metrics to represent urban water demand and supply profiles and assembling a dataset of a large number of international urban cases. Three simple and widely available metrics—population, water use intensity, and climatic water balance—were chosen to represent urban water profiles, as discussed in Section 3.3.1. Since this work was pursued as part of larger efforts to 1. characterize urban sustainability by the UrbMet and 2. compare cities to facilitate adoption of blue-green water infrastructure, the MIT Urban Metabolism Database of 142 cities was chosen as the starting point for this dataset [86, 356]. With 142 cities representing six continents and four of the five major climate types, the UrbMet Database of cities provided a (moderately) large-*n* sample of cities with reasonable international and climatic coverage as a useful starting point for large-*n* comparative water research.

The original population and water use intensity data from UrbMet were extended by location and climatic data, as described in Section 3.3.2. Scripts were written in Python and `R` to promote the comparability of the cases, the replicability of the analysis, and the facility in extending the analysis to include new cases. These scripts linked the UrbMet data to other widely used databases on city location, population, WSP indicators, and climatic information and are described in Section 3.3.2 and Appendix A.1.2.

The results of this second part of research, presented in Section 3.6, provided some general global context and perspective and demonstrated that even basic statistical methods could provide useful global context and perspective when performed more thoroughly—i.e., going beyond summary statistics—on a large-*n*, international set of urban cases. For instance, the results of the univariate analysis lend further support to recent findings from the emerging science of cities on urban scaling [25, 23, 24, 34, 32]. The ordered bar charts suggested that cities tend to fall on a more continuous rather than discrete—spectrum. In other words, even when only 142 cities were compared on only three variables there was a large diversity in urban water contexts. This result reemphasized one of the questions highlighted in Figure 1.2, of whether each different urban water context is uniquely different or if patterns in the diversity can be identified. The histograms presented in Section 3.6.1 highlighted that the data, especially for population, spanned multiple orders of magnitude, suggesting that the data be transformed by a logarithm to improve the symmetry of the distributions and to improve the match between the data and the normal distributions. The results of the $\log_{10}$ transformation further supported the idea that the data would be better approximated by a log-normal distribution rather than a normal distribution. The symmetry of the data distributions was found to be improved by the $\log_{10}$ transformation, as evidenced in the resulting histograms and qq-plots (presented in Sections 3.6.1 and 3.6.1 in comparison with the un-transformed data) and to improve the correlation of the data with the normal distribution, as seen in Figure 3.7. However, while the univariate results were intriguing, the bivariate analysis found no significant correlation between population, water use intensity, and net water balance.

The results of the univariate and bivariate hierarchical clustering were compared visually with statistical quantiles through rank order bar plots and scatterplots. These results visually demonstrated that grouping by statistical quantiles—a not uncommon practice in other large-*n* studies of urban water systems—resulted in groups of cities that were not more similar to each other (statistically speaking) than to cities in other groups. These results imply that statistical quantiles are not necessarily a good basis for making distinctions between urban contexts for the purpose of informing policy. Further supporting this conclusion, none of the univariate or bivariate clusterings were able to provide useful distinctions for the variables excluded for clustering. For instance, the univariate cluster results for population did not result in useful distinctions for water use intensity or net water balance; neither did the results from the bivariate clustering of water use intensity and net water balance provide insight into population. These results of the univariate and bivariate clustering suggest that constricting large-*n* comparison to univariate and bivariate analysis provides limited insight into global water resource challenges.

Hierarchical clustering and t-SNE were then performed on the trivariate combination of population, water use intensity, and net water balance. The results of the trivariate clustering generated groupings that were more intuitive and distinct across all three metrics and led to the identification of six meaningful groups. The results of the trivariate clustering were presented in Section 3.8 and discussed in Section 3.8.5. These were recast as a typology consisting of:

Type 1 Large cities with low w_N, and a range of climatic conditions

Type 2 Medium-sized cities with medium-low w_N, and mid-level q_{Net}

Type 3 Small cities with medium w_N and medium-high–high q_{Net}

Type 4 Large cities with high w_N, medium-high q_{Net}

Type 5 Arid cities: cities with a range of sizes, high w_N, and very low q_{Net}

Type 6 Small cities with high w_N and medium-high q_{Net}

While some of the groups identified by trivariate clustering overlapped on one or two metrics, no groups overlapped on all three metrics—in contrast to the univariate and bivariate clustering. Importantly, while the groups were intuitive *post facto*, they would have been difficult to identify *a priori* and were therefore surprising and intriguing.

Since the clustering depends on relative similarities and differences between cities, the results are sensitive to the set of cities included in the analysis. Expanding the set of cities included would likely lead to different clustering results—perhaps the thresholds between groups would change, or sub-groups within this typology might become more resolved. For this reason, the typology developed here is a creative exercise, rather than a scientific conclusion. Instead, one contribution of this work was the demonstration of a viable approach to structuring the diversity inherent in urban water systems around the world.

3.10 Future Work

Uncovering more nuanced predictive relationships between water supply and demand metrics and cities would require further data and more in-depth analysis at small-N, medium-N, and large-*n* scales of comparative analysis.

High-priority extensions of this analysis would be to expand the set of cities and integrate additional metrics into the analysis. Additional metrics could include a water quality metric, a disaggregation of water use into components (such as municipal/industrial), and/or metrics associated with more socioeconomic aspects of water management (such as GDP, HDI, or average household income). Additionally, data on the financial, energy, material, and land use intensity of urban water supply could provide insight into the urban water-land-energy nexus. The dataset could also be augmented by the inclusion of performance indicators such as leakage rates and cost recovery included in databases such as IBNET. This would allow for a more refined approach to identifying types and target areas to enhance the sustainability of urban water management.

Another attractive step would be to include spatial and temporal variation in the underlying data. For instance, the climate metric could be disaggregated to a monthly water budget. This would allow for distinctions to be made between cities with intra-annual variation in their climatic water budgets since this variation can have important implications for water management. Expanding the dataset to include multiple years of data for cities might allow for identification of types based on variability and trends—for instance, the distinction of a city that was decreasing in water use (w_N).

Just as important as identifying groups of cities with similar resource consumption profiles, it is would also be desirable to examine whether environmental impacts were correlated with metrics for human development, including economic and health indicators. The reason for this is that the ultimate goal for all cities is generally to ensure (at least in theory) a good quality of life for its citizens, with the added caveat that this be achieved within sustainable levels of resource consumption.

3.11 Conclusions

These results provide an initial answer to the question—how similar and how different are cities in terms of their water resource supply and usage patterns?

This work provides a context for assessing water management policies and challenges in a more meaningful way. Both expected and surprising pairings and groupings of cities were identified; these could be explored through further comparative analysis in small-*n* or medium-N studies. For instance, the typology identified unanticipated similarities, most notably among cases of cities that consumed far more than their local supply, even when those supplies varied from arid to humid.

This work is only a first step in characterizing urban water supply and demand patterns around the world. Expanding the set of cities can help better characterize the global spectrum of urban water profiles and be used to test the stability of this typology (i.e., its internal validity). Breaking down the indicators used to represent urban water profiles may resolve or highlight other areas of similarity or difference. For instance, water use intensity can be broken down into contributions from industry versus households, which might distinguish cities with efficient water use from those with a shortfall in infrastructure (which affects w_N). Climate can be better represented by adding other components of water balance, and the clustering analysis might consider monthly variation, rather than aggregating to the annual scale.

The results demonstrated that statistical clustering is a useful method for developing a quantitative basis for small-*n* and large-*n* comparative urban water management case study research. The typology presented here is a significant contribution to this effort, but it is still only a start. It will benefit from future case study research, standardization of data, expansion of the metrics to consider temporal and spatial variation, disaggregation of water consumption and net annual water balance, and the inclusion of more cities.

The next chapter, Chapter 4, demonstrates how the typology and WUCI can be used in the meaningful selection of case studies.

Chapter 4

Small-*n* Comparative Analysis of Historical Data Los Angeles and Singapore

4.1 Introduction

In Chapter 3, a new approach was proposed for identifying a global typology of cities based on three commonly available metrics of water demand and supply—population (N), water use intensity (w_N or WUI, in $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$), and net climatic water balance (q_{Net} or AB, in $\mathrm{m} \cdot \mathrm{yr}^{-1}$). Widely used statistical clustering methods were applied to a large dataset of cities from around the world. In spite of the diversity exhibited by the cities across all three variables, at least one intriguing typology was identified using that approach. Chapter 3 also introduced a new metric for local water availability—the Water Use and Climate Index (i_{UC} or WUCI, in $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$)—was introduced to facilitate intuition and interpretation of the adopted typology. This new metric indexed WUI for an urban population to natural water availability (annual precipitation height or q_P, in $\mathrm{m} \cdot \mathrm{yr}^{-1}$). Cities with high a WUCI or total WUCI ($I_{UC} = N * i_{UC}$) will likely be less able to meet urban water demand from local water resources. While intriguing, the results from Chapter 3 painted the picture of urban water supply and demand in strokes that were too broad to make specific recommendations for any type, let alone for any specific case. The data did not take into account local surface water resources, climate variability, or changing use patterns.

In this chapter, the typology from Chapter 3 is used to guide the choice of two cities for a more in-depth analysis of urban water supply and demand profiles. As in Chapter 3, the analysis in Chapter 4 focuses on simple metrics for urban profiles of supply and demand: population, WUI, and AB; supporting data on annual precipitation; and the derived indicator WUCI. In Chapter 4, annual values for these data are assembled for each case over a longer period of time (1960–2016) and obtained from common datasets, where possible. In addition to these metrics, Chapter 4 also includes data on city area (A_N in $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$), which is used later in the analysis in calculating the potential self-sufficiency ratio, R_{SS}.

4.1.1 Objectives

The primary aim of the case studies is to compare the past and prospects for self-sufficiency of urban water management by two cities to gain insight into the water resource challenges for cities of a similar type; a secondary aim was validating the typology developed in Chapter 3 as well as its utility in framing the selection of case studies in a rigorous way, i.e.:

1. use the typology as a framework for case study selection and
2. analyze urban water availability in greater depth for two cases.

The case-specific objectives of performing in-depth case study analysis are, for each case:

1. to characterize historical water demand and supply;
2. to evaluate the potential for (and constraints to) self-sufficiency from local rainfall;
3. to identify future pathways to greater self-sufficiency and make recommendations for how to balance water supply with demand management.

As was the case in developing the method for typological identification in Chapter 3, methods for analysis developed in Chapter 4 attempt to prioritize the following properties:

1. Developability: Able to accommodate new data points in a time series;

2. Scalability: In theory, able to accommodate growth or decline in key variables;

3. Robustness: Able to handle missing values;

4. Comparability: Allow direct comparison with a different urban system.

4.1.2 Overview

The case study analysis is divided between Chapters 4 and 5: Chapter 4:

1. Case study selection from typology
2. Assessment and comparison of historical self-sufficiency

Chapter 5:

3. Assessment and comparison of simulated self-sufficiency
4. Recommendations for pathways to greater self-sufficiency

Sections 4.2 and 4.3 describes the selection of two cases from the UrbMet database using as a framework the typology developed in Chapter 3. For this analysis, Los Angeles and Singapore are chosen to represent Type 4 cities with different climates. However, the approach could be used with a different set of cities and a different set of requirements for case selection.

Section 4.4 provides an overview of urban water supply management for each of the two cases and defines the system boundary.

Chapter 4.5 describes the data and approach used to characterizing historical water demand and supply and evaluating the potential for (and constraints to) self-sufficiency from local rainfall. Results for each case are presented in Chapter 4.8.

4.2 Selection of a City Type

4.2.1 Logic

It is first necessary to establish a sound logic as a basis for choosing two case studies for comparison. The first step towards this end was the identification of key metrics for urban water demand and supply (as defined in Chapter 3.3.1). Next, a dataset of these metrics was assembled for a large number of international cities (as described in Chapter 3.3.2). With a dataset exhibiting a wide variation across three independent variables (as shown in Chapter 3.4), statistical clustering methods were used to facilitate identification of similar subsets of cities (as described in Chapter 3.5).

Ideally, I desired two cases that with significant water resource challenges. I assumed that such cities were more likely to have large populations, high WUCI (i_{UC}), and high total WUCI (i_{UC}). While it would also have been interesting to look at a rapidly developing city and an affluent city, I decided to choose focus on cities with an average water use intensity that surpassed the 36.5 $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$ recommended by the WHO as a minimum level of water consumption for human development.

In summary, I wanted two cases with:

1. High WUCI i_{UC} or total WUCI i_{UC}
2. Large population N
3. High WUI ($w_N \geq 36.5 m^3 \cdot cap^{-1} \cdot yr^{-1}$)

I used the typology developed in Chapter 3 as the basis for choosing two case studies for comparison. Recall that six different types comprise the typology:

Type 1 Large N, low w_N

Type 2 Medium N, medium w_N, q_{Net} 0.0

Type 3 Small N, medium w_N, high q_{Net}

Type 4 Large N, high w_N, positive q_{Net}

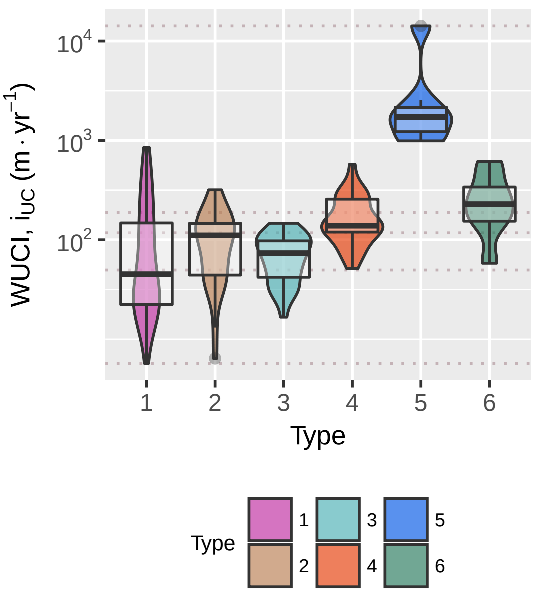
Type 5 High w_N, highly arid q_{Net}

Type 6 Medium N, high w_N, negative q_{Net}

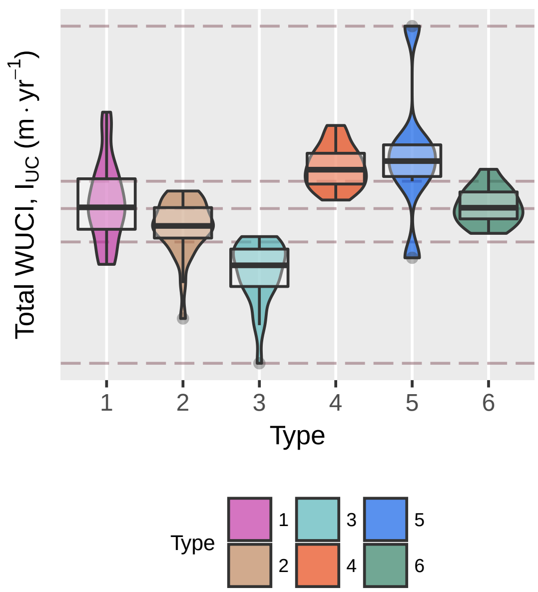
4.2.2 Results

Results for WUCI and total WUCI for each cluster type were presented in Chapter 4 and are shown again in Figure 4.1. Comparing per capita WUCI across clusters, Types 5, 6, and 4 had the highest per capita WUCI (i_{UC}) levels (in descending order) of the six types. With respect to total WUCI (I_{UC}), Types 5, 4, 6, and 1 had the highest levels of the six types, in descending order. Looking then to the independent variables, Types 1 and 4 had the largest populations (descending order), while Types 6, 5, and 4 had the highest w_N (in descending order) (see Figures 3.9a–3.17c). Since only Type 4 had representation across the types with the desired attributes, I decided to choose two cases from Type 4.

Figure 4.1: Review of violin- and box-plot of per capita WUCI and total WUCI, by type (from Chapter 3).

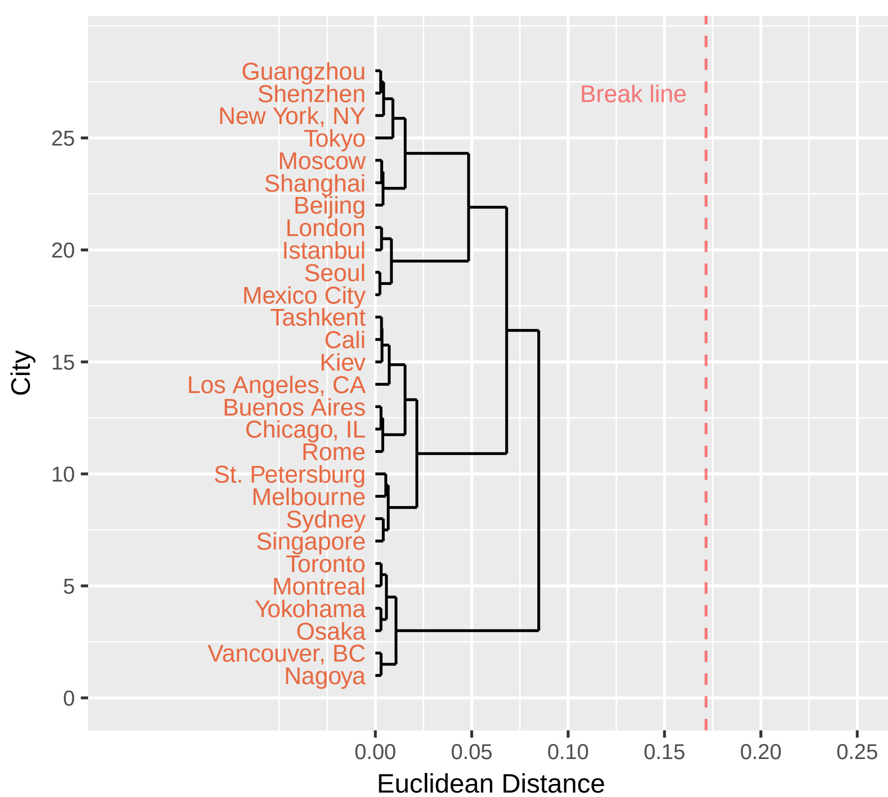


(a) WUCI (i_{UC} in m$^2 \cdot$ cap^{-1}), by type.



(b) Total WUCI ($I_{UC} = N * i_{UC}$ in km^2), by type.

(c) Close up of dendrogram for Type 4 cities.



4.3 Case Study Selection

According to the results of the trivariate clustering, the Type 4 cities were more similar to each other—in terms of their water supply and demand profiles—than to most of the other cities around the world[1]. The null hypothesis was that this similarity implied that the water resource challenges that any two of these cities were facing would likely also be similar. To test the internal validity of the typology, it was necessary to pursue a more detailed comparison of two Type 4 cities that differed in some notable way.

Of the three independent variables—population N, WUI w_N, and water balance q_{net}—Type 4 cities exhibited the greatest variation in q_{net}, spanning all four quartiles (see Figure 3.9d). From Type 4 cities—listed in Figure 4.1c—it was decided to choose two cases with relatively similar levels of N and w_N but differed in climatic water availability, with respect to considering both water balance q_{net} and annual precipitation q_P.

For Type 4, minimum and maximum values for annual precipitation (q_P) were Los Angeles, with 0.273 $\mathrm{m \cdot yr^{-1}}$and Singapore, with 2.621 $\mathrm{m \cdot yr^{-1}}$—a difference of 2.358 $\mathrm{m \cdot yr^{-1}}$. Similarly, net water balance also exhibited a large range over Type 4 cities; while the general trend was that Type 4 cities had a positive net water balance, this was not true for a small fraction of Type 4 member cities—a little over one-fifth of Type 4 cities (6 out of 28) were found to have a negative average net annual water balance[2]. A difference of 0.702 $\mathrm{m \cdot yr^{-1}}$was observed between the minimum and maximum values of q_{Net} within Type 4: -0.577 $\mathrm{m \cdot yr^{-1}}$for Los Angeles and 1.279 $\mathrm{m \cdot yr^{-1}}$in Tokyo, respectively.

Of the Type 4 cities that were at opposite ends of the water availability spectrum, Singapore and Los Angeles were chosen as the two cases for further analysis.

4.4 Overview of Cases

4.4.1 Introduction

The city and metropolitan region of Los Angeles (LA) is located on the southern coast of California, on the west coast of the U.S., at a $Lat/Long$ of $34.05°N, 118.25°W$. Since the 1990 census, Los Angeles has been ranked the second-largest city and metropolitan in the U.S.[3] [324]. The population of LA city was estimated at 3.97 million in 2015, while the county and metropolitan populations were estimated at 10,241,335 and 13,131,431 in 2013, respectively [325, 323].

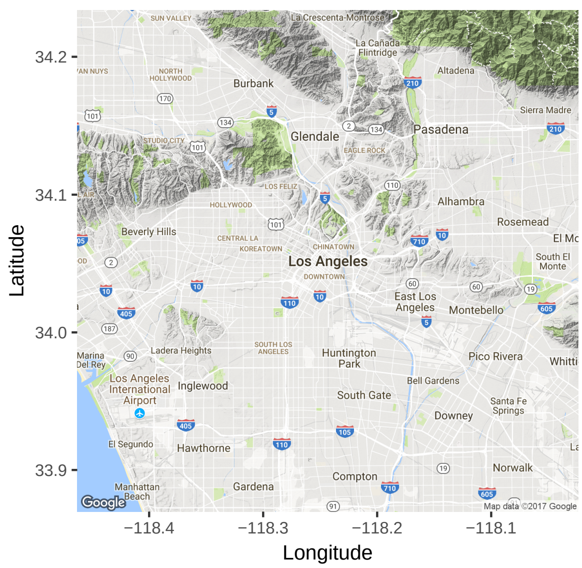
Singapore is an urban nation located on a small archipelago off the coast of Malaysia in the Malacca Straight, at a $Lat/Long$ of $1.28°N, 103.84°E$. The population of Singapore was estimated at 5.61 million in 2015 [292].

[1] At least with respect to those that were included in the clustering analysis
[2] In order of increasing water balance, these cities were: Los Angeles, Cali, Beijing, Tashkent, Istanbul, and Kiev.
[3] New York City is ranked first [324].

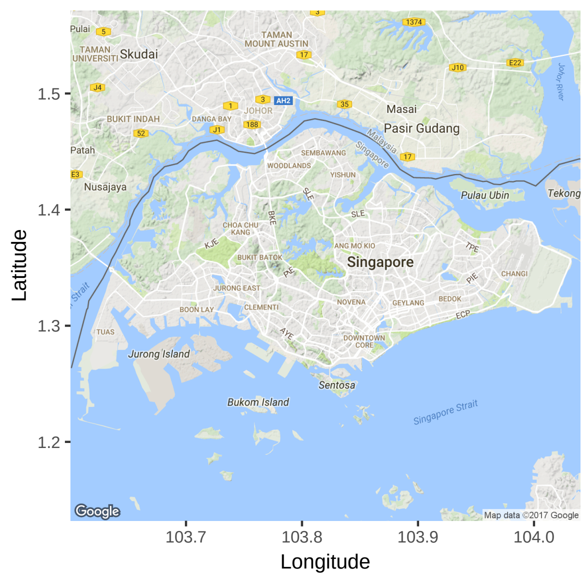
Figure 4.2: Overview of case studies

Maps
Source: Google Maps, *zoom* $= 11$ (produced with the `ggmap` package in `R`).

(a) Los Angeles



(b) Singapore



Overview of cases

Metric	Unit	Case: Los Angeles	Case: Singapore
Lat/Long	degrees	$34.05°N, 118.25°W$	$1.28°N, 103.83°E$
Population	capita	3.97×10^6	5.61×10^6
City area	km^2	1.303×10^3	0.7191×10^3
Population density	$cap \cdot km^{-2}$	3.05×10^3	8.22×10^3
Water use intensity	$m^3 \cdot yr^{-1} \cdot cap^{-1}$	169	76.7
Annual precipitation height	$m \cdot yr^{-1}$	0.48	2.45
WUCI	$m^2 \cdot cap^{-1}$	352	31.3
Total WUCI	km^2	1.40×10^3	0.176×10^3
R_{SS}	unitless	0.932	4.09
R_{SS}^{-1}	dimensionless	1.07	0.244

After a cursory look at the water supply and demand profiles for Los Angeles and Singapore, it was not clear to what extent water resource challenges—and the portfolios of technologies and policies used to meet them—would be similar between the two cases. A look at the average monthly water balance for the two cases was first used to ground expectations about WRM for the two cases, followed by a consideration of additional factors, such as population, water use intensity, and city area.

Figure 4.3: Average monthly temperature profiles for Los Angeles and Singapore (from WebWIMP [191]). The temperature profiles were relatively similar between the two cases throughout the year.

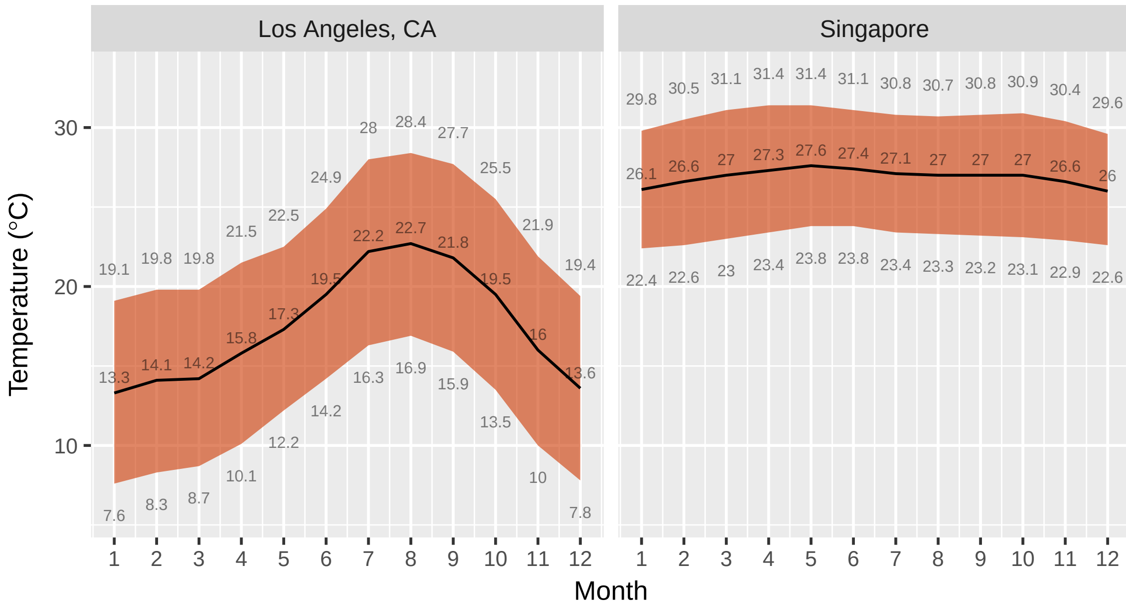
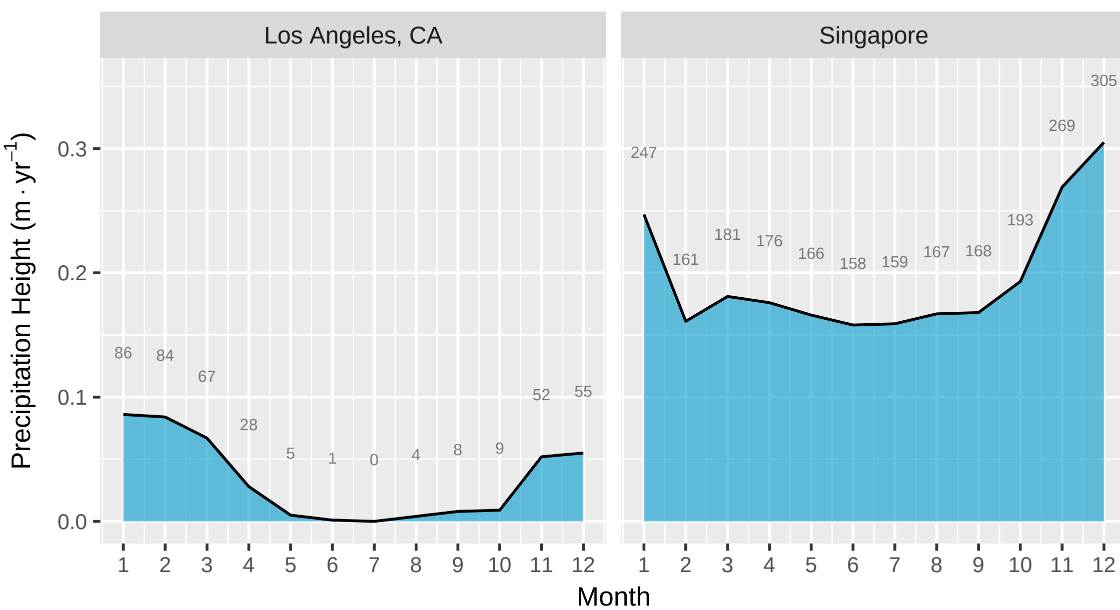


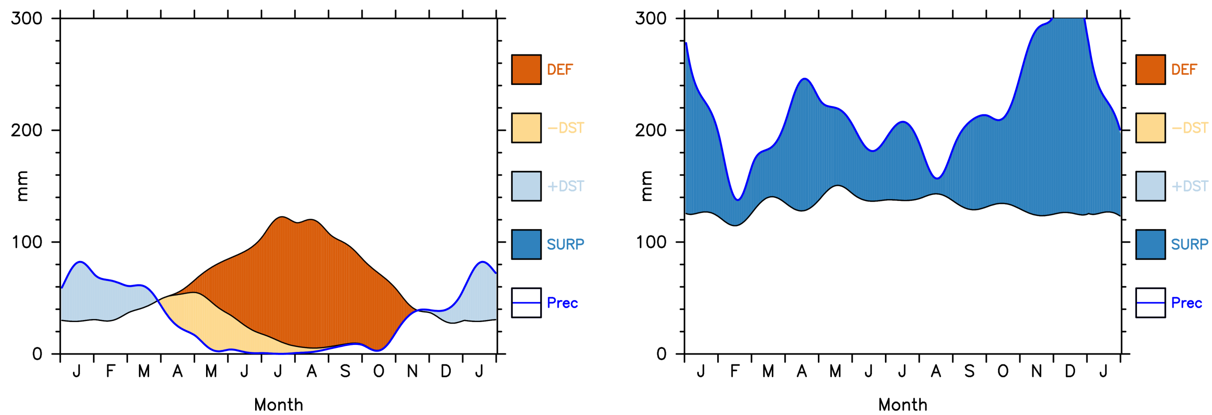
Figure 4.4: Average monthly precipitation profiles for Los Angeles and Singapore (from WebWIMP [191]). While the temperature profiles between the two cases (shown in Figure 4.3) were relatively similar, a notable difference can be observed in precipitation—Singapore has much higher precipitation.



While the average monthly temperatures of the two cases were relatively similar throughout the year, as seen in Figure 4.3, a substantial difference was observed in precipitation profiles, shown in Figure 4.4. In light of that disparity, it was not surprising that the two cases also dif-

Figure 4.5: Average climatic water availability and components for Los Angeles and Singapore (from WebWIMP [191]). Singapore has a positive average monthly surplus throughout the year (though larger is some months than others), as seen in Figure 4.5b. In contrast, Los Angeles experiences a deficit during 7 months of the year, during which storage of water as soil moisture is depleted; during the remaining 4 months of the year, surplus is minimal or non-existent since precipitation goes towards recharging storage that was depleted during the drier period.

(a) Los Angeles (b) Singapore



fered in average monthly water balance components, as shown in Figure 4.5[4] [191]. Singapore has a positive average monthly surplus throughout the year (though larger is some months than others), as seen in Figure 4.5b. In contrast, Los Angeles experiences a deficit during 7 months of the year, during which storage of water as soil moisture is depleted; during the remaining 4 months of the year, surplus is minimal or non-existent since precipitation goes towards recharging storage that was depleted during the drier period.

Water use intensity was relatively similar between the two cases, with $w_N = 156\ \mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$ in Los Angeles and slightly lower in Singapore at $136\ \mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$.

If only monthly water balance was considered, it would be reasonable to expect that water resource challenges would be much more acute in Los Angeles and Singapore would have few issues with water availability. Even with the inclusion of water use intensity, which were similar between the two cases, the differences in climatic water availability are great enough that Singapore would be much more likely to be more self-sufficient that LA, as there would be a greater likelihood that Singapore would be able to obtain a greater quantity of freshwater closer to home. This expectation would be even more reasonable if the WUCI and total WUCI were calculated for the two cases. WUCI was found to be $i_{UC} = 463\ \mathrm{m}^2 \cdot \mathrm{cap}^{-1}$ in Los Angeles, nearly nine times larger than Singapore ($i_{UC} = 51.5\ \mathrm{m}^2 \cdot \mathrm{cap}^{-1}$). Although the population in Los Angeles ($N = 4.03 \cdot 10^6$) was 28% smaller than that of Singapore ($N = 5.61 \cdot 10^6$), its total WUCI ($I_{UC} = 2510\ \mathrm{km}^2$) was also approximately nine times as large as Singapore ($I_{UC} = 289\ \mathrm{km}^2$).

[4]The source for these figures of monthly water balance components was *WebWIMP: The Web-based, Water-Budget, Interactive, Modeling Program*, which was also the source for the data on climatic water availability used in clustering analysis in Chapter 3 [191].

Taking population into account makes it even more likely that Singapore would be closer to self-sufficiency than LA with respect to local natural water resources. However, considered in a slightly different light, at I_{UC} = 2410 km^2, the total WUCI (a type of water footprint) in LA is reasonably large. In some area of the world—Europe, for instance, this water footprint is on the same order as the size of nations. For example, I_{UC} in LA is nearly as large as the country of Luxembourg, which has an area of 2588 km^2[5]. In other words, self-sufficiency might also be an important consideration for LA, since the magnitude of its total WUCI implies that LA likely has to obtain a substantial fraction of its water from areas outside of its borders—and therefore subject to other political jurisdictions.

The self-sufficiency ratio was then calculated for both cities, using Equation 4.14 and data on city area, population, average annual precipitation, and water use intensity. The resulting ratio is, therefore, an upper bound of a *potential for self-sufficiency*, since cities cannot capture 100% of precipitation. Results for this calculation are shown in Table 4.2. These initial estimates of the potential self-sufficiency ratio was found to be nearly 5 times higher in Singapore, with $R_{SS} = 0.52$ for Los Angeles $R_{SS} = 2.49 > 1$ for Singapore.

Taking the inverse of R_{SS} provides an estimate of the fraction of urban area from which water would need to be collected to match water consumption. For Singapore, this was found to be only 0.402, while for LA this was found to be 1.92; in other words, if 100% of precipitation were collected, Singapore could meet its current water needs within less than half of its area, while LA would require an area nearly twice its current size.

While some evidence from the two cases was aligned with the results of this back-of-the-envelope analysis, this initial comparison was—unsurprisingly—unable to fully capture the water resource challenges faced by the cities in question. In practice, Singapore has been collecting stormwater runoff from over two-thirds of its area, with plans to expand this collection to 100% of the main island to accommodate population growth [305, 262]. LA sources its water supply from an area over 22,000 km^2—over fifteen times the city size—which could be considered aligned with the back-of-the-envelope results. However, this rough comparison did not explain why Los Angeles—with its low climatic water availability and a net annual climatic water deficit—only reuses 2% of its wastewater (primarily through groundwater recharge), while Singapore—with its high precipitation and net annual water surplus—has become a global leader in treating seawater and wastewater for water supply, with the reported capacity to meet 65% of its water demand through desalination and wastewater reuse [171, 169, 262].

In Chapter 4.4.2, an overview of water resources management and geographic, political, and socioeconomic considerations provided further insight into contemporary challenges for Los Angeles and Singapore, respectively.

[5] See `https://en.wikipedia.org/wiki/Luxembourg`.

4.4.2 Water Resources Management

Both Los Angeles and Singapore are each covered by a single water utility overseeing both water supply and distribution. These entities—respectively, the Los Angeles Department of Water and Power (LADWP) and the Singapore's water agency, the Public Utilities Board (PUB)—each published long-term water management plans in 2016, and each of these plans stated intended goals to increase the proportion of their water supply met by "local" sources, both of which clearly distinguished between "local"and imported water resources [171, 261].

- In LADWP's long-term plan (entitled *Urban Water Management Plan: 2015* but published in 2015) the stated goal was to increase local production capacity to meet 55.7% of demand in 2040 with local sources (compared to 38% in 2014) [171].

- In PUB's 2016 long-term water plan, *Our Water, Our Future*, Singapore's stated target was increasing local production capacity to meet 100% of water demand with local sources by 2060 [261].

As an island located in a saltwater sea, Singapore's administrative boundary conveniently aligned with local hydrologic units, and its definition of local sources—local catchments ("Reservoirs"), reused/recycled water ("NEWater"), and desalinated water—paired exactly with Equation 4.2's Q_R, Q_{WWR}, and Q_{DS}.

In contrast to Singapore, Los Angeles is not an island, and its administrative boundaries do not align with local hydrologic units. However, *Urban Water Management Plan: 2015* (UWMP 2015) did clearly distinguish between water imports and local water resources [171]. Moreover, these sources were clearly defined within the plan and analysis. Los Angeles had two imported sources, the Metropolitan Water District of Southern California (MWDSC) and LA's very own Los Angeles Aqueduct (LA Aqueduct). Compared to Equation 4.2, LADWP's definition of local sources included groundwater Q_{GW}, in addition to stormwater capture Q_R, and water reuse Q_{WWR}.

Los Angeles

The LADWP, founded in 1902, oversees water supply in LA [167]. In 2016, the LADWP served a population of over 4 million in area of 1,230 square kilometers [168].

The LA region receives just enough precipitation, on average, to avoid the Köppen climate classification as *arid* (*B*); instead, it has a designation as a *Mediterranean* climate with *hot/warm summers* (*Csa inland/Csb coast*). Most of the rainfall in LA occurs, during the winter months; throughout most of the year, LA experiences a net negative water balance. Droughts in the area are common—72 of the 117 years from 1900-2016 experienced lower than average precipitation[6]

[6] In this instance, "average" was defined according to the Western Regional Climate Center (See `http://www.wrcc.dri.edu/cgi-bin/cliMAIN.pl?ca5115`).

Adding to the water management challenge in LA is substantial variability in precipitation patterns from one year—or one decade—to the next. When averaged over 100 years, the average annual precipitation was found to be about 0.381 m/year. Interannual variability is frequently high: for instance, within a five-year period, the actual annual precipitation might range from 0.152 m/year (less than half of the average) to 1.016 m/year (more than twice the average) [140]. Adding to the burden of water resources management in LA is that interannual variability is not consistent from one year to the next. Periods of prolonged drought alternate with those of heavy precipitation. Since its incorporation in 1850, LA has experienced three lengthy droughts (1865–1885, 1928–1934, and 1987–1992) and one decade-long wet spell (1935–1944) [140]. Since 1850, the duration of these periods of extreme wet or dry weather ranged from 5 to 20 years. However, evidence from the last 2000 years indicates that the length of these cycles is unpredictable—and can be significantly longer than what LA has witnessed over the last 166 years. For instance, the two longest periods of extreme dryness lasted 220 years (892 C.E.–1112 C.E.) and 141 years (1209 C.E.–1350 C.E.); these epic droughts were separated by the longest period of extreme wetness which lasted 97 years [140]. No periodicity in these cycles has been observed.

Table 4.1: Overview of Aqueducts Serving LA. Abbreviations for managing entities are as follows: LADWP—Los Angeles Department of Water and Power, MWDSC—Metropolitan Water District of Southern California, and CASWP—California State Water Project.

Aqueduct	**Branch**	**Constructed** Years	**Length** km	**Watershed**	**Entity**
Los Angeles (LA)	1st:	1908–1913	375	Owens River	LADWP
	2nd:	1965–1970	220		
Colorado (CO)		1933–1941	389	Colorado River	MWDSC
California (CA)		1963–1973	489	Sacramento, San Joaquin Rivers	CASWP

Table 4.2: Summary of Contributions to LADWP Water Supply by Source (1960–2015).

Statistic	N	Mean	St. Dev.	Min	Max
MWDSC	46	0.307	0.234	0.032	0.746
LA Aqueduct	46	0.546	0.226	0.097	0.852
Local GW	46	0.144	0.036	0.051	0.230
Recycled	46	0.003	0.005	0.000	0.019

Low precipitation and high variability have caused LA to supplement its water supply with sources imported from watersheds far from the metropolitan region itself. LA's first imported water source was constructed in 1913—the $375km$-long Los Angeles Aqueduct, which sources water from the Sierra Nevadas, typically serves as the LADWP's primary water source. Over 1970–2014, contributions to the LADWP's water supplies from the LA Aqueduct ranged from 10.3%–85.3% and averaged 55.6% ($\sigma = 21.8\%$) (see Table 4.2).

However, significant uncertainty exists in the annual water supply from LA Aqueduct, due to significant variability in the climate throughout the region. The LADWP, therefore, found it prudent to pursue additional, supplemental water sources from other regions to increase the reliability of its supplies and to accommodate additional growth in demand. By the early 1920s (less than a decade after the completion of the LA Aqueduct), LA had begun investigating additional sources of water, In 1921, LA supported an early version of the State Water Project in the CA legislature (which did not pass; the so-called Marshall Plan would have brought an additional four times the volume of the LA Aqueduct at the time [140]. In 1922, LA supported the Colorado River Compact, which allocated water from the Colorado River between seven states and began to pave the way for the Colorado River Aqueduct [140, 321]. Even as Arizona stirred up opposition to the newly-signed bill, delaying further formal action until 1928 (with the passing of the Boulder Canyon Bill), engineers affiliated with LA were sent out to begin surveying in 1923 [170].

In 1941, the 389km-long Colorado River Aqueduct was completed. However, the 489km-long California Aqueduct of the CA State Water Project was not finished until 1972. In the meantime, concerns over water supply prompted LA to pursue an expansion of the LA Aqueduct on the Owens River, the second LA Aqueduct, which was completed in 1970—two years before the California Aqueduct.

The LADWP has access to water from Colorado River and California Aqueducts through the MWDSC [140, 224]. The MWDSC has typically been LA's second-most important water source and becomes a primary source during dry years. Since 1970, water from the MWDSC has ranged from 3.4%–74.6% of the LADWP's supply ($\mu = 29.8\%$, $\sigma = 22.9$).

Groundwater pumping has been the LADWP's third most important source; between 1960–2015 groundwater averaged 14.3% of annual supply ($\sigma = 3.4\%$). And, since being introduced in 1993, recycled water has made a small but increasing contribution to LA's water supply—in 1993, recycled water made up less than 0.1% (.03%), but by 2014 this had risen to 1.7%. Since 1970, water supplies from the MWDSC (which also includes water that the MWDSC purchases from the California State Water Project (CASWP)) averaged 29.8% of LA's water, 55.6% from the LA Aqueduct, 14.3% from local groundwater pumping, and 0.3% from recycled wastewater (See Table 4.2).

The MWDSC was created in 1928 by the California state Legislature to build and operate the Colorado River Aqueduct, which draws water from the Colorado River. The State of California was allotted 4.4 million acre-feet (maf)—58.70% of 7.50 maf—of water from the Colorado River by the Upper Colorado River Basin Compact of 1948 [13]. Within California (CA), the MWDSC was allotted 12.5% (0.55 maf) of the 4.4 maf. The LADWP has preferential rights to 21.97% of the MWDSC's water supply; with respect to water from the Colorado River alone, this would amount to 0.1153 maf[7] [54].

[7] The preferential rights scheme is governed by Section 135 of the MWD Act [54].

Development, climate variability, and politics have increased uncertainty in the LADWP's water supply from the MWDSC. The MWDSC is the largest supplier of treated water in the US and serves 26 members: 14 city-level water utilities (including the LADWP), 12 municipal water districts, and one county water authority [225]. These entities have a combined area of 13,000 km^2 and population of 19 million. Rapid growth in the MWDSC's other members has led to increased—and unwelcome to LA—competition for water supplies.

In addition, the MWDSC itself has increasingly had to compete for water from the California and Colorado Aqueducts. In CA, water from the Colorado was divided amongst four water and irrigation districts; of these four districts, the MWDSC was allotted the smallest quantity and the lowest priority. Therefore, when water supplies from the Colorado are lower than usual, the MWDSC is one of the first districts affected. For instance, for many years Arizona and Nevada took less water than their apportionments, which allowed California to regularly used more than its allotment. However, this changed as the Central Arizona Project began its operation. In 2002, the US Supreme Court upheld the law of the original contract [93]. When the Secretary of the Interior began enforcing that ruling on January 1, 2003, the flow through the Colorado Aqueduct dropped from full capacity operation at 1.2 maf to 0.5 maf [93]. Since 2002, the region experienced numerous dry spells, pressuring the MWDSC to investigate alternative water sources. This has led to deals with the irrigation districts, purchases of water from the State Water Project, and expansion of water recycling and desalination [223].

During periods of "normal" precipitation, pressure from population growth and economic development has been only moderate. However, during dry periods, the pressure for change has often led to policy changes. The most recent drought, beginning in 2011, saw the implementation of numerous emergency measures [50]. While recent rainfall and an increase in regional snowpack in the winter of 2016–2017 may have ameliorated local water scarcity, long-term water management plans created during this period have given greater emphasis to stormwater collection, wastewater recycling, and desalination [171, 223].

For the LADWP, these policies were:

1. expand water conservation,
2. expand water recycling,
3. enhance stormwater capture, and
4. clean up the groundwater basin.

Singapore

Since gaining independence from Great Britain in 1963, Singapore has rapidly transitioned from an outpost of the British Empire into a dominant economic player in Asia and the world. This transition has seen a three-fold increase in population (from 1.64 million people in 1960 to 4.99 million in 2010) and an increase in GDP from 6.59 billion Singapore dollars (SgD) to 251.4 billion

SgD. Singaporean water resources management is centralized in the Ministry of the Environment and Water Resources, which oversees the national water utility, the PUB. Established in 1960 (three years before independence from Great Britain) the PUB is responsible for coordinating Singapore's water provision and management planning [305]. PUB policies and plans incorporate projections for growth in population, economic activity, and water infrastructure; this growth is forecasted several decades into the future [263]. Long-standing concerns about water security have led to Singaporeans prioritizing development of a self-sufficient water supply.

A cursory look at Singapore's climate and annual precipitation suggests abundant natural water supply. Singapore's climate has a designation of *tropical rainforest* (*AF*) in the Köppen Climate Classification [249]. Total annual precipitation in Singapore averaged at $2.34\ \mathrm{m} \cdot \mathrm{yr}^{-1}$over the period from 1960–2010—at least twice the global average—contributes to a positive net water balance throughout the year [306, 191, 190]. However, in spite of abundant precipitation, Singapore is often listed among the world's most water-scarce countries—a ranking it shares with notoriously arid countries like the United Arab Emirates [101]. This apparent conundrum is resolved when one considers Singapore's large population and high water demand in light of limited natural freshwater supply—for which limited storage volume is the primary constraint. Singapore has no natural freshwater aquifer, and land for water storage is limited on the island of roughly 719 km^2, which housed a population of over 5.6 million in 2015 [292].

Due to these natural constraints, Singapore has historically obtained the greatest fraction of its potable water from an international transfer from neighboring state of Johor in Malaysia. Historically, Singapore's water imports have been protected by two treaties with Malaysia, signed in 1961 and 1962, and ensuring water transfers to Singapore at a price of less than 1 cent per 1000 gallons[8] [305]. However, the first treaty expired in 2012, the second will expire in 2060, and thus far the two countries have been unable to negotiate a replacement.

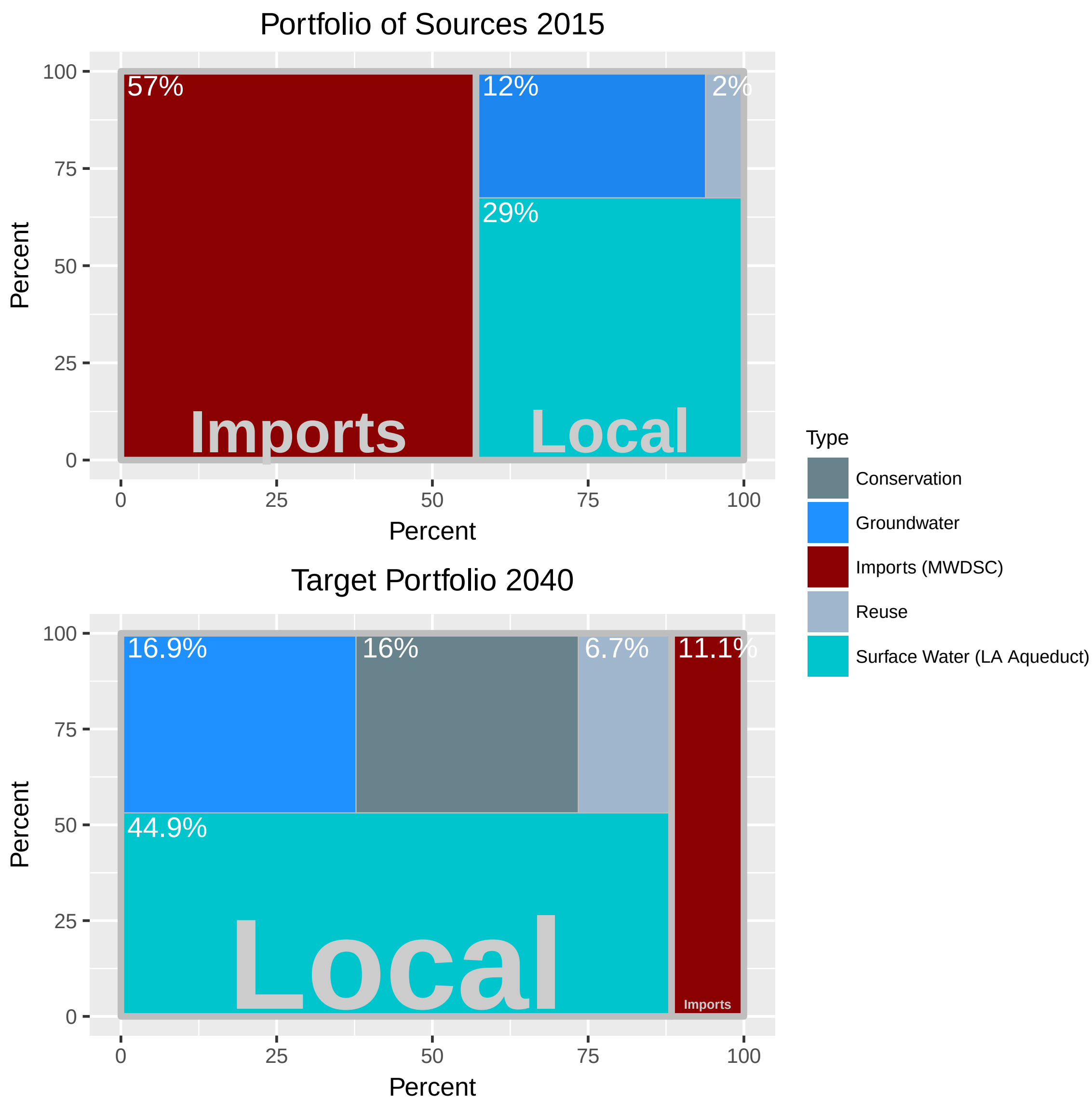
Even in the early 1960s, when the two treaties were signed, Singapore's government was aware of the potential difficulty in extending the arrangement and were also concerned that Malaysia could break off the agreement at any time. Since that time, expansion of on-island (i.e., in-city) production capacity has been a priority for the Singaporean government. Where possible, watersheds of key reservoirs were protected from development, while new reservoirs were built to expand the fraction of island area from which water was collected. Land use policies were adopted to improve quality of water collected from developed areas by preserving a relatively high ratio of vegetated to built area—in spite of a growing population and density. In addition to stormwater runoff, Singapore was an early adopter of desalination and water reuse. In 2012, Singapore reported the capacity to meet 20% of its demand from its own reservoirs, 30% from water reclamation, and 10% from desalination, and has stated the ambitious target of meeting 100% of water demand from on-island sources by 2060—the expiration date of the remaining water treaty.

[8] 1000 gallons = 3.7841 m^3

4.4.3 Recent Supply Portfolios and Targets

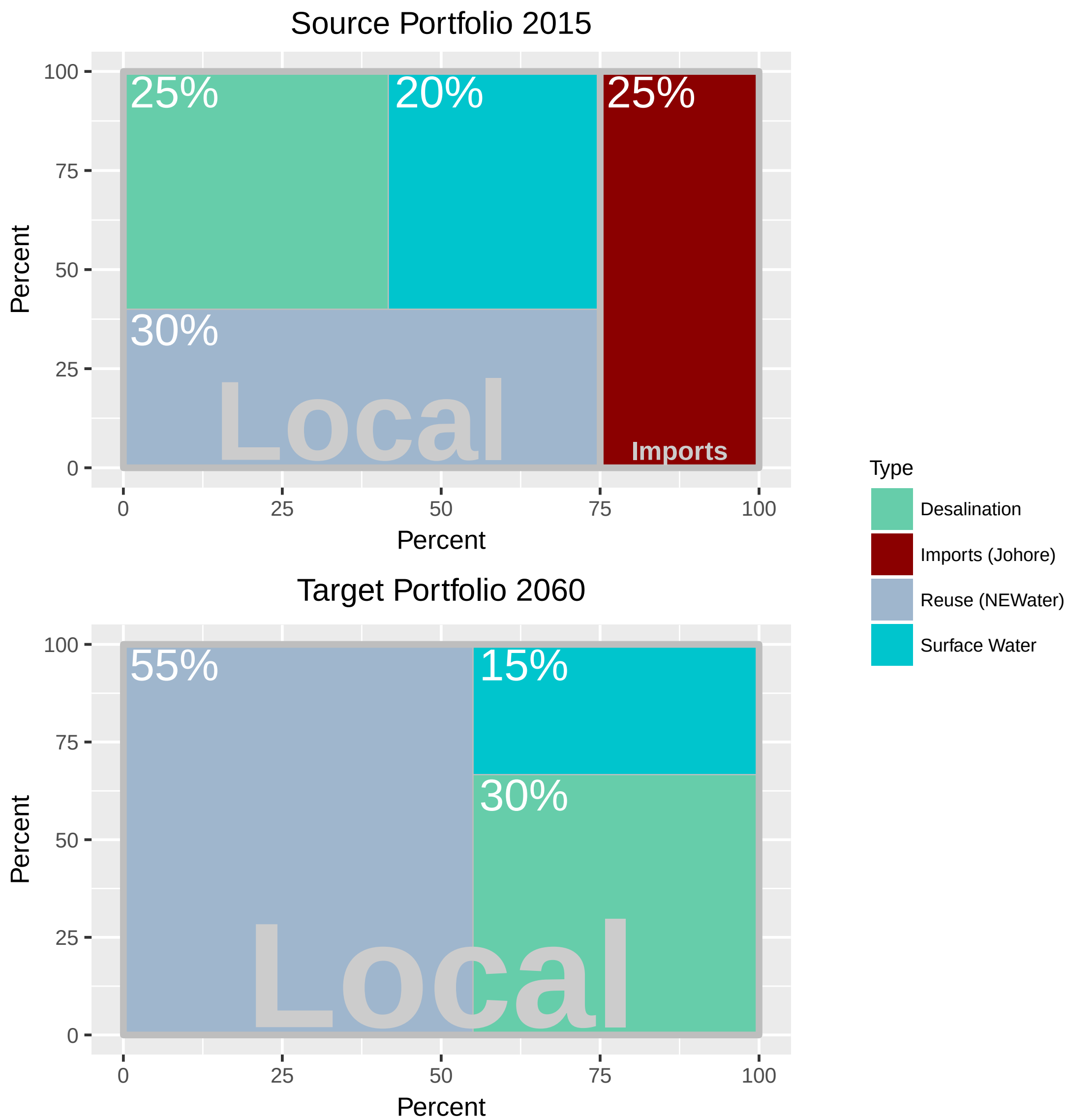
and Supply Portfolios

Figure 4.6: Water supply portfolios for Los Angeles: 2015 versus 2040. Information about target percentages was taken from the LADWP's *Urban Water Management Plan: 2015* (2016) [171]). The primary surface water source for LA is the LA Aqueduct, which transports water from over 100 km away. LA imports water from northern California and from the Colorado River basin through the MWDSC. LA reuses some wastewater through groundwater recharge. LA's water management plan for 2040 also included conservation—the contribution of conservation to total supply is relative to estimates of water demand without such efforts.



Figures 4.6 and 4.7 show the portfolio of sources for water supply in 2015 and the target source portfolios for Los Angeles and Singapore, respectively.

Figure 4.7: Water supply portfolios for Singapore: 2015 versus 2060. Information about target percentages was taken from PUB's *Our Water, Our Future* (2016) [261]. *Johore* refers to water imports from the neighboring state of Johore in Malaysia, while *NEWater* refers to treated wastewater. In contrast to LA, Singapore's future water management plan did not distinguish conservation as contributing to its supply; however, Singapore does have conservation policies.



In 2015, Los Angeles obtained 57% of its water from water imports (sources outside of the service area of the LADWP, excluding supply from the Los Angeles Aqueduct). A further 29% of its supply was imported through the Los Angeles Aqueduct, which is owned and operated by the LADWP, with 12% obtained from local groundwater and 2% from water reuse [171]. By 2040, the LADWP has started targets to meet 42.4% of its projected demand with water sourced from the LA Aqueduct, 11.1% from water imports from the MWDSC, 16.0% from conservation, 16.9% from groundwater pumping, and 13.6% from wastewater reuse and stormwater collection [169, 171]. As of 2016, the LADWP had no intention to introduce desalination Q_{DS} as a potential source[9] [171].

In 2015, Singapore reported the capacity to meet 30% of its water needs through treatment of wastewater (which Singapore's PUB refers to as NEWater), 25% through desalination, and 25% through the collection of stormwater runoff, with the remaining 25% met with imports from the neighboring Malaysian state of Johor [261]. By 2060, Singapore plans to meet as much as 55% of its needs from reclaimed wastewater, 30% from desalination, with the remaining fraction sourced from stormwater capture [261].

4.5 Approach to Assessment and Comparison of Historical Self-Sufficiency

The typology from which case studies were selected (4.2) was developed from profiles of urban water demand and natural climate water availability (Chapter 3). Similarly, case study analysis focused on the potential contribution of stormwater runoff (Q_R in Equation 4.2) to urban self-sufficiency.

Historical data on water demand, water supply, and supportive variables were collected. Data sources and data post-processing is described in 4.6. Water supply and distribution management for both cities were each overseen by a single water utility, which provided data on total water use W_N and water use intensity w_N. Data on population and city area were obtained from local and national data sources. Local climate conditions, including precipitation and net water balance, were obtained from several sources, including WebWIMP, NOAA, and the World Bank [190, 230, 355].

Historical data were then viewed and analyzed. Time series for N and k_N, W_N and w_N, and q_P and A_N were plotted. Descriptive statistics (statistical moments and quantiles) were assessed for "normalized" variables—

Water use intensity w_N (in $\mathrm{m}^3 \cdot \mathbf{yr}^{-1} \cdot \mathbf{cap}^{-1}$) total annual water demand W_N ($\mathrm{m}^3 \cdot \mathrm{yr}^{-1}$) normalized by population N (cap$^)$,

Population proportional growth constant k_N (in $\% \cdot yr^{-1}$) population growth normalized by population N (cap$^)$, and

Precipitation height q_P (in $\mathrm{m} \cdot \mathbf{yr}^{-1}$) conceptually, total precipitation volume Q_P ($\mathrm{m}^3 \cdot \mathrm{yr}^{-1}$) normalized by area A_N (m^2)

The normality of k_N, w_N, and q_P was then assessed using statistical tests and histogram, density, and quantile-quantile plots. The stationarity, auto-correlation, and cross-correlation of these variables were also analyzed.

[9] Though the MWDSC, from which LADWP has obtained as much as 74.6% of its water supply, did have long-term plans to increase desalination [223].

In addition, data were collected to characterize the relationship between stormwater runoff Q_R and precipitation Q_P. World water balance data were available from WebWIMP [190]. These data provided monthly water balance components at a 0.5° resolution from 1901–2010. Annual data were then calculated from monthly data and plotted as time series and cross-correlation.

The analysis performed in 4.5 (results shown in 4.8) were then used in Chapter 5 to establish "realistic" ranges for k_N, w_N, and q_P, which were then used to simulate future water demand and stormwater runoff.

4.5.1 General System Definition

While assessing the self-sufficiency of an urban water system, what sources can be considered local?

What does it mean for a city to be self-sufficient with respect to water supply? Rygaard, Binning, and Albrechtsen defined R_{SS} as the ratio of the volume of water obtained from local sources originating within an area of analysis (Q_{local}) relative to the total demand for water resources in that same area (W_{total}) over a period of time [278, p. 187][10]:

$$R_{SS} = \frac{Q_{\text{local}}}{W_{\text{total}}} \tag{4.1}$$

Local water resources Q_R was defined as "the amount of water sourced from within a given area" and more specifically as the sum of stormwater runoff Q_R, wastewater reuse Q_{WWR}, and desalination Q_{DS} (from a shoreline within the study area) [278, p. 187]:

$$\sum_{i=1}^{12} Q_{local_i} = \sum_{i=1}^{12} Q_{R_i} + Q_{WWR_i} + Q_{DS_i} \tag{4.2}$$

In practice, delineating "water sourced from within a given area" from that sourced externally is more complex than that simple definition of Q_R by Rygaard, Binning, and Albrechtsen might suggest [278, p. 187].

Since all cities require substantial water supplies, it would be reasonable to assume that all cities have freshwater sources—e.g., groundwater, rivers, and lakes—that are located within or immediately adjacent to the urban area. However, it is much less common for the hydrologic unit—from which water drains into said freshwater body—to lie entirely within the municipal boundary. In other words, even if a river or a lake lies within a municipal boundary, much of the water contained therein often originates far outside the urban area. For instance, some mu-

[10]Following the Oxford English Dictionary's definition of the broader concept of *self-sufficient*, self-sufficient urban water management could be defined generally as *Needing no outside help in the production of water to meet urban demand*. Philosophically, self-sufficiency could be interpreted to fit as many definitions as there are ways in which a city might use outside assistance in its production of municipal water supplies. However, the focus of this dissertation was on the built environment. Case study analysis therefore concentrated on the physical capacity to meet urban demand with local water sources.

nicipalities were delineated by hydrologic features such as an ocean, river, or lake. But consider that, if the hydrologic feature delineates a boundary between two political entities, to which of the two does the water in that feature belong?

Because municipal boundaries rarely align with natural hydrologic units, Rygaard, Binning, and Albrechtsen limited consideration of sources for $\dot{Q}_{local}$ to be:

1. Q_R stormwater runoff and

2. Q_{WWR} reuse/recycling of wastewater (collected from the area in question), and

3. Q_{DS} desalination of seawater (collected from the shoreline of the area in question).

However, even Equation 4.2 can be complicated to assess. Consider, for instance, stormwater runoff Q_R. As stated before, hydrologic units often do not neatly align with political and administrative boundaries. For cities for which this is true, some fraction of stormwater runoff that passes through a city will have originated from outside the urban boundary. Is that fraction of water originating outside of the city be considered local—or not?

Another issue with interpreting Q_R in Equation 4.2 is that some cities may own land in the areas in which their water does originate. For instance, New York City owns substantial land rights in the Catskill Mountains, from which it obtains almost all of its municipal supply. If the city obtains water from outside of its boundaries, but happens to own the land from which it is obtained, should that water still be considered an "outside" source? In the case of New York City, it would seem reasonable to consider water from the Catskills to be a "local" water resource in assessing self-sufficiency.

Taking a philosophical view, ownership is essentially a long-term legal status ensuring access to said resource. However, there exist other types of arrangements that also convey similar rights to resources—indeed, there exists a large and varied body of laws, judgments, policies, and treaties governing many different aspects of water use. So, if it is reasonable to include in Q_R water that is obtained from land that is owned by—but lies outside of—a municipality, then it would also seem reasonable to include water resources to which a city has a long-term legal right.

In summary, it is non-trivial to define the boundary of an urban system for the purpose of assessing self-sufficiency of water management. For any single case study, making the delineation between *local* and *non-local* water resources in a meaningful way requires taking into account the socio-politico-bio-geo-physical context of the city in question. Since said context can vary, multiple definitions for the boundary may exist even for a single case.

To simplify the issue, water sources were considered local if the city in question 1. considered the to be local and 2. owned or had long-term access to the sources, even if said sources lay outside of the municipal boundary.

Given the many difficulties involved, identifying a robust, meaningful, one-size-fits-any-and-all definition of local water resources appeared to be an intractable problem. Instead, a pragmatic approach was adopted for delineating local sources for the case studies. Long-term water management plans from each case study were used to guide system definition.

4.5.2 Water Resources

For this study, focus was given primarily to the potential resource of stormwater runoff generated within the municipal boundary[11]. Ignoring contributions from wastewater (Q_{WW}) and desalination (Q_D), Equation 4.2 became:

$$\dot{Q}_{local} \approx \dot{Q}_R \tag{4.3}$$

Annual stormwater runoff is a function of precipitation (Q_P), evapotranspiration (Q_{ET}), interception losses Q_{inter}, and infiltration Q_{infil} [200]:

$$Q_{R_i} = Q_{P_i} - Q_{\mathrm{inter}_i} - Q_{\mathrm{infil}_i} \tag{4.4}$$

Interception losses are a function of evapotranspiration Q_{ET}, and therefore also a function of land use, vegetation cover, vegetation type, temperature, irradiation, and soil moisture. Infiltration is a function of topography/slope, soil type, and soil moisture. In other words, the relationship between Q_P and Q_R is not straightforward, and numerous complex models exist to predict Q_R from Q_P. However, in short, often $Q_R \neq Q_P$.

If interannual variations in soil moisture storage and losses to deep seepage are small relative to the variations in annual fluxes of precipitation, evapotranspiration, and runoff, following McMahon and Price [200], then runoff Q_R is equal to the difference of precipitation Q_P and evapotranspiration Q_{ET} [200, p. 136, Equation 1]:

$$q_R(t) = q_P(t) - q_{ET}(t) \tag{4.5}$$

Q_P, the volumetric flux of precipitation into the urban study area. If $\dot{q}_p(x, y, z)$ is the precipitation height at point (x, y, z) in the study area at an instant in time, then, according to the Kelvin–Stokes theorem, Q_P is equal to the line integral of $\dot{q}_p(x, y, z)$ over the city boundary, $\mathbf{A}_N$.

The general Stokes theorem for a vector field ($\mathbf{F}$) over a surface (Σ) in Euclidean three-dimensional space $\mathbf{r}$ is:

$$\int\int_{\Sigma} \bigtriangledown \times \mathbf{F} \cdot d\Sigma = \oint_{\Sigma} \mathbf{F} \cdot d\mathbf{r} \tag{4.6}$$

$$\int\int_{\Sigma} \left\{ \left(\frac{\partial F_z}{\partial y} - \frac{\partial F_y}{\partial z}\right) \cdot dydz + \left(\frac{\partial F_x}{\partial z} - \frac{\partial F_z}{\partial x}\right) \cdot dz \cdot dx + \left(\frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y}\right) \right\} = \oint_{\Sigma} (P \cdot dx + Q \cdot dy + R \cdot dz) \tag{4.7}$$

[11] Contributions from desalination and wastewater reuse were considered in the analysis of results.

However, if the assumption is made that the vector field $\mathbf{F}$ is a constant F over $\mathbf{r}$, and S is the surface area of Σ, then the right side of Equation 4.7 collapses to:

$$\int\int_{\Sigma} \bigtriangledown \times \mathbf{F} \cdot d\Sigma = F \cdot S \tag{4.8}$$

Applying Equation 4.8 to precipitation, the volumetric flux of precipitation (Q_P) into a city of area A over a period t is given by:

$$Q_P = \int_t \dot{q}_P \cdot A_N = q_P \cdot A_N \tag{4.9}$$

where $\dot{q}_P$ is the unit volumetric flux at a point in space at an instant in time.

4.5.3 Water Use

Water is used in many different activities in a city. Different types of activities have different types of use characteristics. It can also be useful to distinguish how the water is being used, e.g., for toilet flushing, lawn watering, cooking, cleaning, cooling, and so forth. A use-based perspective, rather than a user-based perspective, is particularly important for understanding how changing technological water efficiency could induce changes in water demand.

In industrial ecology, different use types are known as "services", and the *Material Input required Per unit of Service (or good) produced (MIPS)* is the general term for the resource intensity required to produce a good or service. For instance, toilet flushing, lawn watering, etc. could all be considered services. The MIPS for toilet flushing would be the water required to flush the toilet. The total water demand for toilet flushing would then be the number of toilet flushes multiplied by the water used to flush the toilet.

However, while distinguishing between different services and MIPS is useful, data have historically been difficult to acquire. Technological developments such as "smart" sensing, the "Internet of Things", social media, and other technologies have made the acquisition of such data more tractable, in theory. The practice, however, is still out of reach for many cities.

Since data on individual uses of water are difficult to acquire, cities monitor water use at a more aggregate level. The MIPS most commonly used by water managers is "Per Capita Water Consumption". This average use intensity is calculated by dividing total water use by the known population, and is used as representative of normalized water demand in lieu of more detailed analysis.

Instead of services, urban water managers may distinguish between different categories of users, especially *residential*, *commercial/industrial*, *municipal/other*, and sometimes even *agricultural*. These user categories may be used to differentiate between pricing levels or conservation policies.

If $W_{j,k}$ is the total water use of user k, $w_{j,k}$ is the MIPS for user k for service j_k, then the total water use W_{total} for all users and all uses is given by:

$$W_{total} = \sum_j \sum_k W_{j,k} = \sum_j \sum_k w_{j,k} \cdot j_k \tag{4.10}$$

To start, local water demand (W_N) was approximated as the product of the average water use intensity (UI, denoted w_N) and population (N):

$$W_N = W_{\text{tot}} = w_N \cdot N \tag{4.11}$$

4.5.4 The Potential for Self-Sufficiency from Climatic Resources

Substituting Equation 4.1 into Equation 4.3 gives:

$$R_{SS} = \frac{Q_R}{W_{total}} \tag{4.12}$$

Then substituting Equations 4.1 (ignoring evapotranspiration and infiltration and interception losses Q_{ET} and Q_{II}) and 4.11 into Equation 4.12:

$$R_{SS} = \frac{Q_P}{w_N \cdot N} \tag{4.13}$$

Finally, substituting Equation 4.9 into 4.14 yields:

$$R_{SS} = \frac{q_P \cdot A_N}{w_N \cdot N} \tag{4.14}$$

4.5.5 WUCI

Recall that the WUCI (i_{UC}) was introduced in Chapter 3 as an index of relative total and per capita water footprint. WUCI was defined as use intensity (w_N in $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$) divided by local annual precipitation (q_P, in $\mathrm{m} \cdot \mathrm{yr}^{-1}$); WUCI had units of $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$ while total WUCI has units of area (m^2 or km^2).

When compared with the area of a city, total WUCI provides an initial estimation of the potential to supply water locally–i.e., self-sufficiency. WUCI is an extremely rough estimate of the watershed area required to capture enough precipitation to supply a single person with a typical water use pattern over a period of time. The calculation of WUCI does not take into account local surface water or groundwater sources, interception or infiltration losses, collection efficiency, or other important aspects of a true water balance. However, as an index, WUCI provides useful index for comparing cities based on the relative ability of the local climate to support a typical water use pattern.

Recognizing that population density, ρ, is defined as population N divided by the city area q_P:

$$\rho = \frac{N}{A_N} \tag{4.15}$$

Equation 4.14 can be rewritten as:

$$R_{SS} = \frac{q_P}{w_N \cdot \rho} \tag{4.16}$$

Recall that Total WUCI was defined in Chapter 3 as water use intensity (w_N) normalized by local average annual precipitation height, q_P:

$$\text{WUCI} = i_{UC} = \frac{w_N}{q_P} \tag{4.17}$$

While total WUCI i_{UC} is equal to the product of WUCI and population:

$$\text{Total WUCI} = I_{UC} = N \cdot \frac{w_N}{q_P} \tag{4.18}$$

Self-sufficiency R_{SS} can be rewritten in terms of i_{UC} by substituting Equation 4.17 into Equation 4.16:

$$R_{SS} = \frac{1}{i_{UC} \cdot \rho} \tag{4.19}$$

Or in terms of i_{UC} by substituting Equation 4.18 into Equation 4.14:

$$R_{SS} = \frac{A_N}{I_{UC}} \tag{4.20}$$

4.6 Data

4.6.1 Population and Population Growth

Population data were available for Singapore on an annual basis from 1960 through 2015 through yearbooks provided by Singapore's Department of Statistics [292].

For Los Angeles, population data for the city were available on a decadal basis from 1900 through 2010 [324, 323]. Annual estimates of population were available from the US Census Bureau from 1990 through 2010; from 2004 through 2014 from the Los Angeles City Controller; and from 2010-2016 from the Los Angeles Almanac [323, 163, 165].

Average population growth rate (k_N) was estimated from data on annual population (N) over the time period (t_f, t_i), using the exponential function:

$$N_f = N_i \cdot \exp^{k_i \cdot (t_f - t_i)} \tag{4.21}$$

Where data were available, population growth rate was estimated over a yearly period (i.e., $t_f - t_i = 1$). Where there were gaps in annual population data, the population growth rate was estimated by assuming that the average annual growth rate was constant over the time period.

4.6.2 Water Use and Water Use Intensity

Total water use and average water use intensity data were available for Singapore from 1960–2015 in annual reports published by Singapore's water utility, the PUB; annual reports for the fiscal years between 2010/2011 and 2015/2016 were available online, while others were accessed in person from Singapore's National Library [262]. Total water use data were available for the City of Los Angeles from 1970–2016 [164].

Average annual water use intensity w_N was also estimated from total water use w_N and population N by rearranging Equation 4.11:

$$w_N = \frac{W_N}{N} \tag{4.22}$$

4.6.3 City Area

Data on island area were available from Singapore's Department of Statistics for 1960–2015. Data on city area were available on a decadal basis from the LA Almanac (`http://www.laalmanac.com/`) and the US Census Bureau.

Missing data on city area were estimated for LA using both constant interpolation (assuming that city area was constant and equal to the last data observation where data points were missing). Both estimates were performed using the `R` function `approx` from the `stats` package. The year and city area in km^2 were given as inputs to `approx`. For `approx`(`method`=*constant*), the rule `f`=0 was used, specifying a left-continuous step function. The resulting interpolation can be seen in Figure 4.22a, and was then used in calculating R_{SS} for the cases.

4.6.4 Water Balance

Precipitation

Monthly precipitation data from 1901–2009 were available for Singapore from The World Bank Group's Climate Change Knowledge Portal at `http://sdwebx.worldbank.org/climateportal/index.cfm` [355]. For Los Angeles, monthly precipitation data were obtained for the period 1893–2016 from the U.S. National Oceanographic and Atmospheric Administration's National Climatic Data Center through their portal Climate Data Online [230].

Water Balance

In addition, monthly global water balance data were available at a 0.5° resolution from 1900–2014 from the creators of WebWIMP [190]. The data were provided in tabular format for each variable, for each year. The values from each of these tables, for each case study, were extracted. Each of these tables was converted to a raster.

There were 9 data variables, including temperature and precipitation, for 114 years[12]. With $9x114 = 1026$ tables and 2 case studies, it was necessary to automate extraction of values. Since shapefiles of the political boundaries of both cities were easily available, extraction was performed using the `extract` method from the `raster` package in `R`[13]. However, this method could easily be adapted for locations for which shapefiles are not available.

Where necessary, water balance values were averaged over several grid cells. Monthly values for precipitation, evapotranspiration (potential and actual), snowmelt, snowpack, and deficit were summed over full years[14]. Annual precipitation q_P, evapotranspiration (potential q_{ET^0} and actual q_{ET}), and surplus q_S were then used to characterize the relationship between precipitation and runoff.

Evapotranspiration

High resolution global water balance data were obtained from the Consortium of International Agricultural Research Centers (CGIAR) Consortium for Spatial Information. Data were available for monthly effective precipitation and actual evapotranspiration from 1901–2015 [190]:

1. Monthly effective precipitation
2. Mean annual actual evapotranspiration (AET)

Data for Singapore and Los Angeles were extracted from the WebWIMP data series using shapefiles of Singapore and Los Angeles. Monthly values for $q_R(t)$ were calculated from $q_P(t)$ and $q_{ET}(t)$ for each element of the raster. These values were summed over the year and the mean, standard deviation, and quantiles (with probabilities, $p = 0, 0.25, \ldots 1.00$) were found for the annual data. The covariance of precipitation with evapotranspiration was assessed. Distributions for precipitation, evapotranspiration, and estimated runoff were then hypothesized and tested.

4.7 Analysis of Historical Self-Sufficiency

The following descriptive statistics were calculated for the population growth constant (k_N), water use intensity (w_N), and annual precipitation (q_P):

[12] A full listing of WebWIMP water balance data is provided in A.1.2.

[13] The shapefiles of the political boundaries of Los Angeles and Singapore were obtained from the Los Angeles City Controller Control Panel and the GADM database of Global Administrative areas, respectively [166, 132]. More information provided in A.2.1.

[14] No snow was observed in either case study over the study period.

Quantiles: at 0%, 25%, 50%, 75%, and 100% probabilities;

Statistical moments: the mean μ, standard deviation σ, skewness γ, and kurtosis.

These descriptors were then used to assess the normality of the distributions of k_N, w_N, and q_P. These distributions were also examined using visualization methods, particularly histogram, density, and qq-plots. The stationarity, auto-correlation, and cross-correlation of these variables were also analyzed.

4.7.1 Descriptive Statistics and Randomness

Summary statistics (moments and quantiles) were calculated for the population growth constant (k_N), annual precipitation (q_P), and water use intensity (w_N) as described in 3.4.1.

The mean, variance, skewness, and kurtosis were calculated for using the built-in `R` functions `mean`, `sd`, `skewness`, and `kurtosis`. Four quantiles (i.e., at 0%, 25%, 50%, 75%, and 100% probabilities) were calculated for k_N, w_N, and q_P using the built-in `R` function `quantile`. Missing values were removed.

Histograms were plotted using `ggplot` with `geom_histogram` from the `ggplot2` package for `R` as described in 3.4.1 [349]. Densities were estimated for each variable and plotted with the corresponding histogram. Qq-plots were also generated for k_N, w_N, and q_P. Missing values were removed before plotting.

For the density distributions and qq-plots, k_N was compared to the Student's t distribution and w_N and q_P were compared to the Gamma distribution. Since the population growth constant (k_N) had values that were less than and greater than zero, observations were assumed to be drawn from the normal distribution. However, since the sample size of the data was only moderately large, the data for k_N were compared to a Student's t-distribution. Since values of w_N and q_P must be ≥ 0, density and qq-plots for k_N and q_P compared the data to the Gamma distribution.

To plot the estimated density for each variable with its histogram, Equation 3.8 was modified using `stat_function`, and specifying the density function (dFunction) and a list of relevant arguments (arguments):

$$\begin{aligned}
&\texttt{ggplot}(\mathit{data} = \text{dataset}, \texttt{aes}(x = \text{x}))+\\
&\qquad \texttt{geom_histogram}(\texttt{aes}(y = ..\text{density}))+\\
&\qquad \texttt{stat_function}(\mathit{fun} = \text{dFunction}, \mathit{args} = \mathit{list}(\text{arguments}))
\end{aligned} \tag{4.23}$$

To plot the a qq-plot for each variable, Equation 3.9 in 3.4.1 was modified by specifying the density function, dFunction, and related arguments (arguments). A fit line between the observations was also added to the plot by specifying a slope and intercept with `geom_abline`:

$$\begin{aligned}
&\texttt{ggplot}(data = \text{dataset}, \texttt{aes}(sample = \texttt{x})+ \\
&\qquad \texttt{stat_qq}(data = \text{dataset}, \texttt{aes}(sample = \texttt{x}), \\
&\qquad distribution = stats::\text{dFunction}, dparams = list(\text{arguments})) + \\
&\qquad \texttt{geom_abline}(intercept = \text{value3}, slope = \text{value4})
\end{aligned} \tag{4.24}$$

For the Student's t-distribution, dFunction $= dt$, and $list(\text{arguments}) = list(\text{df} = \text{value2}_t)$, where df refers to degrees of freedom. For the Student's t-distribution, $\text{value1}_t = \texttt{quantile}(x)$ (i.e., the quantiles of the data series, x). The remaining values (value2_t, value3_t, and value4_t) were estimated in the following way:

First a fit (fit_v) was estimated using the `survreg` and `Surv` functions from the `R` library `survival` (specifying the distribution as $dist =' t'$). Then the values of x were sorted into a new vector (v_1). Next, a vector of random observations (v_2) was drawn from the Student's t-distribution using the function `rt`, specifying the number of values $n = \texttt{length}(x)$ and the degrees of freedom $df = fit_v\$df$. Then a linear regression was run on v_1 and v_2 using the `lm` function. The following shows the general structure of this process as written for `R`:

$$\begin{aligned}
fit_v &\leftarrow \texttt{survreg}(\texttt{Surv}(x) \approx 1, dist =' t') \\
v_1 &\leftarrow \texttt{sort}(x) \\
v_2 &\leftarrow \texttt{sort}(\texttt{rt}(n = \texttt{length}(x), df = fit_v\$df)) \\
qq_v &\leftarrow \texttt{lm}(v1 \approx\ v2)
\end{aligned} \tag{4.25}$$

Then, value2, value3, and value4 were taken from the fit with value2 $= fit_v\$df$, value3 $= fit_v\$coefficients[1]$, and value4 $= fit_v\$coefficients[2]$.

For the Gamma distribution, dFunction $= dgamma$, and $list(\text{arguments}) = list(shape = \text{value1}_\Gamma, scale = \text{value2}_\Gamma)$. For the Gamma distribution, $\text{value1}_\Gamma = \mu_x^2/\sigma_x^2$ and $\text{value2}_\Gamma = \sigma_x^2/\mu_x$, where μ_x and σ_x refer to the mean and standard deviation, respectively, of the data series x. The process for estimating the slope and intercept of a fit line for the Gamma distribution was similar to that described for the Student's t-distribution, except that it was unnecessary to calculate fit_v since v_2 could be estimated directly using the *rgamma* function and, instead of specifying df, specifying the shape and scale parameters as described earlier in this paragraph.

4.7.2 Time series

While it can be useful to examine the summary statistics of the data, in reality, the data were not randomly drawn from the same sample at an instant in time. Instead, the data for the case studies represented observations of the system state that were collected over five decades or longer. Nor was there any reason to think that the state of the system had not changed over time. A cursory look at local (or even global) history made it clear that substantial changes had occurred to economic development, population growth, demographics, land use, and water use. Therefore, the expectation was that the distributions for population growth k_N, water use intensity w_N, and annual precipitation q_P would not be stationary, i.e., that some changes over time would be observed.

The starting question was whether any trends or patterns could be identified in the observed data through basic statistical and time series methods.

Time Series Plots To begin the time series analysis, data for k_N, w_N, and q_P were plotted as points oriented in time, i.e., as time series (Figures 4.10a, Figures 4.14a, and 4.18a, respectively). While a cursory overview of time series plots k_N and w_N in particular (Figures 4.10a and 4.14a, respectively) suggested that those series were not stationary. However, the null hypothesis was that the data series were random variables and stationary in both μ and σ.

Relative Deviation The first step was to calculate the relative deviation of each series. The relative deviation is one measure of the dispersion of a data series; i.e., the residuals relative to the expected value. The residual for series $\mathbf{x} = x_1, x_2, \cdots, x_n$ at time $t = i$ is defined as $\mathrm{res}(x_i) \equiv x_i - \bar{x}$; the relative deviation is then $\mathrm{rdev}(x_i) \equiv \bar{x}^{-1} \cdot \mathrm{res}(x_i)$.

The mean of the relative deviation approaches zero as the number of observations becomes arbitrarily large. Taking the relative deviation therefore re-centered each series to zero. Since the relative deviation included a factor that normalized deviations by the expected value, the original shape of the series was generally preserved while allowing proportional changes to be more easily identified and compared between the cases. Thus, the relative deviation of k_N, w_N, and q_P were calculated and plotted over time. Histograms, density plots, and qq-plots of the relative deviations were also created and compared with the original series and between cases.

First Time Derivative Results from plotting the original data and relative deviations as histograms, density functions, and qq-plots, and time series suggested that the mean of the series had not been stationary in time. To highlight trends in the series, the first time derivatives of k_N[15] and q_P were taken and normalized by the expected value of the series. If the first time derivative of $\mathbf{x}$ is $\dot{\mathbf{x}}$, the first derivative was defined for element x_i as:

$$\dot{x}_i \equiv \frac{\mathrm{d}x_i}{\mathrm{d}t} = \frac{x_i - x_{i-1}}{t_i - t_{i-1}} \tag{4.26}$$

Since the time derivative was calculated between years, $t_i - t_{i-1} = 1$, so $\dot{x}_i = x_i - x_{i-1}$. Thus, the normalized first derivative was $\bar{x}^{-1}\dot{x} = \bar{x}^{-1} \cdot (x_i - x_{i-1})$.

[15]Since k_N was calculated as the exponential change in population over a year, the first time derivative of k_N was a second time derivative.

For w_N, a distinctive trend was observed in the original data and the relative deviation for both cases. Thus, a proportional growth factor k_w was calculated for w_N:

$$\dot{k}(w_i) = \frac{\ln(w_i) - \ln(w_{i-1})}{t_i - t_{i-1}} = \ln(\frac{w_i}{w_{i-1}}) \tag{4.27}$$

The normalized derivatives with respect to time were then plotted as a time series and as histograms, density functions, and qq-plots and compared with the original data and relative deviation.

While the relative deviation highlighted the shape of each original data series relative to its expected value, the normalized first derivative highlighted *changes between years* (relative to the expected value). These changes were found to be small and usually less than 10%, with the exception of q_P (as seen in Figures 4.10c, 4.14c, and 4.18c).

Second Time Derivative The second time derivative gives information about the change in the change of the series, i.e., about the acceleration. Since k_N was calculated as the proportional growth factor from the first time derivative of population N (assuming exponential growth, as described in Section 4.6.1), the first time derivative of k_N was a quasi-second time derivative of population. Periods where sequential values of $\dot{k}_N \approx 0$ suggested that the shape of the population was not changing much, while the residuals indicated the deviation of population away from average exponential growth.

The second derivatives for w_N and q_P were also found. As with $\dot{k}_N$, values for the second derivative of w_N and q_P indicated changes to the relative shape of the original series.

Moving Average analysis (MAV) Moving Average Analysis was performed using `sma` from the `smooth` package in `R`, which finds an optimal period T_P (in years) for the window used to calculate the moving average. MAVs for k_N, w_N, and q_P are shown in Figure 4.11, 4.16, and 4.19, respectively. At the top of these MAV plots the time period T_P used to calculate the moving average is shown as "SMA(T_P)".

4.8 Historical Self-Sufficiency: Results and Comparison

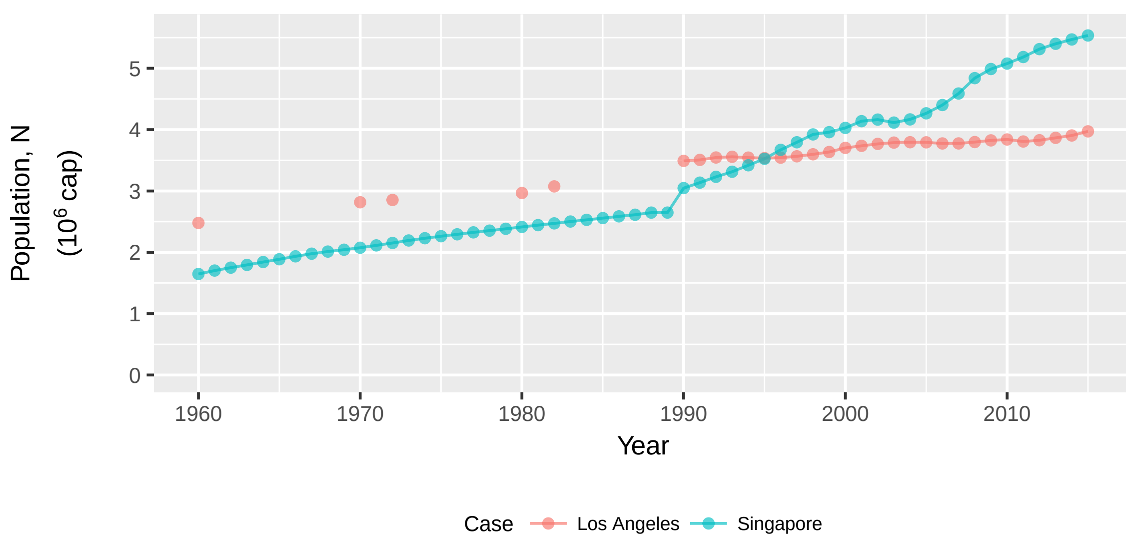
4.8.1 Overview

Distributions of the population growth factor k_N were found to be reasonably approximated by the normal distribution, while distributions for water use intensity w_N and annual precipitation q_P were reasonably approximated by the Gamma distribution. Distributions of k_N overlapped between the two cases. Distributions of historical data for w_N and q_P exhibited no overlap, and theoretical distributions exhibited no overlap of the interquartile range IQR.

While the random distributions were not unreasonable fits to the historical data, viewing the data as time series indicated that the series were likely to not be completely random or stationary. Taking the relative deviations preserved trends in the original series while rescaling data from the two cases to be more easily compared with each other.

4.8.2 Population and Population Growth

Figure 4.8: Population (N) of case studies (1960–2016). While Los Angeles was missing data before 1990, it is clear from the figure that the population of Singapore has grown faster than LA over the study period. Singapore's population surpassed that of LA after lagging for the first three decades, and has continued to grow rapidly in recent years.



The population for the case studies is plotted as a time series in Figure 4.8. Annual time series were obtained for Singapore throughout the study period. In contrast, there were numerous missing observations in the annual data for LA.

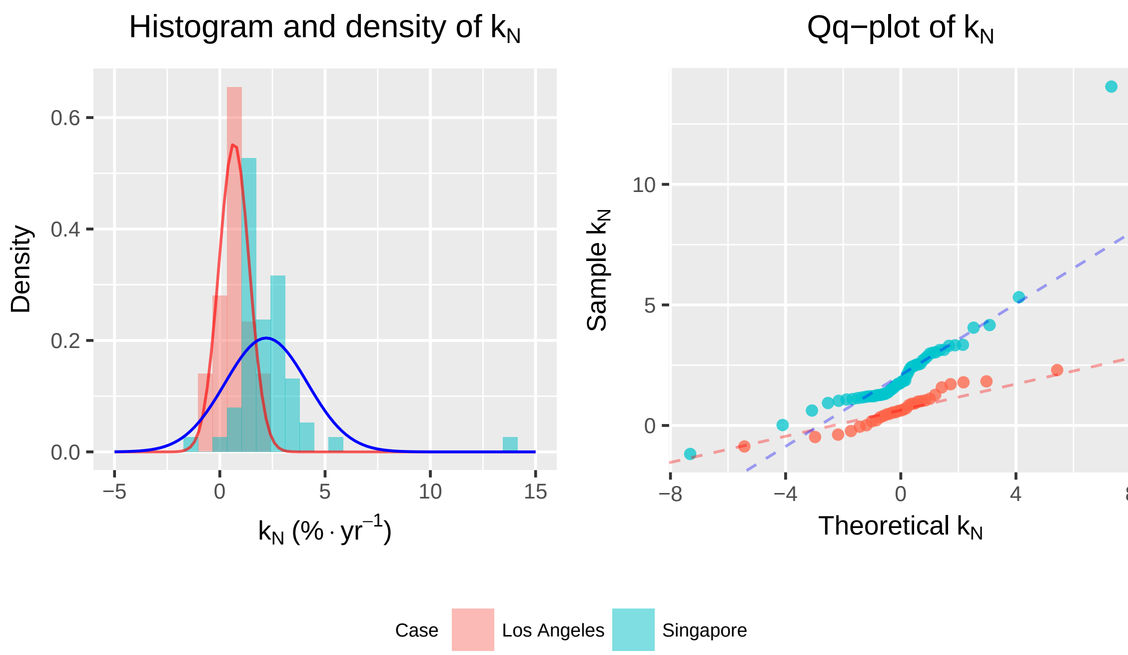
Population in both cities was found to increase over the study period; however, population increased more in Singapore than in Los Angeles. In 1960, the population of LA was 2.48 million, making it the third largest city in the US at the time. The population of Singapore in 1960 was 1.65 million–66% of LA's size at the time.

In 1995, LA's population was 3.54 million, while Singapore's was 3.52 million; within the year, Singapore's population had surpassed that of LA: while LA's population had remained nearly constant—increasing by only 0.2%—Singapore's population had increased by 4.2%.

In 2015, LA's population had only risen to 3.97 million—an increase of only 60% above its size in 1960. In contrast, Singapore's population had more than tripled over this time period, to 5.54 million. Unsurprisingly, the mean of the proportional growth constant ($\mu(k_N)$) over 1960–2015 was lower for LA than Singapore (with $\mu(k_N^{LA}) = 0.68\%$, $\mu(k_N^{SG}) = 2.20\%$, $\sigma(k_N^{LA}) = 0.72\%$, and $\sigma(k_N^{SG}) = 1.95\%$).

Figure 4.9: Case study historical data: annual population growth constant k_N ($\% \cdot yr^{-1}$), 1960–2016. As shown in the figures below, the normal distribution was found to reasonably represent k_N for both cases.

(a) Histogram, density, and qq-plots of the annual population growth constant (k_N)



(b) Summary statistics

Statistic	Los Angeles	Singapore
n	31.00	55.00
μ	0.68	2.20
σ	0.72	1.95
γ	0.11	4.02
Quantile	Los Angeles	Singapore
κ	-0.32	22.08
0%	-0.87	-1.18
25%	0.27	1.26
50%	0.62	1.77
75%	1.02	2.73
100%	2.30	14.05

The proportional growth constant (k_N) associated with this growth is shown as a histogram/density plot and qq-plot in Figure 4.9a. The densities and the theoretical quantiles in Figure 4.9a were both estimated from the normal distribution. While the normal distribution appeared to be a reasonable match for LA, it was not ideal for Singapore. While for LA, the median value for k_N^{LA} was within 10% of the mean, the median value for k_N^{SG} was within 20% of the mean. Both k_N^{LA} and k_N^{SG} exhibited positive skewness (i.e., $\gamma > 0$), indicating distributions skewed to the right. However, the distribution for Singapore exhibited much greater skewness, with $\gamma(k_N^{LA}) = 0.11 < \gamma(k_N^{SG}) = 4.13$. For distributions near the normal distribution, the median value will be close to $\mu - \gamma \cdot \sigma/6$. For LA, $\mu - \gamma \cdot \sigma/6 = 0.67$—close to the median value of 0.62. However, for Singapore, $\mu - \gamma \cdot \sigma/6 = 0.86$, which was much lower than the median value of 1.77.

The time series plot of k_N shown in Figure 4.10a shed additional light on the statistics. Between 1960–1987, k_N^{SG} showed a relatively steady decrease. However, starting around 1988, k_N^{SG} began to exhibit increased volatility. The normalized first time derivative $\bar{k}_N^{-1} \cdot \dot{k}_N$ was calculated and is shown in Figure 4.14b. Yearly changes were generally small relative to the expected value.

The MAVs for k_N for the cases were found and are shown in Figure 4.11, with missing values plotted as zero.

Figure 4.10: Time series plots of k_N: original series, relative deviation, and first time derivative.

(a) k_N: Original data series.

(b) Relative deviation of k_N re-centered the series to zero and rescaled the data by the mean.

(c) Yearly changes in $\dot{k}_N$ over time, normalized by $\bar{k}_N$.

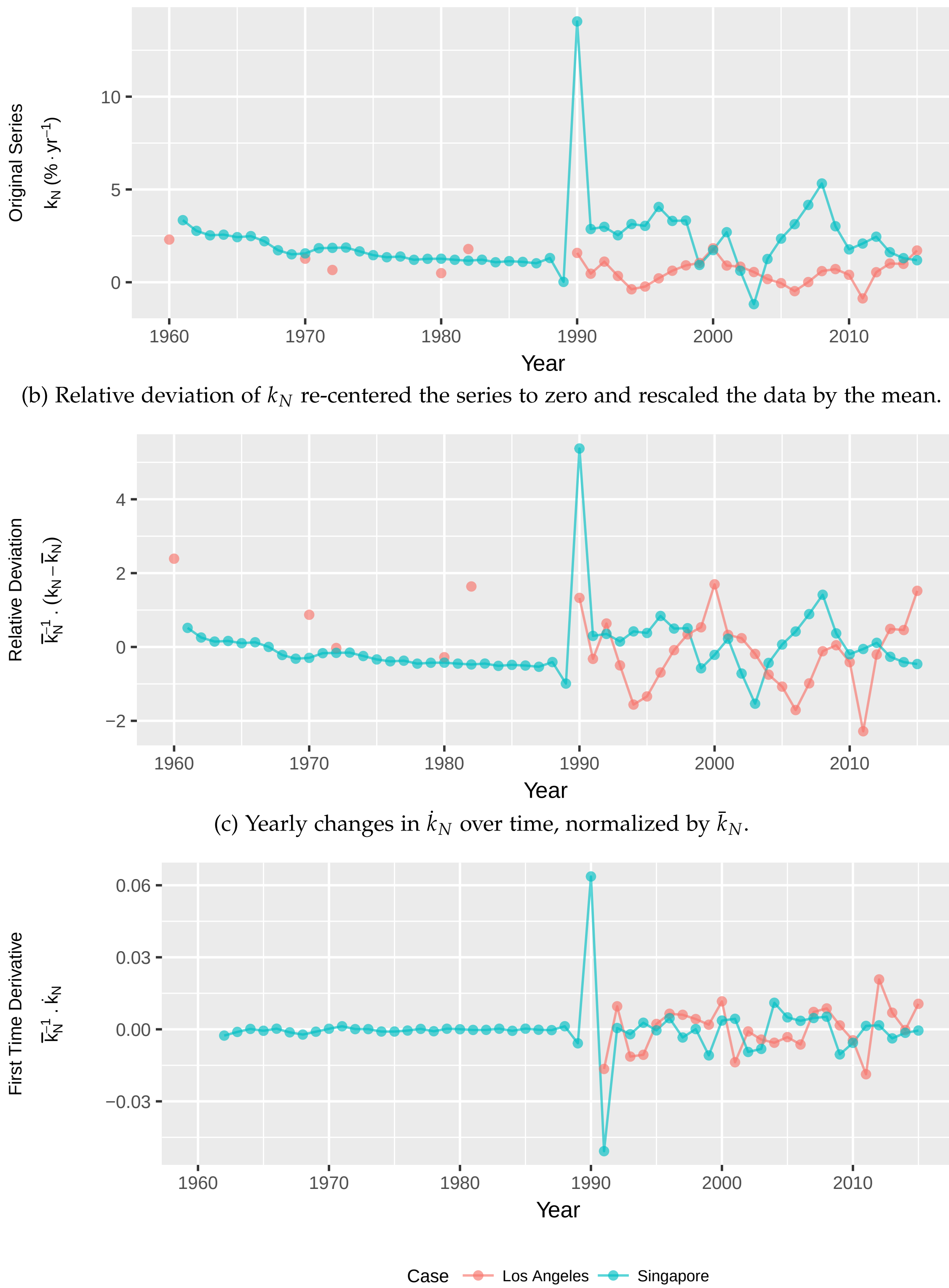
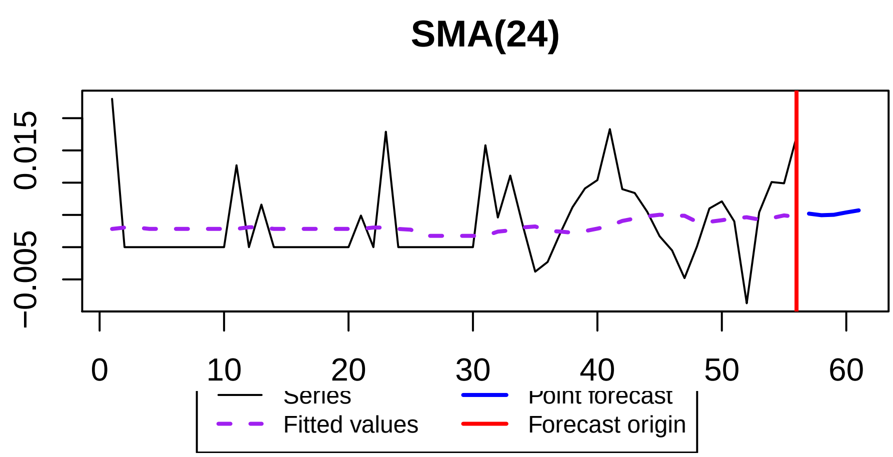
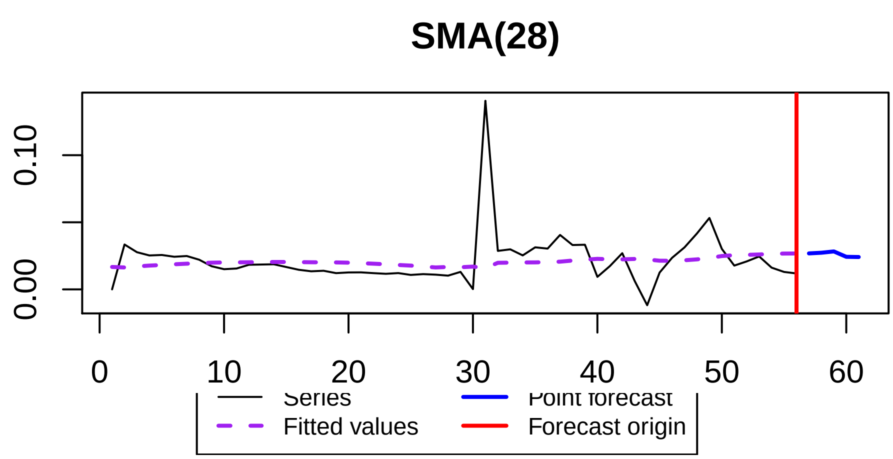


Figure 4.11: Moving average of the proportional growth constant (k_N) for the two cases over the study period (1960–2015). Los Angeles was missing values, which are shown as zero in the plots. The moving average analysis algorithm identified a period for smoothing (T_p), shown in parentheses above the plots; the analysis suggested $T_p = 24$ for LA (slightly shorter than the period for which data was available) and $T_p = 28$ for Singapore. The Moving Average Value analysis (MAV)/SMA results suggests that population growth has increased in the last several decades for both cases.

(a) LA: MAV for k_N^{LA}, with $T_P = 24$.

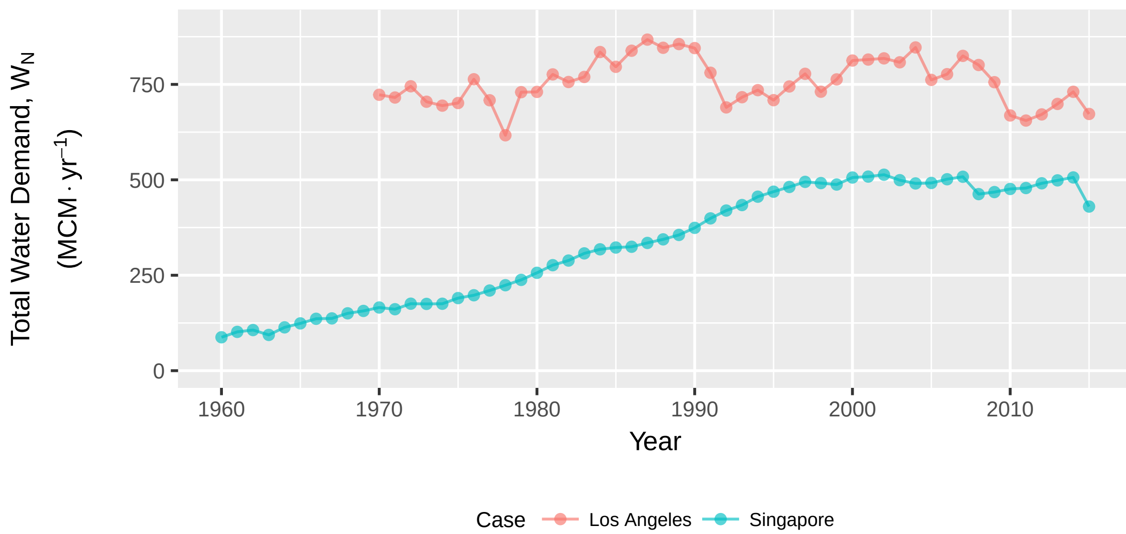


(b) Singapore: MAV for k_N^{SG}, with $T_P = 28$.



4.8.3 Water Use and Water Use Intensity

Figure 4.12: Total water use (W_N) for both cases (1960–2016). While data in Los Angeles over 1960–1970 was missing, it is clear that at the beginning of the study period LA's total water demand was much greater than that of Singapore's, which was less than 50 10^6 $m^3 \cdot yr^{-1}$ in 1960. Over the study period, total water use in LA has remained relatively constant, oscillating around 750 10^6 $m^3 \cdot yr^{-1}$, while that in Singapore has exhibited substantial growth. Around 1995, total water use in Singapore had reached 500 10^6 $m^3 \cdot yr^{-1}$, nearly ten times its value in 1960; however, this growth appears to have leveled off in recent years.



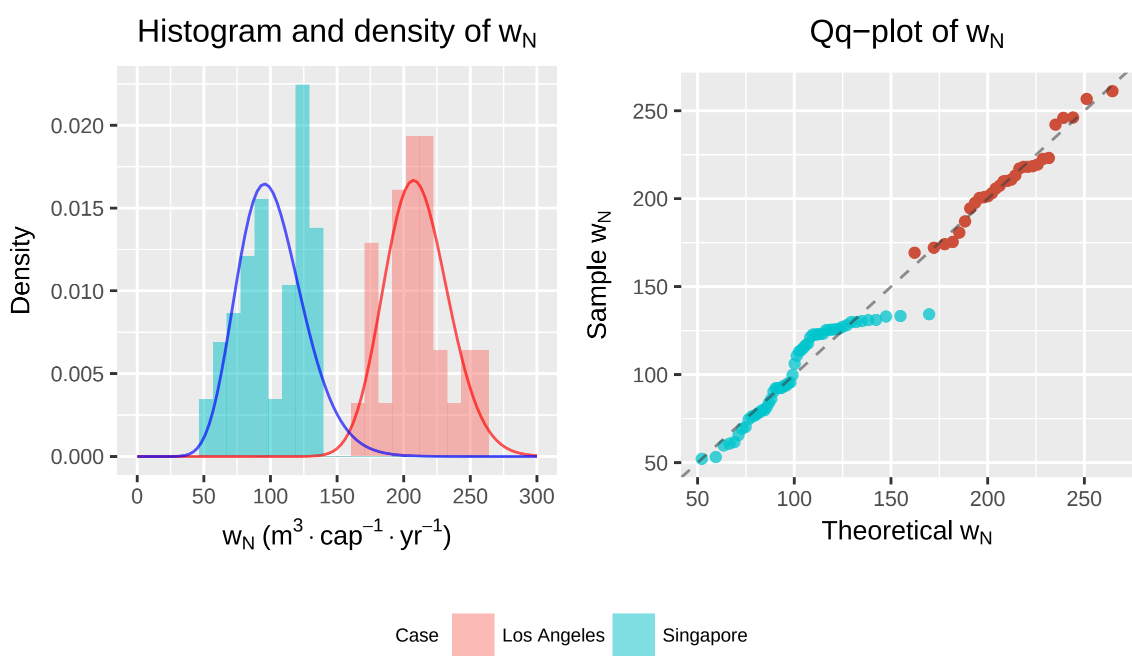
Total water use over 1960–2015 is shown in Figure 4.12. Throughout the study period, total water use (W_N) was much greater in LA than in Singapore. In 1970, total water use in LA was $722.9 \times 10^6 m^3 \cdot yr^{-1}$ ($W_N^{LA} \mid_{t=1970}= 722.9 myr$), while in Singapore $W_N^{SG} \mid_{t=1970}= 165.5 m^3 \cdot yr^{-1}$—nearly four times smaller than $W_N^{LA} \mid_{t=1970}$. However, Singapore's total water use increased significantly over the study period, peaking in 2002 at 513.410^6 $m^3 \cdot yr^{-1}$—a three-fold increase above $W_N^{SG} \mid_{t=1970}$ and nearly six times $W_N^{SG} \mid_{t=1960}$.

In contrast, the maximum W_N observed over the study period was 867.410^6 $m^3 \cdot yr^{-1}$ in 1987 at a value 1.2 times $W_N^{LA} \mid_{t=1970}$. Over this same period (1970–1987), W_N increased to $W_N^{SG} \mid_{t=1987}= 334.7 \times 10^6 m^3 \cdot yr^{-1}$—two times its level in 1970. Singapore's increase in W_N over 1970–1987 was also slightly greater in magnitude than that of LA—16910^6 $m^3 \cdot yr^{-1}$ compared to 144.510^6 $m^3 \cdot yr^{-1}$.

However, Singapore's population also grew more substantially over this time period. Figure 4.14a shows total water use normalized by population (i.e., water use intensity w_N) plotted as a time series. Between 1960–1970, w_N increased in Singapore from $53.2 m^3 \cdot yr^{-1} \cdot cap^{-1}$ to $79.78 m^3 \cdot yr^{-1} \cdot cap^{-1}$. By 1987, w_N had reached $128.1 m^3 \cdot yr^{-1} \cdot cap^{-1}$—2.4 times $w_N^{SG} \mid_{t=1960}$. Water use intensity in Singapore peaked two years later at $134.3 m^3 \cdot yr^{-1} \cdot cap^{-1}$ in 1989, and began a steady

Figure 4.13: Case study historical data: water use intensity w_N ($\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$), 1960–2016. Since $w_N \geq 0$, the density function and theoretical quantiles were generated from the Gamma function, with shape parameter $k = \mu^2/\sigma^2$ and scale parameter $\theta = \sigma^2/\mu$.

(a) Histogram, density, and qq-plots of water use intensity (w_N)



(b) Summary statistics

Statistic	Los Angeles	Singapore
n	30.00	56.00
μ	210.00	102.00
σ	24.00	25.00
γ	0.00	-0.00
Quantile	**Los Angeles**	**Singapore**
κ	-1.00	-1.00
0%	169.00	52.00
25%	198.00	80.00
50%	210.00	103.00
75%	219.00	125.00
100%	261.14	134.32

decline after 1996. By 2015, w_N had decreased to $w_N^{SG}\mid_{t=1987}= 77.7\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$—a value 1.5 times $w_N^{SG}\mid_{t=1960}$, approximately equal to $w_N^{SG}\mid_{t=1970}$, and 0.6 times the maximum observed w_N, $w_N^{SG}\mid_{t=1989}$.

With numerous missing values before 1990, trends in w_N in LA were more difficult to identify. In 1970, $w_N^{LA}\mid_{t=1970}= 256.7\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$—over three times $w_N^{SG}\mid_{t=1970}$. The maximum value for w_N in LA that was observed over the time series was $261.1\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$ in 1972. In 1980,

$w_N^{SG} \mid_{t=1980} = 246.2\text{m}^3 \cdot \text{yr}^{-1} \cdot \text{cap}^{-1}$ and in 1990, $w_N^{LA} \mid_{t=1990} = 242.1\text{m}^3 \cdot \text{yr}^{-1} \cdot \text{cap}^{-1}$. In 2015, $w_N^{LA} \mid_{t=2015} = 169.3\text{m}^3 \cdot \text{yr}^{-1} \cdot \text{cap}^{-1}$—a value 0.7 times $w_N^{LA} \mid_{t=1970}$ and 0.6 times the maximum w_N of $w_N^{LA} \mid_{t=1972}$.

A histogram/density and qq-plot are shown for w_N in Figure 4.13a, with summary statistics provided in Table 4.13b. Since $w_N \geq 0$, the density function and theoretical quantiles were generated from the Gamma function rather than the normal function. The Gamma function took as arguments a shape parameter $k = \mu^2/\sigma^2$ and a scale parameter $\theta = \sigma^2/\mu$.

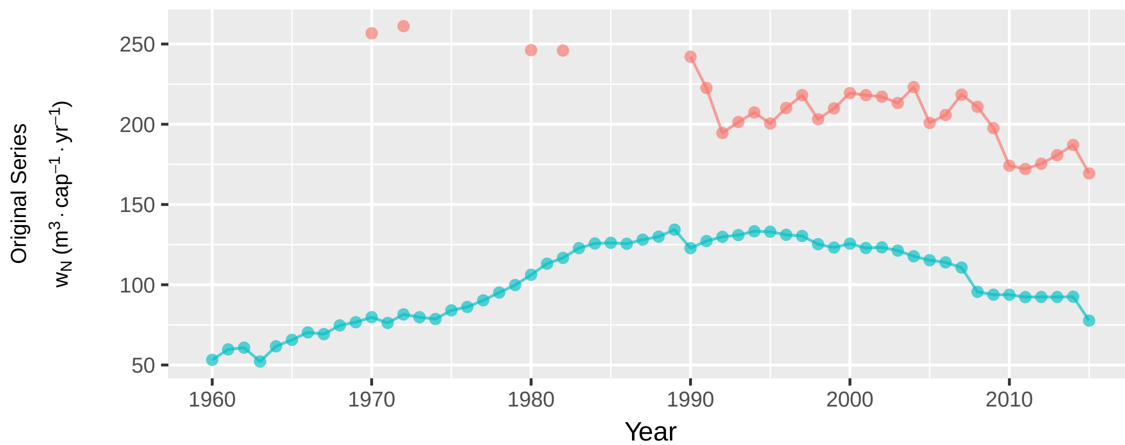
No overlap was observed in historical water use intensity data across the two cases, as seen in the histogram in Figure 4.13a (with only minor overlap observed in the density functions.)

The proportional growth factor for w_N ($k(w_N)$) was found and plotted in Figure 4.14c. $k(w_N^{SG})$ was generally greater than zero before 1995 and generally less than zero post-1990. The expected value of $k(w_N^{LA})$ was slightly less than zero after 1990 (with a larger deviation).

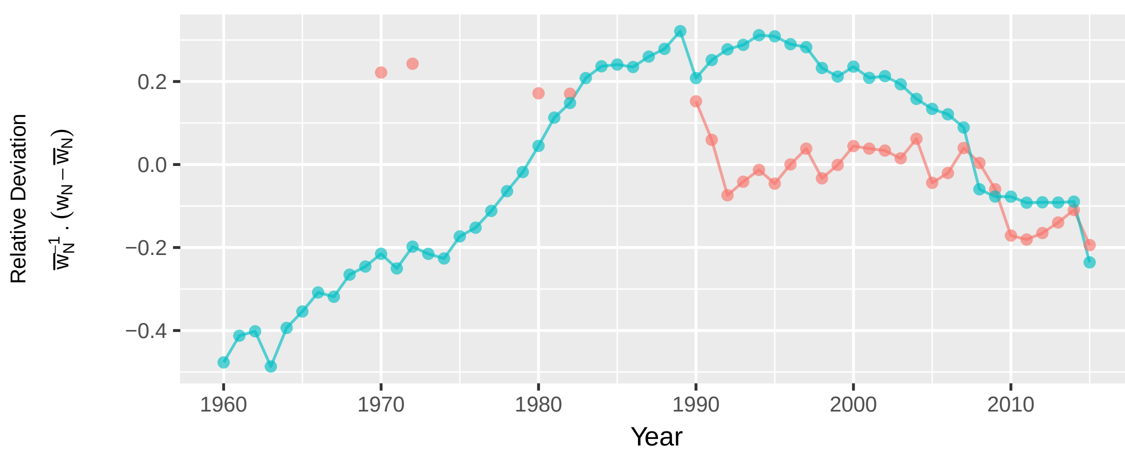
The MAVs for w_N were found and are shown in Figure 4.16, with missing values plotted as zero. As can be seen from these figures (and the figures of relative deviation), the trend in the underlying series was large relative to the noise. Since the series for LA exhibited a general downward trend—and that for Singapore exhibited a smooth increase, peak, and decline—it was decided to look at w_N as a process exhibiting proportional (i.e., exponential) growth. The proportional growth factor was found for the series at one-year intervals.

Figure 4.14: Change in w_N over time, by case.

(a) w_N: Original data series.



(b) Relative deviation of w_N re-centered the series to zero and rescaled the data by the mean.



(c) Proportional growth factor $k(w_N)$ for w_N over time.

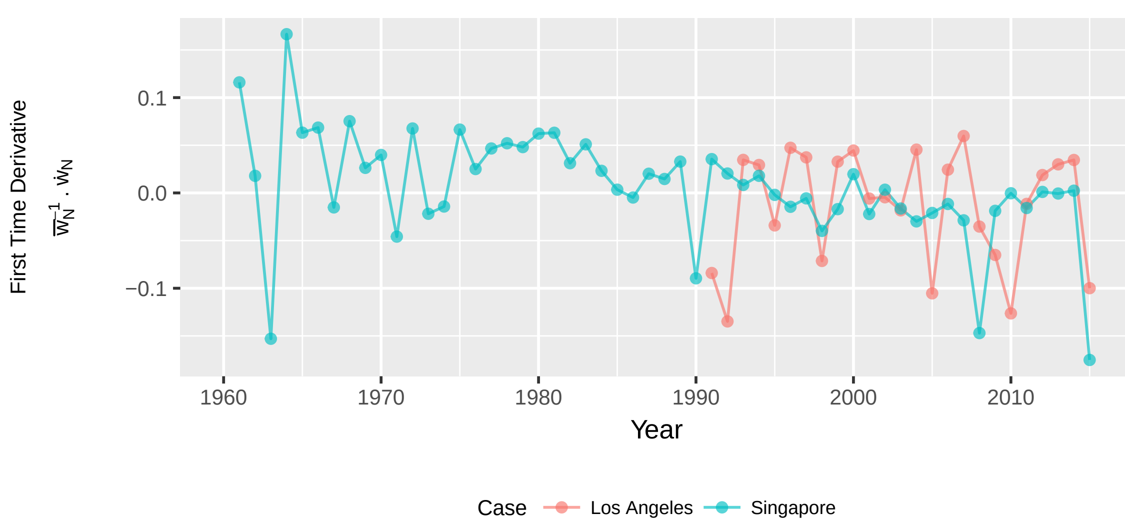
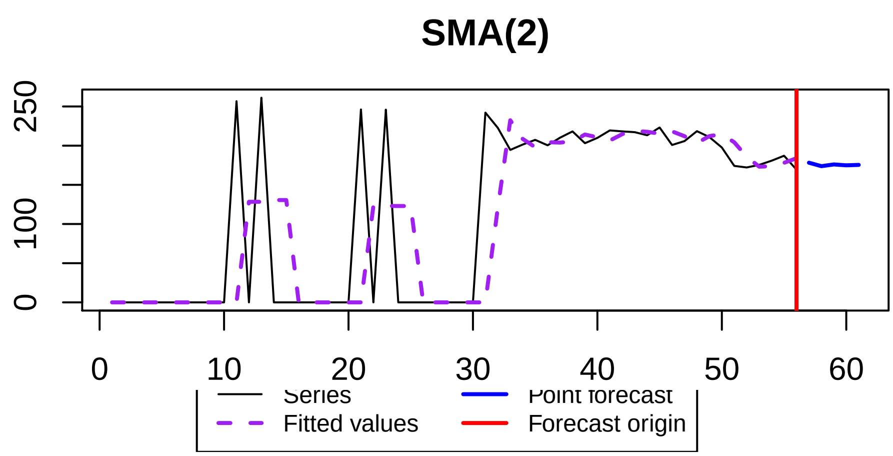


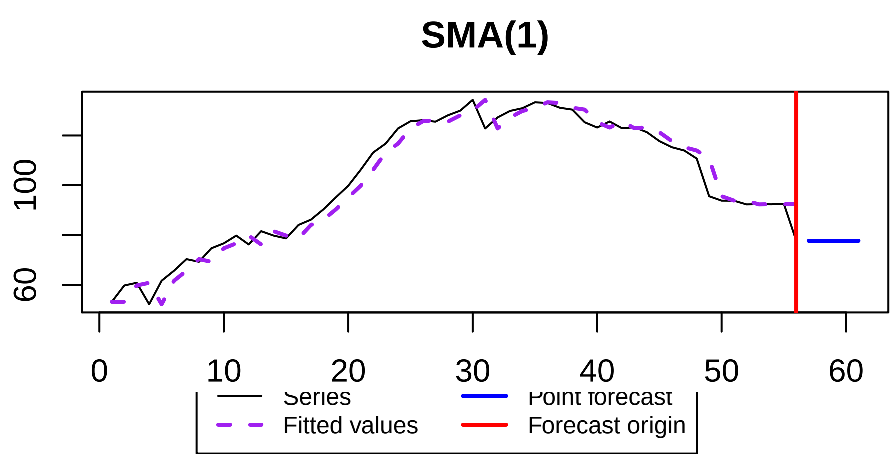
Figure 4.15: Moving average of w_N (missing values shown as zero).

Figure 4.16: Moving average of the water use intensity (w_N) for the two cases over the study period (1960–2015). Los Angeles was missing values, which are shown as zero in the plots. The moving average analysis algorithm identified a suggested period for smoothing (T_p), shown in parentheses above the plots; the analysis suggested a $T_p = 2$ for LA and $T_p = 1$ for Singapore. The moving averages follow w_N closely for both cases, indicating that the trends in w_N are not random.

(a) LA: MAV for w_N^{LA}, with $T_P = 2$



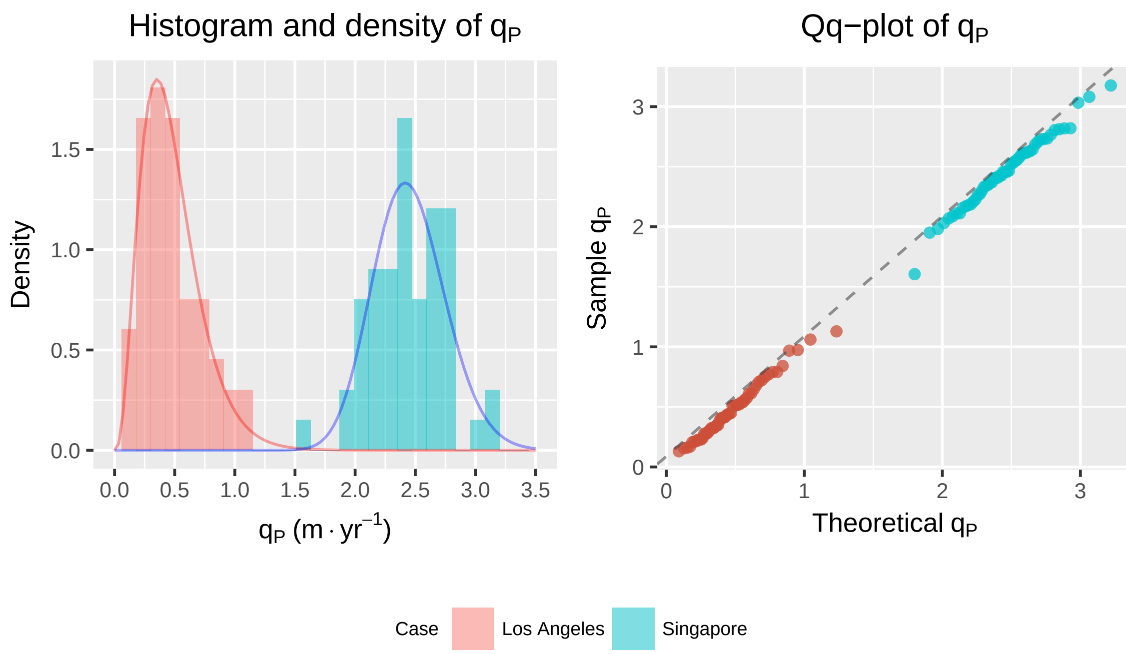
(b) Singapore: MAV for w_N^{SG}, with $T_P = 1$



4.8.4 Precipitation and Surplus

Figure 4.17: Case study historical data: annual precipitation q_P ($\mathrm{m} \cdot \mathrm{yr}^{-1}$), 1960–2016. Since $q_P \geq 0$, the density function and theoretical quantiles were generated from the Gamma function, with shape parameter $k = \mu^2/\sigma^2$ and scale parameter $\theta = \sigma^2/\mu$.

(a) Histogram, density, and qq-plots of annual precipitation height (q_P)



(b) Summary statistics

Statistic	Los Angeles	Singapore
n	55.00	55.00
μ	0.48	2.45
σ	0.24	0.30
γ	0.77	-0.05
Quantile	**Los Angeles**	**Singapore**
κ	-0.11	0.12
0%	0.13	1.61
25%	0.29	2.25
50%	0.43	2.46
75%	0.61	2.64
100%	1.13	3.18

Figure 4.18a shows annual precipitation in LA and Singapore over time. The means for each case have been plotted as horizontal dotted lines.

No overlap was observed in historical precipitation data across the two cases, as seen in the histogram in Figure 4.17a (with only minor overlap observed in the density functions). The mean precipitation for LA over the observed period was 0.477 $\mathrm{m} \cdot \mathrm{yr}^{-1}$($\mu(q_P^{LA}) = 0.477 m \cdot yr^{-1}$), while for Singapore the mean precipitation was much higher, with $\mu(q_P^{SG}) = 2.453 m \cdot yr^{-1} > 5 \cdot \mu(q_P^{LA})$.

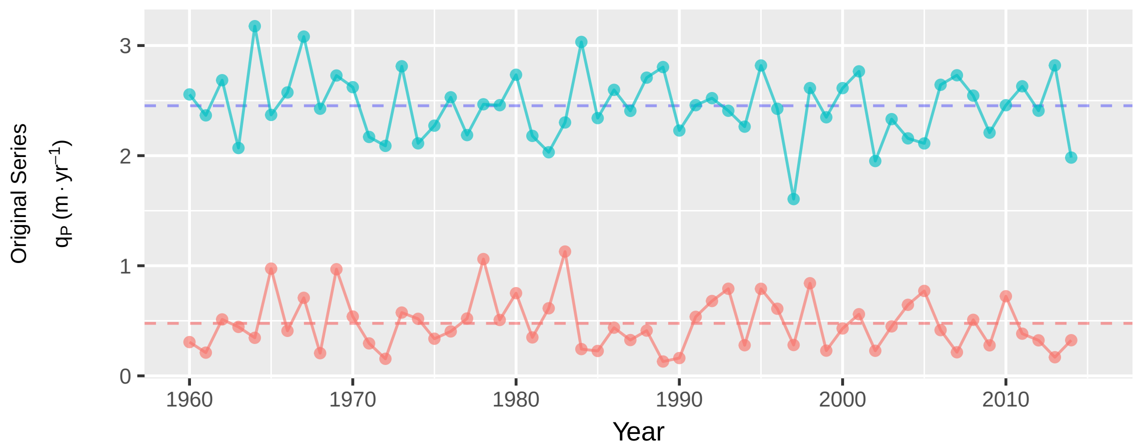
The minimum precipitation observed ($\min(q_P)$, where $\Pr[q_P < \min(q_P)] \leq 0$) was 0.130 m $\cdot$ yr^{-1}for LA and 1.606 for Singapore; $\min(q_P^{LA}) < 12 \cdot \max(q_P^{SG})$. The variation between the upper bounds for the two cases was lower than the mean values, with $\max(q_P)^{LA} < 3 \cdot \max(q_P^{SG})$. However, as noted before, no overlap was observed between precipitation values for the two cases; in other words, $\max(q_P^{LA}) \leftarrow \min(q_P^{SG})$.

Since $q_P \geq 0$, the density function and theoretical quantiles were generated from the Gamma distribution function rather than the normal function, with shape parameter $k_\Gamma = \mu^2/\sigma^2$ and scale parameter $\theta_\Gamma = \sigma^2/\mu$. As seen in the qq-plot of q_P in Figure 4.17a, the Gamma distribution was found to be a reasonable fit to observations from both cases, with points from both lines falling on the 45° line.

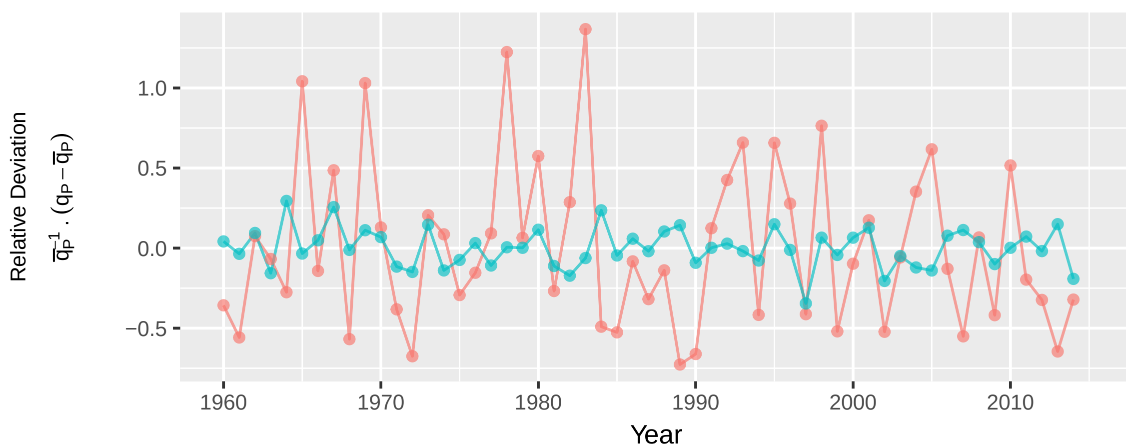
The first time derivative $\dot{q}_P$ was found and plotted in Figure 4.18c. Yearly changes in q_P were generally small relative to the expected value but the relative change was much larger for LA than Singapore. Moving average analysis (shown in Figure 4.19, with missing values plotted as zero) highlighted some periodic behavior in q_P^{LA} and a slight decline in q_P^{SG}. The periodic behavior in q_P^{LA} likely arises from the Southern Oscillation. Further analysis would use these patterns to normalize the residuals of q_P for further time series analysis.

Figure 4.18: Time series plots of q_P: original series, relative deviation, and first time derivative.

(a) q_P: Original data series.



(b) Relative deviation of q_P re-centered the series to zero and rescaled the data by the mean.



(c) Yearly changes in $\dot{q}_P$ over time, normalized by $\bar{q}_P$.

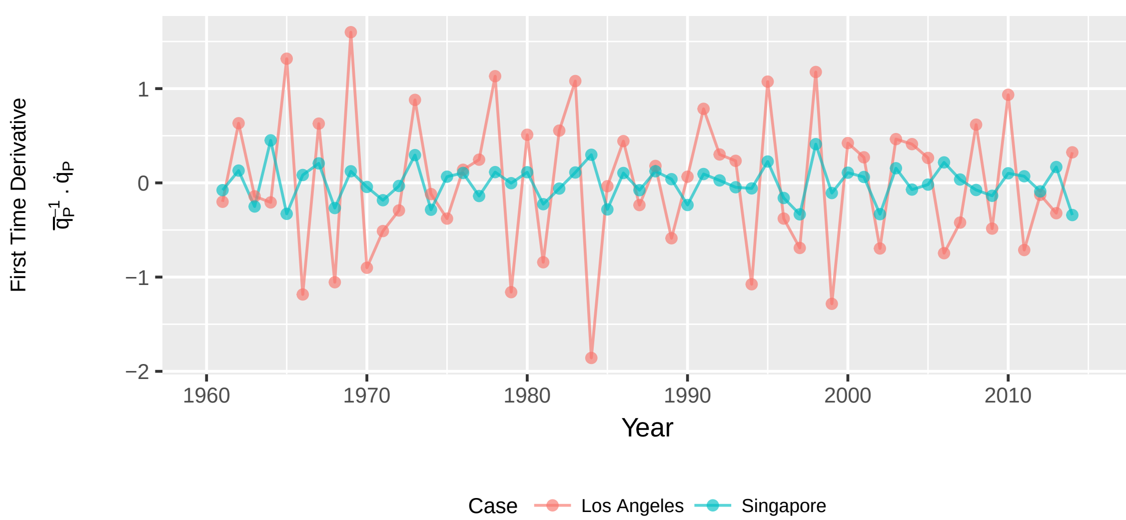
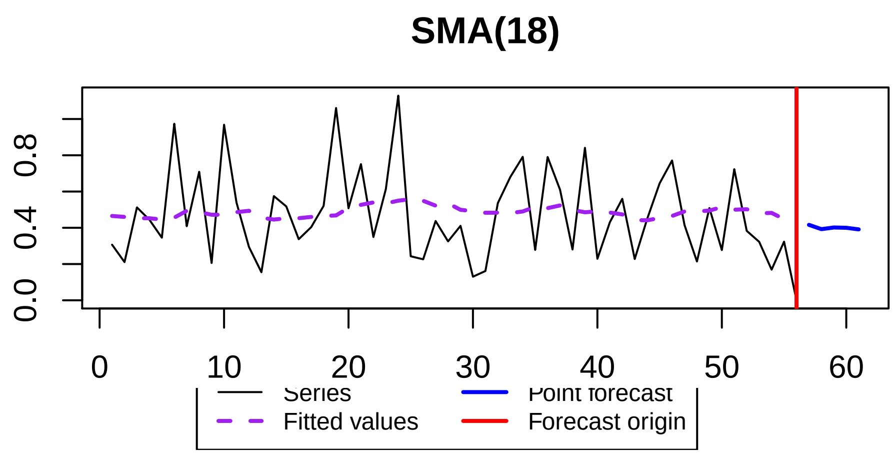


Figure 4.19: Moving average of annual precipitation (q_P) for the two cases over the study period (1960–2015). The moving average analysis algorithm identified a suggested period for smoothing (T_p), shown in parentheses above the plots; the analysis suggested a $T_p = 18$ for LA and $T_p = 22$ for Singapore. The results of the SMA are relatively constant over time, indicating that the model of q_P as a random, stationary process is reasonable for both cases. In Singapore, the SMA is smoother and flatter than in Los Angeles. However, recent years suggest that the mean of the series may be trending downwards, which could have implications for future water supplies. A similar trend can also be observed in the SMA for LA; these data do not take into account rainfall over the 2016–2017 winter season, which was historic. Interestingly, the SMA results for LA show some oscillation; further analysis would most likely link this to the Southern Oscillation.

(a) LA: MAV for q_P^{LA}, with $T_P = 18$.



(b) Singapore: MAV for q_P^{SG}, with $T_P = 22$.

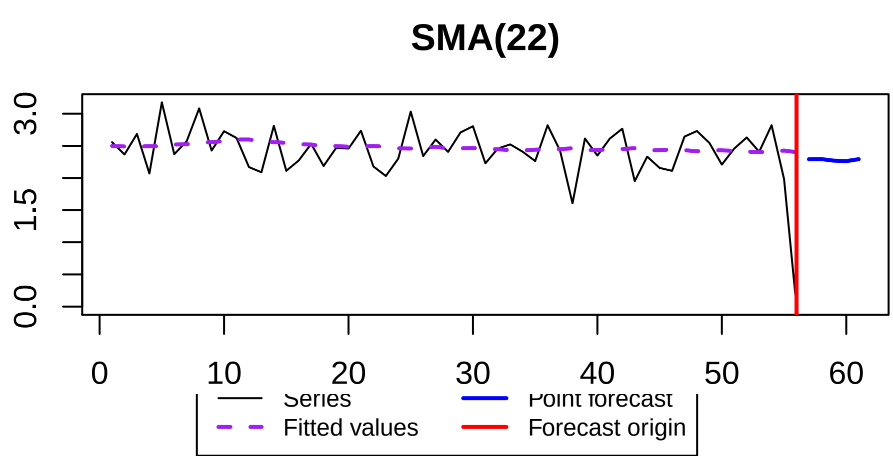
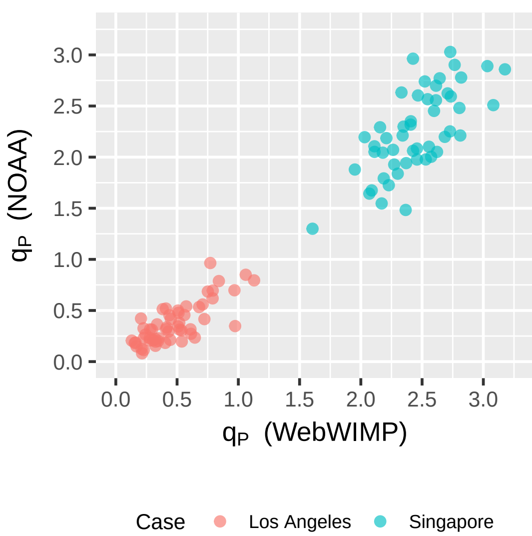


Figure 4.20: Water balance data for cases. Fig. 4.20a plotted q_P ($\mathrm{m \cdot yr^{-1}}$) data from NOAA against data from Matsuura and Willmott, while Fig. 4.20b plotted q_P against q_S (both from Matsuura and Willmott) [190]. The figure on the left, Fig. 4.20a, is clustered along the 45-degree angle; it illustrates the variation in data between different sources for a single metric. The figure on the right, Fig. 4.20b, shows the calculated annual climatic surplus relative to precipitation, and provides a rough index and upper bound of the fraction of precipitation that could be collected.

(a) q_P from two sources.



(b) q_S versus q_P.

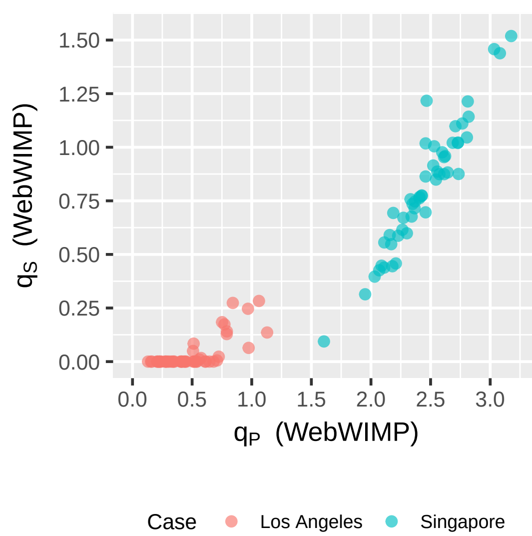
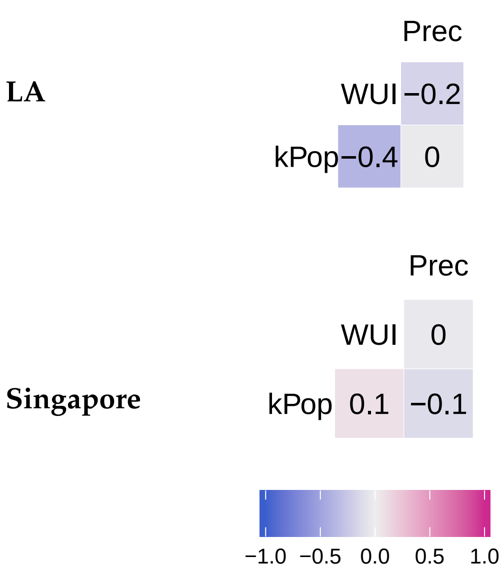
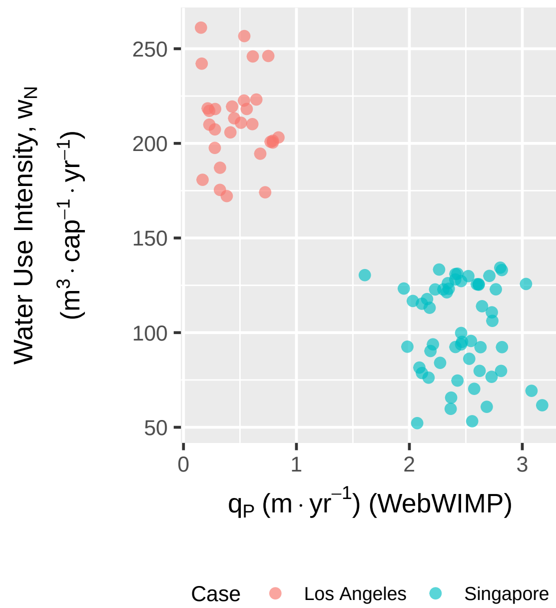


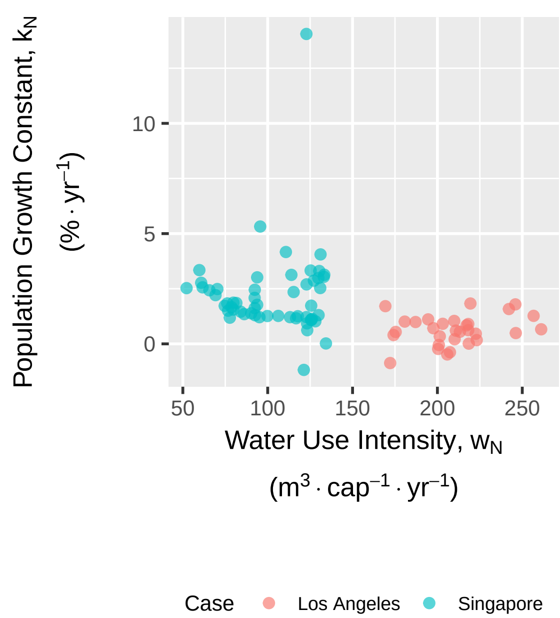
Figure 4.21: Correlation matrices and scatterplots of k_N, w_N, and q_P. Abbreviations: kPop - proportional growth constant (k_N), WUI - water use intensity (w_N), and Prec - annual precipitation (q_P).



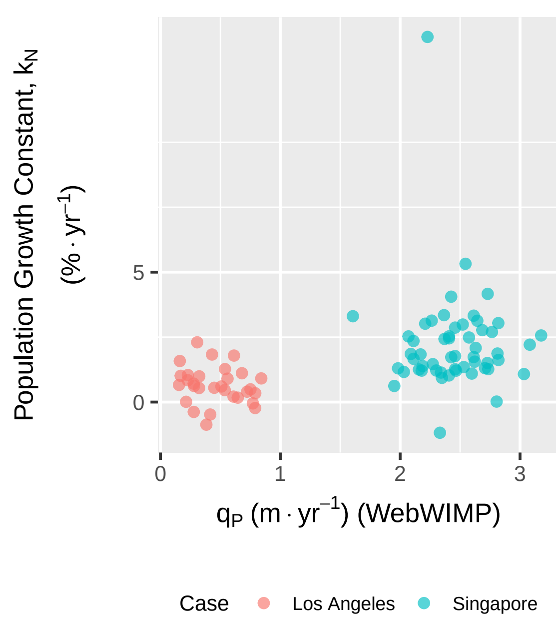
(a) Correlation



(b) Scatterplot of w_N versus q_P



(c) Scatterplot of w_N versus k_N



(d) Scatterplot of k_N versus q_P

4.8.5 Correlation

Figure 4.21a shows the correlation matrices for Los Angeles and Singapore, while Figures 4.21b, 4.21c, and 4.21d show the underlying scatterplot distributions. As seen in the correlation matrices in Figure 4.21a, correlation between k_N, w_N, and q_P for both cases was relatively low. In Singapore, a low negative correlation was observed between k_N and q_P, a low positive correlation was observed between k_N and w_N, and no correlation was observed between w_N and q_P. In LA, no correlation was observed between k_N and q_P; a negative correlation was observed between k_N and w_N, and a low negative correlation was observed between w_N and q_P.

These correlations are not statistically significant, especially since no cross-correlation was identified, but do suggest opportunities for further investigation. For instance, in Singapore, no correlation was observed was observed between w_N and q_P while in LA a low negative correlation was observed—indicating that water use in LA is more sensitive to changes in precipitation. At first approach, this might seem like an indicator of SUWM in LA. However, consider that water use intensity in LA is much higher than in Singapore—there is more room to decrease water use during years of low precipitation. Also, LA has experienced a greater number and severity of dry periods over the study period, which implies that there would be more data points in LA than in Singapore where residents had cut back water use due to drought.

An avenue for future research could be to compare the sensitivity of water use intensity to periods of low precipitation, and to conservation measures versus water pricing. This would require a more detailed breakdown of water use by user categories, as well as information about underlying policy.

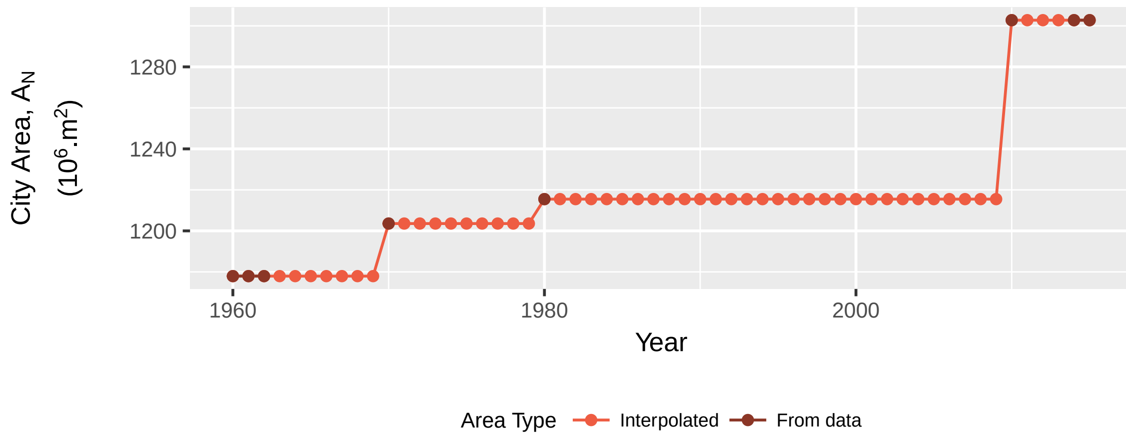
4.8.6 Area and Population Density

Few data points were obtained on city area for Los Angeles so data were interpolated so that R_{SS} could be assessed for LA over the study period, as seen in Figure 4.22a. For the interpolation, it was assumed that city area was constant between data points.

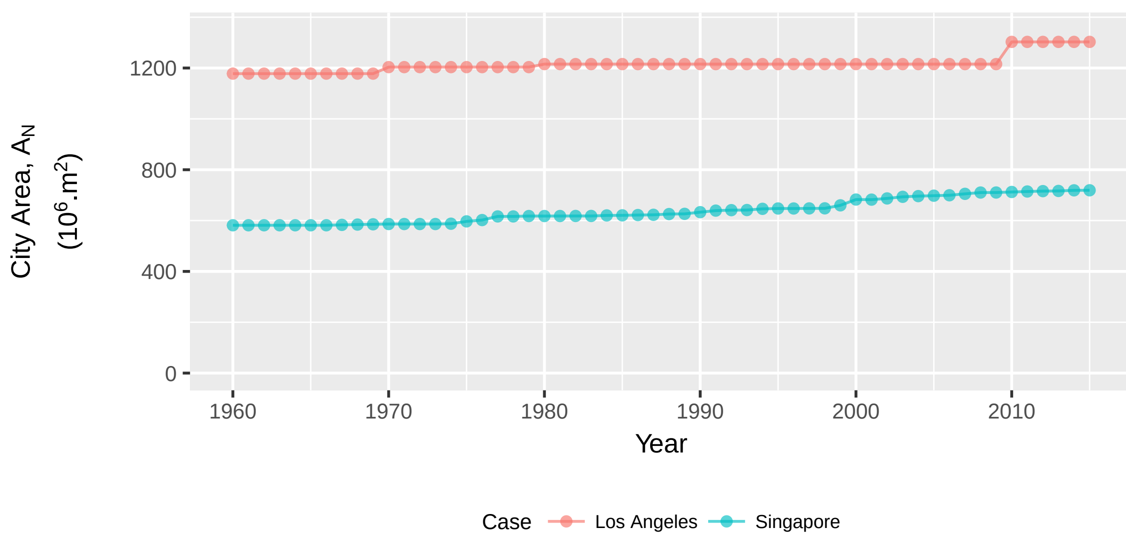
Population density over 1960–2016 is plotted in Figure 4.23. Population density, 1960–2016. For Los Angeles, the density was estimated from interpolated data of city area, due to missing data. Throughout the study period, population density has been higher in Singapore than in LA and has increased over time. In recent years, population density in Singapore has reached nearly 8000 residents per km^2—more than twice that in LA.

Figure 4.22: City area: time series data and interpolation.

(a) Los Angeles city area: data and interpolation. Few data points were obtained on city area for Los Angeles, so data were interpolated so that R_{SS} could be assessed for LA over the study period. For the interpolation, it was assumed that city area was constant between data points.



(b) City area, both cases.

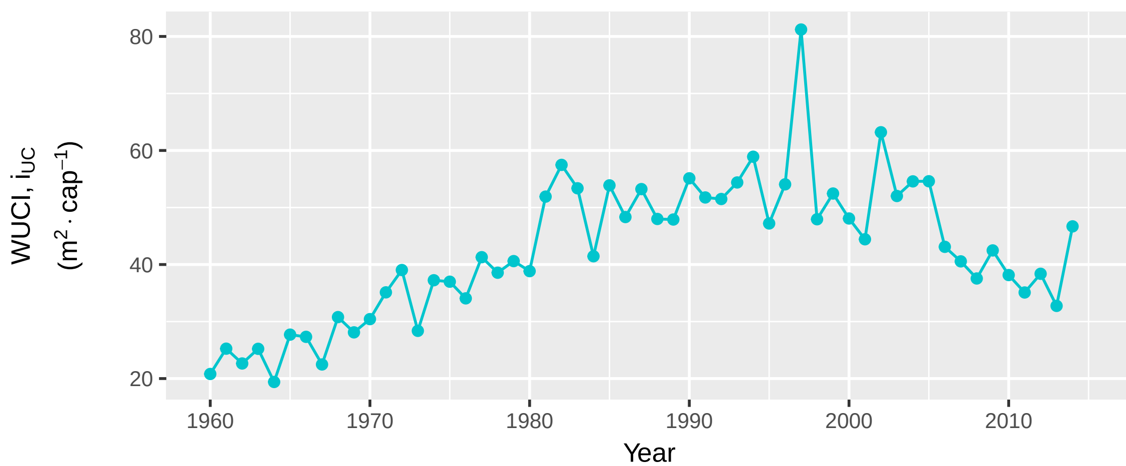



```
## NULL
```

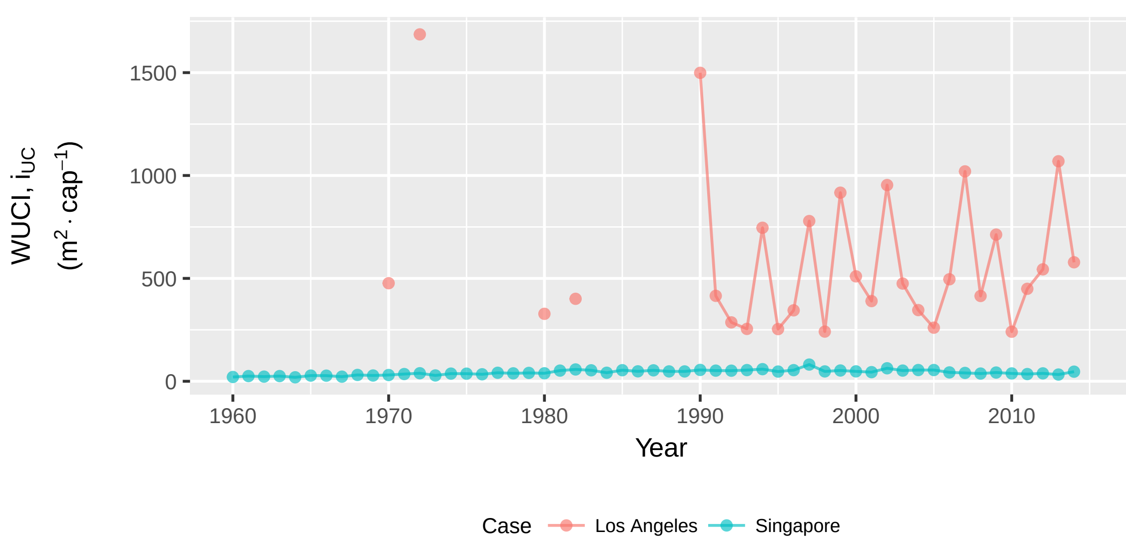
Figure 4.23: Population density, 1960–2016. For Los Angeles, the density was estimated from interpolated data of city area, due to missing data.

Figure 4.24: Water Use and Climate Index (i_{UC}), 1960–2016. Since WUCI in LA dwarfs that of Singapore over the study period, as seen in Figure 4.24b, a close-up of Singapore is shown in Figure 4.24a. In Singapore, the trend in WUCI roughly follows the general trend in water use intensity that was seen in Figure 4.14a. However, the same is not true for LA. While w_N in LA was observed to decrease over the study period, i_{UC} appears to have remained relatively constant or even increased over the past two decades—perhaps an amplification of underlying precipitation trends.

(a) Close-up of WUCI in Singapore.



(b) WUCI i_{UC}, both cases.

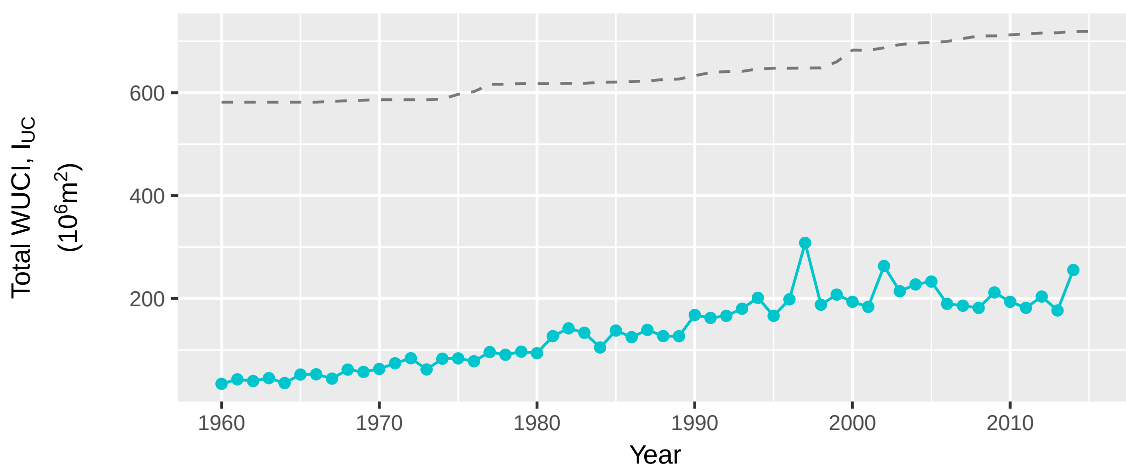


4.8.7 WUCI and Total WUCI

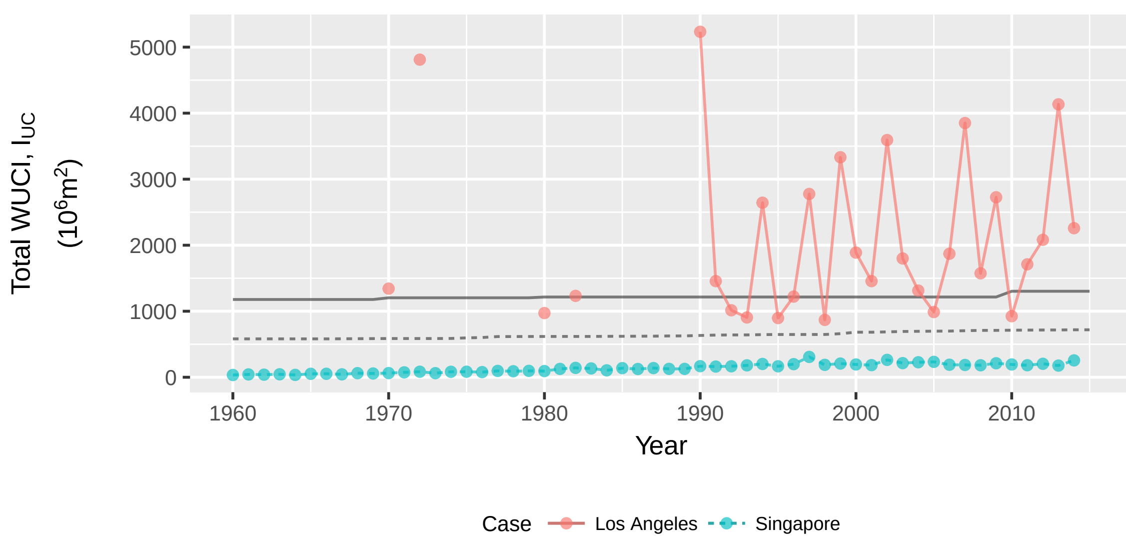
Water Use and Climate Index (i_{UC}), total WUCI ($I_{UC} = i_{UC} * N$), and the potential self-sufficiency ratio R_{SS} were calculated from historical data (or interpolated data, where necessary) and plotted in Figures 4.24–4.26. Figures 4.22–4.26 include a plot of both cases as well as a close-up of Singapore, Singapore, and LA, respectively.

Figure 4.25: Total Water Use and Climate Index I_{UC} over time (1960–2015). Total WUCI is the WUCI multiplied by population ($i_{UC} * N$) in units of area ($1\text{km} = 10^6 \cdot \text{m}$). City area for both cases is shown on the plots as black lines (solid for LA and dashed for Singapore). As seen in Fig. 4.25a, total WUCI has been only a fraction of city area throughout the study period. In contrast, total WUCI in LA was typically greater than city area, with a few exceptions, as observed in the bottom figure.

(a) I_{UC}, close-up of Singapore.



(b) Total Water Use and Climate Index I_{UC}, both cases.



In Figure 4.24a, the shape of per capita WUCI i_{UC} followed the general shape of w_N, increasing before 1990, leveling off around 1995, and then declining. In Figure 4.24b it was emphasized that i^{SG}_{UC} is generally an order of magnitude smaller than i^{LA}_{UC}. A slight upward trend was identified in i^{LA}_{UC} after 1993. Recall that:

$$i_{UC} = \frac{w_N}{q_P} = \frac{W_N}{N} \cdot \frac{1}{q_P} \tag{4.28}$$

Over the period 1993–2015, W_N decreased slightly, though remaining nearly constant, while N increased, resulting in a general decline in w_N. However, a downward trend in q_P was barely evident in the MAV for q_P^{LA} (seen in Figure 4.19a). However, the combined effect is much more noticeable as an upward trend in i_{UC}^{LA}—suggesting an increased per capita environmental footprint over time that was not evident in w_N.

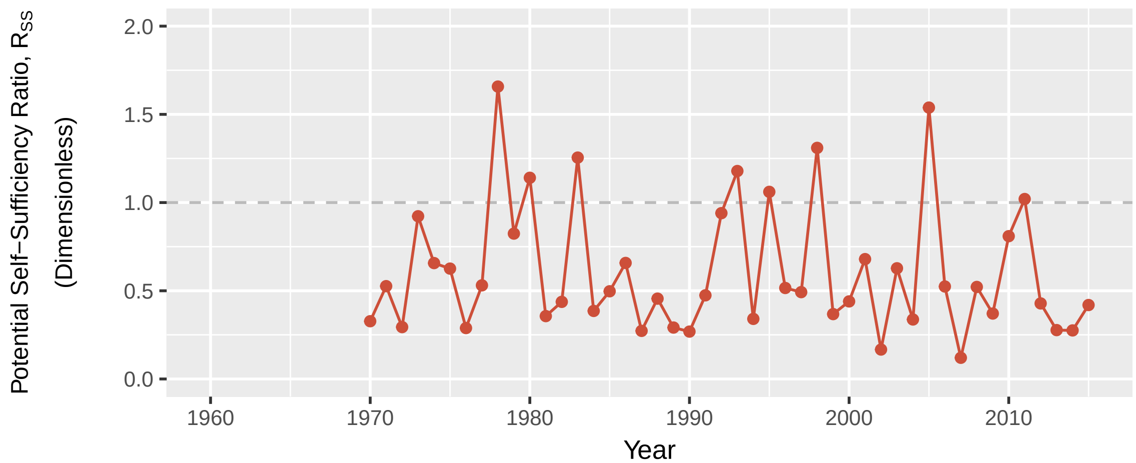
An upward trend was observed in the total WUCI I_{UC} for both cases, as seen in Figure 4.25. I_{UC}^{SG} more than doubled since 1960, and the same may be true for I_{UC}^{LA}. I_{UC}^{SG} in recent years has reached 30% of A_N^{SG}, which is high, especially considering that actual runoff may be only 50% of q_P or lower. While I_{UC}^{SG} has historically been smaller than island area, I_{UC}^{LA} may be as much as four or five times A_N^{LA}. Viewing I_{UC}^{LA} highlighted how much more difficult it would be for LA to become completely sufficient on water resources obtained from within the city boundary, especially considering that actual annual runoff in LA is typically only 25% of q_P or lower.

4.8.8 Potential Self-Sufficiency

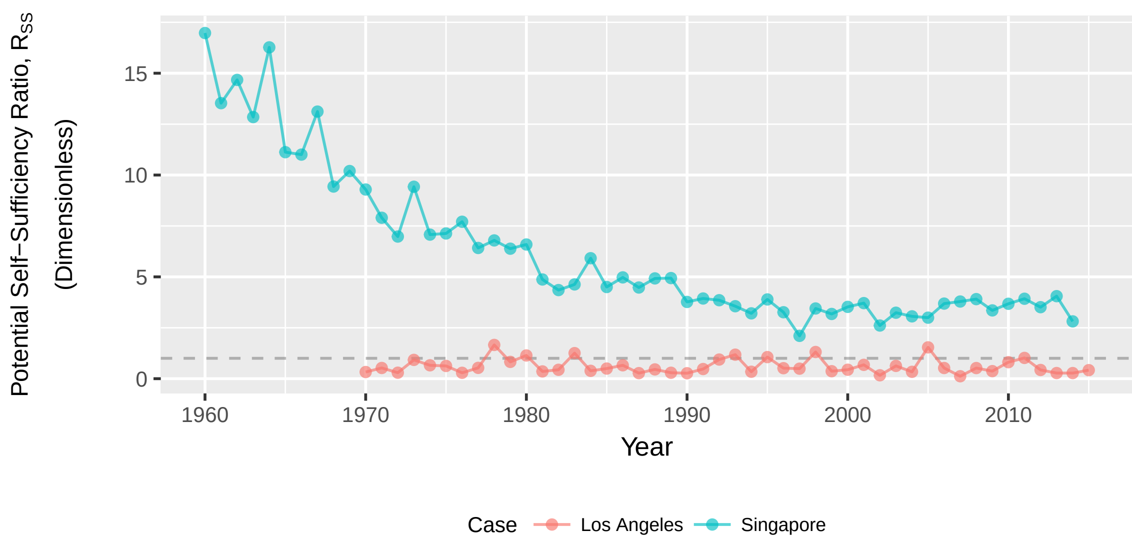
The self-sufficiency ratio R_{SS} shown in Figure 4.26 is equivalent to total WUCI normalized by city area, i.e., $R_{SS} \equiv I_{UC}/A_N$. The potential R_{SS} self-sufficiency of Singapore (R_{SS}^{SG}) decreased from 15 in 1960 to 3 in 2014, while R_{SS}^{LA} remained somewhat constant over this time period. The significant reduction in R_{SS}^{SG} arose from increases in both N_{SG} and w_N^{SG} over the same period; recent self-sufficiency in Singapore would likely be even lower if w_N^{SG} had not begun to decrease after peaking around 1990. Trends in R_{SS}^{LA} were hidden in the plot by the dominant influence of q_P and required further analysis to elucidate.

Figure 4.26: R_{SS}, 1960–2016 (with $R_{SS} = 1$ shown as a dotted line). The close-up of LA in the top figure shows that R_{SS} has typically been less than one throughout the study period, with several exceptions. In the bottom figure, R_{SS} in Singapore has decreased dramatically over time, but is still greater than unity.

(a) R_{SS}, close-up of LA.



(b) R_{SS}, both cases.



4.9 Summary of Work

The results of the analysis in Chapter 4.5 provide specific and general insight into the two cases.

The analysis found that historical population growth (k_N), water use intensity (w_N), and precipitation (q_P) were reasonably represented as random variables. Treating these metrics as random, it was found that the two cases differed greatly, with little overlap. The exception was population growth, in which the upper bound of k_N for LA overlapped the lower part of k_N for Singapore. This indicated that Singapore had grown faster than LA over the study period, which was confirmed by examining population for each case. Population growth was found to be reasonably represented by a normal distribution for both cases, while w_N and q_P were better represented by the Gamma distribution.

Population growth (k_N), water use intensity (w_N), and precipitation (q_P) were also examined as time series. The original data series for each metric were compared with their relative deviations and to proportional changes from year to year. Taking the relative deviation rescaled the data series and allowed relative trends to be compared between the two cases in spite of their different magnitudes. In addition, the proportional changes from year to year were also found.

Analysis of moving averages indicated that the mean values for k_N were not stationary, and indicated an average increase for both cases over the past 25 years. The MAV indicated a decrease in w_N for both cases over the past 25 years. In the first 30 years of analysis, w_N in Singapore increased; data for LA had numerous missing values over that period so no conclusions could be drawn.

The distribution of historical data for w_N was reasonably represented by the Gamma function, although deviations occurred in both cases. Examining the underlying time series shed some light into that deviation. In the original data series for both cases, w_N is much greater in LA than in Singapore. In Singapore, w_N increased before 1990 in Singapore, after which it decreased. The series for LA is missing observations before 1990; after 1990, w_N generally decreased and may have been decreasing as far back as 1970. For both LA and Singapore, MAV of w_N followed the shape of the underlying data series, indicating that these processes were strongly influenced by underlying trends. The relative deviations of w_N rescaled the series but maintained the general trends. However, this rescaling highlighted that the relative decrease in w_N after 1990 was greater in Singapore than in LA. When proportional yearly changes in w_N were examined, this highlighted an inflection point for growth in w_N in Singapore around the year 1995; growth was positive before and negative afterwards. Proportional change in w_N was examined for LA also trended towards negative values in the period for which data were available (after 1990) though with higher year-to-year variation.

The distribution of historical values of q_P was found to be well-represented by the Gamma function for both cases. The original data series indicated variation of a similar magnitude for both cases. However, taking the relative deviation indicated that the relative variability in q_P is much greater in LA than in Singapore. Examining the moving averages indicated that q_P for Singapore may be gradually decreasing. MAV for LA highlighted some oscillatory behavior, which may coincide with the El Niño-Southern Oscillation. However, compared precipitation appears to have been historically a more stochastic process than k_N or q_P.

The q_P data used in the case study analysis were drawn from a global dataset of 0.5-degree gridded data for terrestrial locations [190]. These data were compared with precipitation data from another sources (See Section 4.6.4), which has been shown in Figure 4.20a. The comparison suggested that the data were reasonable representations of annual precipitation. Figure 4.20b shows data for q_P used in the case study analysis with an estimation of annual surplus (q_S) from the same source [190]. As seen in Figure 4.20b, q_S is generally above 0.5 m $\cdot$ yr^{-1}for Singapore and below 0.25 $\mathrm{m} \cdot \mathrm{yr}^{-1}$for LA. Estimating annual runoff from watersheds is complex, and poorly understood for urbanized watersheds [40, 189, 79, 177, 196]; q_S is not necessarily equal to runoff. However, examining q_P relative to q_S highlighted that estimates of water volume using q_P should therefore be taken as an upper bound.

Correlations between historical distributions of k_N, w_N, and q_P were plotted and are shown in Figure 4.21. These plots highlight the historical differences between the two cases in terms of these underlying attributes, suggesting that historical WRM likely differed between the two cases. The results indicate little correlation amongst k_N, w_N, and q_P for Singapore; a slight positive correlation of 0.1 was observed for Singapore and a slight negative correlation was observed between k_N and q_P. In contrast, no correlation between k_N and q_P was observed for LA. However, a slight negative correlation was observed between w_N and q_P in LA, as was a mild correlation between k_N and w_N.

History can provide some hypotheses for the observed correlations. Over the study period, population in LA grew while water use intensity decreased. At the start of the study period, w_N in LA was very high, and had room to shrink. In contrast, w_N in Singapore at the beginning of the study period was very low. However, within the first several decades, Singapore developed rapidly. In the last couple of decades, w_N has begun to decrease; however, there growth in w_N at the beginning of the data series in Singapore still dominate the summary statistics.

A slight negative correlation between water use intensity and precipitation was observed for LA, while no correlation was observed for Singapore. This indicates that water use in LA is more sensitive to precipitation than in Singapore. Why might this be so? To start, w_N in LA is much higher than that in Singapore; in other words, there is more room for water use to decrease in LA than in Singapore, where water use is already more efficient. Additionally, precipitation is much lower in LA than in Singapore and has higher variability; this suggests that droughts occur more frequently and with greater intensity in LA than in Singapore.

The negative correlation between k_N and q_P in Singapore is more surprising. However, over the results from MAV suggested that population growth had increased over the last 25 years, while precipitation might be decreasing. One hypothesis to explore that relationship further could be that when q_P is lower than usual, the government in Singapore increases incentives to reduce population growth, or removes incentives for population growth. Alternatively, population growth might have decreased as precipitation increased; population growth did decrease in the first 30 years of the study period, but there is no indication that precipitation increased.

Interestingly, the greater variability in proportional growth in w_N in LA was similar to that observed before 1977 in Singapore. Examining histories of Singapore's water supply provide additional context for examining this further; before 1977 Singapore was a young country and actively developing its water supply [305]. The climate of Singapore is less variable than that of LA. Water use in LA is much higher, and therefore there is greater room for reduction during a drought.

However, in general, k_N and w_N were not highly correlated with q_P for either case over the study period, potentially highlighting the mediating role of infrastructure.

In addition, historical WUCI (i_{UC}), total WUCI (I_{UC}), and the potential self-sufficiency ratio (R_{SS}) were also assessed from historical values. When the data series for the two cases are plotted together, LA dwarfs Singapore in plots of i_{UC} and I_{UC}, while Singapore dwarfs LA in the plot of R_{SS}. Figure 4.24 shows the plot of i_{UC} (with a close-up of Singapore). WUCI for Singapore followed the general trend for w_N. In contrast, while w_N has decreased in LA, i_{UC} shows greater variability and may even be increasing.

Historical total WUCI is plotted in Figure 4.25. Total WUCI in Singapore follows the general trend observed for total water use, seen in Figure 4.12, and is smaller than the city area. Total WUCI in LA dwarfs that for Singapore, which is unsurprising considering that water use intensity (in the denominator of total WUCI) and in LA is higher than in Singapore, while q_P (in the denominator of I_{UC}) is much lower, and less than one. While total water use in LA remains relatively constant over the study period, I_{UC} exhibits substantial variability, and appears to more closely follow the dynamics of i_{UC}. In contrast to Singapore, I_{UC} for LA is generally greater than city area. While population appears to dominate the trend observed in I_{UC} in Singapore, precipitation appears to dominate the I_{UC} observed in LA.

The potential self-sufficiency ratio is for both cases are plotted as time series in Figure 4.26, with a close-up of LA. The self-sufficiency ratio in Singapore has decreased over time and has been relatively smooth. From 1990 onwards, $R_{SS} < 5$ in Singapore. In contrast, potential R_{SS} in LA is nearly always less than 1, and generally less than 0.75. This indicates that Los Angeles cannot expect to rely as heavily on local resources as in Singapore. Higher variability was observed in R_{SS} in LA, and the mean value appeared more stable, than in Singapore.

4.10 Future Work

After finding the proportional growth in w_N from year to year, the resulting series threw local several local minimums into relief[16]. Interestingly, these low points were typically only 1–2 years in a row. After each low period, w_N was observed to "rebound". This behavior suggests that these low points are responses to sudden changes in external factors, such as low precipitation, that apply an acute pressure on water use intensity but do not actually change water demand (such that, after the pressure is lifted, the latent demand causes w_N to return to its previous state). Future work could test that hypothesis, which could provide useful diagnostics to decision-makers.

In recent years, the self-sufficiency ratio of Singapore has begun to approach that of LA, suggesting that the pressures on existing water resources had increased. In consequence, Singapore's PUB began to increase stormwater collection, water recycling, and seawater desalination—all measures that had previously been considered by water-scarce LA. Additionally, a notable decline in w_N has been observed in Singapore since 1995. In contrast, the self-sufficiency ratio in LA has been low for some time (indicating a relative scarcity in local water resources). While similar

[16]For instance: 1961, 1962, 1967, 1971, 1990, 2008, and 2015 in Singapore; and 1991, 1995, 1998, 2005, 2010, and 2015 in LA.

solutions had been considered and reconsidered by LA throughout the decades, it is only recently that LA has taken actions towards implementation. Perhaps this is because water resources in LA had been reportedly stressed since the early 1900s, at which point in time water recycling and seawater desalination technologies (if one can refer to them as such) were not as developed as today. Because of this, LA had to seek out other solutions to its water scarcity, namely importing water. LA was so successful in such projects that by the 1960s, when desalination technologies were beginning to be viable on a large scale, LA had access to an abundance of freshwater. And so it is only as those sources have become less certain that LA has returned to local resources.

While the historical context in the preceding paragraph is intriguing, the analysis in Chapter 4, there are several steps that would need to be taken to provide a more comprehensive comparison of self-sufficiency for the two cases. The self-sufficiency ratio assessed in this chapter was that of an *ideal*, *potential* R_{SS}: i.e., the R_{SS} each case would have if 100% of precipitation that fell within the city boundaries was collected, treated, and distributed, with zero losses. It would be interesting to further characterize the potential resources from runoff within the city, which would require application of at least a simple hydrological model that distinguishes between pervious and impervious surfaces, topography, and other features [200, 335].

It is clear from the results that LA cannot be self-sufficient with respect to local precipitation alone. The local climatic water availability indicates a larger water footprint. This highlights the need for external water supplies, and could be used as a predictor of transboundary water imports. The next step would be to compare this footprint with the actual catchment area of LA's watershed infrastructure. In other words, in addition to *potential* self-sufficiency, it would also be useful to assess *actual* self-sufficiency, with respect to water sources to which each city already had access—and more specifically, for which infrastructure already existed. Those sources can include those to which the city had long-term access, which are not necessarily local; e.g., for LA those sources would include water from the Owens and Colorado Rivers.

4.11 Conclusions

While the analysis performed in this chapter has its limitations, it was able to highlight that a simple historical analysis of a few common metrics can be performed and provide some insight—at least for two "similar but different" cases. The historical analysis established trends that can be used for future simulation and, with a few tweaks, could be used for scenario analysis. It also highlighted that, while LA and Singapore were not similar in 1960, they were substantially more similar in recent years. This conclusion was supported by similarities between long term visions for urban water management for both cases.

At first glance, Los Angeles and Singapore may not have seemed like the most obvious choice for in-depth, comparative analysis of water resources. Not only do LA and Singapore have very different climates, their approach to water management is widely viewed in popular culture to be completely different. However, these two cities were identified through clustering analysis to be more similar to each other than to many other cities in the UrbMet dataset.

Analysis of historical population growth, water use intensity, and precipitation upheld the expectation that WRM in LA and Singapore would likely exhibit substantial differences. However, while significant differences were observed, analysis of the data as time series was able to identify similarities in the two cases. Therefore, the results demonstrate that these methods can provide useful insight into *how* similar and different cities in Type 4 area, and to highlight specific questions for further analysis and comparison.

While the individual results may not prove surprising to water experts of either city, the insight provided by the comparative analysis might be non-intuitive.

For instance, the role of the constant threat of water scarcity in the history of water resources in LA may be familiar to its residents. However, those in LA may be surprised to learn that similar factors shaped water infrastructure in Singapore, in spite of apparently abundant precipitation. In the last several decades, Singapore has developed a reputation for sustainable and efficient use of water resources. However, residents in Singapore may be surprised to learn that population growth continues to drive their water footprint upwards, or that the ideal, potential R_{SS} has begun to converge with that of LA.

In conclusion, the results, analysis, and discussion illustrated that even this rough analysis using simple performance metrics:

- Provided insight into water resources management in LA and Singapore, individually;
- Demonstrated that the same method can be applied to two, apparently different, case studies from Type 4;
- Provided concrete points for quantitative comparative analysis to assess *how* and *in what ways* water resources management in these cities differs.

The next chapter, Chapter 5, uses the results from Chapter 4 to simulate potential futures for water resources management in the two case studies.

Chapter 5

Small-*n* Comparative Analysis of Simulated Futures Los Angeles and Singapore

5.1 Overview

In Chapter 4, the typology developed in Chapter 3 was used in the selection of Los Angeles and Singapore for case study analyses. These cities were chosen for being "similar but different"—from the same type within the typology but representing different climate regimes.

In Chapter 4, data for both cases were collected from various sources to assess the potential for self-sufficiency over the period 1960–2016 using a simple model of urban water supply and demand. The analysis in Chapter 4 also examined to what extent each of the supporting metrics—the population growth constant (k_N), water use intensity (w_N or WUI, in $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$), and annual precipitation (q_P in $\mathrm{m} \cdot \mathrm{yr}^{-1}$)—could be modeled as stationary random variables. The analyses found that all of the series could be reasonably modeled as random and stationary, but also that all of the series exhibited some degree of non-stationarity, with k_N and w_N exhibiting the greatest deviation. The results of these analyses, when compared across case studies, were able to quantify some of the key differences between the two cases.

Results from Chapter 4 were used in the analyses presented in this chapter to establish "realistic" ranges for k_N, w_N, and q_P, which were used in the simulation of water demand and potential stormwater supply for both cases over two periods:

1. the historical study period, 1960–2016; and

2. the present and near future, 2017–2040.

The results from the simulations over 1960–2016 were compared with actual historical data over the study periods to assess the limitations of the model of k_N, w_N, and q_P as random and stationarity over that period.

The simulations over the period 2017–2040 were compared with long-term water plans published by case study water utilities to assess the degree to which the simple model of urban water supply and demand could provide meaningful insight into portfolios of technology and policy that had been developed through more complicated approaches.

The analysis in this Chapter has four parts:

1. simulation of water supply and demand profiles for both cases over the historical study period and into the future,

2. assessing the potential for self-sufficiency from local supplies over these two periods,

3. comparison of simulations over the historical period with actual trends, and

4. comparison simulations over the future period to actual projected portfolios for each case.

5.2 Simulation

To explore the likelihood of certain scenarios, simulations were run using $k_{runs} = 2000$ trials.

The seed for either `rnorm` or `rlnorm` could be set prior to running the function using the function `set.seed(`n_S`)`, where n_S could be any number between 0 and 3000.

A vector of seed numbers $\mathbf{n}_S$ of length k_{runs} was generated using the `runif` function from the `stats` package in `R`, which creates a random uniform distribution over a specified interval (the interval specified was [0, 3000]).

Before each trial, the seed value for the random distribution was set to the value in $\mathbf{n}_S$ corresponding to the trial number.

Each trial, generated a vector of values of length equal to the number of years in the simulation period T, beginning with initial year t_i and ending with final year t_f. Each trial vector therefore had length equal to $length(T) = t_f - t_i$. These vectors were then combined into a matrix $\mathbf{M}$ of dimensions $length(T) \times k_{runs}$

Trial matrices were generated for the population growth constant k_N, annual precipitation q_P, and water use intensity w_N. For the first-order analysis, water use intensity was assumed to be independent of precipitation; however, this assumption is not realistic in practice.

5.2.1 Generating Random Distributions

Random distribution functions were generated for the population growth constant k_N, annual precipitation q_P, and average annual water use intensity w_N.

For k_N and w_N, a normal distribution was assumed. For each of these variables the mean μ and the standard deviation σ calculated in Chapter 4 were entered as parameters into the `rnorm` function from the `stats` package in `R`. The `rnorm` function also requires that the number of observations n_O be specified.

For q_P and w_N, a Gamma distribution was assumed. for q_P was generated using the `rgamma` function from the `stats` package in `R`. The Gamma function took as arguments a shape and scale parameter (which were calculated from the mean and standard deviation of the historical data series).

Population

Using $\mathbf{M}_h$—the trial matrix for the population growth constant k_N—a simulation matrix $\mathbf{M}_N$ was calculated for population assuming exponential growth as per Equation 4.21 and starting values $N_0 = N(t_i - 1)$ and $h_0 = h(t_i - 1)$.

$$\mathbf{M}_W(t, j) = \mathbf{M}_N(t-1, j) \cdot \mathbf{M}_{w_N}(t-1, j) \tag{5.1}$$

Total Water Demand

Using $\mathbf{M}_N$ (the simulation matrix for population, calculated as described in Chapter 5.2.1) and $\mathbf{M}_{w_N}$ (the simulation matrix for water use intensity) the simulation matrix $\mathbf{M}_W$ was calculated for total water demand using element-wise multiplication:

$$\mathbf{M}_{W_N} = \mathbf{M}_N \cdot \mathbf{M}_{w_N} \tag{5.2}$$

Local Water Supply

Using $\mathbf{M}_{q_P}$ (the simulation matrix for annual precipitation q_P) and area $A_N = A_N(t_i - 1)$ (the simulation matrix for water use intensity) the simulation matrix $\mathbf{M}_{Q_{local}}$ was calculated for total water demand using constant multiplication:

$$\mathbf{M}_{Q_{local}} = A_N \cdot \mathbf{M}_{q_P} \tag{5.3}$$

Self-Sufficiency

The potential for supply from stormwater capture was then compared with surface water resources to which the case studies had long-term access, considered to be land ownership or political or legal agreements extending past 50 years. Using $\mathbf{M}_{W_N}$ (the simulation matrix for total water demand, described in Chapter 5.2.1) and $\mathbf{M}_{Q_{local}}$ (the simulation matrix for local water supply Q_{local}, described in Chapter 5.2.1) the simulation matrix $\mathbf{M}_{R_{SS}}$ was calculated for self-sufficiency using element-wise division:

$$\mathbf{M}_{R_{SS}} = \frac{\mathbf{M}_{Q_{local}}}{\mathbf{M}_{W_N}} \tag{5.4}$$

5.2.2 Replicability and Scalability

To promote the replicability and scalability of these analyses, several data structures and methods were created for the `R` programming environment and are provided in Appendix D.2. These functions were written to take as input either:

- a single time series
- a dataframe of time series data, or
- a matrix of summary statistics,

and produce as output a data object for each case with simulations for each variable and type of projection (historical and future).

5.3 Simulation Results and Discussion

5.3.1 Simulation Results versus Historic Values

Figures 5.1a–5.5 show the results of the simulation over the historic period for both cases. Since the proportional growth constant (k_N), water use intensity (w_N), and annual precipitation (q_P) were all assumed to be random, stationary variables for the purposes of this analysis, they appear as bands in Figures 5.1a–5.1c. The figures provide a rough visual assessment of the limitations of that simple model.

As seen in Figure 5.1c, the simulations that best match historical data are those for q_P, which follows the findings in Chapter 4.8.4. However, the simulations for Singapore do not quite capture the minimum precipitation in Singapore. The series for LA seems a reasonable approximation, but would likely change if the Southern Oscillation were taken into account.

For water use intensity, the simulations are not the best match to observed historical data. Since Chapter 4.8.3 found that the mean values of the series were not stationary in time, the assumption of stationarity leads to simulations that are unlikely to capture future trends, as seen in Figure 5.1b.

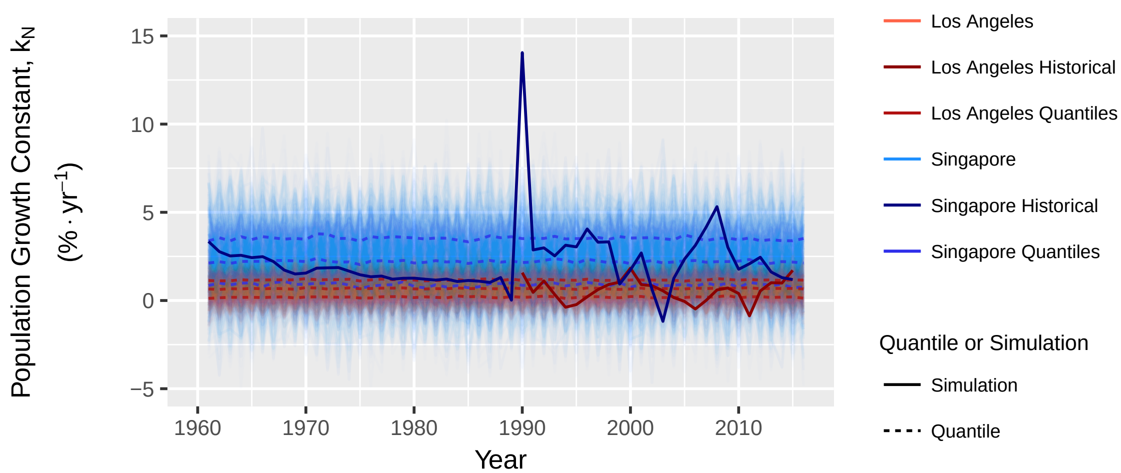
As discussed in Chapter 4, especially 4.8.2, k_N was better represented as a random variable than w_N; as with q_P, the simulations appear to be a reasonable approximation of historical values.

The simulated values for k_N were used in calculating population, shown in Figure 5.2 from a simple exponential growth model as described in Chapter 5.2.1. Interestingly, the observed historical data appear to be within the inter-quartile range for most t. However, the simulations underestimate the population in LA, though this may be because of missing values before 1990; if the growth constant over that period was higher than in recent years, this would explain the discrepancy.

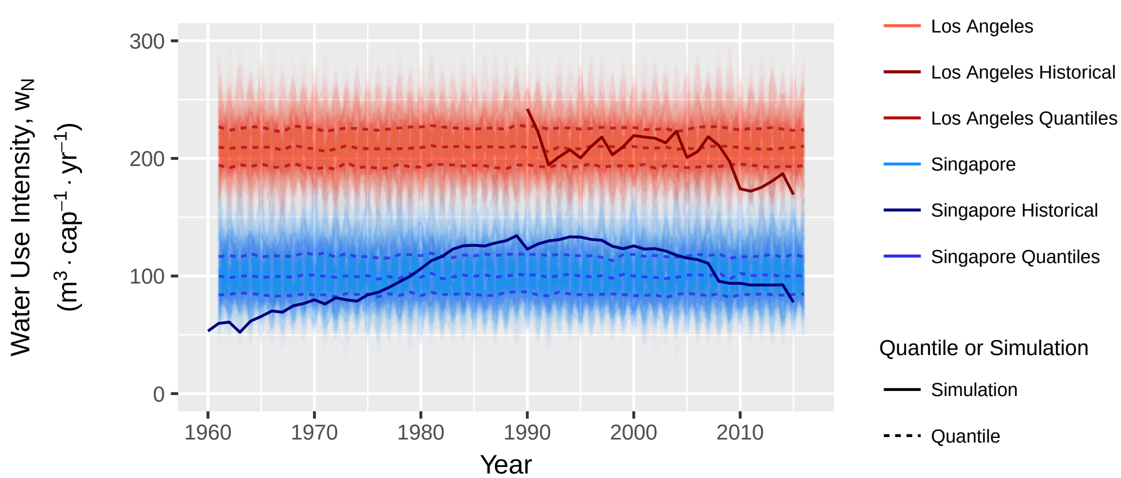
The simulations for w_N were multiplied by those of population to obtain estimates of total water demand (W_N), seen in Figure 5.3. This method leads to an underestimation of W_N before 2008, and an overestimation of demand in recent years. This is likely due to missing values

Figure 5.1: Simulation results over the historical period, 1960–2015: population annual proportional growth constant (k_N, in $\% \cdot yr^{-1}$); water use intensity (w_N in $m^3 \cdot yr^{-1} \cdot cap^{-1}$); and annual precipitation height (q_P, in $m \cdot yr^{-1}$).

(a) Population annual proportional growth constant (k_N, in $\% \cdot yr^{-1}$).



(b) Water use intensity (w_N in $m^3 \cdot yr^{-1} \cdot cap^{-1}$).



(c) Annual precipitation height (q_P, in $m \cdot yr^{-1}$).

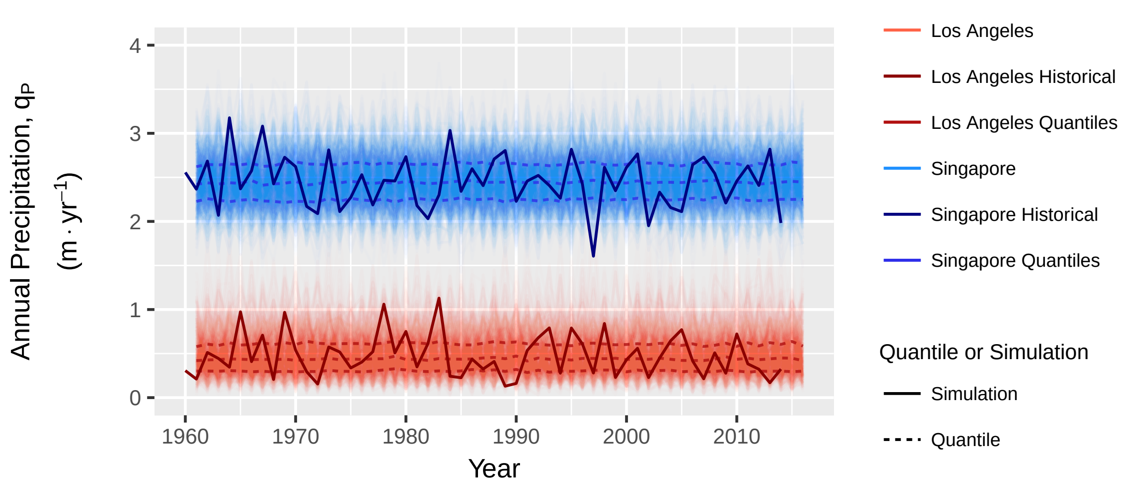


Figure 5.2: Simulation results over the historical period, 1960–2015: population (N in capita).

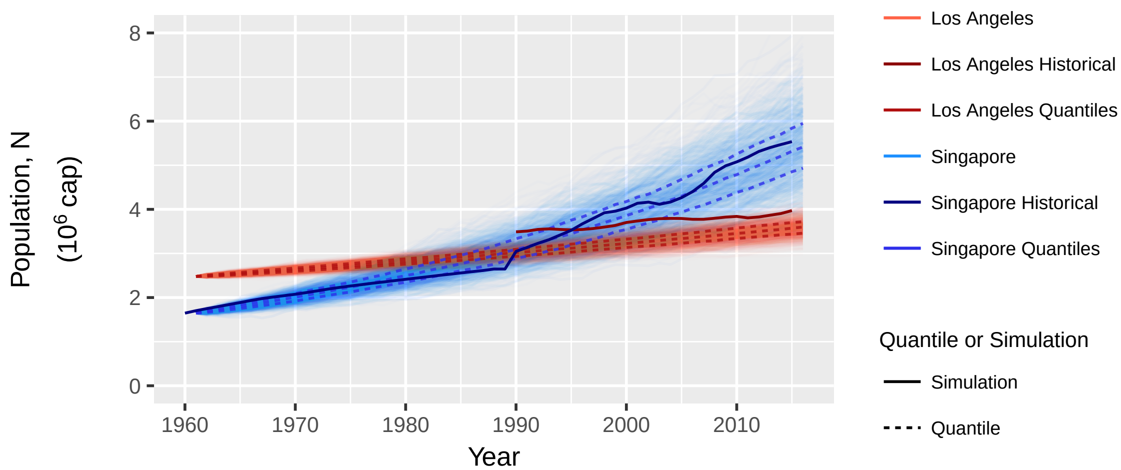
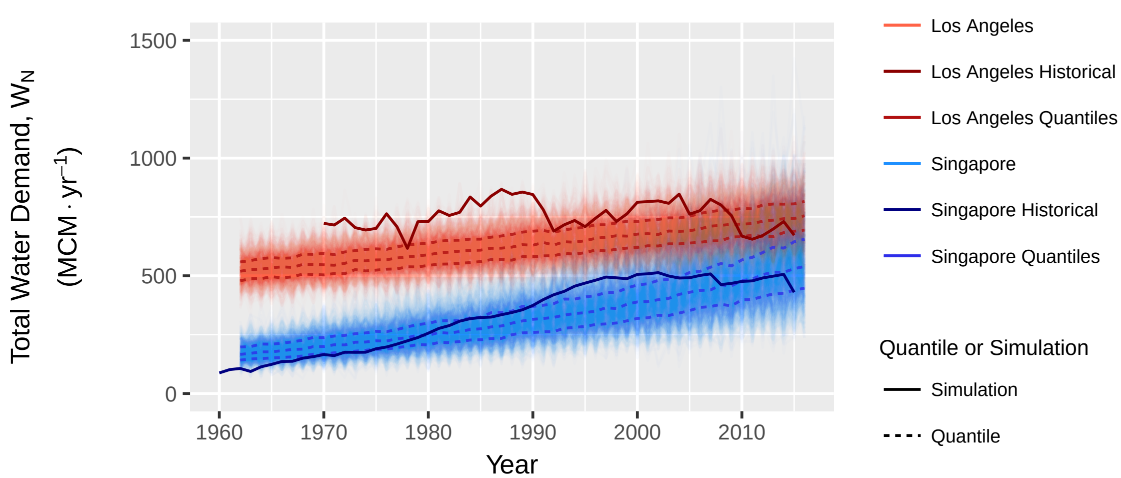


Figure 5.3: Simulation results over the historical period, 1960–2015: total water demand (W_N in 10^6 m$^3 \cdot$ yr^{-1}, where $W_N = N \cdot w_N$).



before 1990 in the underlying data series for w_N in LA, when WUI was on average likely higher than in recent years. In Singapore, the simulations tended to overestimate W_N before 1975 and underestimate it between 1990–2004. If the actual observed trend in w_N continues, this method for estimating W_N will likely begin to overestimate total demand in the near future.

Estimates of the potential total water supply from rainfall, Q_R, were obtained by multiplying simulations of q_P by the observed or interpolated historical city area. Since simulations of q_P were found to be a reasonable match to historical data, as discussed earlier in this section, and since historical data were used to calculate Q_R, the resulting simulations (seen in Figure 5.4) appear to be a reasonable match to the calculated historical values.

Figure 5.4: Simulation results over the historical period, 1960–2015: potential supply from rainfall (Q_R in 10^6 m$^3 \cdot$ yr^{-1}, where $Q_R = q_P \cdot A_N$).

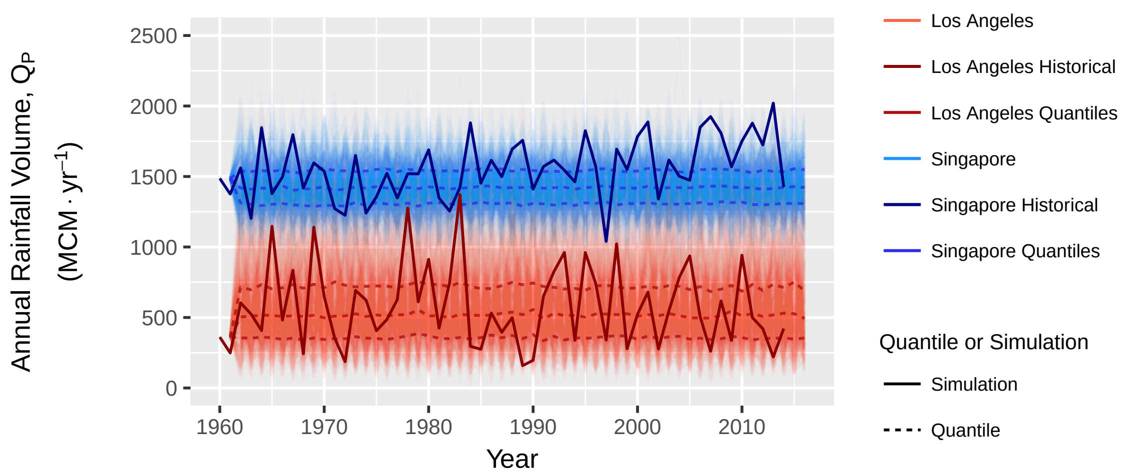
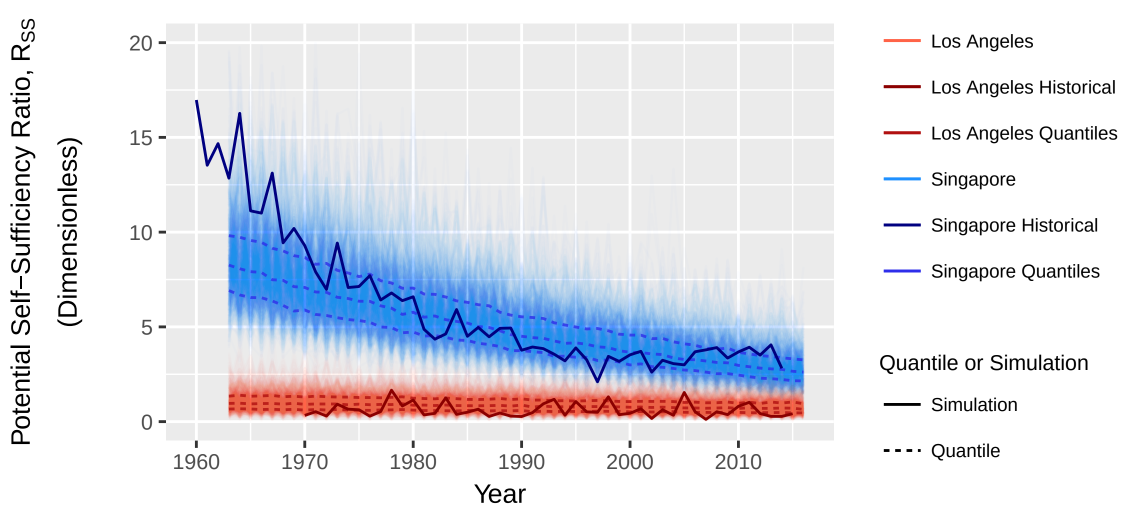
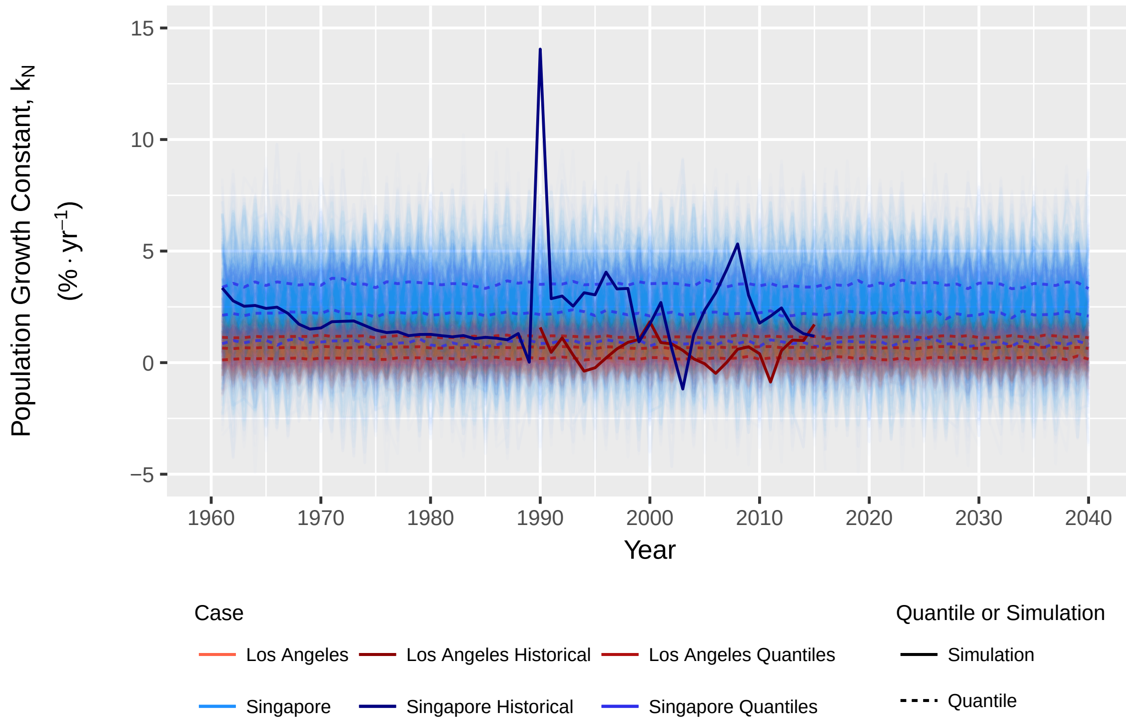


Figure 5.5: Simulation results over the historical period, 1960–2015: self-sufficiency ratio (R_{SS}, dimensionless).



Finally, estimates of R_{SS} by the simulations are shown in Figure 5.5. The simulation method over underestimates self-sufficiency in Singapore before 1970 (likely due to the overestimation of w_N), and overestimates R_{SS} between 1995–2003. The method tends to overestimate R_{SS} in Los Angeles.

Figure 5.6: Simulation results over the projection period, 2016–2040: population annual proportional growth constant (k_N, in $\% \cdot yr^{-1}$).



5.3.2 Future Projections

5.3.3 Discussion

Simulation results are shown in Figures 5.8–5.12. The simulations plotted in these figures were the base simulations—stationary random distributions were assumed for k_N, w_N, and q_P. Because of this assumption, the relative relationships between the cases was preserved in the simulation runs, as seen in Figures 5.8–5.8. For the base runs, it was also assumed that city area A_N constant at $A_N \mid_{t=2015}$.

Figure 5.9 shows the effects of the simulation results for k_N on population projections. Due to the historical data and assumptions about the underlying distributions, population projections simulated for Singapore had a much wider range than those for LA. Since $\bar{k}_N^{SG} > \bar{k}_N^{LA}$ over the observed time period (including recent years), it is likely that the population of Singapore in 2040 will be several times larger than that of LA.

Figure 5.10 shows projections for total water demand (W_N) for the two cases. Due to projected differences in population growth, the inter-quartile ranges (R) of W_N for the two cases overlap beginning in 2025, i.e., within the next decade. The R for the projections of W_N for Singapore could even surpass those of LA beginning around 2035, given business-as-usual.

Figure 5.7: Simulation results over the projection period, 2016–2040: water use intensity (w_N in $m^3 \cdot yr^{-1} \cdot cap^{-1}$).

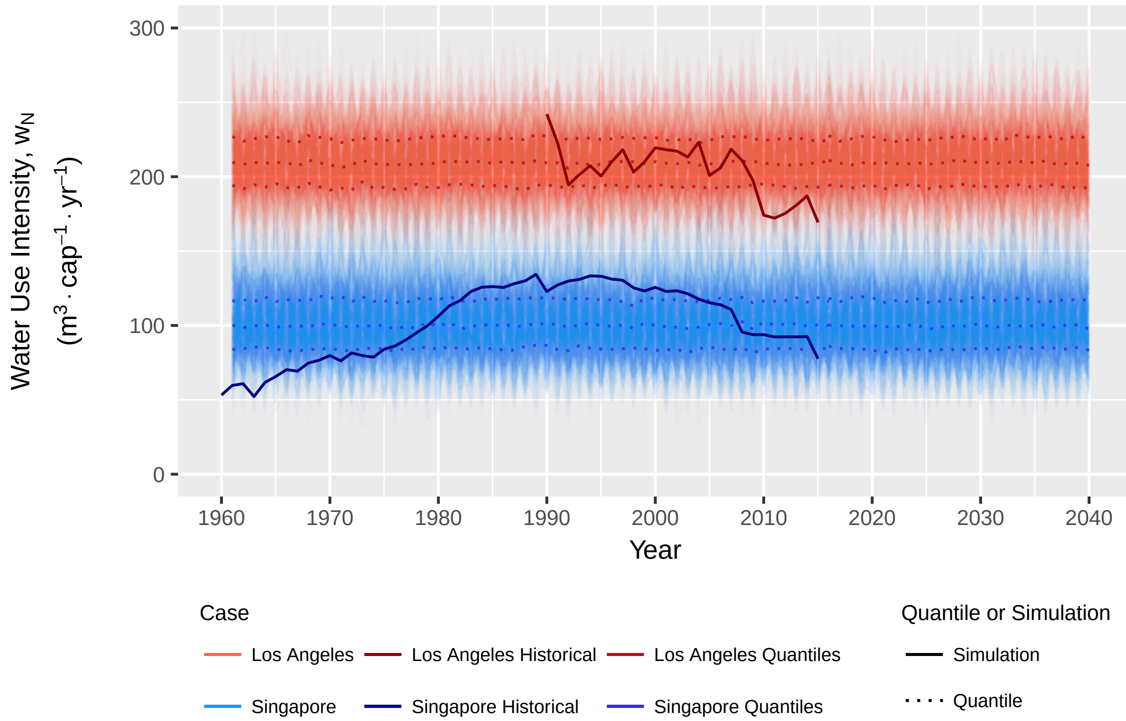


Figure 5.11 shows projections for ideal potential supply from rainfall within the city boundary for each case. Since the distribution of q_P was assumed stationary and A_N constant, the distribution of Q_P (the total rainfall volume over the city) remains constant for the base scenarios over the simulation period.

The results of the simulation for ideal potential R_{SS} are shown in Figure 5.12. These results indicate that R_{SS}^{LA} and R_{SS}^{SG} will likely continue to converge over the simulation period given business-as-usual. The deviation in R_{SS} decreased in Singapore over the simulation period, in contrast to deviation in projected population and water demand, which increased. While total water demand (W_N) was projected to converge for the two cases, no overlap of the IQRs was observed for R_{SS} for the base scenarios, although R_{SS} for Singapore continued to decrease. Projected R_{SS} in LA was not observed to vary significantly aver the simulation period. This suggests that if LA's long-term transboundary imports do not change substantially, the city will experience lower pressure than Singapore to develop alternative resources like desalination or water reuse.

However, the R for R_{SS}^{SG} began to dip below $R_{SS} = 2.5$ in the first five years of the simulation. However, since these numbers were calculated for rainfall and not stormwater runoff, the rainfall available as a freshwater resource would likely be much lower. This further supports the hypothesis that pressures on Singapore's resources may be greater than LA's.

Figure 5.8: Simulation results over the projection period, 2016–2040: annual precipitation height (q_P, in $\mathrm{m} \cdot \mathrm{yr}^{-1}$).

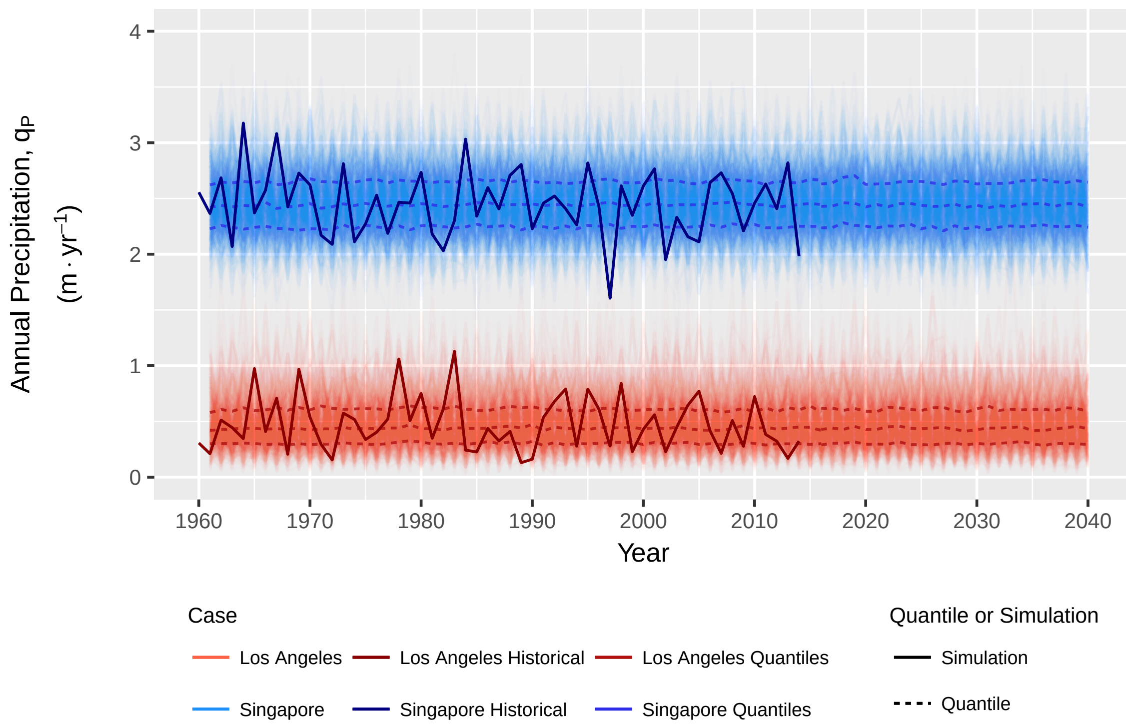


Figure 5.9: Simulation results over the projection period, 2016–2040: population (N in capita).

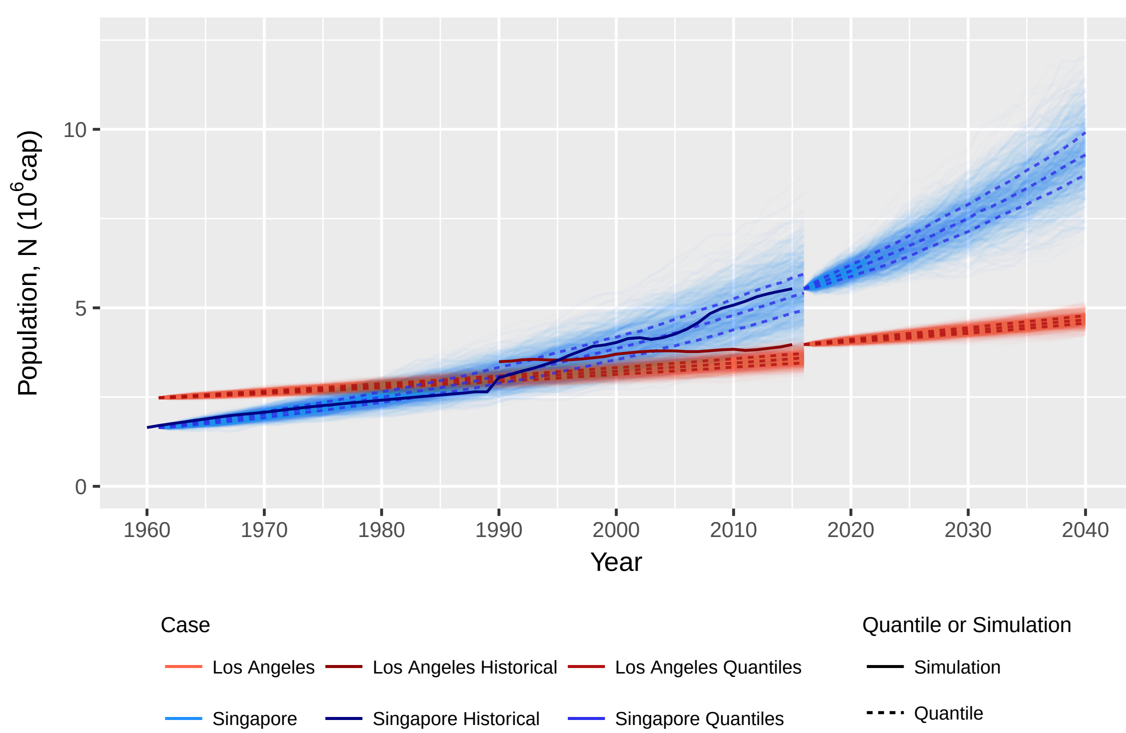


Figure 5.10: Simulation results over the projection period, 2016–2040: total water demand (W_N in 10^6 $\mathrm{m}^3 \cdot \mathrm{yr}^{-1}$, where $W_N = N \cdot w_N$).

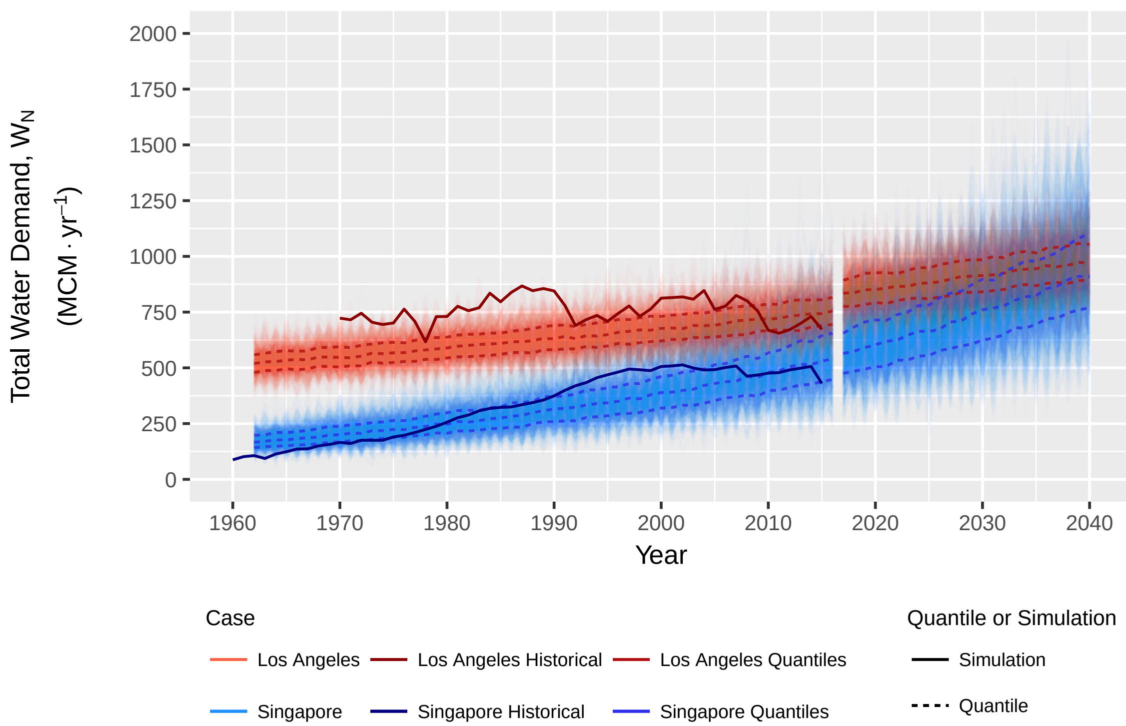


Figure 5.11: Simulation results over the projection period, 2016–2040: potential supply from rainfall (Q_R in 10^6 m$^3 \cdot$ yr^{-1}, where $Q_R = q_P \cdot A_N$).

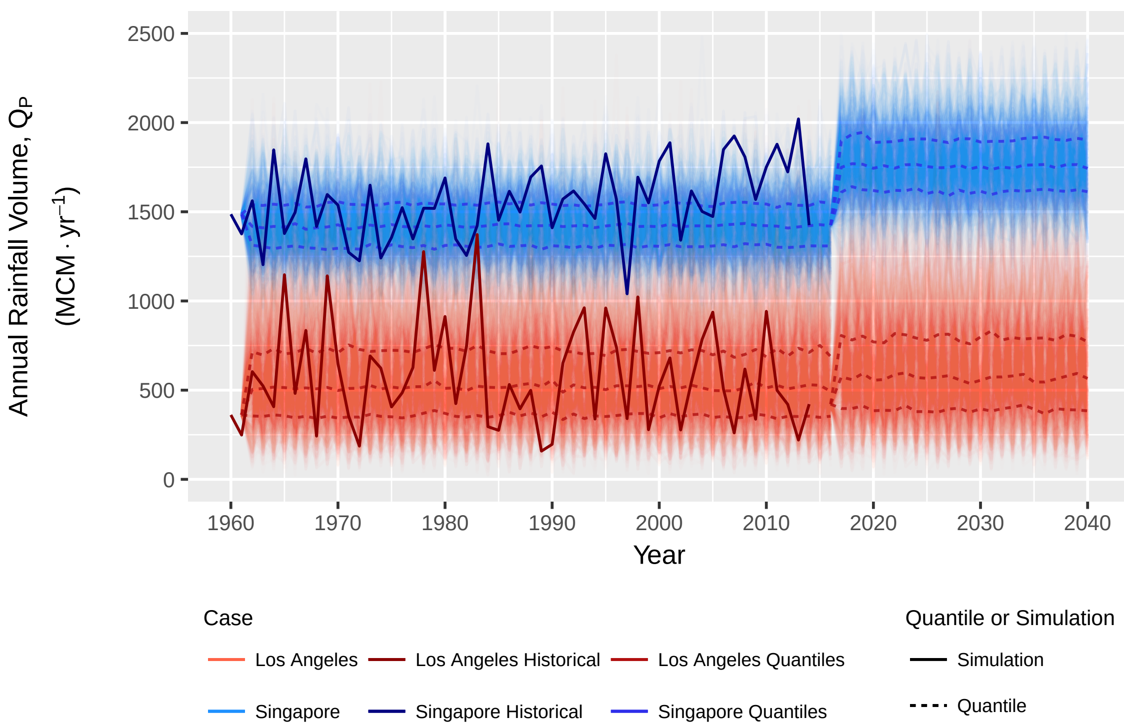
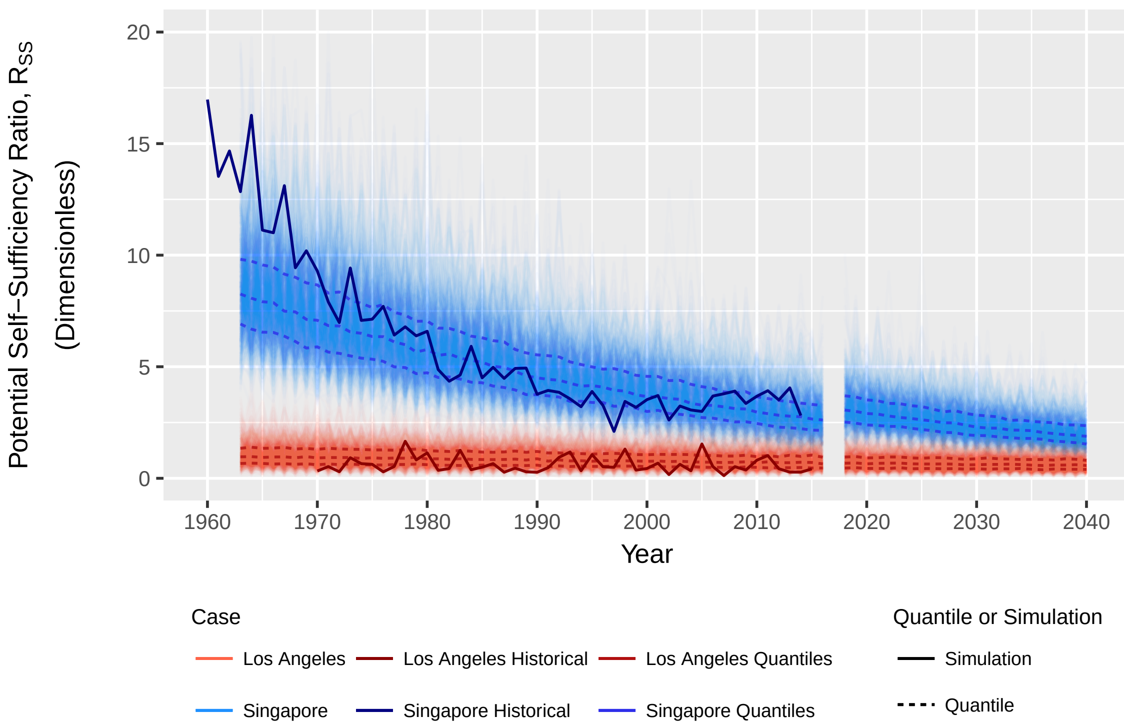


Figure 5.12: Simulation results over the projection period, 2016–2040: self-sufficiency ratio (R_{SS}, dimensionless).



5.4 Summary of Work

In this chapter, I used the results from the case study analysis in Chapter 4 as inputs to simple projections of urban water supply and demand for Los Angeles and Singapore. To do this, I developed methods within `R` to standardize simulation across both cases. I created new `R` functions and methods to facilitate replication of this analysis or its application to new cases. I then applied these methods to the two cases for two study periods: the historical period (1960–2015) and future (2017–2040) and plotted the results with historical data.

5.5 Future Work

More sophisticated scenarios could be formulated to take into account observed trends in k_N, w_N, and q_P and/or to reflect desired policy changes (e.g., for population growth or size or water use intensity). Additional scenarios could be formulated to reflect the desired goals of the two cases, shown in Figures 4.6 and 4.7, or to assess the feasibility of these targets with respect to local resources.

To provide more specific and comprehensive recommendations for each case, the analysis would benefit from at least a simple hydrological model and an assessment of existing water infrastructure. Ideally, hydrological models take into account local topography, impervious surfaces, and storage [137, 94, 200, 335]. In addition to more detailed hydrological analysis, water use intensity could be further disaggregated into contributing uses, such as industrial and municipal versus domestic use. Such an analysis might also consider attributes of urban form and climatic water balance. For instance, Singapore has a greater population density and higher climatic water availability than in LA; therefore, less water would likely be required for landscape irrigation in Singapore. For this reason, drought-resistant landscaping has been increasingly encouraged in LA [171]. Disaggregating water use intensity into contributing components would allow for the exploration of more specific technology and technology portfolios and also be useful in identifying meaningful similarities and differences between cities.

Several data structures and methods were created for the `R` programming environment to facilitate the simulation of historical and future urban water supply and demand conducted in this chapter. These methods, as well as those created for the analyses in preceding chapters, could be combined into a package of code for dissemination. Beyond the methods developed for this dissertation, such a package would ideally include methods for a user to more easily glean supporting data from supporting online databases, beginning with the databases used in this dissertation.

5.6 Conclusions

The results from the base scenarios in Chapter 5 highlighted that the pressures on WRM may likely grow faster in Singapore than in LA. Therefore, while the analysis of historical k_N, w_N, and q_P in Chapter 4 found that the cases were relatively different, the momentum in the systems suggests that WRM could become increasingly similar.

The analysis in Chapters 4 and 5 has numerous limitations. For instance, focusing on q_P at the annual scale does not highlight other issues of climate change and urbanization such as increased flooding.

Chapter 6

Discussion and Final Remarks

6.1 Overview

In this final chapter, I summarize the work performed in this dissertation chapter by chapter. For each chapter, I restate the research question motivating the research for the chapter, followed by an overview of the work performed and a description of the results. I then list the contributions, limitations, and future work.

Following this summary, I consider the contribution of the body of the work to three areas:

- theory,
- the case studies,
- and practice.

Finally, I revisit the framework for advancing knowledge about UWS presented in Chapter 2.

6.2 Summary of Work

6.2.1 Chapter 2

The research question for Chapter 2 was:

> How do I learn about UWS, how does comparative analysis contribute to learning, and how are UWS compared?

Chapter 2 discussed the concept of learning by individuals or institutions as a feedback process. At the heart of the process is the acquisition of raw data from the real world system under consideration, which we assessed using different methods to update our mental model of the state of that system. We use this information to make decisions about what actions to take to effect a desired change in the system. The decisions are structured by decision rules that are guided by our mental model of the actual system. We can also use the information to create, test, and update our mental models. Through this process, we advance our knowledge of the system.

When new information suggests that our mental model of the world no longer sufficiently represents the real system, our mental model may begin to transform to a new paradigm, as is happening with WRM in UWS. The new paradigm of sustainability, SUWM, requires new methods for analyzing data about the world. It has also brought into question many of our previous mental models. to We can also actively seek to test our mental models through a systematic approach to advance our learning and knowledge faster.

The scientific method is the predominant approach for testing formalized theories by designing experiments. For systems where it is difficult to perform experiments, like UWS, we turn to methods like comparative analysis of historical data and computer simulation. As discussed in Chapter 2 literature reviews have suggested that comparative water research has tended towards unstructured analysis and non-standard metrics and methods, which tendency has limited the validity and generalizability of the results.

However, recent research by Gondhalekar, Mollinga, and Saravanan has recommended an approach for systematic design of small-*n* comparative research. This approach consisted of a four-step iterative process whereby cases were chosen on the basis of their similarities and differences in the following way:

Step 1 MSSD + MMD, dealing with differences in similar cases;

Step 2 MSSD + MMA, dealing with similarities in similar cases;

Step 3 MDSD + MMD, dealing with differences in different cases; and

Step 4 MDSD + MMA, dealing with similarities in different cases [116].

We can use Gondhalekar, Mollinga, and Saravanan's iterative approach to create, test, and update our theories about UWS in a way that can be used to enhance both internal and external validity of our models.

However, Gondhalekar, Mollinga, and Saravanan did not offer a way to assess similarities or differences, i.e., to determine whether variance is "minimized" or "maximized" when only a small number of cases is considered. Therefore, in Chapter 2, I explained how this gap might be filled by applying emerging data-mining methods to quantitative data profiling water supply and demand for a large number of cities to establish these similarities and differences. I also synthesized the models of learning and the approach to small-*n* comparative analysis in a new framework that I believe will facilitate the acquisition of knowledge and learning about UWS to enhance SUWM. I then demonstrated the first steps of the approach in the following chapters.

Contributions

The contributions of this work are:

1. a summary of literature that pertains to forming, testing, validating, and updating knowledge of UWS, and

2. a synthesis of relevant concepts into a new framework for learning about UWS

Limitations

Some limitations of this work are:

- that the approach is somewhat high-level, i.e., conceptual, although some of the ideas are flushed out in Chapters 2–5;
- to be must useful, the approach requires the integration of many common databases, which would benefit from buy-in from numerous global organizations;
- this is to say that ideally, this framework would be applied on an international, global scale, which requires substantial coordination and availability of resources;
- this raises general concerns about complexity, data management, and security.

6.2.2 Chapter 3

The research question for Chapter 3 was:

> Can simple profiles of supply and demand provide both global perspective *and* meaningful insight into regional challenges?

The methods presented in Chapter 3 attempted to provide insight into the diverse space of UWS.

In Chapter 3, I assembled a dataset of simple profiles of urban water supply and demand, consisting of three commonly available metrics: population (N in capita), water use intensity (w_N, in $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$), precipitation ($q_P$ in $\mathrm{m} \cdot \mathrm{yr}^{-1}$), climatic water balance (q_{Net} in $\mathrm{m} \cdot \mathrm{yr}^{-1}$), and city area ($A_N$ in km^2). I compiled data for 142 cities from three sources: the UrbMet database of cities, IBNET, and WebWIMP. To do this, I created methods for obtaining relevant data from these databases and also for matching the data against similar data from other databases to facilitate future assessment of data comparability and quality. I designed these methods so that the analysis can be expanded to include new data in the future with relative ease.

I then applied basic methods for exploratory statistical analysis—such as summary statistics, histograms, qq-plots, and scatterplots—to broadly characterize global trends and establish the spectrum of diversity in N, w_N, and q_{Net} for cities around the world. To the author's knowledge this is the first statistical characterization of these urban water profiles. Through the results, I determined that a $\log_{10}$ transformation of these three metrics.

I then applied two statistical clustering algorithms—hierarchical clustering analysis and t-SNE—to N, w_N, and q_{Net} and identified a set of six distinct city types. I found both intuitive and surprising results within each type and in the resulting typology. Importantly, I also assessed that it would have been difficult to identify from simply comparing cities on the basis of one or two indicators alone, as is common in large-n analysis of UWS.

In Chapter 3, we also introduced two new indicators for water footprint: the WUCI (i_{UC}) and total WUCI (I_{UC}). We defined WUCI as the ratio of precipitation (q_P) to water use intensity (w_N), and total WUCI as per capita WUCI multiplied by population (N). These composite indicators had units of $m^2 \cdot cap^{-1}$ and km^2, respectively. WUCI and total WUCI were not included within the clustering analysis. Instead, as composite indices of the indicators used to cluster, we found that they helped to provide visual intuition into the similarities and differences between the types that emerged and further validated the typology.

The analysis in Chapter 3 provided insight into the relative similarity between LA and Singapore at an instant in time, but was not able to provide historical context, i.e., any insight into the trajectory of the UWS. However, since UWS exhibit inertia, which can affect the relative impact of technologies and policies. It is therefore important to consider the historical context of WRM in cities.

Contributions

The contributions of this work are:

- a dataset for 142 cities of N; w_N; q_{Net}, q_P, other water balance metrics, area, and location data, assembled from larger databases;
- methods for linking these to larger databases and for adding new cases to the dataset, including:
- new Python functions for obtaining specific city data from online sources;
- the first characterization of the spectrum of diversity for simple water supply and demand profiles through basic statistical methods;
- a quantification of similarity between cases across a multivariate space;
- a reduction of the multivariate space to a two-dimensional space to facilitate interpretability of results for researchers and decision-makers;
- a typology of cities based on quantified similarity and that can be used as a structured framework for guiding case study selection for small-*n* research.

Limitations

Some limitations of this work are:

- results do not assess whether similar urban water profiles have similar water challenges (an area for future research, as in Chapter 4);
- the results varied with the variables used in clustering, implying that different sets of metrics would likely lead to different typologies (in other words, this is not the *only* typology);

- a different set of cities could lead to different groups, and as a consequence:
- the groups may not be stable if the number of cities were expanded;
- the results do not give any indication of the trajectory of sustainability;
- the results are not based upon a full climate profile and WRM tends to differ depending on the monthly variation in climatic water balance.

Future Work

- assess whether the types identified have similar water resource challenges (e.g., through small-*n* comparative analysis)
- apply classification and regression tree analysis to identify thresholds for the groups;
- expand the set of cities;
- expand the data over time (may require additional data-mining methods);
- expand the clustering to include monthly climatic water balance data (may require additional data-mining methods); and
- make the dataset, results, and methods available to other researchers (e.g., through Github).

6.2.3 Chapter 4

The research question for Chapter 4 was:

> To what extent could similar cities have been self-sufficient in the past with respect to local water resources, and how does this compare to their actual portfolio of water supply? How similar were their portfolios and to what extent could they be similar?

In Chapter 4, I used the typology developed in Chapter 3 as a framework for choosing two case studies for more in-depth comparative analysis over time. I also used the step-wise logic from Chapter 2 to guide the choice. As determined by the analysis in Chapter 3 the members within each type in the typology are more similar to each other than to cities in other types. However, members in each type still exhibit a range in at least one of the underlying indicators. Since the first step in the step-wise logic for small-*n* research is that of "similar but different" I decided to choose two cities from the same type, to increase the likelihood that the cities would have similar water challenges, but cities that differed in some significant way.

I decided to focus on Type 4 cities, which have the highest per capita and total WUCI compared to other types[1]. Type 4 cities are also characterized by large populations (1.838–14.350 million) and high per capita water consumption (82–285 $m^3 \cdot yr^{-1}$;). However, Type 4 exhibit

[1]With respect to median values and IQR, as discussed further in 3.8.3.

substantial climatic water availability, with members ranging from the highest to lowest quartiles (-0.577–1.279 $m \cdot yr^{-1}$). Of the 27 cities in Type 4[2], I chose LA and Singapore for their different climates and water "footprints": these two cities have the lowest and highest annual precipitation ($0.273m \cdot yr^{-1}$and $2.621m \cdot yr^{-1}$, respectively) and highest and lowest per capita WUCI ($575.1m^2$ and $37.77m^2$, respectively).

I then developed an approach to compare these two cities on the basis of historical WUCI, self-sufficiency (R_{SS}), and their underlying attributes—population, water use intensity, and climatic water availability. I first developed methods in `R` to assemble data from a variety of sources. After assembling the data, I developed methods for analyzing the historical data first as random variables and then as time series. I also created code in `R` to plot both cases on the same plots to facilitate visual comparison.

I applied the methods to both cases and found that it was reasonable to model k_N, w_N, and q_P as random variables, but that this model was quite approximate for k_N and w_N, especially since the results of the time series analysis suggested that these variables were not stationary over time. This result suggested that the variables could be modeled as stationary stochastic processes to establish a baseline for comparison, but that making more detailed recommendations on WRM should take into account the recent momentum in the system, rather than relying solely on summary statistics.

The analysis of population growth, per capita water use, and precipitation as random variables highlighted the differences between the two cases, which only overlapped on population growth. In spite of the dissimilarity, however, visualizing the distributions was helpful in characterizing *how* different the case studies were. This difference was also highlighted visually in the time series plots. However, I highlighted similarities in the cases by looking at relative deviations and proportional change of the attributes, which rescaled the values and highlighted similarities in trends and variation.

In general, I found that LA and Singapore historically had differences in k_N, w_N, and q_P, but that there was some convergence, especially in WUI and the potential self-sufficiency ratio. The differences between the two cases were particularly notable for i_{UC} and I_{UC}, i.e., water "footprint". The potential self-sufficiency ratio was historically much higher in Singapore than in LA but has begun to converge in recent years. However, LA could not historically source 100% of their potable water supply exclusively from precipitation falling within the city bounds—in contrast to Singapore, which could (in theory) have done so.

I found that although LA has lower natural climatic water availability than Singapore, it has much higher water use intensity. I also found an order of magnitude difference in WUCI and total WUCI (higher in LA) and the potential self-sufficiency ratio (higher in Singapore, although it has been converging to that of LA's). Looking at the respective portfolios of planned water supply for the two cases further added to the conundrum: Singapore, with its high natural climatic water availability, had much more aggressive targets for self-sufficiency through expansion of

[2]See Figure 3.15 for a full list of cities in Type 4.

urban stormwater collection and desalination and wastewater recycling capacity. One point of similarity was that LA's water plans also highlighted the importance of self-sufficiency. However, LA was much less aggressive in setting their goals, which did not include desalination as a potential supply, and primarily targeted expansion of wastewater recycling to replenish aquifers, rather than treating it for domestic water use.

I gained insight into this conundrum after looking at portfolios of current water supply portfolios and contextual information about the cases—Singapore is an island nation and has strained political relationships with Malaysia and Indonesia (its closest neighbors). Although Singapore has had two long-term agreements governing water transfer from Malaysia to Singapore, one treaty expired in 2011 and the second will expire in 2060. Singapore is planning for the worst-case scenario, where they cannot negotiate any new treaty. Water security has therefore been a paramount issue in Singapore since the 1960s, and they have been seriously pursuing desalination, water recycling, and urban stormwater runoff since this time.

Somewhat similarly, LA has and does obtain a substantial fraction of its urban water supply through water transfers from the Owens River and Colorado Rivers. Less similarly, these water imports are transported from sources that are hundreds of kilometers away rather than tens of kilometers away, as is the case for Singapore. In the 1960s, LA was, like Singapore, also investigating the desalination of seawater as a potential urban water source; however, the completion of the California Aqueduct and other projects made this unnecessary for LA, unlike Singapore. Los Angeles has a very long-term agreement (no expiration date) at the federal level (the US) that guarantees access to water from one of the largest river systems in the US, the Colorado River. In other words, I found unexpected similarities when I looked at contextual information about WRM for the two cases and considered their portfolios of water supply.

Using the methods developed in the chapter, I did not find a clear answer to the question, *to what extent could [the two cases] be similar*? To answer this question would require a better characterization of the relationship between precipitation and runoff for both cases. Ideally, it would also be useful to assess opportunities for water storage within the urban boundary, since this has been a limiting factor in Singapore's use of stormwater runoff for urban supply [305].

I was not able to capture the full picture of historical urban WRM for the two cases using only simple analysis of a small set of indicators, but identified additional avenues for future work, both on the two cases and on new cases. Therefore, I concluded that, though not perfect, it would be interesting to develop these methods further, especially since the approach yielded some interesting results and has other desirable properties, such as continuity with the analysis in Chapter 3.

Contributions

The contributions of this work are:

- I found that the typology helped to structure my choice of two cases for small-*n* comparative analysis;

- I assembled comparable datasets of water supply and demand metrics for two cities over time;
- I created methods for linking these to larger databases and for adding new cases to the dataset, including:
- new functions for extracting historical climatic water balance data from spatial data for the cases;
- I demonstrated that the small-n methods developed in the chapter were applicable to both cases and yielded interesting results.

Limitations

Some limitations of this work are:

- missing data points within the first three decades for two metrics for LA (N and w_N) which may also be an issue in future case studies;
- the methods did not allow us to answer the research question, *to what extent could the portfolios of the two cities be similar?*—an area for future research.

Future Work

- compare the two cases with cities that are more similar (following Step 2 in the inferential logic, as re-summarized in Section 6.2.1)—e.g., Singapore with Sydney and LA with Tashkent, Cali, or Kiev (as suggested by Figure 3.17).
- compare the two cases with cities that are dissimilar (following Step 3 in the inferential logic)—i.e., cities that differ in type and on all three metrics (N, w_N, and q_{Net});
- characterize the relationship between precipitation for urban runoff and for other "local" watersheds (e.g., the Owens River and Colorado River for LA), which would allow us to quantitatively compare self-sufficiency between the two cases; and
- make the dataset, results, and methods available to other researchers (e.g., through Github).

6.2.4 Chapter 5

The research question for Chapter 5 was:

> Using data and results from Chapter 4, to what extent could these cities be self-sufficient in the future, and how does this compare to their projected portfolios of water supply? Does the comparative analysis in Chapters 4 and 5 offer case-specific or general insight?

In Chapter 4, I used the results from Chapter 4 to generate two sets of simulations for Los Angeles and Singapore.: simulations of urban "pasts" (simulations for the period from 1960–2016) and urban "futures" (2017–2040). I used simple formulations in the simulations for both cases. I assumed that k_N, w_N, and q_P were random and used a simple growth model of population to formulate these simple scenarios. I then developed methods in `R` that I used to generate 2000 simulations for both cases. I also developed code in `R` to visualize both sets of simulation runs on the same plot to facilitate comparison. These plots provided some general insights into the future trajectories of the cases, as well as the basis to formulate more specific research questions for comparison.

The simulation results for k_N, w_N, and q_P, which I modeled as random variables, were not surprising. However, due to the growth model for population, results for population, and derived functions—total water demand, and the potential self-sufficiency ratio (R_{SS})—were more interesting. I found that the simple formulations were a reasonable match for population for both cases, although the simulations tended to underestimate the population for LA (likely due to missing values in the early part of the data series). The simulations also overestimated total water use for both cases towards the end of the series, highlighting that water use intensity is not a stationary random variable and, more interestingly for SUWM shows signs of decreasing or leveling off for both cases. However, the simulations appeared to be a better match for R_{SS} over the historical period. Over the future period, the simulation results for total water use suggested that total water use in Singapore might surpass that in LA—however, this result was based on the assumption that w_N was stationary, which appeared from the historical projections to be an inadequate representation of the last decade. As a consequence of the results for total water use, simulation results for projection period suggest that R_{SS} in Singapore will continue to trend downwards towards LA; however, even though the simplistic assumptions may overestimate total water use, the IQR of R_{SS} in Singapore does not intersect that for LA in the future.

Considering the research question,

> Using data and results from Chapter 4, to what extent could these cities be self-sufficient in the future, and how does this compare to their projected portfolios of water supply?

I found that Singapore's potential for self-sufficiency may decrease. This finding is supported by Singapore's prioritization of desalinated seawater and recycled wastewater as sources for urban water supply over stormwater collection. I also found LA's potential self-sufficiency ratio may decrease, though less drastically than Singapore's.

I did not assess LA's potential for self-sufficiency with respect to its long-term access to (a significant volume of) water from the Colorado River and the Owens River. If I included these sources, which could be considered local, LA might look more comparable to Singapore with respect to WUCI, total WUCI, and R_{SS}. This is an area for future research since it requires assembling climatic water balance data for the relevant watersheds.

Does the work provide case-specific or general insight? While the results of the analysis in Chapter 5 are intriguing, they require further research. At the very least, it would be useful to quantitatively compare the results to the planned portfolios, and the water demand projections generated by the water utilities in LA and Singapore that were used in designing these strategies.

Contributions

The contributions of this work are:

- I created simple procedures for simulating water supply and demand for Los Angeles and Singapore;
- I created new `R` functions and objects for simulation that will facilitate replicability of the analysis and scalability to new cases, as well as comparability to those cases;
- I demonstrated that these methods reasonably represented the urban pasts for the two cases; and
- I demonstrated the method on simple scenarios for projected urban futures.

Limitations

Some limitations of this work are:

- the assumption of stationary stochastic randomness is a weak approximation for k_N and w_N (an area for future development);
- I did not consider scenarios such as improved water conservation or efficiency, population stabilization, or climate change (an area for future research);
- I did not consider additional local sources—although these may not be significant in Singapore's future, they will likely continue to be significant in LA (an area for future research);
- since I did not have a formula relating precipitation to runoff volume, I was unable to quantitatively compare the simulation results to their planned portfolios (another area for future research).

Future Work

Avenues for future research suggested by these results are:

- creating scenarios for k_N, w_N, and q_P where the mean changes over time;
- characterizing precipitation to runoff volume;
- characterizing precipitation and runoff in watersheds to which the cases have long-term access; and

- quantitatively comparing the results to actual projections of water demand and the planned portfolios of water supply.

6.3 Future Work: Toward an Integrated Framework

In Figure 2.5 on Page 67 in Chapter 2 (shown again in Figure 6.1 for convenience), I presented a conceptual diagram relating the research pursued in this dissertation to the recommended approach to small-*n* comparative analysis recommended by Levi-Faur and Gondhalekar, Mollinga, and Saravanan. To what extent have I achieved the goals outlined in this figure?

In Chapter 3, I accomplished Step 0—unstructured diversity—consisting of the assembling of the data of urban water supply and demand profiles for the 142 cities in the UrbMet database of cities. I began to take steps toward structuring this data by characterizing the spectrum of profiles using descriptive statistics, a correlation matrix, and a series of univariate and bivariate plots—ordered bar charts, histograms, quantile-quantile plots, and scatterplots. Also in Chapter 3, I completed Step 1: I used the IPAT heuristic to guide my choice of population, water use intensity, and climatic water balance to represent a simple profile of urban water supply and demand. Using statistical data-mining methods, I quantified similarity; reduced the dimensionality of the

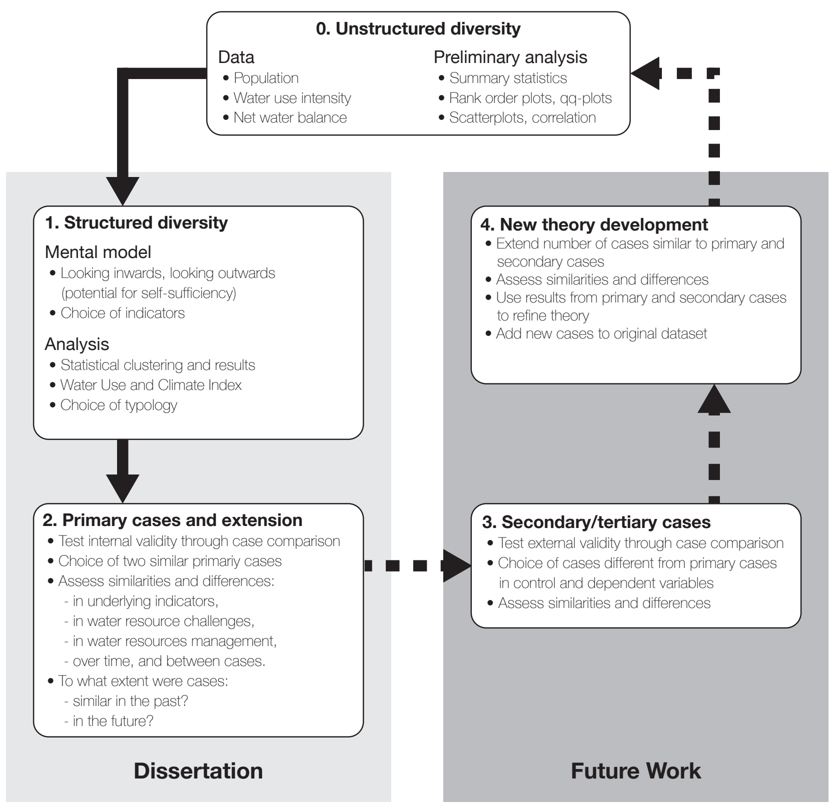


Figure 6.1: The structure of comparative analysis pursued in this dissertation.

data space from three dimensions to two; and identified a meaningful six-type typology. I further structured this diversity by introducing a new urban-scale water footprint indicator—the Water Use and Climate Index (WUCI). I also calculated the total WUCI as the product of WUCI and population and the potential self-sufficiency ratio (R_{SS}), by dividing city area by total WUCI. These derived indicators provided additional intuition into the typology.

In Chapter 4, I pursued an analysis addressing Step 2 of the inferential logic. I chose two "similar but different" cases from Type 4—Los Angeles and Singapore—and compared these cases over a longer period of time. I also considered these results relative to their actual and planned water supply portfolios and found that, in spite of similarities in the dependent variables, the cases were more similar than they appeared. This helps to build the internal validity of the typology, although more research is required.

In Chapter 4, I took another approach to Step 2. I designed simple simulation methods and used the results from Chapter 4 to formulate simple scenarios. I then simulated urban water profiles for the two cases over the historical period and into the future. This provided another opportunity to test the validity of the models (i.e., the models of randomness and growth) that was had assumed in designing the methods for analysis in Chapter 4. I found that ("like with like") these models were reasonable approximations—at least over the historical period.

So, what are the more general next steps for case study analysis, using Figure 6.1 as a reference?

The next step would be to compare both cases with new cases to which they are more similar across all three metrics (as opposed to differing significantly in one). The results of the clustering suggest Singapore be compared with Sydney and LA be compared with Tashkent, Cali, or Kiev—or all three. After that, the results from Chapter 3 would be used to guide the choice of a city dissimilar in type and in N, w_N, and q_{Net} for any or all of these cases. In the last step, these cases would be compared with cases that differed in *type* but were similar in N, w_N, and q_{Net}. For Los Angeles, this might be Paris in Type 2 (as identified from Figure 3.11), while for Singapore the task appears more difficult; further analysis could help to pinpoint whether Santiago, Minsk, Cairo, or Kuala Lumpur might be a better match.

Figure 2.4 on Page 66 in Chapter 2 illustrated a framework for integrating this type of comparative analysis into a larger framework for advancing knowledge about SUWM in UWS. In my design of research throughout the dissertation, I connected existing datasets and develop standard methods for comparison to promote comparability, scalability, and replicability.

I have demonstrated that my approach, and the methods I developed for analysis, have enhanced the comparability of cases and the results. I have also highlighted that, even though my methods were simple, comparing cases in a systematic way can be useful in structuring small-*n* research questions. The replicability of the results has also been demonstrated in the production of this thesis document, which was produced in `R` and `Sweave`, which implement Donald E. Knuth's concept of *literate programming*, in which snippets of code are embedded in environments for documentation, such as $\LaTeX$ or `html`. In other words, almost all of these analyses[3] are rerun every time a new version of this dissertation have been compiled, which is a preliminary demonstration of replicability.

The priority for future work would be to implement these methods in a robust, standardized way that can be easily disseminated, e.g., as an `R` package. This will probably require additional standardization through the creation of new objects and functions. Expanding this work would also benefit from:

1. a structured database, implemented in `JSON` or another database paradigm that can handle a variety of data objects;

2. an online repository of related packages;

3. an online interface, especially one through which decision-makers could view results in a visually-intuitive way.

The datasets for the analyses in Chapters 3 and 4 were constructed largely from global datasets, and where that was not possible, from commonly reported indicators. Therefore, while the simplicity of the methods employed limited the specificity of the recommendations that could be conclusively made for the cases, that simplicity also has several advantages: the analyses can be easily replicated for any of the 142 cities, and the original set of 142 can be easily scaled to accommodate new members. This scalability is particularly important for comparative analysis of UWS, as it allows any city to be accommodated by the approach. The use of commonly available indicators from global datasets has the additional benefit that it can be employed by cities with fewer financial resources, which are often unable to pursue more complex methods. At the same time, both the clustering analysis and the case study analysis could be expanded to include additional indicators or more complex models, based on available resources and user interest, with the caveat that, where possible, standard datasets and freely available software platforms and packages should be prioritized.

Future extension of this work would be to package the code developed for the dissertation into a library for the `R` programming environment. This would facilitate replicability and scalability of the methods and results. In addition, more sophisticated analyses would likely yield greater insight into the comparability of the technology and policy portfolios. Related analyses would be to better quantify actual runoff in urban watersheds, as well as to quantify freshwater resources in sources to which each city has long-term access, such as the Colorado River for LA.

[3] The exceptions are the t-SNE analysis, which finds a local optimization point but not a global optimization point, and the scraping of data from WebWIMP for the 142 cities in Chapter 3.

6.4 Conclusions

In this dissertation, I presented a novel approach to advancing knowledge about sustainable urban water management in urban water systems for comparative analysis of large and small numbers of cases. I illustrated that large-*n* analysis could be conducted in a way that enhances intuition into urban water challenges around the world *and* provides structure to designing small-*n* comparative research. I also demonstrated that small-*n* research can be conducted in a comparable way, and that, when this research design references a broader research framework informed by an appropriate large-*n* analysis, even simple analysis can yield interesting results and open up specific areas for future research.

I also demonstrated one approach to advancing knowledge about urban water systems within the larger framework for integrating data, methods, and research. The analysis pursued in this dissertation was relatively simple; more complex methods exist. I have illustrated by example that enhancing learning about sustainable water management in cities is hopefully a tractable problem, and that we can advance the sharing of information and the creation of new knowledge through thoughtful integration of data and methods, both old and new.

Meeting future water needs for a growing city is a crucial component of an urban development plan for, without sufficient water, both people and industry suffer. However, the extent to which global freshwater resources can support continued growth in water demand has been called into question, and the threat of global water scarcity has emerged as one of the key challenges of our time. Therefore, many growing cities face an uncertain future as water resources dwindle. Amidst this growing crisis, cities must carefully consider their options for future water demand and supply. As competition increases for freshwater sources, it is becoming increasingly important to have control over their water supply. Self-sufficiency—the ratio of local water supply relative to water demand—has emerged as a key metric for urban water management.

The aim of this dissertation was to develop an approach to demonstrate that, while those needs may seem paradoxical, rigorous comparative analysis of UWS can make headway towards integrating those needs. The resulting insights demonstrated the legitimacy of the approach and its ability to provide insight into the self-sufficiency of urban water systems. In other words, although water resources management can be highly specific, it is still possible to gain case-specific insight *and* meaningfully lessons to other cities within a type. These results highlight the saliency of the urban scale to questions of international water policy.

In conclusion, a simple but thoughtful approach to comparative analysis of urban water systems, pursued in an appropriate research framework, can be a powerful tool for adapting innovative portfolios of technologies and policies from one city to another—which will be essential in facilitating transitions to sustainability. However, the approach developed in this dissertation represents only a first step towards this objective. There is a great need for knowledge sharing in this uncertain world, and a huge potential to gain insight from rigorous comparative analysis. In the coming decades, our success in achieving sustainability of urban water resource management depends on a coordinated effort by researchers, practitioners, designers, and other community members to integrate their knowledge in a systematic but adaptive way. Hopefully, the coming decades will see further development of cooperative methods, available databases, sharing of results, and an atmosphere of openness and knowledge sharing.

The coming generations depend on us.

Bibliography

[1] Barry Abrams, Santharajah Kumaradevan, Vasilis Sarafidis, and Frank Spaninks. "An Econometric Assessment of Pricing Sydney's Residential Water Use". In: *Economic Record* 88.280 (2012), pp. 89–105. DOI: 10.1111/j.1475-4932.2011.00779.x.

[2] James M. Acheson. "Institutional Failure in Resource Management". In: *Annual Review of Anthropology* 35.1 (2006), pp. 117–134. DOI: 10.1146/annurev.anthro.35.081705.123238.

[3] Jan Adamowski, Hiu Fung Chan, Shiv O. Prasher, Bogdan Ozga-Zielinski, and Anna Sliusarieva. "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". In: *Water Resources Research* 48.1 (2012), W01528. DOI: 10.1029/2010WR009945.

[4] A. AghaKouchak, L. Cheng, O. Mazdiyasni, and A. Farahmand. "Global warming and changes in risk of concurrent climate extremes: Insights from the 2014 California drought". In: *Geophysical Research Letters* 41 (2014), pp. 8847–8852. DOI: 10.1002/2014GL062308.

[5] Blake Alcott. "Jevons' paradox". In: *Ecological Economics* 54.1 (2005), pp. 9–21. DOI: 10.1016/j.ecolecon.2005.03.020.

[6] Louis Edward Alfeld. "Urban dynamics—the first fifty years". In: *System Dynamics Review* 11.3 (1995), pp. 199–217. DOI: 10.1002/sdr.4260110303.

[7] Myles R. Allen and William J. Ingram. "Constraints on future changes in climate and the hydrologic cycle." In: *Nature* 419.6903 (2002), pp. 224–32. DOI: 10.1038/nature01092.

[8] Massoud Amin. "Toward Secure and Resilient Interdependent Infrastructures". In: *Journal of Infrastructure Systems* 8.September (2002), pp. 67–75.

[9] John M. Anderies, Armando A. Rodriguez, Marco A. Janssen, and Oguzhan Cifdaloz. "Panaceas, uncertainty, and the robust control framework in sustainability science". In: *Proceedings of the National Academy of Sciences of the United States of America* 104.39 (2007), pp. 15194–15199. DOI: 10.1073/pnas.0702655104.

[10] John M. Anderson. "Integrating recycled water into urban water supply solutions". In: *Desalination* 187.1-3 (2006), pp. 1–9. DOI: 10.1016/j.desal.2005.04.062.

[11] José Geraldo Pena de Andrade, Paulo Sergio Franco Barbosa, Luiz Carlos Alves Souza, and Daniel Lucas Makino. "Interbasin Water Transfers: The Brazilian Experience and International Case Comparisons". In: *Water Resources Management* 25.8 (2011), pp. 1915–1934. ISSN: 09204741. DOI: 10.1007/s11269-011-9781-6.

[12] Marta Antonelli and Francesca Greco. *The Water We Eat: Combining Virtual and Water Footprints*. London: Springer, 2015. DOI: 10.1007/978-3-319-16393-2.

[13] State of Arizona, State of Colorado, State of New Mexico, and State of Wyoming. *Upper Colorado River Basin Compact, 1948*. Washington, D.C., 1948.

[14] ASCE. *2013 Report Card For America's Infrastructure: Drinking Water*. Tech. rep. Reston, VA: American Society of Civil Engineers (ASCE), 2013. URL: http://www.infrastructurereportcard.org/a/{\#}p/drinking-water/overview.

[15] ASCE. *Failure to Act: Closing the Infrastructure Investment Gap for America's Economic Future*. Tech. rep. Reston, VA: American Society of Civil Engineers (ASCE), 2016.

[16] Peter Marcus Bach, Wolfgang Rauch, Peter S. Mikkelsen, David T. Mccarthy, and Ana Deletic. "A critical review of integrated urban water modelling — Urban drainage and beyond". In: *Environmental Modelling and Software* 54 (2014), pp. 88–107. DOI: 10.1016/j.envsoft.2013.12.018.

[17] Kenneth D. Bailey. *Typologies and Taxonomies: An Introduction to Classification Techniques*. Vol. 102. Quantitative Applications in the Social Sciences, Vol. 102. Sage, 1994.

[18] Gregory M. Baird. "A game plan for aging water infrastructure". In: *Journal / American Water Works Association* 102.4 (2010), pp. 74–82. ISSN: 0003150X.

[19] Gregory M. Baird. "The Future of Water Infrastructure Asset Management, Part 2: Protect Your Funding...Because All Roads Lead to Finance". In: *American Water Works Association* March (2013), pp. 18–23.

[20] Lawrence A. Baker. *The Water Environment of Cities*. Ed. by Lawrence A. Baker. Boston, MA: Springer US, 2009. DOI: 10.1007/978-0-387-84891-4.

[21] Yaman Barlas. "System Dynamics: Systemic Feedback Modeling for Policy Analysis". In: *Encyclopedia of Life Support Systems, Vol. 2*. Vol. 1. Eolss Publishers, 2009. Chap. 5.12, pp. 1131–1175.

[22] Joseph J. Barsugli, Jason M. Vogel, Laurna Kaatz, Joel B. Smith, Marc Waage, and Christopher J. Anderson. "Two Faces of Uncertainty: Climate Science and Water Utility Planning Methods". In: *Journal of Water Resources Planning and Management* 138.5 (2012), pp. 389–395. DOI: 10.1061/(ASCE)WR.1943-5452.0000188.

[23] Michael Batty. "Building a science of cities". In: *Cities* 29 (2012), S9–S16. DOI: 10.1016/j.cities.2011.11.008.

[24] Michael Batty. *The New Science of Cities*. Cambridge, MA: MIT Press, 2013.

[25] Michael Batty. “The Size, Scale, and Shape of Cities”. In: *Science* 319 (2008), pp. 769–771.

[26] Timothy M. Baynes. “Complexity in Urban Development and Management”. In: *Journal of Industrial Ecology* 13.2 (2009), pp. 214–227. DOI: 10.1111/j.1530-9290.2009.00123.x.

[27] Morgan Bazilian, Holger Rogner, Mark Howells, Sebastian Hermann, Douglas Arent, Dolf Gielen, Pasquale Steduto, Alexander Mueller, Paul Komor, Richard S. J. Tol, and Kandeh K. Yumkella. “Considering the energy, water and food nexus: Towards an integrated modelling approach”. In: *Energy Policy* 39.12 (2011), pp. 7896–7906. ISSN: 03014215. DOI: 10.1016/j.enpol.2011.09.039.

[28] Michael Bruce Beck and Rodrigo Villarroel Walker. “On water security, sustainability, and the water-food-energy-climate nexus”. In: *Frontiers of Environmental Science & Engineering* 7.5 (2013), pp. 626–639. DOI: 10.1007/s11783-013-0548-6.

[29] David Benson, Animesh K Gain, and Josselin J Rouillard. “Water Governance in a Comparative Perspective: From IWRM to a’Nexus’ Approach?” In: *Water Alternatives* 8.1 (2015), pp. 756–773. ISSN: 19650175.

[30] Sanford Berg. *Water Utility Benchmarking: Measurement, Methodologies, and Performance Incentives.* London: IWA Publishing, 2010.

[31] Thomas Bernauer, Halvard Buhaug, Nils Petter, Theresa Tribaldos, and Gerdis Wischnath. “Demand, supply, and restraint: Determinants of domestic water conflict and cooperation”. In: *Global Environmental Change* 29 (2014), pp. 337–348. DOI: 10.1016/j.gloenvcha.2013.11.018.

[32] Luis M. A. Bettencourt. “The origins of scaling in cities”. In: *Science* 340.6139 (2013), pp. 1438–41. DOI: 10.1126/science.1235823.

[33] Luis M. A. Bettencourt and Jasleen Kaur. “Evolution and structure of sustainability science”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108.49 (2011), pp. 19540–5. DOI: 10.1073/pnas.1102712108.

[34] Luis M. A. Bettencourt and Geoffrey B. West. “A unified theory of urban living”. In: *Nature* 467.7318 (2010), pp. 912–3. DOI: 10.1038/467912a.

[35] Luis M. A. Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West. “Growth, innovation, scaling, and the pace of life in cities”. In: *Proceedings of the National Academy of Sciences of the United States of America* 104.17 (2007), pp. 7301–6. DOI: 10.1073/pnas.0610172104.

[36] Asit K. Biswas and Cecilia Tortajada. *Water Security, Climate Change and Sustainable Development*. Singapore: Springer, 2016. URL: www.springer.com/series/7009.

[37] Michael Blackhurst, Chris Hendrickson, and Jordi Sels I. Vidal. “Direct and indirect water withdrawals for U.S. industrial sectors.” In: *Environmental science & technology* 44.6 (2010), pp. 2126–30. DOI: 10.1021/es903147k.

[38] Jane M. Blackmore and Roel A. J. Plant. “Risk and Resilience to Enhance Sustainability with Application to Urban Water Systems”. In: *Journal of Water Resources Planning and Management* 134 (2008), pp. 224–233.

[39] T. Bond, E. Roma, K. M. Foxon, M. R. Templeton, and C. A. Buckley. “Ancient water and sanitation systems—applicability for the contemporary urban developing world”. In: *Water Science & Technology* 67.5 (2013), p. 935. ISSN: 0273-1223. DOI: 10.2166/wst.2013.628. URL: http://www.iwaponline.com/wst/06705/wst067050935.htm.

[40] Derek B. Booth, David Hartley, and Rhett Jackson. “Forest Cover, Impervious Surface Area, and the Mitigation of Stormwater Impacts”. In: *Journal of the American Water Works Association* 38.3 (2002), pp. 835–845.

[41] J. J. Bos and Rebekah R. Brown. “Governance experimentation and factors of success in socio-technical transitions in the urban water sector”. In: *Technological Forecasting & Social Change* 79 (2012), pp. 1340–1353. DOI: 10.1016/j.techfore.2012.04.006.

[42] Christopher T. Boyko, Mark R. Gaterell, Austin R. G. Barber, Julie Brown, John R. Bryson, David Butler, Silvio Caputo, Maria Caserio, Richard Coles, Rachel Cooper, Gemma Davies, Raziyeh Farmani, James Hale, A. Chantal Hales, C. Nicholas Hewitt, Dexter V. L. Hunt, Lubo Jankovic, Ian Jefferson, Joanne M. Leach, D. Rachel Lombardi, A. Robert MacKenzie, Fayyaz A. Memon, Thomas A. M. Pugh, John P. Sadler, Carina Weingaertner, J. Duncan Whyatt, and Christopher D. F. Rogers. “Benchmarking sustainability in cities: The role of indicators and future scenarios”. In: *Global Environmental Change* 22.1 (2012), pp. 245–254. ISSN: 09593780. DOI: 10.1016/j.gloenvcha.2011.10.004.

[43] David J. Bradley and J. K. Bartram. “Domestic water and sanitation as water security: monitoring, concepts and strategy”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371 (2013), p. 20120420. DOI: 10.1098/rsta.2012.0420.

[44] Stefan Bringezu, Helmut Schütz, Sören Steger, and Jan Baudisch. “International comparison of resource use and its relation to economic growth: The development of total material requirement, direct material inputs and hidden flows and the structure of TMR”. In: *Ecological Economics* 51.1-2 (2004), pp. 97–124. DOI: 10.1016/j.ecolecon.2004.04.010.

[45] John Broich. *Water and the Making of a Modern City*. Pittsburgh, PA: University of Pittsburgh Press, 2013.

[46] David B. Brooks. “An Operational Definition of Water Demand Management”. In: *International Journal of Water Resources Development* 22.4 (2006), pp. 521–528. DOI: 10.1080/07900620600779699.

[47] David B. Brooks. “Beyond Greater Efficiency: The Concept of Water Soft Path”. In: *Canadian Water Resources Journal* 30.1 (2005), pp. 83–92. DOI: 10.4296/cwrj300183.

[48] Rebekah R. Brown and M. A. Farrelly. "Delivering sustainable urban water management : a review of the hurdles we face". In: *Water Science & Technology* 59.5 (2009), pp. 839–846. DOI: 10.2166/wst.2009.028.

[49] Rebekah R. Brown, N. Keath, and T. H. F. Wong. "Urban water management in cities: historical, current and future regimes". In: *Water Science & Technology* 59.5 (2009), pp. 847–856. DOI: 10.2166/wst.2009.029.

[50] Edmund G. Brown Jr. *Executive Order B-29-15*. Sacramento, CA, 2014.

[51] Anne Browning-Aiken, Kerri Jean Ormerod, and Christopher a. Scott. "Testing the Climate for Non-Potable Water Reuse: Opportunities and Challenges in Water-Scarce Urban Growth Corridors". In: *Journal of Environmental Policy & Planning* 13.3 (2011), pp. 253–275. DOI: 10.1080/1523908X.2011.594597.

[52] Matthew J. Burns, Tim D. Fletcher, Christopher J. Walsh, Anthony R. Ladson, and Belinda E. Hatt. "Setting objectives for hydrologic restoration: from site-scale to catchment-scale". In: *NOVATECH* (2013), pp. 1–9.

[53] Ximing Cai and Dingbao Wang. "Calibrating Holistic Water Resources–Economic Models". In: *Journal of Water Resources Planning and Management* 132.December (2006), pp. 414–423.

[54] California. *The Metropolitan Water District Act*. 2008.

[55] Antonio Candelieri and Francesco Archetti. "Identifying Typical Urban Water Demand Patterns for a Reliable Short-term Forecasting – The Icewater Project Approach". In: *Procedia Engineering* 89 (2014), pp. 1004–1012. DOI: 10.1016/j.proeng.2014.11.218.

[56] CH-CDM. *City of Los Angeles: Recycled Water Master Plan*. Report December. Los Angeles: Prepared for the City of Los Angeles, Department of Public Works, Bureau of Sanitation, Department of Water, and Power (LADWP), 2006.

[57] Marian R. Chertow. "The IPAT Equation and Its Variants: Changing Views of Technology and Environmental Impact". In: *Journal of Industrial Ecology* 4.4 (2001), pp. 13–29.

[58] Elliot Cohen, Karen M. Noiva, and Jonathan S. Krones. "Visualizing the spatiotemporal domain of water resource constraints to urban energy systems". In: June (2012), pp. 18–20.

[59] H. K. Colebatch. "Governing the use of water: The institutional context". In: *Desalination* 187.1-3 (2006), pp. 17–27. DOI: 10.1016/j.desal.2005.04.064.

[60] John Conlisk. "Why Bounded Rationality?" In: *Journal of Economic Literature* 34.2 (1996), pp. 669–700.

[61] Heather Cooley, Juliet Christian-Smith, Peter H. Gleick, Lucy Allen, and Michael Cohen. *Understanding and Reducing the Risks of Climate Change for Transboundary Waters*. Tech. rep. December. Oakland, California: Pacific Institute, 2009, p. 43. URL: www.pacinst.org.

[62] Andrew Cottam, Noel Gorelick, Alan S. Belward, and Jean-François Pekel. "High-resolution mapping of global surface water and its long-term changes". In: *Nature* (2016), pp. 1–19. ISSN: 0028-0836. DOI: 10.1038/nature20584.

[63] G. F. Craun and R. L. Calderon. "Waterborne disease outbreaks caused by distribution system deficiencies". In: *Journal / American Water Works Association* 93.9 (2001), pp. 64–75. ISSN: 0003150X.

[64] Felix Creutzig. "Towards typologies of urban climate and global environmental change". In: *Environmental Research Letters* 10 (2015), p. 101001. ISSN: 1748-9326. DOI: 10.1088/1748-9326/10/10/101001.

[65] David Cutler and Grant Miller. "The role of public health improvements in health advances: the twentieth-century United States." In: *Demography* 42.1 (2005), pp. 1–22.

[66] Simon J. Dadson, Michael Acreman, and Richard Harding. "Water security, global change and land-atmosphere feedbacks". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371 (2013).

[67] Peter L. Daniels, Manfred Lenzen, and Steven J. Kenway. "The Ins and Outs of Water Use – a Review of Multi-Region Input – Output Analysis and Water Footprints for Regional Sustainability Analysis and Policy". In: *Economic Systems Research* 23.4 (2011), pp. 353–370. ISSN: 0953-5314. DOI: 10.1080/09535314.2011.633500.

[68] Debra J. Davidson, Jeffrey Andrews, and Daniel Pauly. "The effort factor: Evaluating the increasing marginal impact of resource extraction over time". In: *Global Environmental Change* 25 (2014), pp. 63–68. DOI: 10.1016/j.gloenvcha.2014.02.001.

[69] Evan G. R. Davies and Slobodan P. Simonovic. "Global water resources modeling with an integrated model of the social-economic-environmental system". In: *Advances in Water Resources* 34.6 (2011), pp. 684–700. DOI: 10.1016/j.advwatres.2011.02.010.

[70] Kenyon B. De Greene and Jay W. Forrester. *A Systems-Based Approach to Policy-Making*. Ed. by Kenyon B. De Greene. Vol. 1. New York, NY: Springer Science and Business Media, 1993. DOI: 10.1017/CBO9781107415324.004.

[71] Rudolf S. De Groot, Matthew A. Wilson, and Roelof M. J. Boumans. "A typology for the classification, description and valuation of ecosystem functions, goods and services". In: *Ecological Economics* 41.3 (2002), pp. 393–408. ISSN: 09218009. DOI: 10.1016/S0921-8009(02)00089-7.

[72] Ethan H. Decker, Andrew J. Kerkhoff, and Melanie E. Moses. "Global patterns of city size distributions and their fundamental drivers." In: *PloS one* 2.9 (2007), e934. DOI: 10.1371/journal.pone.0000934.

[73] K. Diao, R. Farmani, G. Fu, M. Astaraie-Imani, S. Ward, and D. Butler. "Clustering analysis of water distribution systems: identifying critical components and community impacts." In: *Water Science & Technology* 70.11 (2014), pp. 1764–73. ISSN: 0273-1223. DOI: 10.2166/wst.2014.268.

[74] Sara Dolnicar, Anna Hurlimann, and Bettina Grün. "What affects public acceptance of recycled and desalinated water?" In: *Water Research* 45.2 (2011), pp. 933–943. ISSN: 00431354. DOI: 10.1016/j.watres.2010.09.030.

[75] Emmanuel A. Donkor, Thomas A. Mazzuchi, Refik Soyer, and J. Alan Roberson. "Urban Water Demand Forecasting: A Review of Methods and Models". In: *Journal of Water Resources Planning and Management* (2012), p. 120913012720004. DOI: 10.1061/(ASCE)WR.1943-5452.0000314.

[76] David Drakakis-Smith, Tanya Bowyer-Bower, and Dan Tevera. "Urban poverty and urban agriculture: An overview of the linkages in Harare". In: *Habitat International* 19.2 (1995), pp. 183–193. DOI: 10.1016/0197-3975(94)00065-A.

[77] Rosa Duarte, Vicente Pinilla, and Ana Serrano. "Looking backward to look forward: water use and economic growth from a long-term perspective." In: *Applied Economics* 46.2 (2014), pp. 212–224. DOI: 10.1080/00036846.2013.844329.

[78] Kimberly Duong and Jean-Daniel M. Saphores. "Obstacles to wastewater reuse: an overview". In: *WIREs Water* 2 (2015), pp. 199–214. ISSN: 20491948. DOI: 10.1002/wat2.1074. URL: http://doi.wiley.com/10.1002/wat2.1074.

[79] Jochen Eckart. "Flexible Urban Drainage Systems in New Land-Use Areas". Doctor of Philosophy. University of South Florida, 2012.

[80] EEA. "Performance of water utilities beyond compliance: Sharing knowledge bases to support environmental and resource-efficiency policies and technical improvements". Luxembourg, 2014.

[81] David Ellison, Martyn N. Futter, and Kevin Bishop. "On the forest cover-water yield debate: from demand- to supply-side thinking". In: *Global Change Biology* 18.3 (2012), pp. 806–820. DOI: 10.1111/j.1365-2486.2011.02589.x.

[82] A. Ertug Ercin, Maite Martinez Aldaya, and Arjen Y. Hoekstra. "Corporate Water Footprint Accounting and Impact Assessment: The Case of the Water Footprint of a Sugar-Containing Carbonated Beverage". In: *Water Resources Management* 25.2 (2011), pp. 721–741. DOI: 10.1007/s11269-010-9723-8.

[83] A. Etter, C.A. McAlpine, L. Seabrook, and K.A. Wilson. "Incorporating temporality and biophysical vulnerability to quantify the human spatial footprint on ecosystems". In: *Biological Conservation* 144.5 (2011), pp. 1585–1594. DOI: 10.1016/j.biocon.2011.02.004.

[84] J. S. Famiglietti, A. Cazenave, A. Eicker, J. T. Reager, M. Rodell, and I. Velicogna. "Watching water: From sky or stream?" In: *Science* 349.6249 (2015), pp. 684–685.

[85] Alexander E. Farrell, Jill Jäger, and Stacy D. VanDeveer. "Overview: Understanding Design Choices". In: *Assessments of Regional and Global Environmental Risks: Designing Processes for the Effective Use of Science in Decisionmaking*. Washington, D.C.: Routledge, 2006. Chap. Chapter 1. ISBN: 1-933115-05-X.

[86] John E. Fernández, Karen M. Noiva, Yuliya Kalmykova, and Artessa Niccola D. Saldivar-Sali. *Internal Research Report City Typologies: Derivation and Results from Phases I and II*. Tech. rep. Cambridge, MA, USA: Massachusetts Institute of Technology, 2013.

[87] Paulo Ferrão and John E. Fernández. *Sustainable Urban Metabolism*. Cambridge, MA: MIT Press, 2013, p. 264. ISBN: 026231696X.

[88] Joseph Fiksel. "Sustainability and resilience: toward a systems approach". In: *Sustainability: Science, Practice, & Policy* 2.2 (2006), p. 8.

[89] Walter Leal Filho. *Climate Change Management: Climate Change and the Sustainable Use of Water Resources*. Ed. by Walter Leal Filho. Hamburg: Springer, 2012.

[90] Jonathan Fink, Glen T. Daigger, and Hillary Brown. *Urban Sustainability*. Ed. by George Bugliarello and Carol A. Arenberg. Washington, D.C.: The National Academy of Engineering, 2011.

[91] Marina Fischer-Kowalski and Christof Amann. "Beyond IPAT and Kuznets Curves: Globalization as a Vital Factor in Analysing the Environmental Impact of Socio-Economic Metabolism". In: *Population and Environment* 23.1 (2001), pp. 7–47.

[92] Thomas W. Fitzhugh and Brian D. Richter. "Quenching Urban Thirst: Growing Cities and Their Impacts on Freshwater Ecosystems". In: *BioScience* 54.8 (2004), p. 741.

[93] John Fleck. *What Seven States Can Agree to Do: Deal-Making on the Colorado River*. Palo Alto, 2012. URL: `http://web.stanford.edu/group/ruralwest/cgi-bin/drupal/content/what-seven-states-can-agree-do-deal-making-colorado-river`.

[94] Tim D. Fletcher and Ana Deletic. *Data Requirements for Integrated Urban Water Management: Urban Water Series - UNESCO-IHP (Vol. 1)*. Ed. by Tim Fletcher and Ana Deletic. CRC Press, 2007.

[95] C. M. Fontanazza, V. Notaro, V. Puleo, and G. Freni. "Multivariate Statistical Analysis for Water Demand Modeling". In: *Procedia Engineering* 89 (2014), pp. 901–908. ISSN: 1877-7058. DOI: `10.1016/j.proeng.2014.11.523`.

[96] Andrew W. Ford. *Modeling the Environment*. 2nd. Washington, D.C.: Island Press, 2009.

[97] Jay W. Forrester. "Counterintuitive behavior of social systems". In: *Technological Forecasting and Social Change* 3 (1971), pp. 1–22. ISSN: 00401625. DOI: `10.1016/S0040-1625(71)80001-X`.

[98] Xavier Gabaix. "Zipf's Law for Cities: An Explanation". In: *The Quarterly Journal of Economics* (1999).

[99] Veerender Garg. "Forecasting the Water Demand Using Regression and Cluster Analysis for Salt Lake City". Master's Thesis. Utah State University, 2007. ISBN: 0549530444.

[100] Dustin Garrick and Jim W. Hall. "Water Security and Society: Risks, Metrics, and Pathways". In: *Annual Review of Environment and Resources* 39 (2014), pp. 611–639. DOI: 10.1146/annurev-environ-013012-093817.

[101] Francis Gassert, Paul Reig, Tianyi Luo, and Andrew Maddocks. "A Weighted Aggregation of Spatially Distinct Hydrological Indicators". Washington, D.C., 2013. URL: wri.org/publication/aqueduct-country-river-basin-rankings.

[102] Frank W. Geels. "Co-evolution of technology and society: The transition in water supply and personal hygiene in the Netherlands (1850–1930)–a case study in multi-level perspective". In: *Technology In Society* 27 (2005), pp. 363–397. DOI: 10.1016/j.techsoc.2005.04.008.

[103] Frank W. Geels and René Kemp. "Dynamics in socio-technical systems: Typology of change processes and contrasting case studies". In: *Technology in Society* 29.4 (2007), pp. 441–455. ISSN: 0160791X. DOI: 10.1016/j.techsoc.2007.08.009.

[104] Dieter Gerten, Jens Heinke, Holger Hoff, H. Biemans, M. Fader, and K. Waha. "Global Water Availability and Requirements for Future Food Production". In: *Journal of Hydrometeorology* 12.5 (2011), pp. 885–899. DOI: 10.1175/2011JHM1328.1.

[105] Peter H. Gleick. "Basic Water Requirements for Human Activities: Meeting Basic Needs". In: *Water International* 21.2 (1996), pp. 83–92.

[106] Peter H. Gleick. "Global Freshwater Resources : Soft-Path Solutions for the 21st Century". In: 1524.2003 (2008). DOI: 10.1126/science.1089967.

[107] Peter H. Gleick. "The Changing Water Paradigm: A Look at Twenty-first Century Water Resources Development". In: *Water International* 25.1 (2000), p. 1270138.

[108] Peter H. Gleick. "The human right to water". In: *Water Policy* 1.1998 (1998), pp. 487–503.

[109] Peter H. Gleick. "Water and Conflict: Fresh Water Resources and International Security". In: *International Security* 18.1 (1993), pp. 79–112.

[110] Peter H. Gleick. *Water in Crisis: A Guide to the World's Fresh Water Resources*. Oxford: Oxford University Press, 1993, pp. 13–24. ISBN: 0-19-507628-1.

[111] Peter H. Gleick. "Water Use". In: *Annual Review of Environment and Resources* 28.1 (2003), pp. 275–314. DOI: 10.1146/annurev.energy.28.040202.122849.

[112] Peter H. Gleick and Meena Palaniappan. "Peak water limits to freshwater withdrawal and use". In: *Proceedings of the National Academy of Sciences of the United States of America* 107.25 (2010), pp. 11155–62. DOI: 10.1073/pnas.1004812107.

[113] Patricia Gober. "Getting Outside the Water Box: The Need for New Approaches to Water Planning and Policy". In: *Water Resources Management* 27.4 (2013), pp. 955–957. DOI: 10.1007/s11269-012-0222-y.

[114] Alireza Gohari, Saeid Eslamian, Ali Mirchi, Jahangir Abedi-Koupaei, and Alireza Massah. "Water transfer as a solution to water shortage: A fix that can Backfire". In: *Journal of Hydrology* 491 (2013), pp. 23–39. DOI: 10.1016/j.jhydrol.2013.03.021.

[115] M. Gómez, A. de la Rua, G. Garralón, F. Plaza, E. Hontoria, and M. A. Gómez. "Urban wastewater disinfection by filtration technologies". In: *Desalination* 190.1-3 (2006), pp. 16–28. DOI: 10.1016/j.desal.2005.07.014.

[116] Daphne Gondhalekar, Peter P. Mollinga, and V. S. Saravanan. "Towards systematic comparative water and health research". In: *Water International* 38.7 (2013), pp. 967–976. DOI: 10.1080/02508060.2013.857141.

[117] Simon N. Gosling and Nigel W. Arnell. "A global assessment of the impact of climate change on water scarcity". In: *Climatic Change* 134 (2016), pp. 371–385. DOI: 10.1007/s10584-013-0853-x.

[118] D. Grey, D. Garrick, D. Blackmore, J. Kelman, M. Muller, and C. Sadoff. "Water security in one blue planet: twenty-first century policy challenges for science". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371 (2013), p. 20120406. DOI: 10.1098/rsta.2012.0406.

[119] Neil S. Grigg. *Urban Water Infrastructure: Planning, Management, and Operations*. Malabar, Florida: Krieger Publishing Company, 1992.

[120] Veera Gnaneswar Gude. "Desalination and sustainability–An appraisal and current perspective". In: *Water Research* 89 (2016), pp. 87–106. DOI: 10.1016/j.watres.2015.11.012.

[121] J. W. Hall, D. Grey, Dustin Garrick, F. Fung, C. Brown, and Simon J. Dadson. "Coping with the curse of freshwater variability". In: *Science* 346.6208 (2014), pp. 429–430.

[122] M. C. Hamilton, W. Goldsmith, R. Harmon, D. Lewis, B. Srdjevic, M. Goodsite, J. H. Lambert, and M. Macdonell. "Sustainable Water Resources Management: Challenges and Methods". In: *Sustainable Cities and Military Installations*. Dordrecht: Springer Science, 2014. Chap. 6, pp. 133–144. ISBN: 9789400771611. DOI: 10.1007/978-94-007-7161-1.

[123] Luoheng Han. "Applying Geographic Information Techniques to Study Water Resources for the Next 20 Years". In: *Journal of Contemporary Water Research & Education* 142.1 (2009), pp. 25–27. ISSN: 1936-704X. DOI: 10.1111/j.1936-704X.2009.00048.x.

[124] N. Hanasaki, S. Fujimori, T. Yamamoto, S. Yoshikawa, Y. Masaki, and Y. Hijioka. "A global water scarcity assessment under shared socio-economic pathways–Part 2: Water availability and scarcity". In: *Hydrology and Earth System Sciences Discussions* 9 (2012), pp. 13933–13994. DOI: 10.5194/hessd-9-13933-2012.

[125] Mona Hanna-Attisha, Jenny LaChance, Richard Casey Sadler, and Allison Champney Schnepp. "Elevated blood lead levels in children associated with the flint drinking water crisis: A spatial analysis of risk and public health response". In: *American Journal of Public Health* 106.2 (2016), pp. 283–290. ISSN: 15410048. DOI: 10.2105/AJPH.2015.303003.

[126] Nilgun B. Harmancioglu, Filiz Barbaros, and Cem P. Cetinkaya. "Sustainability Issues in Water Management". In: *Water Resources Management* 27.6 (2013), pp. 1867–1891. ISSN: 0920-4741. DOI: 10.1007/s11269-012-0172-4.

[127] Sasha R. Harris-Lovett, Christian Binz, David L. Sedlak, Michael Kiparsky, and Bernhard Truffer. "Beyond User Acceptance: A Legitimacy Framework for Potable Water Reuse in California". In: *Environmental Science and Technology* 49.13 (2015), pp. 7552–7561. ISSN: 15205851. DOI: 10.1021/acs.est.5b00504.

[128] Tsuyoshi Hashimoto, Jery R. Stedinger, and Daniel P. Loucks. "Reliability, Resiliency, and Vulnerability Criteria for Water Resource System Performance Evaluation". In: *Water Resources Research* 18.1 (1982), pp. 14–20.

[129] Mohamad Hejazi, James Edmonds, Leon Clarke, Page Kyle, Evan Davies, Vaibhav Chaturvedi, Marshall Wise, Pralit Patel, Jiyong Eom, Katherine Calvin, Richard Moss, and Son H. Kim. "Long-term global water projections using six socioeconomic scenarios in an integrated assessment modeling framework". In: *Technological Forecasting & Social Change* 81 (2014), pp. 205–226. DOI: 10.1016/j.techfore.2013.05.006.

[130] Janet G. Hering, T. David Waite, Richard G. Luthy, Jörg E. Drewes, and David L. Sedlak. "A changing framework for urban water systems". In: *Environmental Science & Technology* 47.19 (2013), pp. 10721–6. DOI: 10.1021/es4007096.

[131] Janet G. Hering, David L. Sedlak, Cecilia Tortajada, and Asit K. Biswas. "What scale for water governance? Local perspectives on water". In: *Science* 349.6247 (2015), pp. 479–480.

[132] Robert Hijmans, Julian Kapoor, Wieczorek, Nel Garcia, Aileen Maunahan, Arnel Rala, and Mandel Alex. *GADM database of Global Administrative Areas*. Shapefile. 2015. URL: http://www.gadm.org/.

[133] Arjen Y. Hoekstra and Ashok K. Chapagain. "Water footprints of nations: Water use by people as a function of their consumption pattern". In: *Water Resources Management* 21.1 (2007), pp. 35–48. ISSN: 0920-4741. DOI: 10.1007/s11269-006-9039-x.

[134] Arjen Y. Hoekstra and Mesfin M. Mekonnen. "The water footprint of humanity". In: *Proceedings of the National Academy of Sciences* 109.9 (2012). DOI: 10.1073/pnas.1109936109.

[135] Holger Hoff, P. Döll, M. Fader, Dieter Gerten, S. Hauser, and S. Siebert. "Water footprints of cities – indicators for sustainable consumption and production". In: *Hydrology and Earth System Sciences* 18.1 (2014), pp. 213–226. DOI: 10.5194/hess-18-213-2014.

[136] R. Hope and M. Rouse. "Risks and responses to universal drinking water security". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371 (2013), p. 20120417. DOI: 10.1098/rsta.2012.0417.

[137] George M. Hornberger, Jeffrey P. Raffensperger, Patricia L. Wiberg, and Keith N. Eshleman. *Elements of Physical Hydrology*. Baltimore: John Hopkins University Press, 1998.

[138] Lily A. House-Peters and Heejun Chang. "Urban water demand modeling: Review of concepts, methods, and organizing principles". In: *Water Resources Research* 47.5 (2011), W05401. DOI: 10.1029/2010WR009624.

[139] Shu-Li Huang, Chia-Tsung Yeh, and Li-Fang Chang. "The transition to an urbanizing world and the demand for natural resources". In: *Current Opinion in Environmental Sustainability* 2.3 (2010), pp. 136–143. ISSN: 18773435. DOI: 10.1016/j.cosust.2010.06.004.

[140] Norris Hundley, Jr. *The Great Thirst: Californians and Water: A History*. Revised. Berkeley, CA: University of California Press, 2001.

[141] Anna Hurlimann and Sara Dolnicar. "When public opposition defeats alternative water projects - The case of Toowoomba, Australia". In: *Water Research* 44.1 (2010), pp. 287–297. ISSN: 00431354. DOI: 10.1016/j.watres.2009.09.020. URL: http://dx.doi.org/10.1016/j.watres.2009.09.020.

[142] Charles F. Hutchinson, Robert G. Varady, and Sam Drake. "Old and New: Changing Paradigms in Arid Lands Water Management". In: *Water and Sustainability in Arid Regions*. Ed. by G. Schneier-Madanes and M.-F. Courel. Springer Science and Business Media, 2010. Chap. 19, pp. 139–149. DOI: 10.1007/978-90-481-2776-4.

[143] Guy Hutton. "Global costs and benefits of drinking-water supply and sanitation interventions to reach the MDG target and universal coverage". Geneva, 2012.

[144] IBNET. *IB-NET Database*. 2015. URL: https://database.ib-net.org.

[145] Yannis M. Ioannides and Henry G. Overman. "Zipf's law for cities: an empirical examination". In: *Regional Science and Urban Economics* 33 (2003), pp. 127–137.

[146] IPCC. *Climate Change 2014 Synthesis Report. Contribution of Working Groups I, II and III to the Fifthe Assessment Report of the Intergovernmental Panel on Climate Change*. Geneva, Switzerland: Intergovernmental Panel on Climate Change (IPCC), 2014.

[147] Giorgos Kallis. "Droughts". In: *Annual Review of Environment and Resources* 33.1 (2008), pp. 85–118. DOI: 10.1146/annurev.environ.33.081307.123117.

[148] Melinda Kane and Jon D. Erickson. *Urban Metabolism and Payment for Ecosystem Services: History and Policy Analysis of the New York City Water Supply*. 2007. DOI: 10.1016/S1569-3740(07)07013-7.

[149] Tapio S. Katko, Petri S. Juuti, and Riikka P. Rajala. "Writing the history of water services". In: *Physics and Chemistry of the Earth* 34.3 (2009), pp. 156–163. DOI: 10.1016/j.pce.2008.06.033.

[150] James J. Kay, Henry A. Regier, Michelle Boyle, and George Francis. "An ecosystem approach for sustainability: addressing the challenge of complexity". In: *Futures* 31.7 (1999), pp. 721–742. DOI: 10.1016/S0016-3287(99)00029-4.

[151] Jason P. Kaye, Peter M. Groffman, Nancy B. Grimm, Lawrence A. Baker, and Richard V. Pouyat. “A distinct urban biogeochemistry?” In: *Trends in ecology & evolution* 21.4 (2006), pp. 192–9. DOI: 10.1016/j.tree.2005.12.006.

[152] Christopher A. Kennedy, John Cuddihy, and Joshua Engel-yan. “The Changing Metabolism of Cities”. In: *Journal of Industrial Ecology* 11.2 (2007), pp. 43–59.

[153] Emily H. Kennedy. “Reclaiming Consumption: Sustainability, Social Networks, and Urban Context”. Doctoral Dissertation. University of Alberta, 2011. ISBN: 9780494908112. URL: https://era.library.ualberta.ca/downloads/br86b4754.

[154] Steven J. Kenway. “The Water-Energy Nexus and Urban Metabolism - Connections in Cities”. Queensland, 2013.

[155] Steven J. Kenway, Alan Gregory, and Joseph Mcmahon. “Urban Water Mass Balance Analysis”. In: *Journal of Industrial Ecology* 15.5 (2011). DOI: 10.1111/j.1530-9290.2011.00357.x.

[156] Steven J. Kenway, P. A. Lant, A. Priestley, and P. Daniels. “The connection between water and energy in cities: a review”. In: *Water Science & Technology* 63.9 (2011), p. 1983. DOI: 10.2166/wst.2011.070.

[157] Adam Khamis. “Developing a Typology of Urban Resource Consumption”. In: *1st Civil and Environmental Engineering Student Conference*. London: Imperial College, London, 2012.

[158] Jeong-Sam Kim. “The differentiation of South Korean cities: a multivariate analysis”. English. Master’s thesis. Calgary: University of Calgary, 1997, p. 144. ISBN: 0612245969.

[159] M. Konar, C. Dalin, N. Hanasaki, A. Rinaldo, and I. Rodriguez-Iturbe. “Temporal dynamics of blue and green virtual water trade networks”. In: *Water Resources Research* 48.7 (2012), W07509. DOI: 10.1029/2012WR011959.

[160] Markus Kottek, Jürgen Grieser, Christoph Beck, Bruno Rudolf, and Franz Rubel. “World Map of the Köppen-Geiger climate classification updated”. In: *Meteorologische Zeitschrift* 15.3 (2006), pp. 259–263. DOI: 10.1127/0941-2948/2006/0130.

[161] Matti Kummu, Philip J. Ward, Hans De Moel, and Olli Varis. “Is physical water scarcity a new phenomenon? Global assessment of water shortage over the last two millennia”. In: *Environmental Research Letters* 034006 (2010). DOI: 10.1088/1748-9326/5/3/034006.

[162] Ellen Kynast, Ilona Ba, Stephanie Eisner, Florian Wimmer, Joseph Alcamo, and Martina Flo. “Domestic and industrial water uses of the past 60 years as a mirror of socio-economic development: A global simulation study”. In: *Global Environmental Change* 23 (2013), pp. 144–156. DOI: 10.1016/j.gloenvcha.2012.10.018.

[163] LA. *Demographic and Economic Statistics*. Los Angeles, 2015.

[164] LA. *LADWP Water Supply in Acre Feet*. Los Angeles, 2017. URL: https://catalog.data.gov/dataset/ladwp-water-supply-in-acre-feet-91497.

[165] LA-Almanac. *General Population by City, Los Angeles County. Estimated Populations, 2011-2016 (on January 1).* Los Angeles, 2017. URL: http://www.laalmanac.com/population/po24a.php.

[166] LACounty. *City Boundaries for Los Angeles County*. Shapefile. Los Angeles, CA, 2011. URL: https://egis3.lacounty.gov/dataportal/drp{_}county{_}boundary/.

[167] LADWP. *Annual Report of the Board of Water Commissioners of the Domestic Water Works System of the City of Los Angeles, for the Fiscal Year Ending November 30, 1902*. Tech. rep. Los Angeles, CA: City of Los Angeles Department of Water and Power (LADWP), 1902.

[168] LADWP. *LADWP Water Facts & Figures*. 2016. (Visited on 01/27/2017).

[169] LADWP. *Los Angeles Department of Water and Power: 2015 Briefing Book*. Tech. rep. Los Angeles: Los Angeles Department of Water and Power (LADWP), 2015.

[170] LADWP. *Twenty-Second Annual Report of the Board of Public Services Commissioners of the City of Los Angeles, for the Fiscal Year Ending June 30, 1923*. Tech. rep. Los Angeles, CA: City of Los Angeles Department of Water and Power (LADWP), 1923.

[171] LADWP. *Urban Water Management Plan: 2015*. Report. 2016.

[172] Andrew C. Lemer. "Infrastructure obsolescence and design service life". In: *Journal of Infrastructure Systems* 2.December (1996), pp. 153–161.

[173] Robert J. Lempert, S. Popper, and S. Bankes. "Confronting Surprise". In: *Social Science Computer Review* 20.4 (2002), pp. 420–440. DOI: 10.1177/089443902237320.

[174] David Levi-Faur. "A Question of Size? A Heuristics for Stepwise Comparative Research Design". In: *Innovative Comparative Methods for Policy Analysis: Beyond the Quantitative-Qualitative Divide*. Ed. by Benoit Rihoux and Heike Grimm. New York: Springer, 2006. Chap. 3, pp. 43–67.

[175] Enjie Li, Joanna Endter-Wada, and Shujuan Li. "Characterizing and Contextualizing the Water Challenges of Megacities". In: *Journal of the American Water Resources Association* 51.3 (2015). DOI: 10.1111/1752-1688.12310.

[176] Chuan-Yao Lin, Wan-Chin Chen, Pao-Liang Chang, and Yang-Fan Sheng. "Impact of the Urban Heat Island Effect on Precipitation over a Complex Geographic Environment in Northern Taiwan". In: *Journal of Applied Meteorology and Climatology* 50 (2011), pp. 339–353. DOI: 10.1175/2010JAMC2504.1.

[177] An Liu, Ashantha Goonetilleke, and Prasanna Egodawatta. "Inadequacy of Land Use and Impervious Area Fraction for Determining Urban Stormwater Quality". In: *Water Resources Management* 26.8 (2012), pp. 2259–2265. ISSN: 09204741. DOI: 10.1007/s11269-012-0014-4.

[178] Hugo A. Loaiciga, Juan B. Valdes, Richard M. Vogel, Jeff Garvey, and Harry Schwarz. "Global warming and the hydrologic cycle". In: *Journal of Hydrology* 174.1-2 (1996), pp. 83–127. DOI: 10.1016/0022-1694(95)02753-X.

[179] Philippe Loubet, Philippe Roux, Eleonore Loiseau, and Veronique Bellon-Maurel. "Life cycle assessments of urban water systems: A comparative analysis of selected peer-reviewed literature". In: *Water Research* 67 (2014), pp. 187–202. DOI: `10.1016/j.watres.2014.08.048`.

[180] Daniel P. Loucks. "Sustainable Water Resources Management". In: *Water International* 25.1 (2000), pp. 3–10. DOI: `10.1080/02508060008686793`.

[181] Daniel P. Loucks, Eelco van Beek, Jery R. Stedinger, Jos P. M. Dijkman, and Monique T. Villars. *Water Resources Systems Planning and Management: An Introduction to Methods, Models, and Applications*. Delft: UNESCO, 2005. ISBN: 9231039989.

[182] L.J.P. van der Maaten and G. E. Hinton. "Visualizing High-Dimensional Data Using t-SNE". In: *Journal of Machine Learning Research* 9 (2008), pp. 2579–2605. URL: `http://lvdmaaten.github.io/tsne/`.

[183] Darla Hatton Macdonald, Sébastien Lamontagne, and Jeffery Connor. "The Economics of Water: Taking Full Account of First Use, Reuse, and the Return to the Environment". In: *Irrigation and Drainage* 102.August 2004 (2005), pp. 93–102.

[184] M. Mair, W. Rauch, and Robert Sitzenfrei. "Improving incomplete water distribution system data". In: *Procedia Engineering* 70 (2014), pp. 1055–1062. ISSN: 1877-7058. DOI: `10.1016/j.proeng.2014.02.117`.

[185] Mrinmoy Majumder. *Impact of Urbanization on Water Shortage in Face of Climatic Aberrations*. London: Springer, 2015. DOI: `10.1007/978-981-4560-73-3`.

[186] E. L. Malone. "Vulnerability and Resilience in the Face of Climate Change: Current Research and Needs for Population Information". Richland, Washington, 2009.

[187] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008. URL: `http://www-nlp.stanford.edu/IR-book/`.

[188] David R. Marlow, Magnus Moglia, Stephen Cook, and David J. Beale. "Towards sustainable urban water management: A critical reassessment". In: *Water Research* 47.20 (2013), pp. 7150–7161. DOI: `10.1016/j.watres.2013.07.046`.

[189] Jiri Marsalek, Blanca Jiménez-Cisneros, Mohammad Karamouz, Per-Arne Malmquist, Joel Goldenfum, and Bernard Chocat. "Urban Water Cycle Processes and Interactions". Paris, 2006. URL: `www.unesco.org/water/ihp`.

[190] Kenji Matsuura and Cort J. Willmott. *Terrestrial Water Budget Data Archive*. Data tables. Newark, Delaware, 2015. URL: `climate.geog.udel.edu/{~}climate/html{\_}pages/Global2014/`.

[191] Kenji Matsuura, Cort J. Willmott, D. Legates, and Clinton M. Rowe. *WebWIMP: The Web-based, Water-Budget, Interactive, Modeling Program*. 2009. URL: `climate.geog.udel.edu/{~}wimp/`.

[192] Alex Mayer, Richelle Winkler, and Lauren Fry. "Classification of watersheds into integrated social and biophysical indicators with clustering analysis". In: *Ecological Indicators* 45 (2014), pp. 340–349. ISSN: 1470160X. DOI: 10.1016/j.ecolind.2014.04.030.

[193] Larry W. Mays. *Ancient Water Technologies*. New York: Springer, 2010. ISBN: 9789048186310.

[194] Larry W. Mays. *Water Resources Engineering*. Tempe, AZ: Wiley, 2011.

[195] Larry W. Mays, D. Koutsoyiannis, and Andreas N. Angelakis. "A brief history of urban water supply in antiquity". In: *Water Science & Technology: Water Supply* 7.1 (2007), p. 1. DOI: 10.2166/ws.2007.001.

[196] Robert I. McDonald and Daniel Shemie. *Urban Water Blueprint*. Washington, D.C.: The Nature Conservancy, 2014.

[197] Robert I. McDonald, Pamela A. Green, Deborah Balk, Balázs M. Fekete, Carmen Revenga, Megan Todd, and Mark R. Montgomery. "Urban growth, climate change, and freshwater availability." In: *Proceedings of the National Academy of Sciences of the United States of America* 108.15 (2011), pp. 6312–7. ISSN: 1091-6490. DOI: 10.1073/pnas.1011615108.

[198] Robert I. McDonald, Katherine Weber, Julie Padowski, Martina Flörke, Christof Schneider, Pamela A. Green, Thomas Gleeson, Stephanie Eckman, Bernhard Lehner, Deborah Balk, Timothy Boucher, Günther Grill, and Mark R. Montgomery. "Water on an urban planet: Urbanization and the reach of urban water infrastructure". In: *Global Environmental Change* 27.1 (2014), pp. 96–105. ISSN: 09593780. DOI: 10.1016/j.gloenvcha.2014.04.022.

[199] Gordon McGranahan, Peter J. Marcotullio, Xuemei Bai, Deborah Balk, Tania Braga, Ian Douglas, Thomas Elmqvist, William E. Rees, Jacob Songsore, and Hania Zlotnik. "Urban Systems". In: *Ecosystems and Human Well-being: Current State and Trends*. Ed. by Jerry Eades and Exequiel Ezcurra. Washington, D.C.: Island Press, 2005. Chap. 27.

[200] James E. McMahon and Sarah K. Price. "Water and Energy Interactions". In: *Annual Review of Environment and Resources* 36.1 (2011), pp. 163–191. DOI: 10.1146/annurev-environ-061110-103827.

[201] Thomas A. McMahon, Geoffrey G.S. Pegram, Richard M. Vogel, and Murray C. Peel. "Revisiting reservoir storage-yield relationships using a global streamflow database". In: *Advances in Water Resources* 30.8 (2007), pp. 1858–1872. DOI: 10.1016/j.advwatres.2007.02.003.

[202] MEA. *Ecosystems and Human Well-Being: Wetlands and Water Synthesis*. Washington, D.C.: Water Resources Institute, 2005. DOI: 10.1017/CBO9781107415324.004.

[203] Wietske Medema, Brian S. Mcintosh, and Paul J. Jeffrey. "From Premise to Practice: A Critical Assessment of Integrated Water Resources Management and Adaptive Management Approaches in the Water Sector". In: *Ecology and Society* 13.2 (2008), p. 29.

[204] M. H. Mehaffey, M. S. Nash, T. G. Wade, D. W. Ebert, K. B. Jones, and A. Rager. "Linking land cover and water quality in New York City's water supply watersheds." In: *Environmental Monitoring and Assessment* 107.1-3 (2005), pp. 29–44. DOI: 10.1007/s10661-005-2018-5.

[205] Ruth Meinzen-Dick. "Beyond panaceas in water institutions". In: *Proceedings of the National Academy of Sciences of the United States of America* 104.39 (2007), pp. 15200–5. DOI: 10.1073/pnas.0702296104.

[206] Mesfin M. Mekonnen and Arjen Y. Hoekstra. "Four billion people facing severe water scarcity". In: *Science Advances* 2.February (2016), e1500323.

[207] Mesfin M. Mekonnen and Arjen Y. Hoekstra. "National Water Footprint Accounts: The Green, Blue, and Grey Water Footprint of Production and Consumption: Volume 1: Main Report". Delft, 2011.

[208] Assefa M. Melesse and Wossenu Abtew. "The Nile River Basin". In: *Nile River Basin: Ecohydrological Challenges, Climate Change and Hydropolitics*. Switzerland: Springer International Publishing, 2014. Chap. Chapter 2, pp. 7–16. DOI: 10.1007/978-3-319-02720-3.

[209] John Stuart Mill. *A System of Logic Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. Collected. Toronto: University of Toronto Press, 1843.

[210] Cass T. Miller, Clint N. Dawson, Matthew W. Farthing, Thomas Y. Hou, Jingfang Huang, Christopher E. Kees, C. T. Kelley, and Hans Petter. "Advances in Water Resources Numerical simulation of water resources problems: Models, methods, and trends". In: *Advances in Water Resources* 51 (2013), pp. 405–437. ISSN: 0309-1708. DOI: 10.1016/j.advwatres.2012.05.008.

[211] Grant Miller and David Cutler. "Water, Water Everywhere: Municipal Finance and Water Supply in American Cities". In: *Corruption and Reform: Lessons from America's Economic History*. March. University of Chicago Press, 2006. URL: http://www.nber/org/chapters/c9982.

[212] P. C. D. Milly, Julio Betancourt, Malin Falkenmark, Robert M. Hirsch, W. Zbigniew, Dennis P. Lettenmaier, and Ronald J. Stouffer. "Stationarity Is Dead: Whither Water Management ?" In: *Science* 319.February (2008), pp. 573–574.

[213] Caroline Mini. *Residential water use and landscape vegetation dynamics in Los Angeles*. Ph.D. Los Angeles, CA, 2013.

[214] Ali Mirchi, Kaveh Madani, David Watkins, and Sajjad Ahmad. "Synthesis of System Dynamics Tools for Holistic Conceptualization of Water Resources Problems". In: *Water Resources Management* 26.9 (2012), pp. 2421–2442. ISSN: 09204741. DOI: 10.1007/s11269-012-0024-2.

[215] V. G. Mitchell, H. Duncan, M. Inman, M. Rahilly, J. Stewart, A. Vieritz, P. Holt, A. Grant, Tim D. Fletcher, J. Coleman, S. Maheepala, A. Sharma, Ana Deletic, and P. Breen. "State of the art review of integrated urban water models". In: *NOVATECH*. 2007, pp. 507–514.

[216] Christine L. Moe and Richard D. Rheingans. "Global challenges in water, sanitation and health". In: *Journal of Water and Health* 4.SUPPL. 1 (2006), pp. 41–58. ISSN: 14778920. DOI: 10.2166/wh.2005.039.

[217] François Molle and Peter P. Mollinga. "Water poverty indicators: Conceptual problems and policy issues". In: *Water Policy* 5.5-6 (2003), pp. 529–544. ISSN: 13667017.

[218] Peter P. Mollinga and Daphne Gondhalekar. "Finding Structure in Diversity: A Stepwise Small-N/Medium-N Qualitative Comparative Analysis Approach for Water Resources Management Research". In: *Water Alternatives* 7.1 (2014), pp. 178–198. ISSN: 19650175.

[219] Peter P. Mollinga and Daphne Gondhalekar. "Theorising Structured Diversity. An approach to comparative research on water resources management". London, 2012.

[220] Jason Morrison, Peter Schulte, and Rita Schenck. *Corporate Water Accounting: An Analysis of Methods and Tools for Measuring Water Use and Its Impacts*. Tech. rep. Oakland, CA: United Nations Environment Programme's Division of Technology, INdustry, and Economics (UNEP DTIE), 2009.

[221] Erling Moxnes. "Misperceptions of basic dynamics: the case of renewable resource management". In: *System Dynamics Review* 20.2 (2004), pp. 139–162. ISSN: 0883-7066. DOI: 10.1002/sdr.289. URL: http://doi.wiley.com/10.1002/sdr.289.

[222] Mike Muller, Asit K. Biswas, Roberto Martin-Hurtado, and Cecilia Tortajada. "Built infrastructure is essential". In: *Science* 1594.2014 (2015), pp. 2014–2016.

[223] MWDSC. *2015 Urban Water Management Plan*. Tech. rep. Los Angeles, CA: Metropolitan Water District of Southern Callifornia (MWDSC), 2016.

[224] MWDSC. *The Metropolitan Water District Act & Code*. 2015. URL: http://www.mwdh2o.com/WhoWeAre/MWDAct/Pages/default.aspx (visited on 01/17/2017).

[225] MWDSC. *The Metropolitan Water District Administrative Code (from the Board of Directors Meeting, August 16, 2016)*. Los Angeles, CA, 2016.

[226] Norman Myers and Jennifer Kent. "New consumers: The influence of affluence on the environment". In: *Proceedings of the National Academy of Sciences of the United States of America* 100.8 (2003), pp. 4963–8. DOI: 10.1073/pnas.0438061100.

[227] NACWA, WERF, and WEF. *The Water Resources Utility of the Future: A Blueprint for Action*. Tech. rep. Washington, D.C.: National Association of Clean Water Agencies (NACWA), Water Environment Research Foundation (WERF); Water Environment Federation (WEF), 2013.

[228] A. F. Danil de Namor. “Water purification: from ancient civilization to the XXI Century”. In: *Water Science & Technology: Water Supply* 7.1 (2007), p. 33. ISSN: 16069749. DOI: 10.2166/ws.2007.004.

[229] Richard de Neufville and Stefan Scholtes. *Flexibility in engineering design*. September. Cambridge, MA: MIT Press, 2011.

[230] NOAA. *Climate Data Online: Hourly Precipitation CSV, 1893-01-01–2013-12-22*. Silver Spring, MD, 2016.

[231] Karen Noiva, John E. Fernández, and James L. Wescoat. “Cluster Analysis of Urban Water Supply and Demand: Toward Large-Scale Comparative Sustainability Planning”. In: *Sustainable Cities and Society* (2016). ISSN: 22106707. DOI: 10.1016/j.scs.2016.06.003.

[232] Karen Noiva Welling. “Modeling the water consumption of Singapore using system dynamics”. Master of Science in Building Technology. Massachusetts Institute of Technology, 2011.

[233] Vladimir Novotny. “Urban Water and Energy Use: From Current U.S. Use to Cities of the Future”. In: *Cities of the Future/Urban River Restoration 2010*. Boston: Water Environment Federation, 2010, pp. 118–140.

[234] OECD. *Benefits of Investing in Water and Sanitation: An OECD Perspective*. Paris: OECD Publishing, 2011. DOI: 10.1787/9789264100817-en.

[235] OED-Online. *Complex, n. 1*. Oxford, 2017. URL: www.oed.com.

[236] OED-Online. *Positivism, n. 1*. Oxford, 2017.

[237] OED-Online. *Type, n.1*. Oxford, 2017.

[238] OED-Online. *Typology, n. 1.1*. Oxford, 2017.

[239] Taikan Oki and Shinjiro Kanae. “Global hydrological cycles and world water resources”. In: *Science* 313.5790 (2006), pp. 1068–72. DOI: 10.1126/science.1128845.

[240] Sheila M. Olmstead. “The Economics of Managing Scarce Water Resources”. In: *Review of Environmental Economics and Policy* 4.2 (2010), pp. 179–198. DOI: 10.1093/reep/req004.

[241] Sheila M. Olmstead. “The Economics of Water Quality”. In: *Review of Environmental Economics and Policy* 4.1 (2009), pp. 44–62. DOI: 10.1093/reep/rep016.

[242] Sheila M. Olmstead and Robert N. Stavins. *Managing Water Demand: Price vs. Non-Price Conservation Programs*. Tech. rep. Boston, MA: Pioneer Institute for Public Policy Research, 2006.

[243] B. Orlowsky, Arjen Y. Hoekstra, L. Gudmundsson, and Sonia I. Seneviratne. “Today’s virtual water consumption and trade under future water scarcity”. In: *Environmental Research Letters* 9.7 (2014), p. 074007. DOI: 10.1088/1748-9326/9/7/074007.

[244] Elinor Ostrom. “A general framework for analyzing sustainability of social-ecological systems.” In: *Science* 325.5939 (2009), pp. 419–22. DOI: 10.1126/science.1172133.

[245] Elinor Ostrom, Marco A. Janssen, and John M. Anderies. "Going beyond panaceas". In: *Proceedings of the National Academy of Sciences of the United States of America* 104.39 (2007), pp. 15176–8. DOI: 10.1073/pnas.0701886104.

[246] Claudia Pahl-Wostl. "Towards sustainability in the water sector–The importance of human actors and processes of social learning". In: *Aquatic Sciences* 64.4 (2002), pp. 394–411. DOI: 10.1007/PL00012594.

[247] Claudia Pahl-Wostl. "Transitions towards adaptive management of water facing climate and global change". In: *Water Resources Management* 21.1 (2006), pp. 49–62. DOI: 10.1007/s11269-006-9040-4.

[248] Mi-Hyun Park, Michael Stenstrom, and Stephanie Pincetl. "Water quality improvement policies: lessons learned from the implementation of Proposition O in Los Angeles, California." In: *Environmental management* 43.3 (2009), pp. 514–22. ISSN: 1432-1009. DOI: 10.1007/s00267-008-9262-8. URL: http://www.ncbi.nlm.nih.gov/pubmed/19184186.

[249] Murray C. Peel, Brian L. Finlayson, and Thomas A. McMahon. "Updated world map of the Köppen-Geiger climate classification". In: *Hydrology and Earth System Sciences* 11 (2007), pp. 1633–1644.

[250] Murray C. Peel, Thomas A. McMahon, and Brian L. Finlayson. "Vegetation impact on mean annual evapotranspiration at a global catchment scale". In: *Water Resources Research* 46.9 (2010), pp. 1–16. DOI: 10.1029/2009WR008233.

[251] W. L. Pelton, K. M. King, and C. B. Tanner. "An evaluation of the Thornthwaite and mean temperature methods for determining potential evapotranspiration". In: *Agronomy Journal* 52.7 (1960), pp. 387–395. ISSN: 0002-1962. DOI: 10.2134/agronj1960.00021962005200070006x.

[252] Lina Perelman and Avi Ostfeld. "Topological clustering for water distribution systems analysis". In: *Environmental Modelling and Software* 26.7 (2011), pp. 969–972. ISSN: 13648152. DOI: 10.1016/j.envsoft.2011.01.006.

[253] S. T. A. Pickett, M. L. Cadenasso, J. Morgan Grove, C. H. Nilon, Richard V. Pouyat, W. C. Zipperer, and Robert Costanza. "Urban Ecological Systems: Linking Terrestrial Ecological, Physical, and Socioeconomic Components of Metropolitan Areas". In: *Annual Review of Ecological Systems* 32 (2001), pp. 127–157.

[254] Mark Pires. "Watershed protection for a world city: The case of New York". In: *Land Use Policy* 21 (2004), pp. 161–175. DOI: 10.1016/j.landusepol.2003.08.001.

[255] Anand K. Plappally and John H. Lienhard V. "Costs for water supply, treatment, end-use and reclamation". In: *Desalination and Water Treatment* 51.1-3 (2013), pp. 200–232. ISSN: 1944-3994. DOI: 10.1080/19443994.2012.708996.

[256] Anand K. Plappally and John H. Lienhard V. "Energy requirements for water production, treatment, end use, reclamation, and disposal". In: *Renewable and Sustainable Energy Reviews* 16.7 (2012), pp. 4818–4848. DOI: 10.1016/j.rser.2012.05.022.

[257] Casey Vander Ploeg. *New Tools for New Times: A Sourcebook for the Financing, Funding and Delivery of Urban Infrastructure*. September. Alberta, Canada: Canada West Foundation, 2006. ISBN: 1894825888. URL: www.cwf.ca.

[258] Ryan Plummer, Rob de Loë, and Derek Armitage. "A Systematic Review of Water Vulnerability Assessment Tools". In: *Water Resources Management* 26.15 (2012), pp. 4327–4346. ISSN: 09204741. DOI: 10.1007/s11269-012-0147-5. arXiv: arXiv:1011.1669v3.

[259] Sandra Postel, Gretchen C. Daily, and Paul R. Ehrlich. "Human Appropriation of Renewable Fresh Water". In: *Science* 271 (1996), p. 785.

[260] Adam Przeworski and Henry Teune. *The Logic of Comparative Social Inquiry*. New York: Wiley-Interscience, 1970. DOI: 10.2307/2094059.

[261] PUB. *Our Water, Our Future*. Tech. rep. Singapore: Singapore Public Utilities Board (PUB), 2016, pp. 1–20. DOI: 10.1007/978-3-319-01457-9_1.

[262] PUB. *PUB Annual Report*. Singapore, 2016.

[263] PUB. *PUB Annual Report 2009/2010*. Tech. rep. Singapore: Public Utilities Board (PUB), 2010.

[264] David James Quinn. "Estimating material and energy intensities of urban areas". PhD thesis. Massachusetts Institute of Technology, 2012. URL: http://hdl.handle.net/1721.1/72819.

[265] Bianca Rahill-Marier and Upmanu Lall. *America's Water: An exploratory analysis of Municipal Water Survey Data*. Tech. rep. New York, NY: Earth Institute, 2013, pp. 1–45. URL: http://water.columbia.edu/2013/10/16/americas-water-an-exploratory-analysis-of-municipal-water-survey-data/.

[266] R. Rajagopal. "Integration of Water Data for Decision-Making and Research". In: *Journal of Contemporary Water Research & Education* 142.1 (2009), pp. 28–35. ISSN: 19367031. DOI: 10.1111/j.1936-704X.2009.00049.x.

[267] Anu Ramaswami, Christopher Weible, Deborah Main, Tanya Heikkila, Saba Siddiki, Andrew Duvall, Andrew Pattison, and Meghan Bernard. "A Social-Ecological-Infrastructural Systems Framework for Interdisciplinary An Integrative Curriculum Across Seven Major Disciplines". In: *Journal of Industrial Ecology* 16.6 (2012), pp. 801–813. DOI: 10.1111/j.1530-9290.2012.00566.x.

[268] Wolfgang Rauch and Eberhard Morgenroth. "Urban water management to increase sustainability of cities." In: *Water Research* 47.20 (2013), p. 7149. ISSN: 1879-2448. DOI: 10.1016/j.watres.2013.10.042.

[269] Rashid Rehan, M. A. Knight, C. T. Haas, and A. J. A. Unger. “Application of system dynamics for developing financially self-sustaining management policies for water and wastewater systems.” In: *Water research* 45.16 (2011), pp. 4737–50. DOI: 10.1016/j.watres.2011.06.001.

[270] Rashid Rehan, Andre Unger, Mark A. Knight, and Carl Haas. “Strategic Water Utility Management and Financial Planning Using a New System Dynamics Tool”. In: *Journal American Water Works Association* 107.1 (2015), pp. 22–36. DOI: 10.5942/jawwa.2015.107.0006.

[271] Brian D. Richter, David Abell, Emily Bacha, Kate Brauman, Stavros Calos, Alex Cohn, Carlos Disla, Sarah Friedlander O’Brien, David Hodges, Scott Kaiser, Maria Loughran, Cristina Mestre, Melissa Reardon, and Emma Siegfried. “Tapped out: how can cities secure their water future?” In: *Water Policy* 15.3 (2013), p. 335. DOI: 10.2166/wp.2013.105.

[272] Bradley G. Ridoutt and Stephan Pfister. “A revised approach to water footprinting to make transparent the impacts of consumption and production on global freshwater scarcity”. In: *Global Environmental Change* 20.1 (2010), pp. 113–120. DOI: 10.1016/j.gloenvcha.2009.08.003.

[273] Eliza Roberts and Brooke Barton. *Feeding Ourselves Thirsty. How the Food Sector is Managing Global Water Risks: A Benchmarking Report for Investers*. Tech. rep. Boston, MA, 2015.

[274] Michael T. Rock. “The Dewatering of Economic Growth What Accounts for the Declining Water-Use Intensity of Income ?” In: *Journal of Industrial Ecology* 4.1 (2000), pp. 57–73.

[275] Evangelos Rozos and Christos Makropoulos. “Source to tap urban water cycle modelling”. In: *Environmental Modelling and Software* 41 (2013), pp. 139–150. DOI: 10.1016/j.envsoft.2012.11.015.

[276] Franz Rubel and Markus Kottek. “Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification”. In: *Meteorologische Zeitschrift* 19.2 (2010), pp. 135–141. DOI: 10.1127/0941-2948/2010/0430.

[277] Matthias Ruth and Dana Coelho. “Understanding and managing the complexity of urban systems under climate change”. In: *Climate Policy* 7.4 (2007), pp. 317–336. DOI: 10.3763/cpol.2007.0716.

[278] Martin Rygaard, Philip J. Binning, and Hans-Jørgen Albrechtsen. “Increasing urban water self-sufficiency: new era, new challenges.” In: *Journal of Environmental Management* 92.1 (2011), pp. 185–94. DOI: 10.1016/j.jenvman.2010.09.009.

[279] Artessa Niccola D. Saldivar-Sali. “A Global Typology of Cities: Classification Tree Analysis of Urban Resource Consumption”. PhD thesis. MIT, 2010.

[280] Marie Sanderson. “The Classification of Climates from Pythagoras to Koeppen”. In: *Bulletin of the American Meteorological Society* 80.4 (1999), pp. 669–673. ISSN: 00030007. DOI: 10.1175/1520-0477(1999)080<0669:TCOCFP>2.0.CO;2.

[281] S. Sandoval-Solis, D.C. McKinney, and Daniel P. Loucks. "Sustainability Index for Water Resources Planning and Management". In: *Journal of Water Resources Planning and Management* 137.5 (2011), pp. 381–390. DOI: 10.1061/(ASCE)WR.1943-5452.0000134..

[282] David Satterthwaite. "Missing the Millenium Development Goal targets for water and sanitation in urban areas". In: *Environment and Urbanization* 28.March (2016), pp. 1–20. ISSN: 0956-2478. DOI: 10.1177/0956247816628435.

[283] David Sauri. "Water Conservation: Theory and Evidence in Urban Areas of the Developed World". In: *Annual Review of Environment and Resources* 38.1 (2013), pp. 227–248. DOI: 10.1146/annurev-environ-013113-142651.

[284] Heinz Schandl, Franzi Poldy, Graham M. Turner, Thomas G. Measham, Daniel H. Walker, and Nina Eisenmenger. "Australia's Resource Use Trajectories". In: *Journal of Industrial Ecology* 12.5-6 (2008), pp. 669–685. DOI: 10.1111/j.1530-9290.2008.00075.x.

[285] Mario Schirmer, Sebastian Leschik, and Andreas Musolff. "Current research in urban hydrogeology A review". In: *Advances in Water Resources* 51 (2013), pp. 280–291. DOI: 10.1016/j.advwatres.2012.06.015.

[286] Jeremy J. Schmidt. "Historicising the hydrosocial cycle". In: *Water Alternatives* 7.1 (2014), pp. 220–234. ISSN: 19650175.

[287] David Seckler, Upali Amarasinghe, David Molden, Radhika de Silva, and Randolph Barker. "World Water Demand and Supply, 1990 to 2025: Scenarios and Issues". Colombo, Sri Lanka, 1990.

[288] Ariamalar Selvakumar and Anthony N .Tafuri. "Rehabilitation of Aging Water Infrastructure Systems : Key Challenges and Issues". In: 18.September (2012), pp. 202–209. DOI: 10.1061/(ASCE)IS.1943-555X.0000091..

[289] David Simon. "Urban Environments: Issues on the Peri-Urban Fringe". In: *Annual Review of Environment and Resources* 33.1 (2008), pp. 167–185. DOI: 10.1146/annurev.environ.33.021407.093240.

[290] Slobodan P. Simonovic. "Assessment of Water Resources Through System Dynamics Simulation: From Global Issues to Regional Solutions". In: *Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS'03)*. 1974. Honolulu: IEEE Computer Society, 2002, p. 9. ISBN: 0769518745.

[291] Slobodan P. Simonovic. *Managing Water Resources: Methods and Tools for a Systems Approach*. London: Earthscan, 2009. ISBN: 978-1-84407-554-6, 2009.

[292] Singapore. *Annual Indicators on Population*. Singapore, 2016. URL: http://www.singstat.gov.sg/statistics/browse-by-theme/population-and-population-structure-tables.

[293] Vijay P. Singh. *The Security of Water, Food, Energy and Liveability of Cities: Challenges and Opportunities for Peri-Urban Futures*. Ed. by Basant Maheshwari, Ramesh Purohit, Hector Malano, Vijay P. Singh, and Priyanie Amerasinghe. Vol. 71. Water Science and Technology Library. Dordrecht: Springer Netherlands, 2014. ISBN: 978-94-017-8877-9. DOI: 10.1007/978-94-017-8878-6.

[294] Robert Sitzenfrei and W. Rauch. "Investigating transitions of centralized water infrastructure to decentralized solutions — an integrated approach". In: *Procedia Engineering* 70 (2014), pp. 1549–1557. ISSN: 1877-7058. DOI: 10.1016/j.proeng.2014.02.171.

[295] Murugesu Sivapalan, M. Konar, V. Srinivasan, A. Chhatre, Amber Wutich, C. A. Scott, James L. Wescoat, Jr., and I. Rodriguez-Iturbe. "Socio-hydrology: Use-inspired water sustainability science for the Anthropocene". In: *Earth's Future* 2 (2014), pp. 225–230. DOI: 10.1002/2013EF000164.

[296] Adrian Smith, Jan-Peter Voss, and John Grin. "Innovation studies and sustainability transitions: The allure of the multi-level perspective and its challenges". In: *Research Policy* 39.4 (2010), pp. 435–448. DOI: 10.1016/j.respol.2010.01.023.

[297] David Soll. *Empire of Water: An Environmental and Political History of the New York City Water Supply*. Cornell University Press, 2013, p. 296. ISBN: 080146806X.

[298] Krystyna A. Stave. "A system dynamics model to facilitate public understanding of water management options in Las Vegas, Nevada". In: *Journal of Environmental Management* 67.4 (2003), pp. 303–313. DOI: 10.1016/S0301-4797(02)00205-0.

[299] John D. Sterman. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston: Irwin McGraw-Hill, 2006. ISBN: 0072311363.

[300] John D. Sterman. "Learning in and About Complex Systems". In: *System Dynamics Reivew* 10.2-3 (1994), pp. 291–330. ISSN: 08837066. DOI: 10.1002/sdr.4260100214.

[301] Jon Strand. "Inertia in Infrastructure Development: Some Analytical Aspects, and Reasons for Inefficient Infrastructure Choices". In: *Journal of Infrastructure Development* 2.1 (2010), pp. 51–70. DOI: 10.1177/097493061100200103.

[302] Kenneth Strzepek and Brent Boehlert. "Competition for water for the food system." In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 365.1554 (2010), pp. 2927–40. DOI: 10.1098/rstb.2010.0152.

[303] Matthew D. Stuber, Christopher Sullivan, Spencer A. Kirk, Jennifer A. Farrand, Philip V. Schillaci, Brian D. Fojtasek, and Aaron H. Mandell. "Pilot demonstration of concentrated solar-powered desalination of subsurface agricultural drainage water and other brackish groundwater sources". In: *Desalination* 355 (2015), pp. 186–196. DOI: 10.1016/j.desal.2014.10.037.

[304] S. Suweis, M. Konar, C. Dalin, N. Hanasaki, A. Rinaldo, and I. Rodriguez-Iturbe. "Structure and controls of the global virtual water trade network". In: *Geophysical Research Letters* 38.10 (2011), pp. 1–5. ISSN: 00948276. DOI: 10.1029/2011GL046837. arXiv: 1207.2306.

[305] Yong Soon Tan, Tung Jean Lee, and Karen Tan. *Clean, Green, and Blue: Singapore's Journey Towards Environmental and Water Sustainability*. Singapore: Institute of Southeast Asian Studies, 2009.

[306] Cecilia Tortajada. "Water Management in Singapore". In: *International Journal of Water Resources Development* 22.2 (2006), pp. 227–240. ISSN: 0790-0627. DOI: 10.1080/07900620600691944.

[307] Simon Toze. "Water reuse and health risks - Real vs. perceived". In: *Desalination* 187.1-3 (2006), pp. 41–51. DOI: 10.1016/j.desal.2005.04.066.

[308] Graham M. Turner. "A comparison of The Limits to Growth with 30 years of reality". In: *Global Environmental Change* 18.3 (2008), pp. 397–411. DOI: 10.1016/j.gloenvcha.2008.05.001.

[309] UN-DESA. "World Urbanization Prospects: The 2014 Revision". New York, NY, 2015.

[310] UN-Habitat. *State of the World's Cities 2012/2013: Prosperity of Cities*. New York, NY: Earthscan for UN Habitat, 2013. ISBN: 9780415838887.

[311] UN-WWAP. *The United Nations World Water Development Report 2003: Water for People, Water for Life*. Paris: UNESCO, 2003. ISBN: 9231038818.

[312] UN-WWAP. *The United Nations World Water Development Report 2006: Water: A Shared Responsibility*. Paris: UNESCO, 2006. DOI: 10.7748/nm.21.4.12.s12.

[313] UN-WWAP. *The United Nations World Water Development Report 2012: Managing Water under Uncertainty and Risk*. Vol. 1. Paris, France: UNESCO, 2012, pp. 1–909. ISBN: 9789231042355.

[314] UN-WWAP. *The United Nations World Water Development Report 2015: Water for a Sustainable World*. Paris, France: UNESCO, 2015. ISBN: 9789231000713. URL: http://unesdoc.unesco.org/images/0023/002318/231823E.pdf.

[315] UN-WWAP. *The United Nations World Water Development Report 2016: Water and Jobs*. Tech. rep. Paris: UNESCO, 2016. URL: http://unesdoc.unesco.org/images/0024/002439/243938e.pdf.

[316] UN-WWAP. *The United Nations World Water Development Report 2017. Wastewater: The Untapped Resource.* Paris: UNESCO, 2017. ISBN: 9789231002014.

[317] UN-WWAP. "The United Nations World Water Development Report 3: Water in a Changing World". In: box 1. London: UNESCO, 2009, pp. 9–12. ISBN: 978-9-23104-095-5.

[318] UNESCO. *World Urbanization Prospects: The 2007 Revision*. New York, NY: UNESCO, 2008.

[319] Christian Urich and Wolfgang Rauch. "Exploring critical pathways for urban water management to identify robust strategies under deep uncertainties". In: *Water Research* 66 (2014), pp. 374–389. ISSN: 0043-1354. DOI: 10.1016/j.watres.2014.08.020.

[320] Christian Urich and Wolfgang Rauch. "Modelling the urban water cycle as an integrated part of the city: a review". In: *Water Science & Technology* 70.11 (2014), pp. 1857–1873. DOI: 10.2166/wst.2014.363.

[321] US-BR. *Colorado River Compact*. Washington, D.C., 1922.

[322] US-BR. *Reclamation: Managing Water in the West. Colorado River Basin Water Supply and Demand Study.* 2012.

[323] US-Census. *Alternative Vintage 2010 Total Population Estimates and 2010 Census Counts for the United States, States, Counties, and Subcounty Areas: Subcounty Totals*. Washington, D.C., 2013. URL: https://www.census.gov/programs-surveys/popest/technical-documentation/research/evaluation-estimates.html.

[324] US-Census. *Population of the 100 Largest Urban Places: 1790–1990*. Washington, D.C., 1998.

[325] US-Census. *Ten U.S. Cities Now have 1 Million People or More; California and Texas Each Have Three of These Places (Press Release CB15-89)*. Washington, D.C., 2015. URL: https://www.census.gov/newsroom/press-releases/2015/cb15-89.html.

[326] USNRC. *Adaptive Management for Water Resources Project Planning*. Tech. rep. Washington, D.C.: Panel on Adaptive Management for Resource Stewardship, Committee to Assess the U.S. Army Corps of Engineers Methods of Analysis and Peer Review for Water Resources Project Planning, 2004, p. 138. URL: http://www.nap.edu/catalog/10972.html.

[327] G. Venkatesh and Helge Brattebo. "Energy consumption, costs and environmental impacts for urban water cycle services: Case study of Oslo (Norway)". In: *Energy* 36.2 (2011), pp. 792–800. ISSN: 03605442. DOI: 10.1016/j.energy.2010.12.040.

[328] G. Venkatesh, Arthur Chan, and Helge Brattebo. "Understanding the water-energy-carbon nexus in urban water utilities: Comparison of four city case studies and the relevant influencing factors". In: *Energy* 75 (2014), pp. 153–166. ISSN: 03605442. DOI: 10.1016/j.energy.2014.06.111.

[329] Mateus Ricardo Nogueira Vilanova, Paulo Magalhães Filho, and José Antônio Perrella Balestieri. "Performance measurement and indicators for water supply management: Review and international cases". In: *Renewable and Sustainable Energy Reviews* 43 (2015), pp. 1–12. DOI: 10.1016/j.rser.2014.11.043.

[330] Rodrigo Villarroel Walker, Michael Bruce Beck, and Jim W. Hall. "Water–and nutrient and energy–systems in urbanizing watersheds". In: *Frontiers of Environmental Science & Engineering* 6.5 (2012), pp. 596–611. DOI: 10.1007/s11783-012-0445-4.

[331] Coleen Vogel, Susanne C. Moser, Roger E. Kasperson, and Geoffrey D. Dabelko. "Linking vulnerability, adaptation, and resilience science to practice: Pathways, players, and partnerships". In: *Global Environmental Change* 17.3-4 (2007), pp. 349–364. DOI: 10.1016/j.gloenvcha.2007.05.002.

[332] Charles J. Vörösmarty, Arjen Y. Hoekstra, and S. E. Bunn. "What scale for water governance? Fresh water goes global". In: *Science* 349.6247 (2015), pp. 478–479.

[333] Charles J. Vörösmarty, Michel Meybeck, and Christopher L. Pastore. "Impair-then-Repair: A Brief History & Global-Scale Hypothesis Regarding Human-Water Interactions in the Anthropocene". In: *Daedalus* 144.3 (2015), pp. 94–109.

[334] Charles J. Vörösmarty, P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, S. E. Bunn, C. A. Sullivan, C. Reidy Liermann, and P. M. Davies. "Global threats to human water security and river biodiversity". In: *Nature* 467.7315 (2010), pp. 555–561. DOI: 10.1038/nature09549.

[335] Christopher J. Walsh, Tim D. Fletcher, and Matthew J. Burns. "Urban stormwater runoff: a new class of environmental flow problem." In: *PloS one* 7.9 (2012), e45814. DOI: 10.1371/journal.pone.0045814.

[336] Olivier L. de Weck, Daniel Roos, and Christopher L. Magee. *Engineering Systems: Meeting Human Needs in a Complex Technological World*. Cambridge, MA: MIT Press, 2011.

[337] WEF. *The Global Risks Report 2015: 10th Edition*. Tech. rep. Geneva: World Economic Forum, 2015.

[338] WEF. *The Global Risks Report 2016: 11th Edition*. Tech. rep. Geneva: World Economic Forum, 2016.

[339] Yigang Wei, Cui Huang, Patrick T. I. Lam, and Zhiyang Yuan. "Sustainable urban development: A review on urban carrying capacity assessment". In: *Habitat International* 46 (2015), pp. 64–71. ISSN: 0197-3975. DOI: 10.1016/j.habitatint.2014.10.015. URL: http://dx.doi.org/10.1016/j.habitatint.2014.10.015.

[340] James L. Wescoat, Jr. "Comparative International Water Research". In: *Journal of Contemporary Water Research & Education* 142.1 (2009), pp. 61–66. DOI: 10.1111/j.1936-704X.2009.00055.x.

[341] James L. Wescoat, Jr. "Searching for Comparative International Water Research: Urban and Rural Water Conservation Research in India and the United States". In: *Water Alternatives* 7.1 (2014), pp. 199–219. ISSN: 19650175.

[342] Dave D. White, Amber Wutich, Kelli L. Larson, Patricia Gober, Timothy Lant, and Clea Senneville. "Credibility, salience, and legitimacy of boundary objects: water managers' assessment of a simulation model in an immersive decision theater". In: *Science and Public Policy* 37.3 (2010), pp. 219–232. DOI: 10.3152/030234210X497726.

[343] P. G. Whitehead, J. Crossman, B. B. Balana, M. N. Futter, S. Comber, L. Jin, D. Skuras, A. J. Wade, M. J. Bowes, and D. S. Read. "A cost-effectiveness analysis of water security and water quality: impacts of climate and land-use change on the River Thames system". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371 (2013), p. 20120413. DOI: 10.1098/rsta.2012.0413.

[344] WHO. "Domestic Water Quantity, Service Level and Health". Geneva, 2003.

[345] WHO. "How much water is needed in emergencies". In: *WHO technical notes on drinking-water, sanitation and hygiene in emergencies*. July. Geneva, Switzerland: World Health Organization (WHO), 2013. Chap. 9, pp. 1–4.

[346] WHO. *Progress on Sanitation and Drinking-Water: 2015 Update and MDG Assessment*. Tech. rep. Geneva: WHO/UNICEF Joint Monitoring Programme for Water Supply and Sanitation, 2015, p. 90.

[347] WHO. *WHO guidelines for drinking-water quality.* Fourth. Vol. 38. 3. Geneva, Switzerland: WHO, 2011, pp. 104–108. DOI: 10.1016/S1462-0758(00)00006-6.

[348] WHO and UNICEF. *Progress on Sanitation and Drinking-Water: 2014 Update*. Tech. rep. 2014, pp. 1–78. DOI: 9789241507240.

[349] Hadley Wickham, Winston Chang, and RStudio. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. Version 2.1.1. 2016. URL: https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf.

[350] Cort J. Willmott and Johannes Feddema. "A More Rational Climatic Moisture Index". In: *The Professional Geographer* 44.1 (1992), pp. 84–88. DOI: 10.1111/j.0033-0124.1992.00084.x.

[351] Cort J. Willmott, Clinton M. Rowe, and Yale Mintz. "Climatology of the terrestrial seasonal water cycle". In: *International Journal of Climatology* 5 (1985), pp. 589–606.

[352] Ines Winz, Gary Brierley, and Sam Trowsdale. "The Use of System Dynamics Simulation in Water Resources Management". In: *Water Resources Management* 23.7 (2008), pp. 1301–1323. ISSN: 0920-4741. DOI: 10.1007/s11269-008-9328-7. URL: http://link.springer.com/10.1007/s11269-008-9328-7.

[353] Abel Wolman. "The Metabolism of Cities". In: *Scientific American* 213 (1965), pp. 179–90.

[354] Melanie Kathleen Wong. "Flexible Design: An Innovative Approach for Planning Water Infrastructure Systems Under Uncertainty". Master of Science in Technology and Policy. MIT, 2013.

[355] WorldBank. *Climate Change Knowledge Portal, Historical Temperature and Rainfall Data, 1900-2012*. Washington, D.C., 2015. URL: http://sdwebx.worldbank.org/climateportal/index.cfm?page=downscaled{_}data{_}download{\&}menu=historical.

[356] Matthias Wörlen, Bettina Wanschura, Herbert Dreiseitl, Manfred Moldaschi, James L. Wescoat, Jr., and Karen Noiva. *Strengthening Blue-Green in Our Cities: Enhancing Blue-Green and Social Performance in High Density Urban Environments: Final Report*. Ramboll Foundation, 2016.

[357] Shuangli Wu, Mou Lv, Shen Dong, Jianjun Wang, and Hui Xu. "Classification calculation on water consumption of urban water supply network based on clustering analysis". In: *World Automaton Congress*. 2012, pp. 1–4.

[358] Ningjie Yu and Rui Chen. "Research on the Development of Urban Infrastructure in China Based on Cluster Analysis". English. In: *2010 International Conference on Management and Service Science*. IEEE, 2010, pp. 1–5. DOI: 10.1109/ICMSS.2010.5577626.

[359] Yang Yu, Keisuke Kojima, Kyoungjin An, and Hiroaki Furumai. "Cluster analysis for characterization of rainfalls and CSO behaviours in an urban drainage area of Tokyo". In: *Water Science and Technology* 68.3 (2013), pp. 544–51. ISSN: 0273-1223. DOI: 10.2166/wst.2013.253. URL: http://www.ncbi.nlm.nih.gov/pubmed/23925181.

[360] David Jason Zetland. "Rational waste: the political-economy of desalination". In: *Browser Download This Paper* (2014). URL: https://papers.ssrn.com/sol3/papers.cfm?abstract{_}id=2498938.

[361] Jakob Zscheischler, Miguel D. Mahecha, and Stefan Harmeling. "Climate classifications: The value of unsupervised clustering". In: *Procedia Computer Science*. Vol. 9. 2012, pp. 897–906. DOI: 10.1016/j.procs.2012.04.096.

Appendix A

Appendix: Supplementary Methods

A.1 Supplementary Methods for Chapter 3

A.1.1 UrbMet Data Methods

Table A.1: Variables from the UrbMet database of cities used in this dissertation [279].

Metric	Symbol	Unit
City	—	name
Continent	—	name
Country	—	name
Population	N	capita
City area	A_N	km^2
Water use intensity	w_N	$m^3 \cdot yr^{-1} \cdot cap^{-1}$

Water use intensity and population data were obtained from the UrbMet database of cities. The data used in the clustering started from the data collected for a 2010 master's thesis by Saldivar-Sali [279]. These metrics used for clustering and supporting analysis and discussion are summarized in Table A.1, and raw data is shown in Tables C.1 in C.1.1 on Page 289.

Saldivar-Sali's original dataset included 155 cities, which was reduced to 141 cities due to missing values [279]. In addition to these base cities, the set was expanded to include Hanover[1].

For the most part, data on water use intensity (w_N) for cities in the UrbMet database was obtained from IBNET, the International Benchmarking Network for Water and Sanitation Utilities, an international organization that collects data through standardized surveys from water utilities around the world. Online access to this data is available at `www.ib-net.org`.

[1]This clustering analysis was funded as part of a larger research grant by the Ramboll Foundation. The larger research grant, Enhancing Blue-Green and Social Performance in High Density Urban Environments, consisted primarily of several in-depth case studies on factors contributing to the success or failure of adoption of BGI infrastructure in high-density urban environments [356]. The clustering analysis supported these case studies by providing quantitative context. With the exception of Hanover, the cities in the Ramboll case studies were already in Saldivar-Sali's original dataset. To support the Ramboll case studies, Hanover was therefore added to Saldivar-Sali's original dataset.

A.1.2 Climate Data Methods

Data Overview

In addition to the UrbMet city data, there were two additional kinds of data needed for every city: 1. location data (i.e. city latitude and longitude), and 2. water balance (especially q_P and q_{Net}).

In the interest of comparability, it was desirable for the data to be obtained from common sources; in the interest of scalability and extendibility, it was desirable that the datasets extend far beyond the set of cities in UrbMet. Location data for the UrbMet cities were obtained and verified from two freely available sources: the Google Maps API and the `world.cities` dataset (from the `maps` library in `R`). The location data were then submitted to WebWIMP—an online interface that provided access to "climatically-averaged" monthly and annual water balance data (including temperature and precipitation), at a 0.5° resolution.

Since there were over one hundred cities for which climate data were needed, it was decided to automate data acquisition in the interest of veracity, replicability, and scalability. Scripts were written in `R` and Python for the location data and water balance data. These methods have been described in A.1.2 and A.1.2, respectively. The methods used in `R` and the Python scripts could be easily repeated by any researcher around the world. The packages used were open source and freely available for download for common operating systems.

City Location

Location data was required as an input to obtain city-specific water balance data from WebWIMP, which required that the location be terrestrial (i.e. that the location fell on land, rather than water) and that the data be formatted as a latitude (Φ) and longitude (Λ), rounded to the nearest 0.5° [191]. Scripts were written to use city and country names to extract the associated latitudes and longitudes for each city from two datasets:

`world.cities` from the `R` library, `maps`; and

`Google Maps API` , using the `geocode`(Φ, Λ) function from the `R` library, `ggmaps`.

Both `world.cities` and `Google Maps API` had data that extended far beyond the UrbMet cities; therefore, the methods developed for the UrbMet cities could be easily applied for other sets of cities.

world.cities The dataset `world.cities` consisted of 43,645 cities with a minimum population of zero[2], and a maximum population of 15,018,783, with the city name ("name"), country name ("country.etc"), approximate population[3] ("pop"), latitude and longitude ("lat" and "long"), and bi-

[2]There were 17 cities with a population of zero, primarily in the Marshall Islands and Greenland, and Svalbard, with one each in Sudan and Montserrat.

[3]Population data was circa January 2006, copyrighted to Stefan Helders, and freely available from his website at `world-gazetteer.com`.

nary indicating whether or not the city was a capital ("capital"). Observations from `world.cities` were selected if they matched the city and country information for cities in UrbMet. It was required that alternative options be provided for 38 city names and/or or country names, shown in Table C.2. The observation attributes matched with cities in the UrbMet dataset on two conditions:

Condition 1: if name from world.cities matched either the *original* **or** *alternate* city name; *and*

Condition 2: if country.etc from world.cities matched either the *original* **or** *alternate* country name.

Those two conditions were found to return unique instances from `world.cities` for each city in the UrbMet dataset. The attributes for latitude, longitude, and population were then appended as new data columns to the UrbMet dataset, and provided in Table C.3 as Φ_1, Λ_1, and N_1. Since the data source and time period for population values in the UrbMet dataset (N_0) and that for population from `world.cities` was not the same, $N_0 \neq N_1$. However, visual inspection of the population data from the two sources found the values to roughly correspond, which helped provide validation of the resulting locations.

Further verification of city locations was sought from Google Maps, for which the `geocode` function from the library `ggmap` was used. For each city in the UrbMet dataset, the command `geocode`(*city*, *country*) submitted a query to the Google Maps API and returned a latitude and longitude. These latitudes and longitudes were appended to the UrbMet dataset and shown in Table C.3 as Φ_2, Λ_2.

The fractional differences for latitude and longitude between the two sources (δ_Φ and δ_Λ, respectively) were calculated as:

$$\delta_\Phi = \frac{\Phi_2 - \Phi_1}{\Phi_1}$$
$$\delta_\Lambda = \frac{\Lambda_2 - \Lambda_1}{\Lambda_1}$$

This ratio was rounded to the nearest 0.1°, and shown in Table C.3 as δ_Φ, δ_Λ. Only six cities had non-zero values for δ_Φ and/or δ_Λ: these cities were (in alphabetical order): Accra, Doha, London, Quezon City, Quito, and Tripoli. Of these cities, the fractional difference was found to be less than 0.5° (i.e. less than the WebWIMP map resolution). The exception of Tripoli, Algeria, which had a fractional difference of $\delta_\Lambda = -0.8$; this difference was considered to be reasonable.

The latitudes and longitudes from `world.cities` were then used as the location values submitted to WebWIMP, as described in the following section, A.1.2.

Water Balance

WebWIMP provided average monthly and annual temperature and precipitation data, based on averaged and spatially-interpolated from historical data, for any user-specified terrestrial location rounded to the nearest 0.5° [191, 190]. WebWIMP also estimated the average monthly and annual water balance components for the specified location using a modified Thornthwaite approach, as described in A.1.2. The online interface was developed by Matsuura et al. at the University of Delaware in 2003 with updates made in 2009 [191]. Thornthwaite's approach to estimating water balance components was itself developed in 1960, 1985, with modifications made in the 1980s by Willmott, Rowe, and Mintz, Willmott and Feddema while [251, 351, 350]. More information can be found at the WebWIMP portal at `climate.geog.udel.edu/~wimp/`. Equations and definitions associated with the modified Thornthwaite approach are presented in Section A.1.2 Table A.2.

A Python script was written to automate the process of submitting location data to WebWIMP and obtaining the resulting average annual and monthly temperature, precipitation, and water balance data. An overview of the general process is described in A.1.2. The Python script itself is provided in A.1.2.

Thornthwaite's Water Balance and Moisture Index

Thornthwaite's method for estimating water balance was developed for classification and assessment of global climatology. The governing equation defined the time rate of change of soil moisture available in the root zone (w) as a function of precipitation (q_P) and evapotranspiration (q_{ET}) [251, 351, 350]:

$$\partial w/\partial t = q_P - -q_{ET} - -q_S \tag{A.1}$$

where S was the local surplus: both surface and subsurface runoff (from the root zone). According to Thornthwaite's approach, E can be modeled as a function of precipitation, potential evapotranspiration (q_{ET^0}), and the soil moisture holding capacity (w^*). Potential evapotranspiration was modeled as a function of temperature (T) and time (h). The equations for evapotranspiration were given by:

$$q_{ET} = \begin{cases} q_P + \beta(w, w^*)[q_{ET^0}(T,h) - -q_P] & q_P < q_{ET^0}(T,h) \\ q_{ET^0}(T,h) & q_P \geq q_{ET^0}(T,h) \end{cases} \tag{A.2}$$

In particular, h refers to the duration of daylight during the day. β is a monotonically increasing function on $[0,1]$ relating $[(q_{ET} - q_P)/(q_{ET^0} - q_P)]$ to $[w/w^*]$. S is then defined in terms of precipitation, evapotranspiration, and soil moisture as:

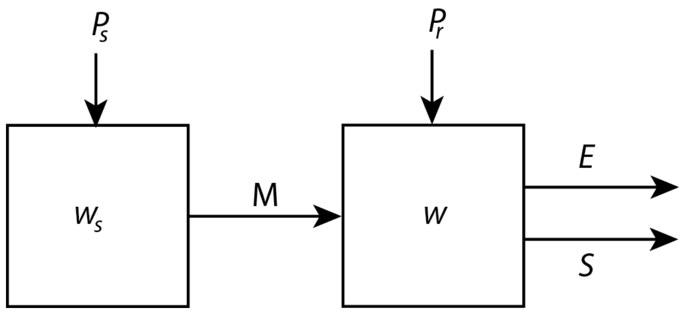
$$q_S = \begin{cases} q_P - -[q_{ET} + (w^* - w)] & q_P > [q_{ET} + (w^* - w)] \\ 0 & q_P \leq [q_{ET} + (w^* - w)] \end{cases} \tag{A.3}$$

The net water balance, q_{Net}, was then defined as:

$$q_{net} = q_P + q_M - q_{ET^0} \tag{A.4}$$

To account for the role of snow as a source of water in higher latitudes and elevations, Willmott,

Figure A.1: Water balance schematic, relating water stored in snow pack (w_s) to water stored in soil moisture (w) through snow melt.



Rowe, and Mintz modified Thornthwaite's approach by making a distinction between precipitation falling as rain versus snow (q_P^r and q_P^s, respectively). Precipitation was considered to fall as rain (P^r) if $T \geq c$ and as snow (P^s) if $T < c$ (with $c \equiv$) a constant, indexed temperature in °C; Willmott, Rowe, and Mintz defined $c = -1$°C [351]. This modified approach also allowed for distinction to be made between moisture stored in the soil (w) versus snow pack (denoted w^s). With q_M denoting snow melt and q_P^s denoting precipitation that falls as snow (in terms of its equivalent volume as liquid water), the time rate of change for w and w^s were defined as functions of snow melt (q_M), actual evapotranspiration (q_{ET}), runoff (q_S), and precipitation (both snowfall, q_P^s and rainfall, (q_P^r):

$$\partial w^s / \partial t = q_P^s - q_M$$
$$\partial w / \partial t = P^r + M - E - S$$

The β function was then modified to also take into account snow melt, by relating $[(q_{ET} - q_P^r - M)/(q_{ET^0} - q_P^r - M)]$ to (w/w^*). Then, Equation A.2 becomes:

$$q_{ET} = \begin{cases} q_P^r + q_M + \beta(w, w^*)[q_{ET^0}(T,h) - -P^r - -M] & (P^r + M) < q_{ET^0}(T,h) \\ q_{ET^0}(T,h) & (P^r + M) \geq q_{ET^0}(T,h) \end{cases} \tag{A.5}$$

and surplus:

$$S = \begin{cases} q_P^r + M - -[q_{ET} + (w^* - w)] & q_P^r + M) > [q_{ET} + (w^* - w)] \\ 0 & (q_P^r + M) \leq [q_{ET} + (w^* - w)] \end{cases} \tag{A.6}$$

Making snow-melt q_M a function of q_P^r and T, and constrained by w^s (i.e. $q_M = 0$ if $w^s = 0$. Evapotranspiration over month i ($q_{ET^0}(t_i)$) was defined as a function of potential evapotranspiration (q_{ET^0}), mean surface air temperature ($\bar{T}_i$), the number of days (θ_i), and average day-length h over month i^4:

$$q_{ET^0}\prime(t_i) = q_{ET}(t_i, \theta, h) = \begin{cases} 0 & \bar{T}_i < 0°\text{C} \\ 16 \cdot (\frac{10\bar{T}_i}{I_i})^a & 0 \leq \bar{T}_i < 26.5°\text{C} \\ -415.85 + 32.24\bar{T}_i - 0.43\bar{T}_i^2 & \bar{T}_i \geq 26.5°\text{C} \end{cases} \tag{A.7}$$

With constants I and a, determined by:

$$I = \sum^{12}(\frac{\bar{T}_i}{5})^{1.514} a = 6.75 \times 10^{-7} \cdot I^3 - 7.71 \times 10^{-5} \cdot I^2 + 1.79 \times 10^{-1} \cdot I + 0.49$$

Evapotranspiration was then adjusted to the length of the month and day-length using the following:

$$q_{ET^0}(t_i) = q_{ET^0}\prime(t_i) \cdot (\frac{\theta}{30}) \cdot (\frac{h}{12}) \tag{A.8}$$

Overview of WebWIMP

To obtain the values for any location required specifying longitude (Λ) and latitude (Φ) in degrees (rounded to the nearest tenth of a degree). Once the location was entered and submitted into the WebWIMP interface, a new webpage was opened. This second webpage returned either: 1. an error message, 'located on a large body of water,' and the option to revise the location[5]; or 2. the elevation for the chosen location, a table of average monthly and annual air temperature and precipitation, and an opportunity to specify parameters for use in the modified Thornthwaite equation. If the first query of latitude and longitude successfully identified a terrestrial location, the option was given for the user to alter parameters that affected the water balance in the modified-Thornthwaite approach. These parameters included:

Soil water-holding capacity (w^*): with a default of w^* =150.0 mm,

Declining availability (β) function: a monotonic function relating actual evapotranspiration (q_{ET}, or AE) as a percentage of potential evapotranspiration (q_{ET^0}, or PE) to soil moisture (w) a fraction of the soil water holding capacity (w/w^*), with a default of β = curve G in Figure A.2;

[4]Average day length over month i was defined as equal to the duration of days in hours on the fifteenth of the month.

[5]The error message, 'located on a large body of water,' was also given if the latitude and longitude were incorrectly specified, i.e. the entry was not rounded to the nearest half a degree. For instance, the latitude and longitude for Massachusetts Institute of Technology (MIT) in Cambridge, MA are $42.36°N, 71.10°W$; the aforementioned error message was returned if submitted to WebWIMP as $\Lambda = -71.10, \Phi = 42.36$. The correct format was $\Lambda = -71, \Phi = 42.5$. However, since Cambridge, MA is near the eastern coast of the US, and since the resolution of WebWIMP was (at 0.5°) relatively coarse, submission of the correctly-formatted location for MIT still returned an error message. A successful submission required the location be revised to $\Lambda = -71, \Phi = 42$.

Hypothetical climate change: an option that allowed the user to specify a hypothetical change in monthly *or* annual air temperature (in °C) and/or precipitation (as a percent change); the default was no change.

For the purposes of the analysis, the parameters were set to their defaults—which is to say that the dialogue box for each parameter was left blank. After submission of the defaults for w^*, β, and temperature and precipitation, a third webpage was opened that returned three tables:

1. Prescribed air temperature changes (°C),

2. Prescribed precipitation changes (%), and

3. Monthly and annual climatic water balance.

Since no changes in air temperature or precipitation was the default assumption, the third table (of monthly and annual climatic water balance components) was the only one of interest. The monthly water balance components returned in the third table have been summarized in Table A.2. An overview of these variables has been provided in A.1.2.

Figure A.2: The β function: WebWIMP required an assumption of the relationship between actual evapotranspiration (q_{ET}, or AE) and potential evapotranspiration (q_{ET^0}, or PE) in its calculation of the average monthly water balance for any global location (to the nearest half-degree). The modified Thornthwaite procedure employed by WebWIMP assumed a monotonic increasing relationship (the so-called β function) between the ratio q_{ET}/q_{ET^0} and the ratio of soil moisture (w) and soil water-holding capacity (w^*). WebWIMP provided eight options for β (A—H) as shown in the figure, with the default set to curve *G*.

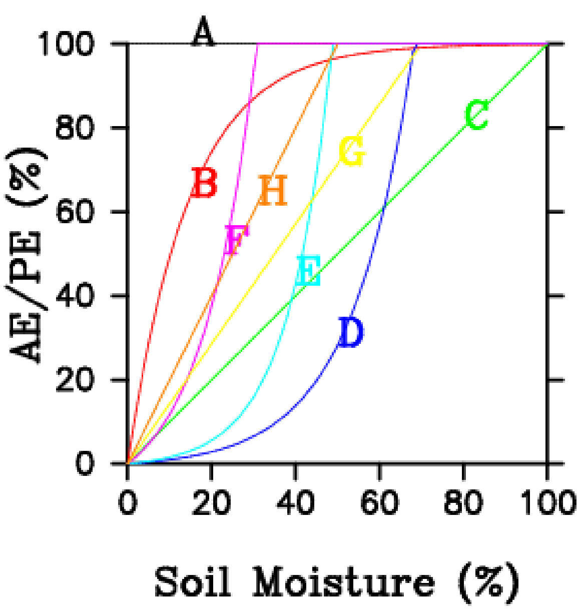


Table A.2: WebWIMP Variables

Abbr.	Symbol	Unit	Description
MON	t	month	Month
TEMP	T	°C	Temperature
UPE		mm/month	Estimated unadjusted potential ET
APE	q_{ET^0}	mm/month	Estimated, adjusted potential evapotranspiration (adjusted for day and month length)
PREC	q_P	mm/month	Average monthly total precipitation
DIFF	q_{Net}	mm/month	Rainfall plus estimated snowmelt minus estimated adjusted potential evapotranspiration
ST	w	mm	Estimated soil moisture at the end of the month
DST	δw	mm/month	Estimated change in soil moisture from the preceding month to the end of the current month
AE	q_{ET}	mm/month	Estimated, actual evapotranspiration
DEF	q_{ET^0}- q_{ET}	mm/month	Estimated deficit or unmet atmospheric demand for moisture
SURP	q_S	mm/month	Estimated surplus (surface runoff and percolation below the plant root zone)
SMT	q_M	mm/month	Estimated snowmelt
SST	w^*	mm	Estimated water retained in the snow pack at the end of the month

Web-Scraping

A Python script was written to submit queries to the online WebWIMP interface and to save results to file. The function, `water_balance_scraper`, and its supporting functions, required the following Python packages: `Browser` from `mechanize`, `BeautifulSoup` from `bs4`, `math`, `pandas`, `time`, and `xlrd`. The Python script consisted of 10 individual functions: the main function, `water_balance_scraper`, and nine smaller functions. The main function integrates the smaller functions so that a user only needs to input the filename of a `.csv file` with the appropriate format and `water_balance_scraper` outputs three new `.csv` files: `city_status`, `temp_precip`, and `water_balance`.

Format for Python input The required format for `water_balance_scraper` is that of a `.csv` file consisting of four columns: c_id, city, latitude, and longitude; these four columns should contain unique integer IDs for each city, the city names, and associated latitude and longitude[6], respectively. An example of five cities is provided in Table A.3.

The main function integrates the smaller functions so that a user only needs to input the filename of a `.csv file` with the appropriate format and `water_balance_scraper` outputs three new `.csv` files: `city_status`, `temp_precip`, and `water_balance`.

[6]Methods for obtaining latitudes and longitudes from city names and country associations were described in A.1.2.

Table A.3: Data format for original and alternative city and country names used in web-scraping.

c_id	**city**	**latitude**	**longitude**
1	Abu Dhabi	24.48	54.37
2	Abuja	9.18	7.17
3	Accra	5.56	-0.20
4	Addis Ababa	9.03	38.74
5	Amman	31.95	35.93

City status file The `city_status` dataframe and file produced by `water_balance_scraper` were introduced so that the web-scraping process would be robust to errors when submitting queries to WebWIMP. The `city_status` file/dataframe has the following columns: cs_id, city, latitude, longitude, round_lat, round_lon, delta_lat, delta_lon, new_lat, and new_lon. These columns hold the following values:

cs_id an integer ID unique to each city;

city the city name

latitude the original latitude, Φ_1

longitude the original longitude, Λ_1

round_lat the original latitude rounded to the nearest half-degree, $\Phi_1\prime$

round_lon the original longitude rounded to the nearest half-degree, $\Lambda_1\prime$

delta_lat the change in latitude (relative to the original) associated with a successful WebWIMP query ($\Delta\Phi$)

delta_lon the change in longitude (relative to the original) associated with a successful WebWIMP query ($\Delta\Lambda$)

new_lat the new latitude associated with a successful WebWIMP query ($\Phi\prime$)

new_lon the new longitude associated with a successful WebWIMP query ($\Lambda\prime$)

As mentioned earlier in this section, there were two cases in which an error message would be returned for some locations: in the first case, the error arose from an improperly formatted submission; in the second case, the error arose due to a mismatch in resolution between WebWIMP's world map and other sources.

First error: In the first case, city latitudes and longitudes were rounded to the nearest half a degree. This was accomplished by: 1. multiplying the original latitude and longitude (Φ_1 and Λ_1) by two; then, 2. round the product of step 1 to the nearest whole number; 3. dividing by two;

and finally, rounding the result to the nearest tenth of a degree to obtain a latitude and longitude formatted for WebWIMP ($\Phi_1\prime$ and $\Lambda_1\prime$):

$$\Phi_1\prime = \texttt{round}(\texttt{round}(\Phi_1 * 2, 0)/2, 1)$$
$$\Lambda_1\prime = \texttt{round}(\texttt{round}(\Lambda_1 * 2, 0)/2, 1)$$

$\Phi_1\prime$ and $\Lambda_1\prime$ were saved to `city_status` columns `round_lat` and `round_lon`, respectively.

Second error: `round_lat` were the values submitted to WebWIMP. Since the latitudes and longitudes had been formatted to ensure that they were appropriate for WebWIMP, if the web query returned an error message then it was an error of the second type. To deal with the second type of error, a two-step process was performed. In the first step, the Python script submitted $\Phi_1\prime$ and $\Lambda_1\prime$ to WebWIMP, as described, and parsed the `html` that was returned. If no error message was returned, a value of *True* was saved to the status column of `city_status` and $\Delta\Phi$ and $\Delta\Lambda$ were both set equal to zero and $\Phi\prime$ and $\Lambda\prime$ were set equal to the original latitude and longitude (Φ_1 and Λ_1), respectively. However, if an error message was returned, the status was set to *False* and no other values were updated. To reduce the likelihood of an interruption, the results were saved to file after each city. In the second step of the process (for dealing with the second type of error), the script identified cities which, from the previous query, had a status=*False*. For these cases, the script produced a temporary Python dictionary for different latitude and longitude combinations associated with 0.5 degree steps from the original observation. For each combination in said Python dictionary, the original latitude and longitude were adjusted by these small changes ($\Delta\Phi$ and $\Delta\Lambda$) and submitted to WebWIMP. Of these observations, the first observation with the smallest distance to the original location (if there was more than one observation) was returned, and the `city_status` status was set to *True*. Additionally, delta_lat, delta_lon, new_lat, and new_lat were updated to associated successful $\Delta\Phi$, $\Delta\Lambda$, $\Phi\prime$, and $\Lambda\prime$ (respectively). For all of the cities in UrbMet, $\Phi\prime$, and $\Lambda\prime$ were each within one degree of the original location, and for most cases were within 0.5 degrees. For most cities, this was likely a sufficient degree of accuracy for the method.

For the cities with a `city_status` status equal to *True*, the new latitude and longitude ($\Phi\prime$ and $\Lambda\prime$, i.e. columns new_lat and new_lat[7]) were then submitted to WebWIMP to obtain the temperature and precipitation data and the water balance data. These steps are summarized in Table A.4.

In its full form, the main function, `water_balance_scraper` allows the following options:

filename The filename (with extension) of the file containing cities and original location information (format=string, *no default*);

[7]Note that the new latitude and longitudes were the same as the original latitudes and longitudes for cities for which a successful WebWIMP query was obtained on the first pass.

Table A.4: Overview of Python script for scraping WebWIMP

Step	Action
1	Read in `.csv` data of cities
2	Convert latitude and longitude to WebWIMP resolution of 0.5 degrees
3	Submit the location data to WebWIMP
4	Parse the results and to identify whether an error occurred
5	Save the results to `city_status.csv`
6	If the option *failed=True* was given as an argument to the function, `water_balance_scraper`, attempt to locate coordinates that are near to the original latitude and longitude and return a successful WebWIMP query
7	Read in `city_status.csv` and obtain subset of values for which status=*True*
8	If *"tp"* was passed as an arguments to *data_types*, scrape the first and only table from the first WebWIMP page and save the results to `.csv`
9	If *"tp"* was passed as an arguments to *data_types*, scrape the third table from the second WebWIMP page and save the results to `.csv`

date — Any string passed as an argument to date will be prepended to the filenames of outputs, with the default prefix being a string with the date on which the script is run formatted as 'YYYmmdd' (format = string, *default=None*).

drange — Should the script only run a subset of the input data file? If so, the arguments to *drange* should be passed as an integer range with format of `drange` $= [i_{first}, i_{last}]$, (the *default=None* returns results for the entire cities file);

failed — Should the script look for new locations for cities in the `cities_status` file with status=*False*? (format = True/False, *default=False*);

silent — Should the script output a progress report to the user? (format = True/False, *default=False*);

data_types — This option provides the user with the option to run the script for only one of the file types (format = list of one or more of the following options: 'status', 'tp', 'wb', with *default=['status', 'tp', 'wb']*). Note: if 'tp' and/or 'wb' are included in the arguments passed to data_types, then either 1. the `cities_status` file must already exist (i.e. option prev=*True*) **or** 2. 'status' must also be included.

prev — Do previous versions of the output files exist? If prev=*True*, then the previous files must have been produced on the same day that the script is being run **or** the appropriate date must be passed as an argument to option prevDate (format = True/False, *default=False*);

prevDate The previous date (as a string) on which previous output files were produced if prev=*True* (format = string with format 'YYYmmdd' *default=None*).

The Python script is provided below for the sake of replicability, and is also available on Github. The user should install the requisite Python packages. The default is for the script to look for input files and produce output files in the same directory as the Python script, but the user can bypass this default by including an alternative directory path (absolute or relative) with the options filename, date, and prevDate.

To use the Python script, the user should save the script (e.g. as 'water_balance_scraper.py') and then pass the filename as an argument to the command `execfile`(). The user can then run `water_balance_scraper` from the command line (or from a text editor) as (e.g. on a cities dataset called '`urbmet_cities.csv`':

$$water\_balance\_scraper(filename =' urbmet\_cities.csv', failed = True, prev = False) \tag{A.9}$$

For the 143 cities in the UrbMet dataframe, the Python script finished all operations within 11.2 minutes (672.0 seconds) (including obtaining new locations for 38 cities whose original locations fell on bodies of water).

A.2 Supplementary Methods for Chapters 4 and 5

A.2.1 Shapefile Data and Methods

Shapefiles for Los Angeles and Singapore were obtained for use in the extraction of climate data. These shapefiles were:

Los Angeles *city-level*, from `controllerdata.lacity.org/` [166]

Singapore *country-level*, from `www.gadm.org/country`[8] [132].

Shapefiles were read into `R` with the `readOGR` function from the `rgdal` package, creating a `SpatialPolygons` object that was used in the extraction of WebWIMP monthly historical data.

Spatial data requires projection of three-dimensional data onto a two-dimensional surface. There are multiple common schemes for projection. For this reason, spatial data commonly includes information about the projection system used. After reading the shapefiles into `R` using the `readOGR` function, projection attributes for the downloaded shapefiles were then obtained using the function `proj4string` from the `sp` package in `R`. The Coordinate Reference System (CRS) of the Singapore and LA city-level data were in longitude/latitude using the `WGS84` datum. This was consistent with the `CRS` of the WebWIMP historical monthly data. However, the function `proj4string()` was also used to re-project the data to the `CRS` of longitude/latitude, using the `proj4string()` from Singapore as reference.

[8]Country-level political boundaries were available from Global Administrative Areas (GADM) as spatial data in multiple formats, including `R SpatialPolygonsDataFrame`, ESRI geodatabase file `.gdb`, Google Earth `.kmz`, and shapefiles (consisting of `.shp`, `.shx`, `.dbf`, and .prj). [132].

For Los Angeles, the feature type was constrained to (data column `feat_type` == *Land*) for consistency with the Singapore boundary.

A.2.2 Climate Data

Monthly climate data—namely precipitation, evapotranspiration, and surplus—were obtained from three sources:

NOAA Precipitation data from 1893–2016 for Los Angeles [230]

World Bank Precipitation data from 1901–2009 for Singapore [355]

WebWIMP Precipitation, temperature, and water balance data for both LA and Singapore [190]

Between the three sources, climate data for both LA and Singapore were obtained from *Terrestrial Water Budget Data Archive*, which has a similar source as the data used in Chapter 3 [191, 190]. Global gridded data of historical monthly water balance components were available for download for the years 1900-2010 with 0.5-degree resolution at `climate.geog.udel.edu/~climate/html_pages/Global2014/`. Datasets on monthly deficit (`DEF`, or q_{def}), actual evapotranspiration (`AE`, E, or q_{ET}), potential evapotranspiration (`APE`, or q_{ET} 0), snowmelt (`M`, or q_M), surplus (`S`, or q_S), mid-monthly snow-cover (`ws`, w_S), and mid-monthly soil moisture (`w`, or w) were obtained.

Data for each metric consisted of 110 files (one for each year) in tabular format. Each row in the table provided a latitude, longitude, and monthly data. These data were read into `R` using the command `as.data.frame(read.fwf(`*`filename`*`))` and then converted to raster using the function `rasterFromXYZ` from the `raster` package. These data had a `CRS` of longitude/latitude, with the `WGS84` datum.

Data for each case were extracted from the raster files using the function `extract`[9] on `SpatialPolygon` objects of the case shapefiles. The function `extract` was used with the options *method* = `"bilinear"`, *fun*=*mean*, and *sp*=**T**. These options, respectively:

method = `"bilinear"` For each cell in the range, returned a value interpolated from the values of the four nearest raster cells;

fun = *mean* Took the mean of the values returned by *method* = `"bilinear"`, returning a single value;

sp = *T* Added the extracted values to the dataframe of the `SpatialPolygons` object.

After the data were extracted, they were plotted along with the shapefiles to visually confirm compatibility.

For Singapore, the given latitude and longitude was found to fall on a body of water. For this reason, the extent of the shapefile for Singapore was expanded to four grid cells, which was similar to the extent of LA. While this assumption may be questioned, the resulting precipitation was found to match other precipitation data, as seen in Figure 4.20a.

These historical climate data used primarily in the analysis were developed by *Terrestrial Water Budget Data Archive*, the same authors who developed the WebWIMP interface used to obtain water balance data for the analysis in Chapter 3 (and described in A.1.2). These historical data are provided at the global level,

[9] The function `extract` was available through the `raster` package in `R`.

and are used by the developers in updating WebWIMP [190]. Therefore, this data provided consistency between analyses in Chapters 3 and 4. Since this data is available at the global scale, the approach could be easily replicated for any other location.

Appendix B

Appendix: Supplementary Results

B.1 Supplementary Results for Chapter 3

B.1.1 Additional Scatterplots

Figures B.1–B.3 are larger versions of the scatterplots of population (N, in capita), water use intensity (w_N, in $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$), and net annual climatic water balance (q_{Net}, in $\mathrm{m} \cdot \mathrm{yr}^{-1}$) from Chapter 3:

- Test
- Figure B.1 on Page 270 shows is a close-up of w_N plotted against q_{Net} (as in Figure 3.8a on Page 100)
- Figure B.2 on Page 271 shows is a close-up of N plotted against q_{Net} (as in Figure 3.8b on Page 100)
- Figure B.3 on Page 272 shows is a close-up of w_N plotted against N (as in Figure 3.8c on Page 100)

B.1.2 Precipitation and Annual Surplus

Figure B.4 shows climatic water balance (q_{Net} in $\mathrm{m} \cdot \mathrm{yr}^{-1}$) plotted against precipitation for data in the UrbMet cities database (q_P in $\mathrm{m} \cdot \mathrm{yr}^{-1}$) as a scatterplot. The data for the plot was obtained from WebWIMP, as described in Appendix 3.3.2. The plot shows that q_{Net} and q_P are highly correlated. The plot is of relevance since q_{Net} was used in the clustering, while q_P was used in calculating WUCI (i_{UC} in $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$), total WUCI ($I_{UC}$ in km^2), and the self-sufficiency ratio (R_{SS}).

B.1.3 Supplementary Results for WUCI, Total WUCI, and Self-Sufficiency

Figure B.1: Scatterplot of water use intensity (w_N, in $m^3 \cdot yr^{-1} \cdot cap^{-1}$) versus net annual balance (q_{Net}, in $m \cdot yr^{-1}$) for cities in the Urban Metabolism database, with quantiles plotted as dotted lines. This is a larger version Figure 3.8a on Page 100 in Chapter 3, with city names.

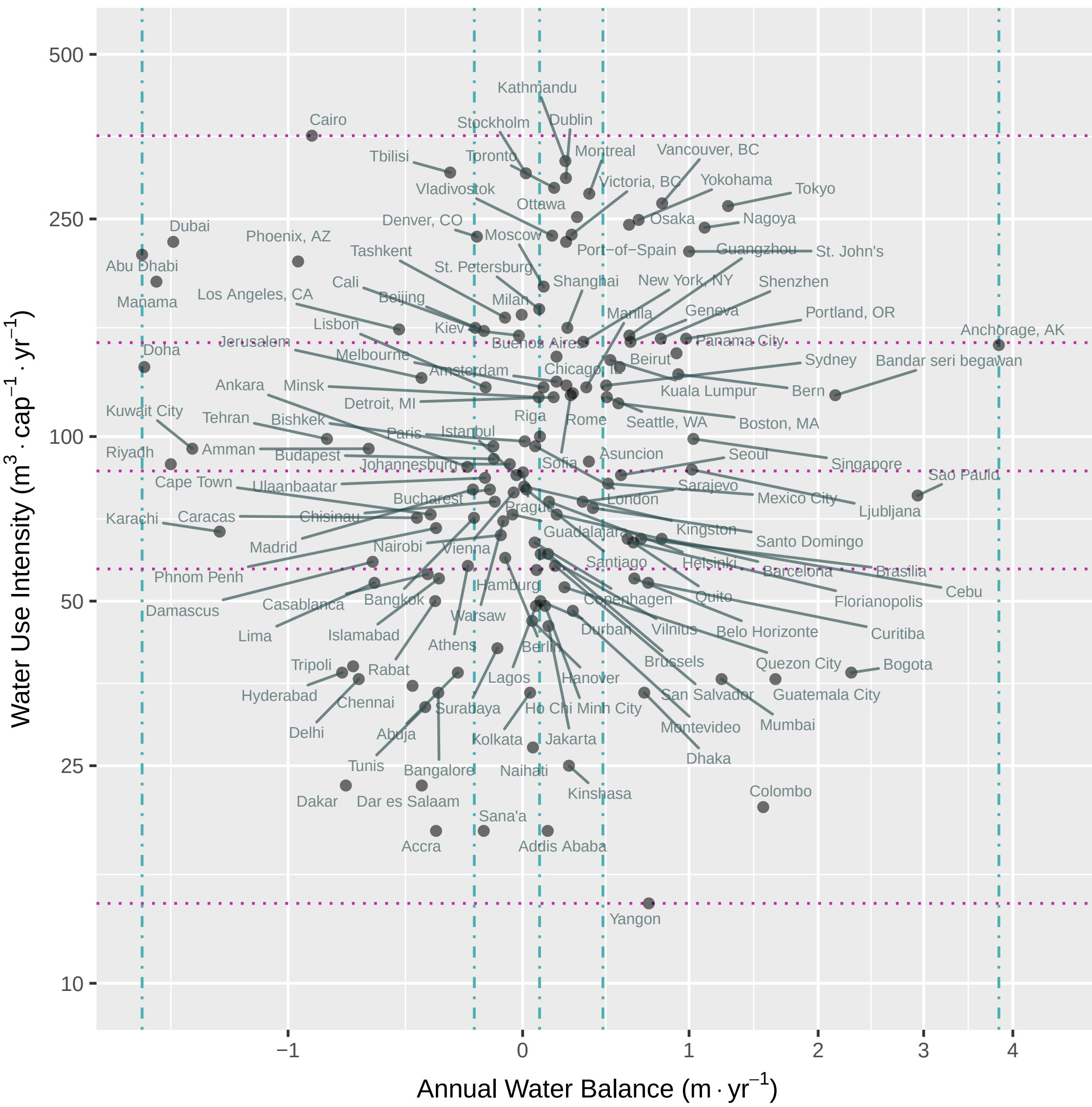


Figure B.2: Scatterplot of population (N, in capita) versus net annual balance (q_{Net}, in $\mathrm{m} \cdot \mathrm{yr}^{-1}$) for cities in the Urban Metabolism database, with quantiles plotted as dotted lines. This is a larger version Figure 3.8b on Page 100 in Chapter 3, with city names.

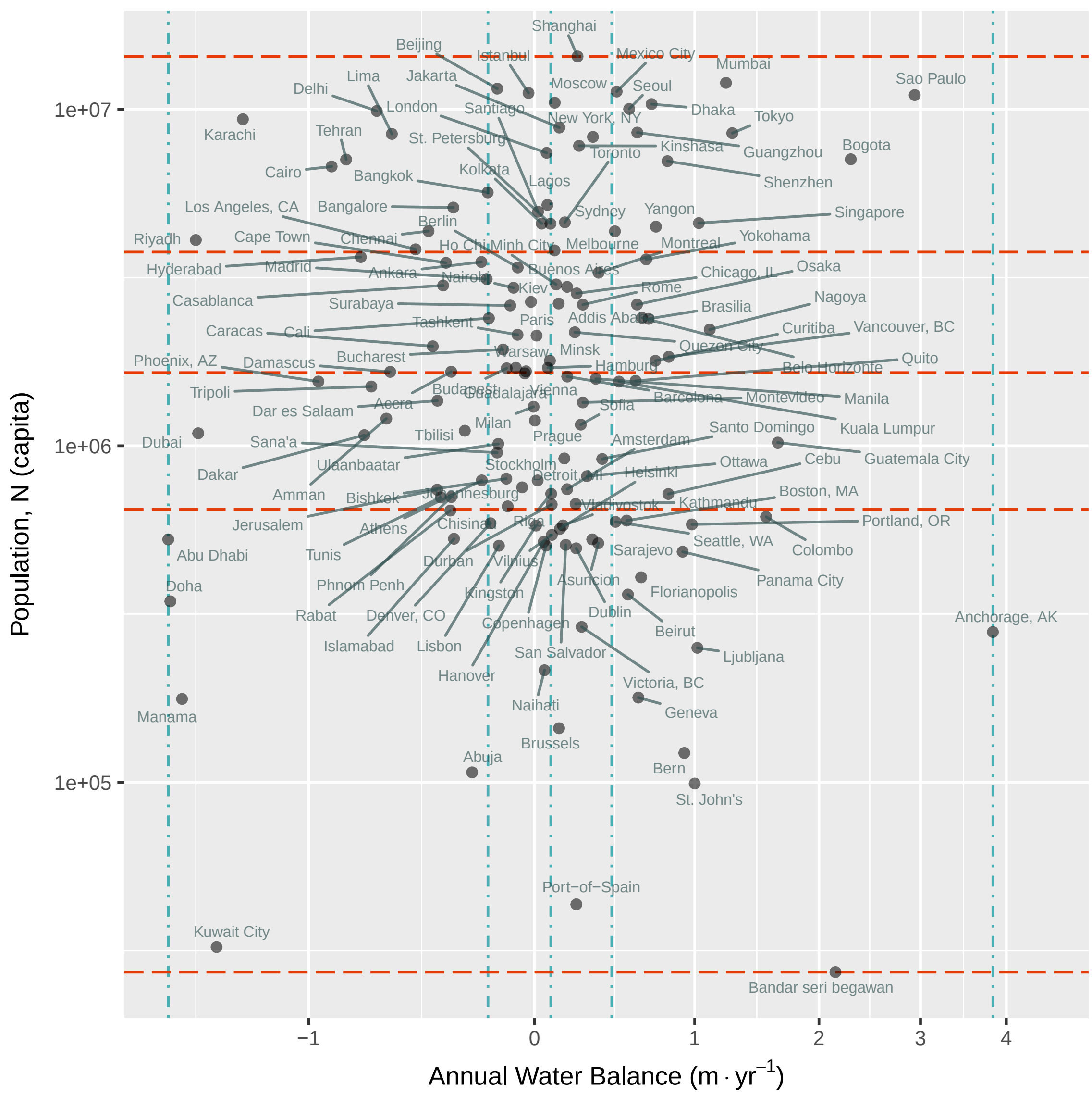


Figure B.3: Scatterplot of water use intensity (w_N, in $\mathrm{m}^3 \cdot \mathrm{yr}^{-1} \cdot \mathrm{cap}^{-1}$) versus population ($N$, in capita) for cities in the Urban Metabolism database, with quantiles plotted as dotted lines. This is a larger version Figure 3.8c on Page 100 in Chapter 3, with city names.

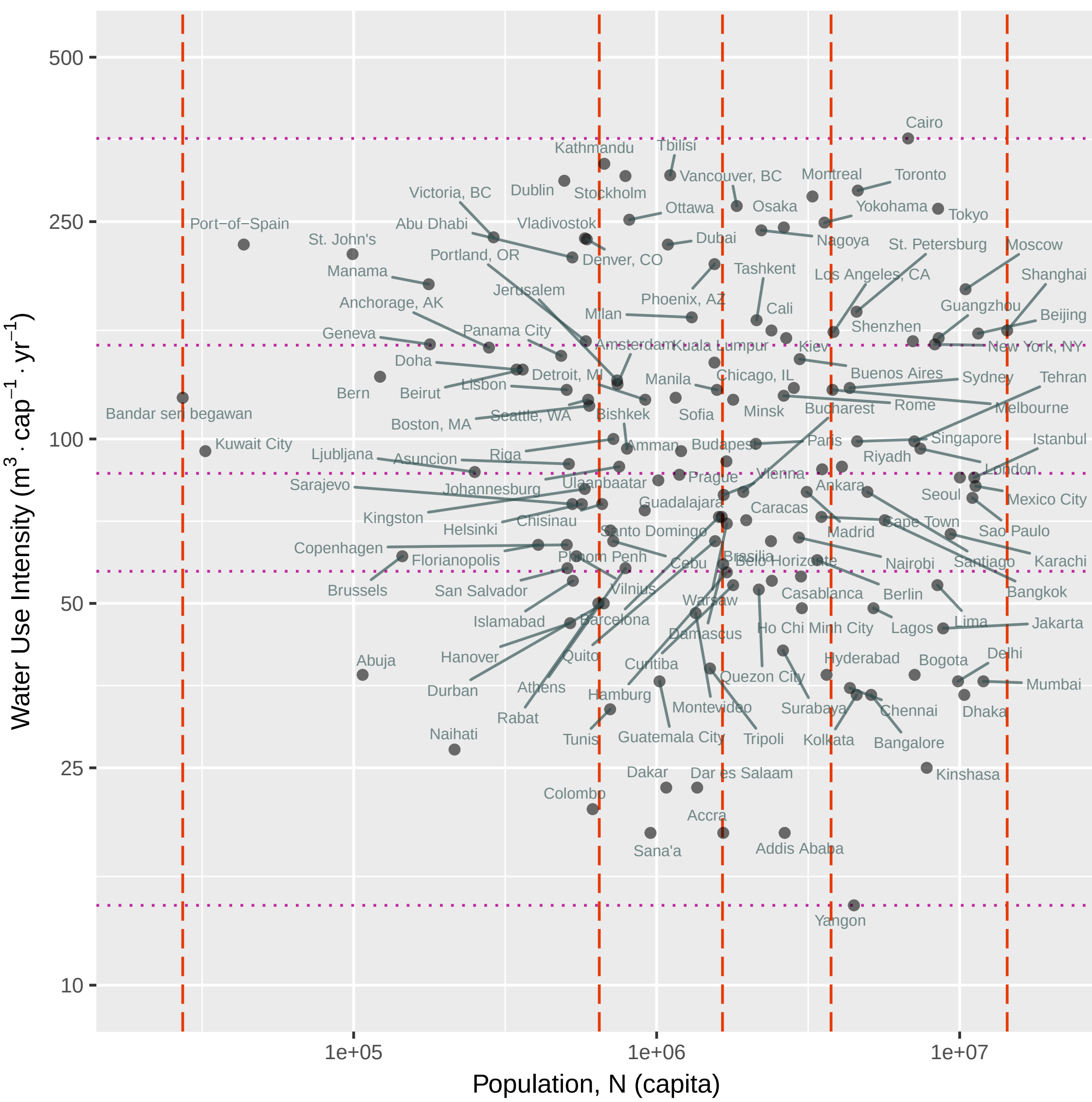


Figure B.4: Scatterplot of net annual balance (q_{Net}, in $\mathrm{m} \cdot \mathrm{yr}^{-1}$) versus precipitation ($q_P$, in $\mathrm{m} \cdot \mathrm{yr}^{-1}$) for cities in the Urban Metabolism database, with quantiles plotted as dotted line.

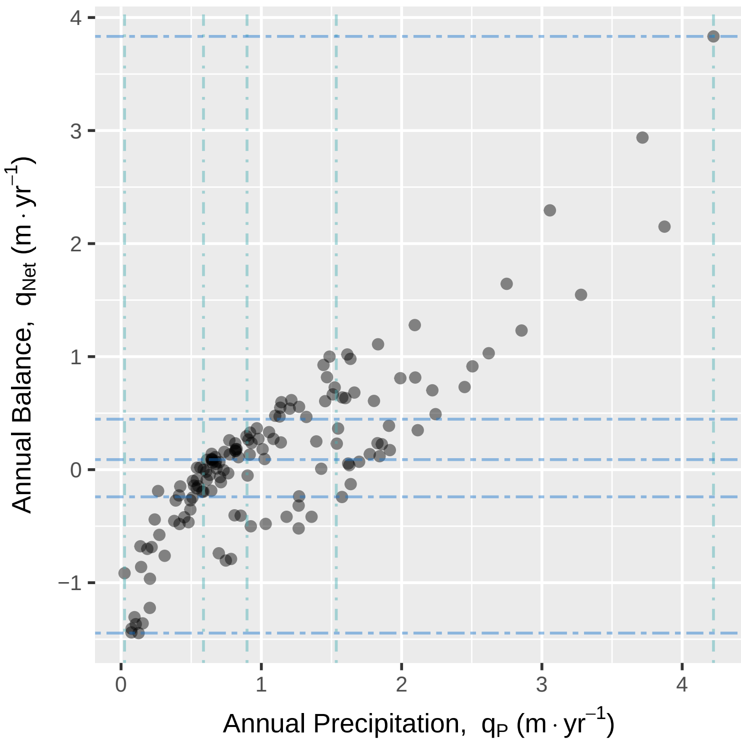
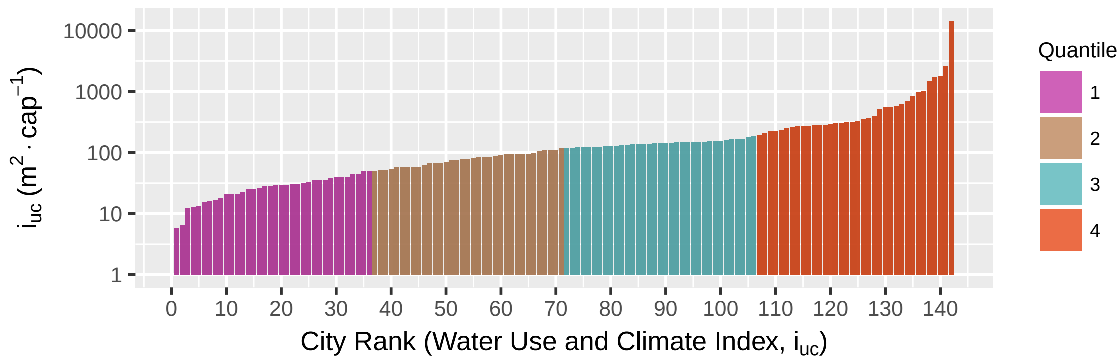


Table B.1: Supplementary summary statistics for the UrbMet database of cities. Summary statistics for annual precipitation height (q_P in $\mathrm{m} \cdot \mathrm{yr}^{-1}$), the Water Use and Climate Index (i_{UC} in $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$), the total Water Use and Climate Index ($I_{UC} = N \cdot i_{UC}$ in km^2), and the potential self-sufficiency ratio (R_{SS}, dimensionless).

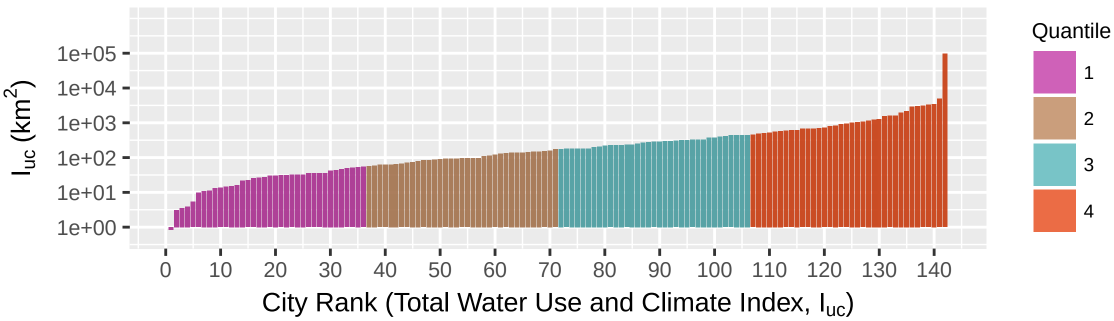
Quantile	q_P $\mathrm{m} \cdot \mathrm{yr}^{-1}$	WUCI $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$	Total WUCI km^2	R_{SS} *Unitless*
μ	1	303	1120893587	10.882
σ	1	1224	8052504077	42.240
γ	1	10	12	8.437
κ	2	114	133	80.471
Statistic	q_P $\mathrm{m} \cdot \mathrm{yr}^{-1}$	WUCI $\mathrm{m}^2 \cdot \mathrm{cap}^{-1}$	Total WUCI km^2	R_{SS} *Unitless*
0%	0	6	838130	0.002
25%	1	50	55453879	0.745
50%	1	118	175789092	1.941
75%	2	189	451846473	5.353
100%	4	14200	95971850200	448.873

Figure B.5: Ordered bar chart and table of Water Use and Climate Index (i_{UC} in $m^2 \cdot cap^{-1}$) for UrbMet cities, colored by quantile (see Table B.1). To identify the position for a particular city in this plot, first find its rank in the table.



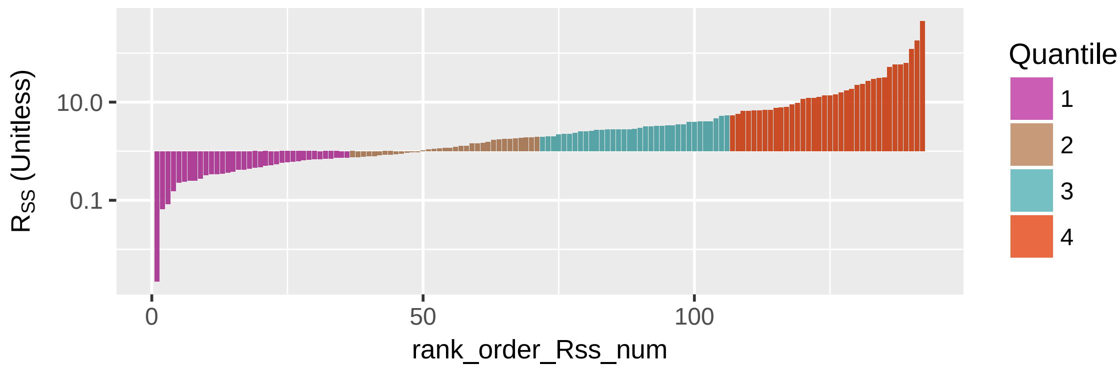
n	City	*n*	City	*n*	City	*n*	City
1	Accra	41	Yangon	81	St. Petersburg	121	Istanbul
2	Naihati	42	Minsk	82	Panama City	122	Toronto
3	Durban	43	Chisinau	83	Tokyo	123	Beirut
4	Rabat	44	Osaka	84	Madrid	124	Rome
5	Lagos	45	Singapore	85	Guangzhou	125	Santiago
6	Hyderabad	46	Kathmandu	86	Quezon City	126	Port-of-Spain
7	Colombo	47	Bogota	87	Cairo	127	Denver, CO
8	Delhi	48	Manila	88	Kinshasa	128	Copenhagen
9	Bangalore	49	Quito	89	Paris	129	Buenos Aires
10	Dakar	50	Hamburg	90	Kiev	130	Victoria, BC
11	Surabaya	51	Sofia	91	Moscow	131	Chicago, IL
12	Asuncion	52	Kolkata	92	Lisbon	132	Ho Chi Minh City
13	Dhaka	53	Mexico City	93	Riyadh	133	Vancouver, BC
14	Islamabad	54	Guadalajara	94	Detroit, MI	134	Beijing
15	Casablanca	55	Ulaanbaatar	95	Nagoya	135	Jerusalem
16	Damascus	56	Johannesburg	96	Cebu	136	Geneva
17	Jakarta	57	Nairobi	97	Tehran	137	London
18	Vilnius	58	Guatemala City	98	Montreal	138	Seattle, WA
19	Athens	59	Prague	99	Abu Dhabi	139	Boston, MA
20	Lima	60	Dublin	100	Cali	140	Tbilisi
21	Tunis	61	Chennai	101	Riga	141	Bishkek
22	Karachi	62	Vienna	102	New York, NY	142	Portland, OR
23	Dubai	63	Ottawa	103	Kuala Lumpur		
24	Sao Paulo	64	Cape Town	104	Anchorage, AK		
25	Dar es Salaam	65	Tashkent	105	Phoenix, AZ		
26	Addis Ababa	66	Santo Domingo	106	Milan		
27	Brasilia	67	Manama	107	Yokohama		
28	Florianopolis	68	Berlin	108	Hanover		
29	Ljubljana	69	Caracas	109	Amsterdam		
30	Helsinki	70	Melbourne	110	Shenzhen		
31	Abuja	71	Warsaw	111	Seoul		
32	Barcelona	72	Ankara	112	Mumbai		
33	San Salvador	73	Stockholm	113	Doha		
34	Kingston	74	Belo Horizonte	114	Tripoli		
35	Brussels	75	Vladivostok	115	Sarajevo		
36	Curitiba	76	Sana'a	116	Bangkok		
37	Bucharest	77	Bandar seri begawan	117	Amman		
38	Shanghai	78	Phnom Penh	118	Kuwait City		
39	Montevideo	79	Budapest	119	Los Angeles, CA		
40	Sydney	80	Bern	120	St. John's		

Figure B.6: Ordered bar chart of total Water Use and Climate Index (I_{UC} in km^2, where $I_{UC} = i_{UC} \cdot N$ and 1km^2 = 10^{6}m^2) for UrbMet cities, colored by quantile (see Table 3.6). To identify the position for a particular city in this plot, first find its rank in the table.



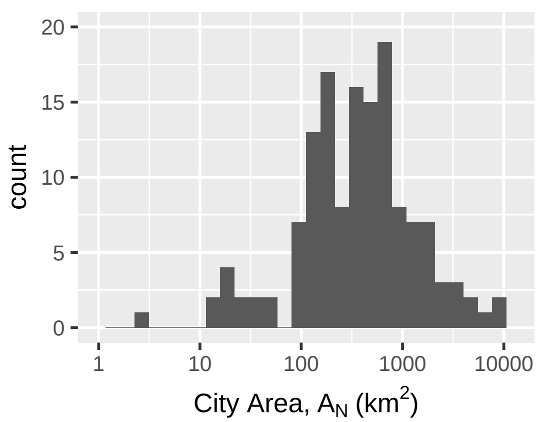
n	City	n	City	n	City	n	City
1	Dubai	41	Florianopolis	81	Toronto	121	Amsterdam
2	Lima	42	Islamabad	82	Los Angeles, CA	122	Sofia
3	Delhi	43	Ankara	83	Asuncion	123	Boston, MA
4	Naihati	44	Berlin	84	Madrid	124	St. John's
5	Bandar seri begawan	45	Doha	85	Shanghai	125	Guangzhou
6	Addis Ababa	46	Damascus	86	Kingston	126	Istanbul
7	Dublin	47	Brasilia	87	Vienna	127	St. Petersburg
8	Guadalajara	48	Durban	88	Kuala Lumpur	128	Kinshasa
9	Lagos	49	Chisinau	89	Tehran	129	Melbourne
10	Sydney	50	Osaka	90	Panama City	130	Amman
11	Detroit, MI	51	Warsaw	91	Phnom Penh	131	London
12	Karachi	52	Helsinki	92	Seoul	132	Seattle, WA
13	Barcelona	53	Bern	93	Kiev	133	Port-of-Spain
14	Sana'a	54	Kathmandu	94	Vladivostok	134	Quezon City
15	Sao Paulo	55	Surabaya	95	Rome	135	Ho Chi Minh City
16	Accra	56	Yokohama	96	Milan	136	Kuwait City
17	Colombo	57	Riyadh	97	Chicago, IL	137	Santiago
18	Yangon	58	Cali	98	Phoenix, AZ	138	Tripoli
19	Singapore	59	Montreal	99	Caracas	139	Copenhagen
20	Dhaka	60	Jakarta	100	Tokyo	140	Jerusalem
21	Dar es Salaam	61	Ulaanbaatar	101	Ottawa	141	Beijing
22	Tunis	62	Moscow	102	Cairo	142	Portland, OR
23	Geneva	63	Hanover	103	Anchorage, AK		
24	Brussels	64	Bangalore	104	Victoria, BC		
25	Minsk	65	Prague	105	Cebu		
26	Bucharest	66	Nairobi	106	Bangkok		
27	Ljubljana	67	Athens	107	Bishkek		
28	Hamburg	68	Rabat	108	Riga		
29	Quito	69	Mexico City	109	Montevideo		
30	Stockholm	70	Hyderabad	110	Kolkata		
31	Johannesburg	71	Lisbon	111	Nagoya		
32	Bogota	72	Abu Dhabi	112	Sarajevo		
33	Santo Domingo	73	Curitiba	113	New York, NY		
34	Cape Town	74	Denver, CO	114	Buenos Aires		
35	Chennai	75	Shenzhen	115	Tbilisi		
36	Dakar	76	Belo Horizonte	116	Manila		
37	Tashkent	77	Beirut	117	Mumbai		
38	San Salvador	78	Abuja	118	Vancouver, BC		
39	Vilnius	79	Budapest	119	Paris		
40	Manama	80	Casablanca	120	Guatemala City		

Figure B.7: Ordered bar chart and table of the dimensionless self-sufficiency ratio (R_{SS}) for Urb-Met cities, colored by quantile (see Table B.1). To identify the position for a particular city in this plot, first find its rank in the table.

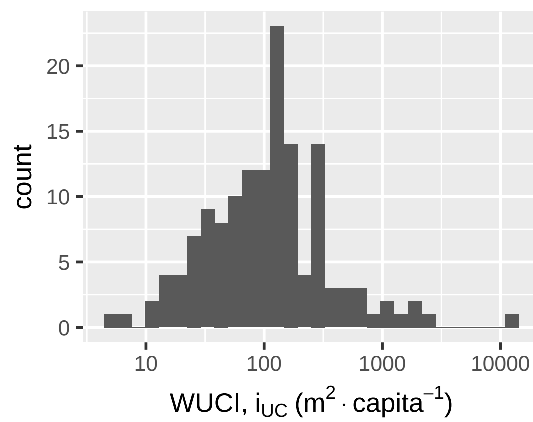


n	City	n	City	n	City	n	City
1	Portland, OR	41	Seattle, WA	81	Ljubljana	121	Sydney
2	Bishkek	42	Tehran	82	Moscow	122	Los Angeles, CA
3	Geneva	43	London	83	Cali	123	Sao Paulo
4	Beijing	44	Montevideo	84	Kingston	124	Brussels
5	Jerusalem	45	Yokohama	85	Buenos Aires	125	Nagoya
6	Copenhagen	46	Manila	86	Shanghai	126	Guadalajara
7	Tbilisi	47	Milan	87	Vilnius	127	Tunis
8	Doha	48	Curitiba	88	Ankara	128	Lagos
9	Istanbul	49	Quezon City	89	Mexico City	129	Durban
10	Riga	50	Paris	90	Casablanca	130	Prague
11	Phoenix, AZ	51	Ho Chi Minh City	91	Amsterdam	131	Accra
12	Kiev	52	Panama City	92	Delhi	132	Barcelona
13	Hanover	53	Chicago, IL	93	Johannesburg	133	Ottawa
14	Kuwait City	54	Ulaanbaatar	94	Beirut	134	Detroit, MI
15	St. John's	55	Santiago	95	Quito	135	Lima
16	Warsaw	56	Mumbai	96	Bern	136	Dhaka
17	Victoria, BC	57	Riyadh	97	Abuja	137	Karachi
18	Tripoli	58	Bangkok	98	Manama	138	Bangalore
19	Stockholm	59	Jakarta	99	Yangon	139	Helsinki
20	Amman	60	Montreal	100	Singapore	140	Dubai
21	Anchorage, AK	61	Melbourne	101	Rabat	141	Naihati
22	Sarajevo	62	Tokyo	102	Brasilia	142	Addis Ababa
23	Toronto	63	Berlin	103	Vladivostok		
24	St. Petersburg	64	Damascus	104	Dublin		
25	Bucharest	65	Sofia	105	Nairobi		
26	Seoul	66	Santo Domingo	106	Guatemala City		
27	Rome	67	Phnom Penh	107	Islamabad		
28	Denver, CO	68	Kuala Lumpur	108	Hamburg		
29	Minsk	69	Hyderabad	109	Asuncion		
30	Kinshasa	70	Chisinau	110	Athens		
31	Guangzhou	71	Surabaya	111	Osaka		
32	Shenzhen	72	Belo Horizonte	112	Florianopolis		
33	Lisbon	73	Boston, MA	113	Colombo		
34	Cebu	74	Budapest	114	Chennai		
35	New York, NY	75	Tashkent	115	Abu Dhabi		
36	Kolkata	76	Madrid	116	Cape Town		
37	Sana'a	77	Bandar seri begawan	117	Dar es Salaam		
38	Bogota	78	Caracas	118	San Salvador		
39	Port-of-Spain	79	Kathmandu	119	Dakar		
40	Vancouver, BC	80	Vienna	120	Cairo		

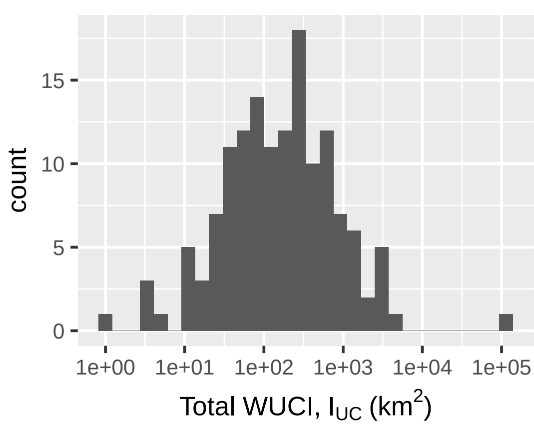
Figure B.8: Histograms of city area (A_N, in km^2), the Water Use and Climate Index (i_{UC}, in m$^2 \cdot$ cap^{-1}), I_{UC} (I_{UC}, in km^2), and the Potential Self-Sufficiency Ratio (R_{SS}, unitless), for cities in the UrbMet database.



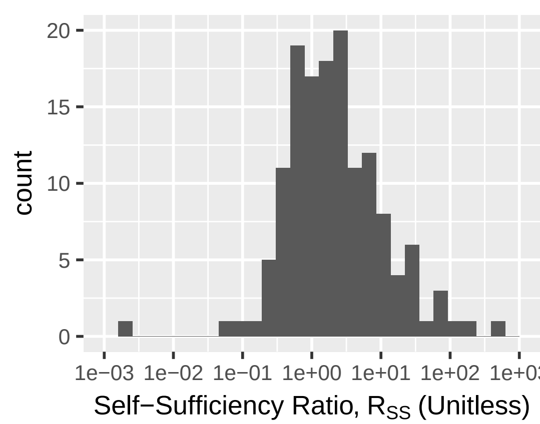
(a) Histogram of city area (A_N).



(b) Histogram of WUCI (i_{UC}).

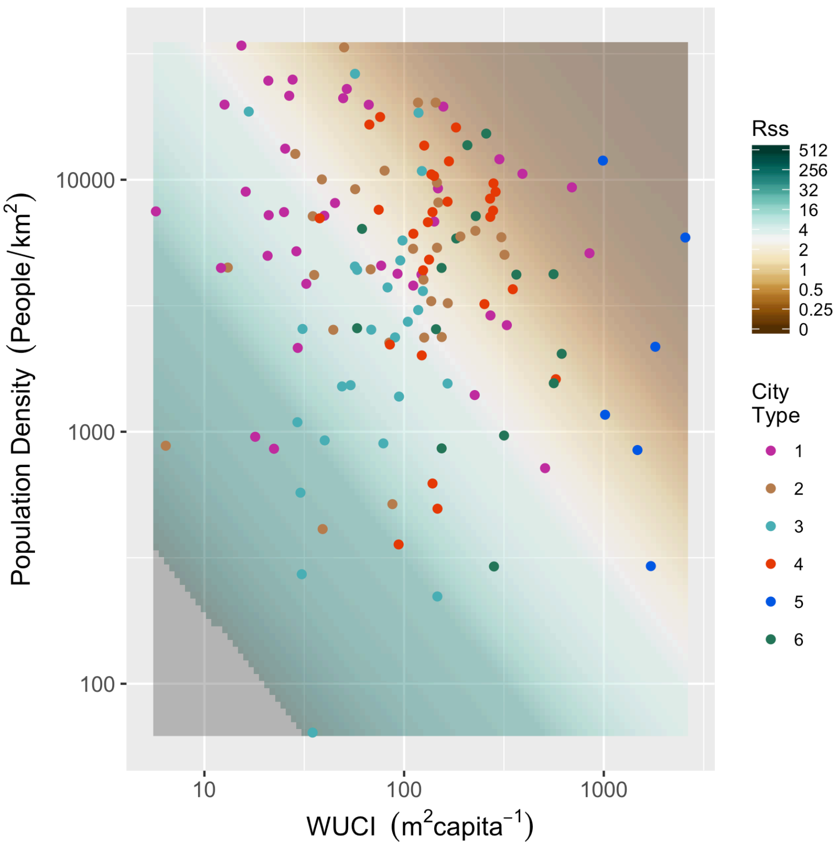


(c) Histogram of total WUCI (I_{UC}).



(d) Histogram of self-sufficiency (R_{SS}).

Figure B.9: Self-Sufficiency R_{SS} plotted against population density ρ and per capita WUCI i_{UC}.



B.1.4 Supplementary Clustering Results

Univariate

The univariate hierarchical clustering results demonstrate that while it is useful to describe the data in terms of statistical quantiles, the quantiles do not necessarily correspond to the results of hierarchical clustering. In other words, cities within the same quantile are not necessarily more similar, in terms of quantitative predictor variables, to other cities within the same quantile than to those in other quantiles.

The simplest application of hierarchical clustering is to observations of a single variable, i.e. 'univariate' clustering. While the results of a univariate clustering are only so interesting, it was decided that it was a useful place to start data exploration to provide a visual insight into the results of the hierarchical clustering algorithm. The univariate results were visually compared with the statistical quantiles (for which membership was the same regardless of whether the log-10 transformation was applied to the data or not).

Univariate hierarchical clustering was performed on each of the three predictor variables: 1. water use intensity, w_N (in $m^3 \cdot yr^{-1} \cdot cap^{-1}$wui); 2. net annual water balance, q_{Net} (in $m \cdot yr^{-1}$); and population, N. Previously, ordered bar charts were produced for each of these three metrics, as seen in Figures 3.4, 3.5, and 3.3, respectively.

1. Water use intensity
2. Net annual water balance
3. Population

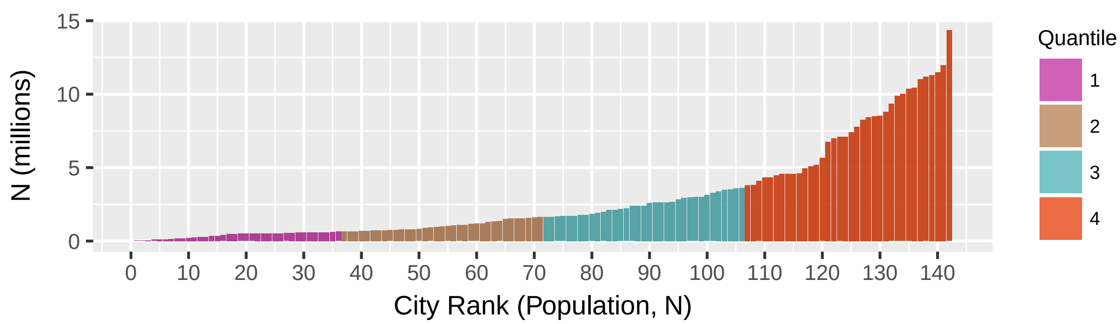
For each of these three metrics, the hierarchical clustering algorithm was applied to the univariate series of either the data after a log-10 transformation (as described in 3.6.1) or the data prior to transformation. As for the trivariate clustering, described in 3.5, the `Euclidean` distance formula was used to calculate the distance matrix and the `hclust` algorithm from the `stats` package in `R` was used to cluster with the `complete` linkage criterion[1]. The Euclidean distance was chosen as the distance formula as there were no particular reasons to prefer the other methods.

The resulting dendrograms were then cut to four clusters to facilitate comparison with the quantiles for each metric. The results for each metric have been provided as ordered bar charts, shown in Figures B.10b–B.12b. In each of these figures, the top plot depicts the ordered bar chart colored by quantile (with probabilities at 0%, 25%, 50%, 75%, and 100%), while the middle and bottom plots show the same underlying ordered bar chart colored by the univariate clustering of either the un-transformed data (middle) or the data after a log-10 transformation (bottom).

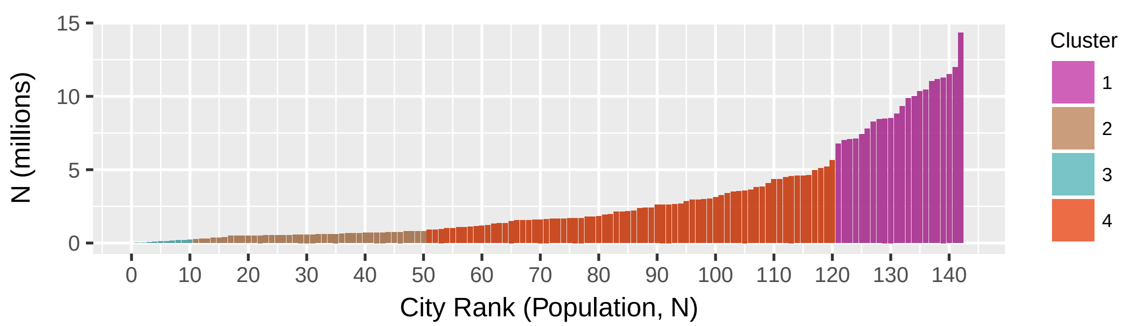
[1]Mathematical formulas for distances and methods for clustering are provided in Tables 3.4–3.5.

Figure B.10: Supplementary clustering results: univariate clustering of population (N in 10^6 cap). Comparison of quantiles (top) with the univariate clustering results of the population data after a log-10 transformation (bottom). To identify the position for a particular city in the ordered bar charts, see Figure 3.3.

(a) Ordered bar chart of population, colored by quantile.



(b) Ordered bar chart of population, colored by univariate clustering.



(c) Scatterplot of WUI and Annual Balance, colored by univariate clustering of population.

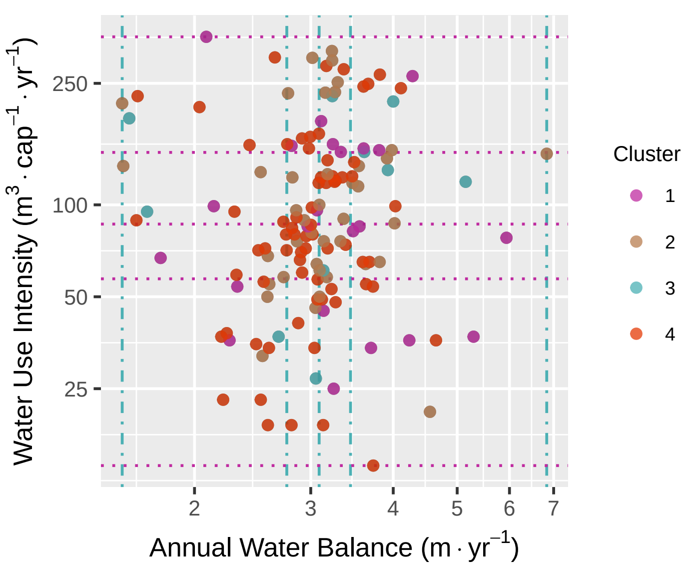
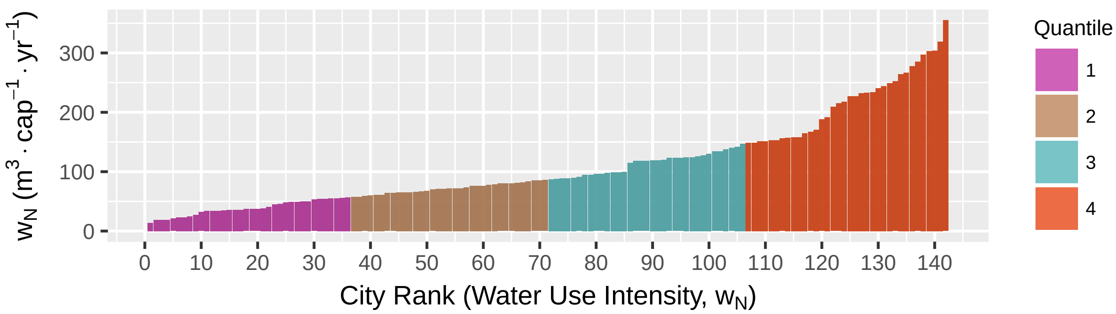
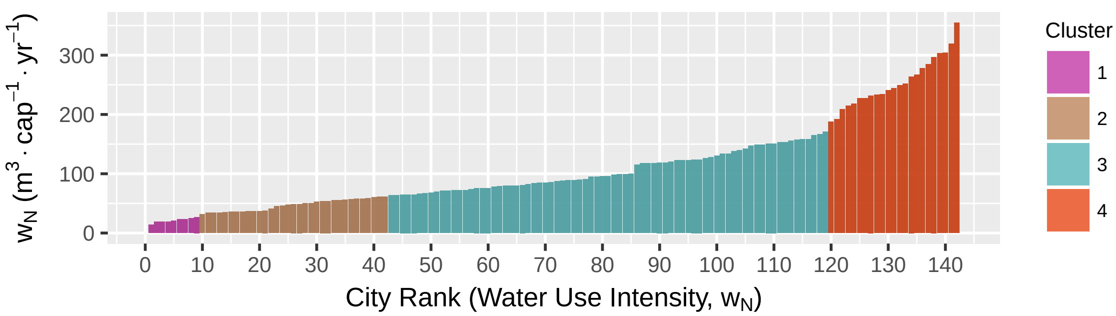


Figure B.11: Supplementary clustering results: univariate clustering of water use intensity. Comparison of quantiles (top) with the univariate clustering results of the water use intensity data after a log-10 transformation (bottom). To identify the position for a particular city in the ordered bar charts, see Figure 3.4.

(a) Ordered bar chart of WUI, colored by quantile.



(b) Ordered bar chart of WUI, colored by univariate clustering.



(c) Scatterplot of population and Annual Balance, colored by univariate clustering of WUI.

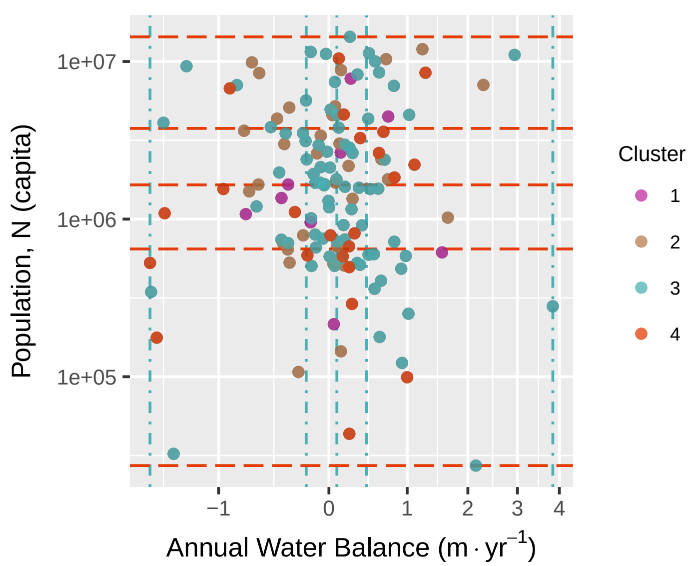
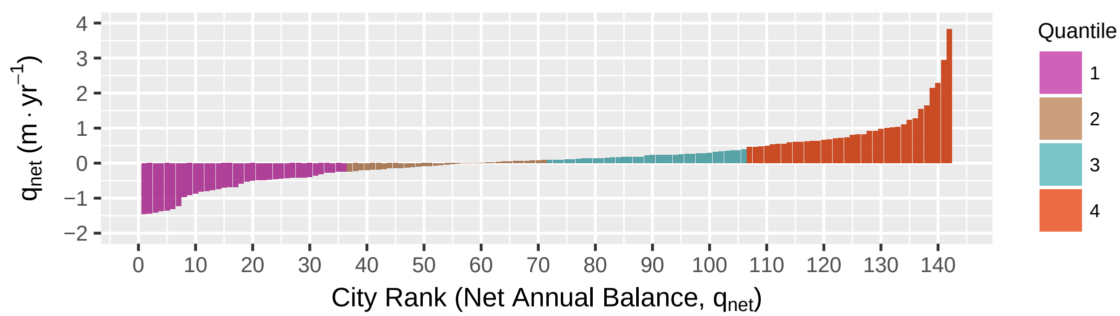
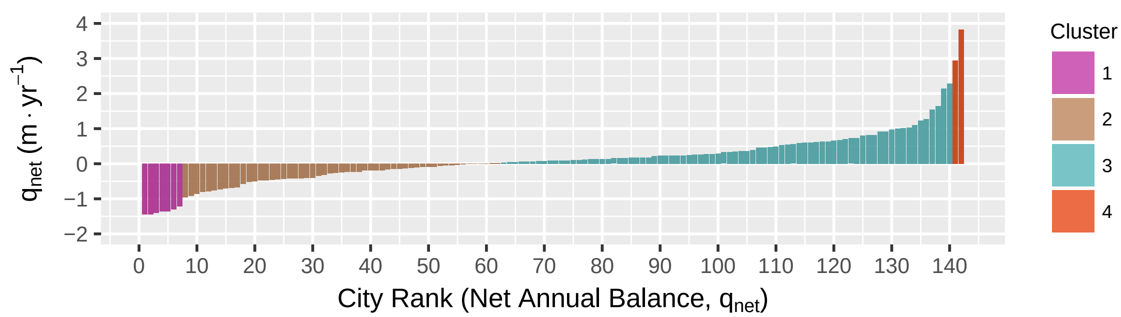


Figure B.12: Supplementary clustering results: univariate clustering of net annual water balance. Comparison of quantiles (top) with the univariate clustering results of the water balance data after a log-10 transformation (bottom). To identify the position for a particular city in the ordered bar charts, see Figure 3.5.

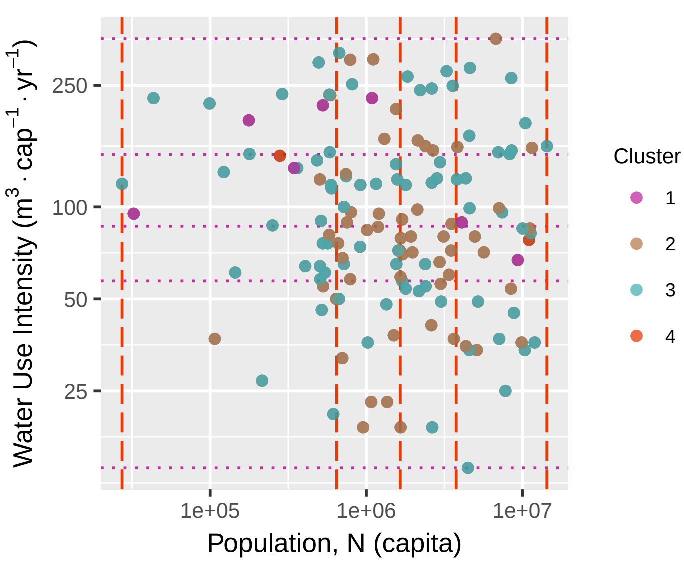
(a) Ordered bar chart of AB, colored by quantile.



(b) Ordered bar chart of AB, colored by univariate clustering.



(c) Scatterplot of WUI (w_N in $m^3 \cdot yr^{-1} \cdot cap^{-1}$) and population ($N$ in capita), colored by univariate clustering of CWB (q_{Net} in $m \cdot yr^{-1}$).



Bivariate

Hierarchical clustering was performed on the following bivariate combinations:

- water use intensity and climatic water balance (w_N versus q_{Net}),
- population and climatic water balance (N versus q_{Net}), and
- water use intensity and population (w_N versus q_{Net}).

The results of bivariate hierarchical clustering were plotted as scatterplots and histograms, and correlations between variables were found for each cluster type using the `ggpairs` function from the `GGally` package for `R`, as seen in Figures B.13–B.15. The results of the bivariate clustering of w_N and q_{Net} are shown in Figure B.13 on Page 284; those for N and q_{Net} are shown in Figure B.14 on Page B.14, and those for w_N and N are shown in Figure B.15 on Page B.15.

Figure B.13: Supplementary clustering results: bivariate clustering of WUI (w_N) and net annual balance (q_{Net}). The plot matrix shows the results as histograms (diagonal), scatterplots (below diagonal), and bivariate correlations (above diagonal). The clustering is seen most clearly in the two-dimensional space formed by w_N and q_{Net} (bottom row, middle).

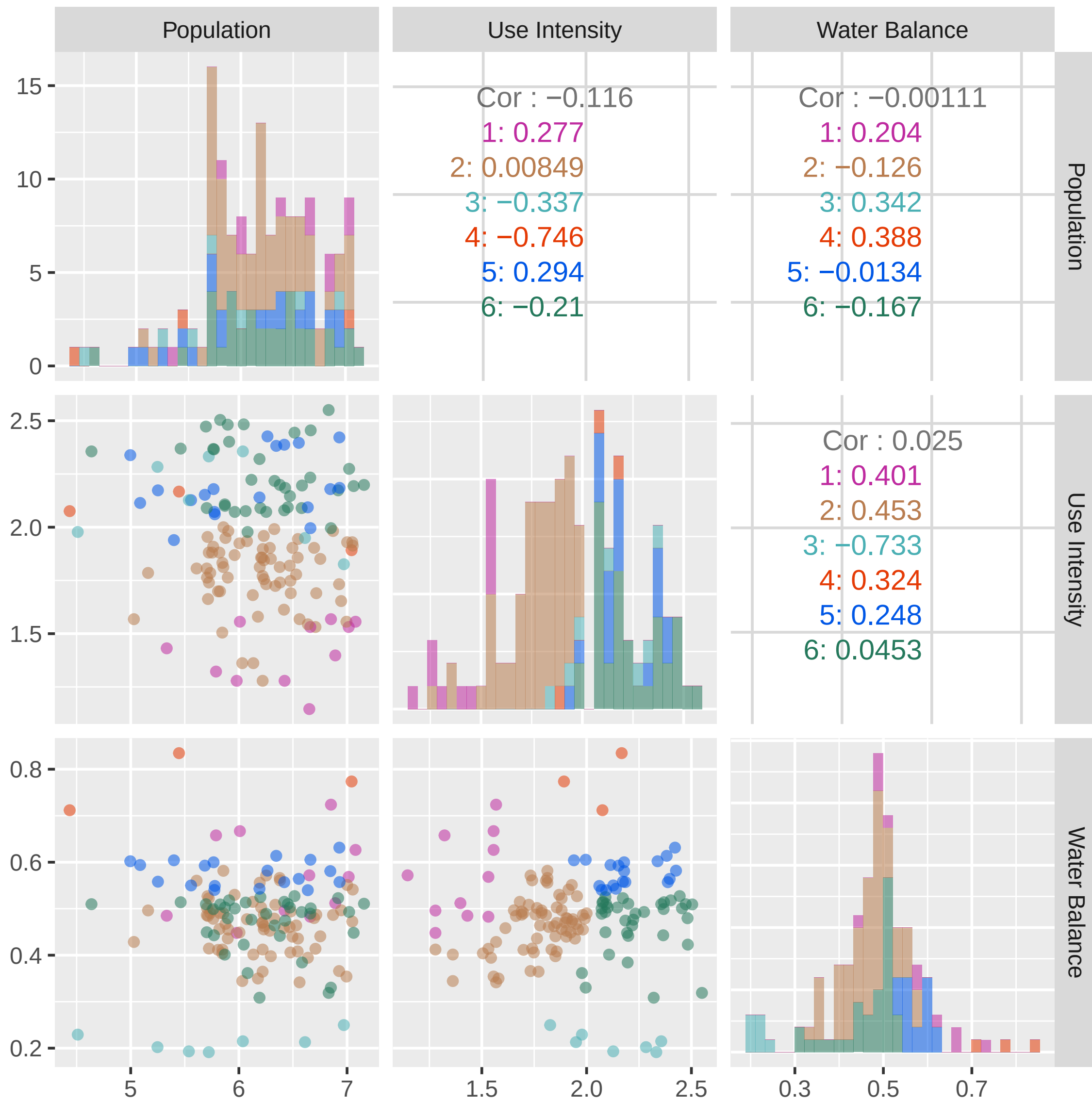


Figure B.14: Supplementary clustering results: bivariate clustering of population (N) and net annual balance (q_{Net}). The plot matrix shows the results as histograms (diagonal), scatterplots (below diagonal), and bivariate correlations (above diagonal). The clustering is seen most clearly in the two-dimensional space formed by N and q_{Net} (bottom row, left).

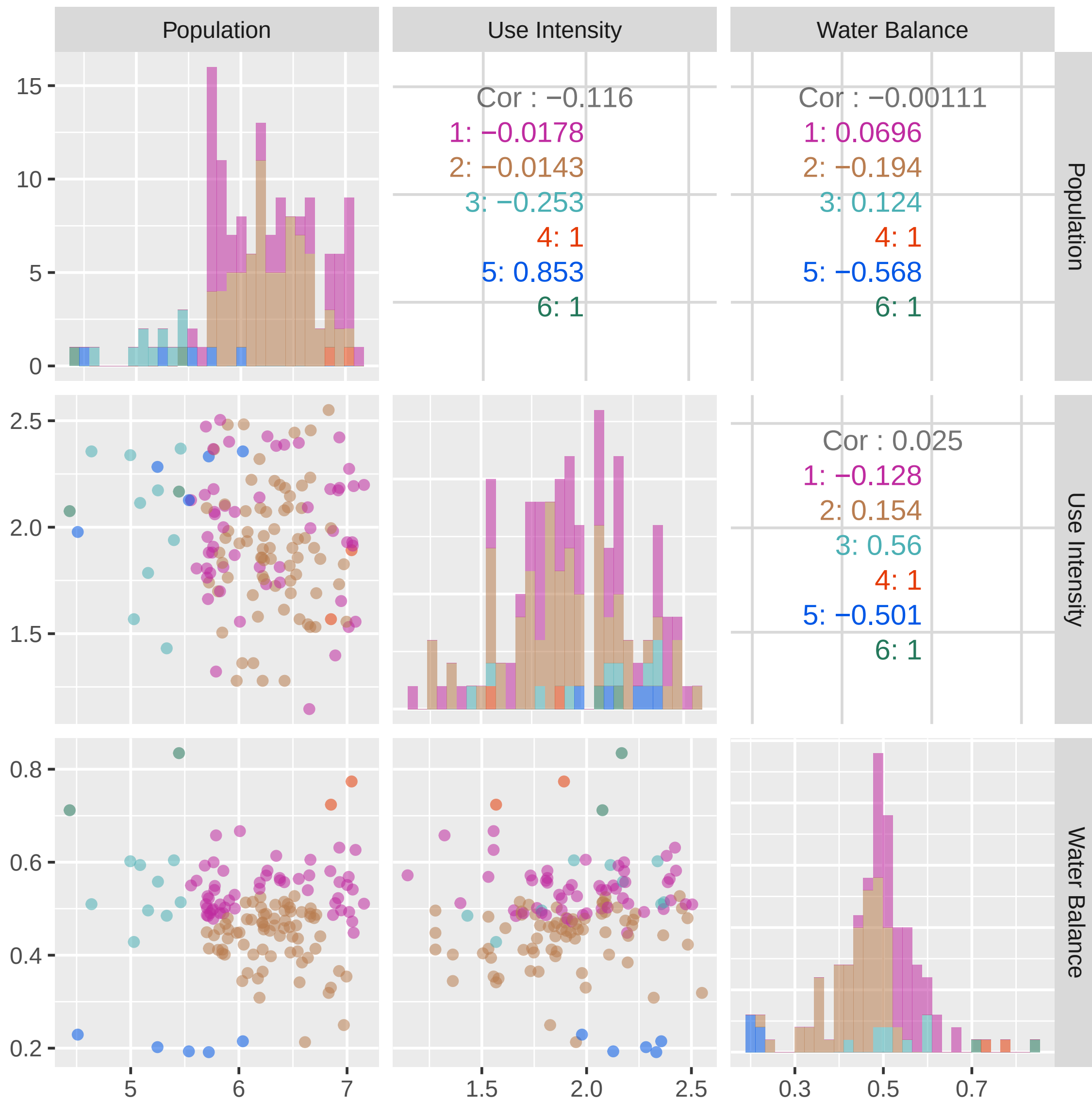
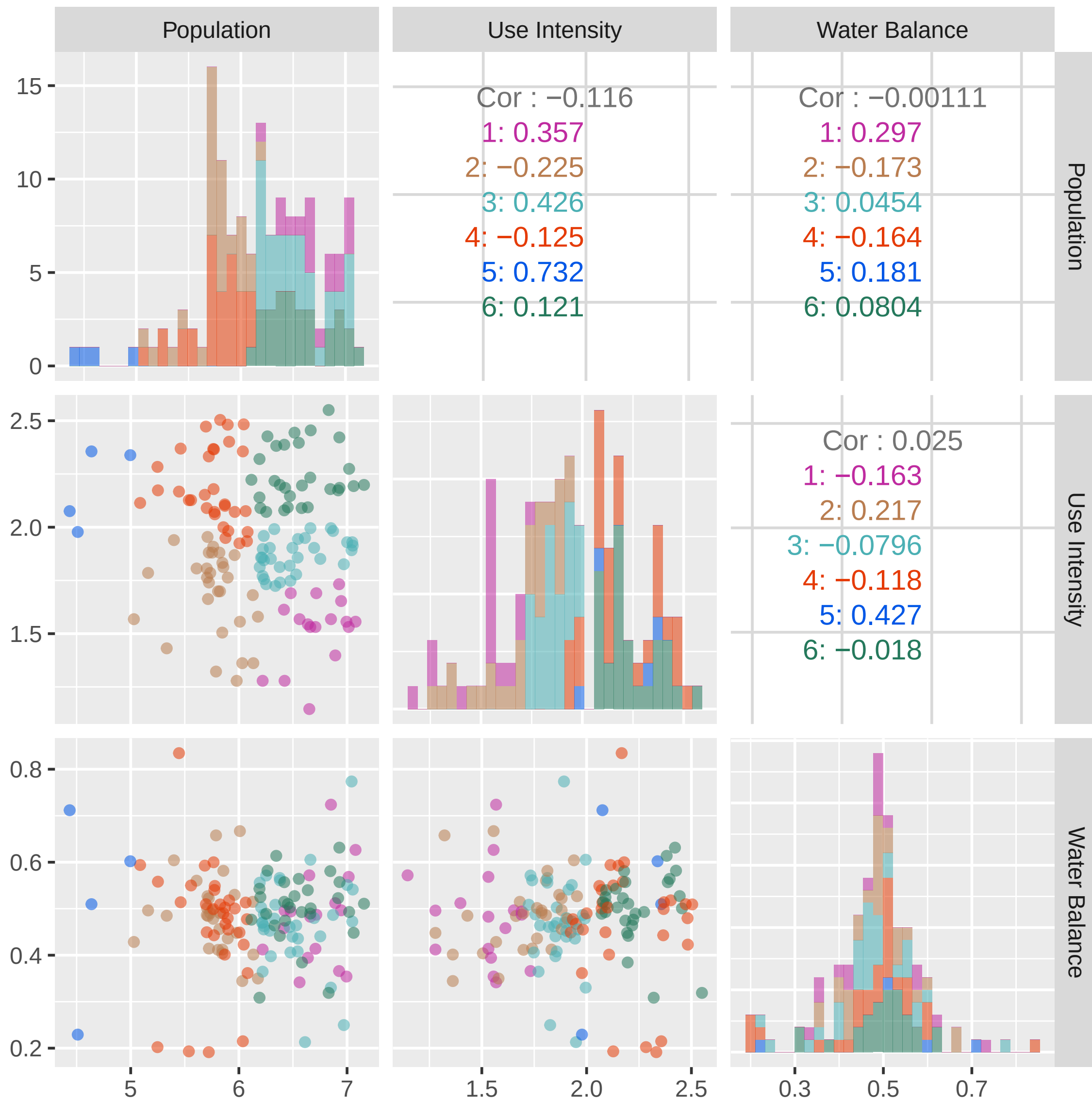


Figure B.15: Supplementary clustering results: bivariate clustering of WUI (w_N) and population (N). The plot matrix shows the results as histograms (diagonal), scatterplots (below diagonal), and bivariate correlations (above diagonal). The clustering is seen most clearly in the two-dimensional space formed by N and q_{Net} (middle row, left).



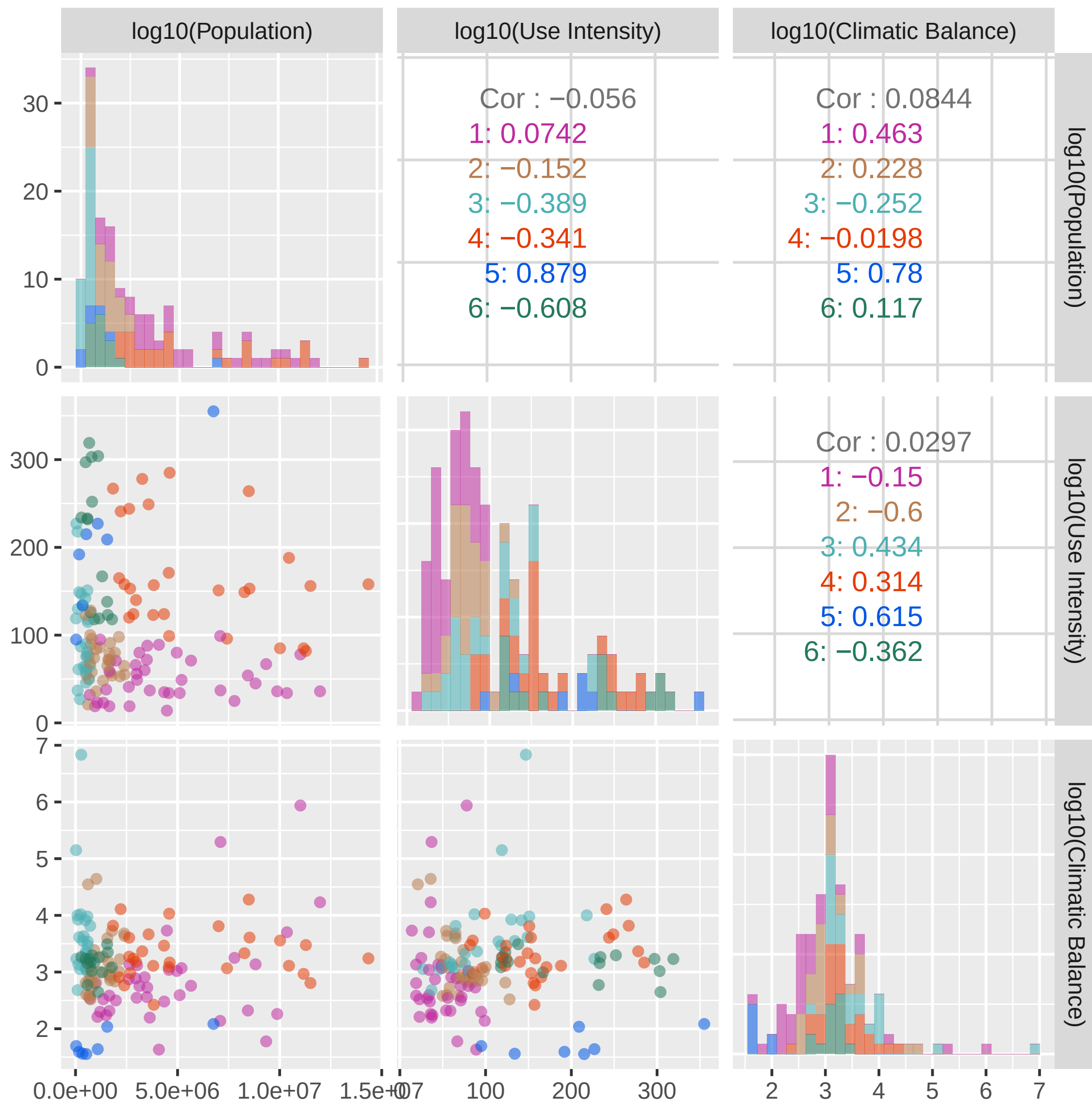
Trivariate

This section contains supplementary results for the trivariate clustering in Chapter 3. Table B.2 on Page 287 shows a full listing of the membership associated with each type. These can also be seen in the dendrograms associated with each type, Figures 3.12a–3.17a on Pages 3.12a–3.17a. The six types from the trivariate clustering have been compared in Figure B.16 on Page B.16 as scatterplots and histograms, with bivariate correlations calculated. Figure B.16 was produced using the `ggpairs` function in the `GGally` library in `R`.

Type					
1	**2**	**3**	**4**	**5**	**6**
Type1	Type2	Type3	Type4	Type5	Type6
Accra	Athens	Abuja	Beijing	Abu Dhabi	Amsterdam
Addis Ababa	Barcelona	Anchorage, AK	Buenos Aires	Cairo	Denver, CO
Amman	Belo Horizonte	Asuncion	Cali	Doha	Detroit, MI
Ankara	Bishkek	Bandar seri begawan	Chicago, IL	Dubai	Dublin
Bangalore	Brasilia	Beirut	Guangzhou	Kuwait City	Kathmandu
Bangkok	Bucharest	Bern	Istanbul	Manama	Kuala Lumpur
Berlin	Budapest	Boston, MA	Kiev	Phoenix, AZ	Manila
Bogota	Chisinau	Brussels	London		Milan
Cape Town	Colombo	Cebu	Los Angeles, CA		Minsk
Caracas	Curitiba	Copenhagen	Melbourne		Ottawa
Casablanca	Guadalajara	Durban	Mexico City		Sofia
Chennai	Guatemala City	Florianopolis	Montreal		Stockholm
Dakar	Hamburg	Geneva	Moscow		Tbilisi
Damascus	Islamabad	Hanover	Nagoya		Victoria, BC
Dar es Salaam	Jerusalem	Helsinki	New York, NY		Vladivostok
Delhi	Johannesburg	Kingston	Osaka		
Dhaka	Lisbon	Ljubljana	Rome		
Ho Chi Minh City	Montevideo	Naihati	Seoul		
Hyderabad	Paris	Panama City	Shanghai		
Jakarta	Phnom Penh	Port-of-Spain	Shenzhen		
Karachi	Prague	Portland, OR	Singapore		
Kinshasa	Quezon City	San Salvador	St. Petersburg		
Kolkata	Quito	Sarajevo	Sydney		
Lagos	Rabat	Seattle, WA	Tashkent		
Lima	Riga	St. John's	Tokyo		
Madrid	Santo Domingo	Vilnius	Toronto		
Mumbai	Ulaanbaatar		Vancouver, BC		
Nairobi	Vienna		Yokohama		
Riyadh	Warsaw				
Sana'a					
Santiago					
Sao Paulo					
Surabaya					
Tehran					
Tripoli					
Tunis					
Yangon					

Table B.2: Table of trivariate cluster members, by type.

Figure B.16: Scatterplots, histograms, and correlation for population (N in capita), water use intensity (w_N in $m^3 \cdot yr^{-1} \cdot cap^{-1}$), and climatic water balance (q_{Net} in $m \cdot yr^{-1}$), colored by trivariate clustering results.



Appendix C

Appendix: Data

C.1 Data from Chapter 3

C.1.1 UrbMet Data

Table C.1: Data from UrbMet dataframe for variables used in analysis: population (N), city area (A_N in km^2), and water use intensity (w_N in $m^3 \cdot yr^{-1} \cdot cap^{-1}$).

c_id	City	Country	N capita	A_N km^2	w_N $m^3 \cdot yr^{-1} \cdot cap^{-1}$
1	Abu Dhabi	United Arab Emirates	527000	1798.63	215
2	Abuja	Nigeria	107069	98.14	37
3	Accra	Ghana	1658937	185.00	19
4	Addis Ababa	Ethiopia	2646000	530.05	19
5	Amman	Jordan	1204110	1679.37	95
6	Amsterdam	Netherlands	742884	166.30	126
7	Anchorage, AK	USA	279671	4369.86	147
8	Ankara	Turkey	3517182	2515.87	88
9	Asuncion	Paraguay	513399	117.00	90
10	Athens	Greece	789166	39.00	58
11	Bandar seri begawan	Brunei	27285	100.31	119
12	Bangalore	India	5104047	708.99	34
13	Bangkok	Thailand	5658953	700.02	71
14	Barcelona	Spain	1605602	3111.63	72
15	Beijing	China	11509595	1368.07	156
16	Beirut	Lebanon	361366	19.60	134
17	Belo Horizonte	Brazil	2399920	335.00	55
18	Berlin	Germany	3386667	890.99	60
19	Bern	Switzerland	122256	51.63	130
20	Bishkek	Kyrgyzstan	798300	127.30	96
21	Bogota	Colombia	7102602	1590.02	37
22	Boston, MA	USA	599351	125.31	115
23	Brasilia	Brazil	2383784	5799.96	65

Continued on the next page

Table C.1: *Cont.*

c_id	**City**	**Country**	**N** capita	$\mathbf{A_N}$ km^2	$\mathbf{w_N}$ $m^3 \cdot yr^{-1} \cdot cap^{-1}$
24	Brussels	Belgium	144784	161.05	61
25	Bucharest	Romania	1931838	237.88	80
26	Budapest	Hungary	1699213	524.93	91
27	Buenos Aires	Argentina	2965403	4758.00	140
28	Cairo	Egypt	6758581	214.00	355
29	Cali	Colombia	2392877	547.32	158
30	Cape Town	South Africa	3497097	290.00	72
31	Caracas	Venezuela	1975294	432.99	71
32	Casablanca	Morocco	2995000	323.99	56
33	Cebu	Philippines	718821	281.01	65
34	Chennai	India	4343645	174.00	35
35	Chicago, IL	USA	2836658	588.64	124
36	Chisinau	Moldova	660726	122.99	76
37	Colombo	Sri Lanka	615000	698.86	21
38	Copenhagen	Denmark	505141	88.00	64
39	Curitiba	Brazil	1788559	426.97	54
40	Dakar	Senegal	1075582	500.04	23
41	Damascus	Syria	1658000	572.91	59
42	Dar es Salaam	Tanzania	1360850	1589.78	23
43	Delhi	India	9879172	430.99	36
44	Denver, CO	USA	588349	377.63	232
45	Detroit, MI	USA	916952	359.31	118
46	Dhaka	Bangladesh	10356500	304.00	34
47	Doha	United Arab Emirates	344939	158.67	134
48	Dubai	United Arab Emirates	1089000	1287.23	227
49	Dublin	Ireland	495781	117.99	297
50	Durban	South Africa	669242	442.33	50
51	Florianopolis	Brazil	406564	440.00	64
52	Geneva	Switzerland	178574	16.49	149
53	Guadalajara	Mexico	1640589	151.00	72
54	Guangzhou	China	8524826	3843.47	153
55	Guatemala City	Guatemala	1022001	228.02	36
56	Hamburg	Germany	1704735	754.98	57
57	Hanover	Germany	518386	204.01	46
58	Helsinki	Finland	566526	186.36	76
59	Ho Chi Minh City	Viet Nam	3015743	140.00	49
60	Hyderabad	India	3637483	172.68	37
61	Islamabad	Pakistan	529180	120.00	55
62	Istanbul	Turkey	11174257	1830.94	85
63	Jakarta	Indonesia	8820603	664.00	45
64	Jerusalem	Israel	740475	125.21	128
65	Johannesburg	South Africa	752349	318.25	89

Continued on the next page

Table C.1: *Cont.*

c_id	City	Country	N capita	A_N km^2	w_N $m^3 \cdot yr^{-1} \cdot cap^{-1}$
66	Karachi	Pakistan	9339023	3526.82	67
67	Kathmandu	Nepal	671846	49.00	319
68	Kiev	Ukraine	2676789	833.89	153
69	Kingston	Jamaica	579137	22.00	81
70	Kinshasa	Democratic Republic of Congo	7785965	8152.84	25
71	Kolkata	India	4572876	185.00	34
72	Kuala Lumpur	Malaysia	1551306	243.00	138
73	Kuwait City	Kuwait	32403	2.72	95
74	Lagos	Nigeria	5195247	999.09	49
75	Lima	Peru	8445211	799.96	54
76	Lisbon	Portugal	504726	84.77	123
77	Ljubljana	Slovenia	250953	164.02	87
78	London	United Kingdom	7421209	706.45	96
79	Los Angeles, CA	USA	3834340	2372.74	157
80	Madrid	Spain	3128600	160.37	80
81	Manama	Bahrain	176909	30.00	192
82	Manila	Philippines	1581082	614.01	123
83	Melbourne	Australia	3806092	7689.07	123
84	Mexico City	Mexico	11285654	1484.95	82
85	Milan	Italy	1306086	182.06	167
86	Minsk	Belarus	1789098	305.99	118
87	Montevideo	Uruguay	1345010	529.95	48
88	Montreal	Canada	3268513	364.99	278
89	Moscow	Russia	10456490	1081.00	188
90	Mumbai	India	11978450	602.99	36
91	Nagoya	Japan	2215062	326.46	241
92	Naihati	India	215303	11.55	27
93	Nairobi	Kenya	2948109	695.97	66
94	New York, NY	USA	8274527	799.47	149
95	Osaka	Japan	2628811	222.10	244
96	Ottawa	Canada	812129	2781.26	252
97	Panama City	Panama	484261	107.00	142
98	Paris	France	2125017	105.00	98
99	Phnom Penh	Cambodia	703963	21.00	68
100	Phoenix, AZ	USA	1552259	1328.99	209
101	Port-of-Spain	Trinidad and Tobago	43396	12.00	227
102	Portland, OR	USA	583776	375.78	151
103	Prague	Czech Republic	1188126	121.97	86
104	Quezon City	Philippines	2173831	171.70	53
105	Quito	Ecuador	1559295	170.01	65
106	Rabat	Morocco	642000	120.65	50
107	Riga	Latvia	719928	303.00	100

Continued on the next page

Table C.1: *Cont.*

c_id	City	Country	N	A_N	w_N
			capita	km^2	$m^3 \cdot yr^{-1} \cdot cap^{-1}$
108	Riyadh	Saudi Arabia	4087152	799.99	89
109	Rome	Italy	2626640	1307.44	120
110	San Salvador	El Salvador	507665	885.98	58
111	Sana'a	Yemen	954448	247.01	19
112	Santiago	Chile	4960815	727.07	80
113	Santo Domingo	Dominican Republic	913540	91.08	74
114	Sao Paulo	Brazil	11016703	1522.91	78
115	Sarajevo	Bosnia	527049	141.00	76
116	Seattle, WA	USA	594210	217.42	118
117	Seoul	South Korea	10020123	605.34	85
118	Shanghai	China	14348535	1928.05	158
119	Shenzhen	China	7008831	395.00	151
120	Singapore	Singapore	4588600	652.75	99
121	Sofia	Bulgaria	1155403	1345.06	119
122	St. John's	Canada	99182	446.77	218
123	St. Petersburg	Russia	4569616	605.97	171
124	Stockholm	Sweden	789024	187.02	303
125	Surabaya	Indonesia	2611506	351.01	41
126	Sydney	Australia	4336374	12146.71	124
127	Tashkent	Uzbekistan	2137218	300.00	165
128	Tbilisi	Georgia	1108600	543.96	304
129	Tehran	Iran	7088287	759.98	99
130	Tokyo	Japan	8489653	621.36	264
131	Toronto	Canada	4612191	1254.34	285
132	Tripoli	Algeria	1500000	356.72	38
133	Tunis	Tunisia	702330	35.39	32
134	Ulaanbaatar	Mongolia	1012733	201.02	84
135	Vancouver, BC	Canada	1837969	114.00	267
136	Victoria, BC	Canada	289625	19.00	234
137	Vienna	Austria	1664146	414.69	79
138	Vilnius	Lithuania	543494	394.12	61
139	Vladivostok	Russia	579811	600.22	233
140	Warsaw	Poland	1704717	517.05	70
141	Yangon	Burma	4477638	597.98	14
142	Yokohama	Japan	3579628	437.39	249

C.1.2 Alternate City Data

C.1.3 Location Data

Table C.2: City and country names and alternative spellings for UrbMet cities and countries, which were used in data verification and scripting.

	Original		**Alternative**	
c_id	**City**	**Country**	**City**	**Country**
4	Addis Ababa	Ethiopia	Addis Abeba	
5	Amman	Jordan	Amman	
7	Anchorage, AK	USA	Anchorage	
11	Bandar seri begawan	Brunei	Bandar Seri Begawan	
16	Beirut	Lebanon	Bayrut	
20	Bishkek	Kyrgyzstan	Biskek	Kyrgyz Republic
22	Boston, MA	USA	Boston	
34	Chennai	India	Madras	
35	Chicago, IL	USA	Chicago	
44	Denver, CO	USA	Denver	
45	Detroit, MI	USA	Detroit	
47	Doha	United Arab Emirates	Doha	Qatar
55	Guatemala City	Guatemala	Guatemala	
58	Ho Chi Minh City	Viet Nam	Ho Chi Minh City	Vietnam
59	Honolulu, HI	USA	Honolulu	
70	Kinshasa	Democratic Republic of Congo	Kinshasa	Congo Democratic Republic
71	Kolkata	India	Calcutta	
73	Kuwait City	Kuwait	al-Kuwayt	
78	London	United Kingdom	London	UK
79	Los Angeles, CA	USA	Los Angeles	
90	Mumbai	India	Bombay	
94	New York, NY	USA	New York	
97	Panama City	Panama	Panama	
99	Phnom Penh	Cambodia	Phnum Penh	
100	Phoenix, AZ	USA	Phoenix	
101	Port-of-Spain	Trinidad and Tobago	Port of Spain	
103	Quezon City	Philippines	Quezon	
110	Sana'a	Yemen	San'a	
114	Sarajevo	Bosnia	Sarajevo	Bosnia and Herzegovina
115	Seattle, WA	USA	Seattle	
116	Seoul	South Korea	Soul	Korea South
121	St. John's	Canada	Saint John's	
122	St. Petersburg	Russia	Saint Petersburg	
129	Tel Aviv	Israel	Tel Aviv-Yafo	
132	Tripoli	Algeria	Tripoli	Libya
135	Vancouver, BC	Canada	Vancouver	
136	Victoria, BC	Canada	Victoria	
141	Yangon	Burma	Rangoon	Myanmar

Table C.3: Latitude and longitudes for UrbMet cities from two sources: the first, from the `world.cities` dataset from the `maps` library; and the second, from the Google Maps API (queried using the `geocode` function from the `ggmap` library in `R`.

	`UrbMet`		`world.cities`			`WebWIMP`			
c_id	**City**	$\mathbf{N_0}$	$\mathbf{N_1}$	Φ_1	Λ_1	Φ_2	Λ_2	δ_Φ	δ_Λ
1	Abu Dhabi	527000	619316	24.48	54.37	24.45	54.38	0.0	0.0
2	Abuja	107069	178462	9.18	7.17	9.08	7.40	0.0	0.0
3	Accra	1658937	2029143	5.56	-0.20	5.60	-0.19	0.0	-0.1
4	Addis Ababa	2646000	2823167	9.03	38.74	8.98	38.76	0.0	0.0

Continued on next page

Table C.3: *Continued*

	UrbMet		world.cities			WebWIMP			
c_id	**City**	$\mathbf{N_0}$	$\mathbf{N_1}$	Φ_1	Λ_1	Φ_2	Λ_2	δ_Φ	δ_Λ
5	Amman	1204110	1303197	31.95	35.93	31.95	35.93	0.0	0.0
6	Amsterdam	742884	744159	52.37	4.89	52.37	4.90	0.0	0.0
7	Anchorage, AK	279671	279428	61.18	-149.19	61.22	-149.90	0.0	0.0
8	Ankara	3517182	3579706	39.93	32.85	39.93	32.86	0.0	0.0
9	Asuncion	513399	507574	-25.30	-57.63	-25.26	-57.58	0.0	0.0
10	Athens	789166	725049	37.98	23.73	37.98	23.73	0.0	0.0
11	Bandar seri begawan	27285	67077	4.93	114.95	4.90	114.94	0.0	0.0
12	Bangalore	5104047	5104047	12.97	77.56	12.97	77.59	0.0	0.0
13	Bangkok	5658953	4935988	13.73	100.50	13.76	100.50	0.0	0.0
14	Barcelona	1605602	1591485	41.40	2.17	41.39	2.17	0.0	0.0
15	Beijing	11509595	7602069	39.93	116.40	39.90	116.41	0.0	0.0
16	Beirut	361366	1273440	33.88	35.50	33.89	35.50	0.0	0.0
17	Belo Horizonte	2399920	2399577	-19.92	-43.94	-19.92	-43.94	0.0	0.0
18	Berlin	3386667	3378275	52.52	13.38	52.52	13.40	0.0	0.0
19	Bern	122256	120596	46.95	7.44	46.95	7.45	0.0	0.0
20	Bishkek	798300	915625	42.87	74.57	42.87	74.57	0.0	0.0
21	Bogota	7102602	7235084	4.63	-74.09	4.71	-74.07	0.0	0.0
22	Boston, MA	599351	567759	42.34	-71.02	42.36	-71.06	0.0	0.0
23	Brasilia	2383784	2260541	-15.78	-47.91	-15.79	-47.88	0.0	0.0
24	Brussels	144784	1031925	50.83	4.33	50.85	4.35	0.0	0.0
25	Bucharest	1931838	1862930	44.44	26.10	44.43	26.10	0.0	0.0
26	Budapest	1699212	1700019	47.51	19.08	47.50	19.04	0.0	0.0
27	Buenos Aires	2965403	11595183	-34.61	-58.37	-34.60	-58.38	0.0	0.0
28	Cairo	6758581	7836243	30.06	31.25	30.04	31.24	0.0	0.0
29	Cali	2392877	2445713	3.44	-76.52	3.45	-76.53	0.0	0.0
30	Cape Town	3497097	3546429	-33.93	18.46	-33.92	18.42	0.0	0.0
31	Caracas	1975294	1808937	10.54	-66.93	10.48	-66.90	0.0	0.0
32	Casablanca	2995000	3177281	33.60	-7.62	33.57	-7.59	0.0	0.0
33	Cebu	718821	814296	10.32	123.90	10.61	123.89	0.0	0.0
34	Chennai	4343645	4352932	13.09	80.27	13.08	80.27	0.0	0.0
35	Chicago, IL	2836658	2830144	41.84	-87.68	41.88	-87.63	0.0	0.0
36	Chisinau	660726	623671	47.03	28.83	47.01	28.86	0.0	0.0
37	Colombo	615000	649496	6.93	79.85	6.93	79.86	0.0	0.0
38	Copenhagen	505141	1091978	55.68	12.57	55.68	12.57	0.0	0.0
39	Curitiba	1788559	1746484	-25.42	-49.29	-25.43	-49.27	0.0	0.0
40	Dakar	1075582	2406598	14.72	-17.48	14.76	-17.37	0.0	0.0
41	Damascus	1658000	1580909	33.50	36.32	33.51	36.28	0.0	0.0
42	Dar es Salaam	1360850	2805523	-6.82	39.28	-6.79	39.21	0.0	0.0
43	Delhi	9879172	11215130	28.67	77.21	28.70	77.10	0.0	0.0
44	Denver, CO	588349	556575	39.77	-104.87	39.74	-104.99	0.0	0.0
45	Detroit, MI	916952	871789	42.38	-83.10	42.33	-83.05	0.0	0.0
46	Dhaka	10356500	6724976	23.70	90.39	23.81	90.41	0.0	0.0
47	Doha	344939	351381	25.30	51.51	25.29	55.39	0.0	0.1
48	Dubai	1089000	1182439	25.27	55.33	25.20	55.27	0.0	0.0
49	Dublin	495781	1030431	53.33	-6.25	53.35	-6.26	0.0	0.0
50	Durban	669242	3244028	-29.87	30.99	-29.86	31.02	0.0	0.0
51	Florianopolis	406564	431029	-27.60	-48.54	-27.60	-48.55	0.0	0.0
52	Geneva	178574	179426	46.21	6.14	46.20	6.14	0.0	0.0

Continued on next page

Table C.3: *Continued*

	UrbMet		world.cities			WebWIMP			
c_id	**City**	**N_0**	**N_1**	Φ_1	Λ_1	Φ_2	Λ_2	δ_Φ	δ_Λ
53	Guadalajara	1640589	1637213	20.67	-103.35	20.66	-103.35	0.0	0.0
54	Guangzhou	8524826	3158125	23.12	113.25	23.13	113.26	0.0	0.0
55	Guatemala City	1022001	1010253	14.63	-90.55	14.63	-90.51	0.0	0.0
56	Hamburg	1704735	1743891	53.55	10.00	53.55	9.99	0.0	0.0
57	Helsinki	566526	558341	60.17	24.94	60.17	24.94	0.0	0.0
58	Ho Chi Minh City	3015743	3496586	10.78	106.69	10.82	106.63	0.0	0.0
59	Honolulu, HI	375571	386345	21.32	-157.80	21.31	-157.86	0.0	0.0
60	Hyderabad	3637483	3632094	17.40	78.48	17.39	78.49	0.0	0.0
61	Islamabad	529180	794431	33.72	73.06	33.73	73.09	0.0	0.0
62	Istanbul	11174257	10034830	41.10	29.00	41.01	28.98	0.0	0.0
63	Jakarta	8820603	8556798	-6.18	106.83	-6.17	106.82	0.0	0.0
64	Jerusalem	740475	731731	31.78	35.22	31.77	35.21	0.0	0.0
65	Johannesburg	752349	2091491	-26.19	28.04	-26.20	28.05	0.0	0.0
66	Karachi	9339023	11969284	24.86	67.01	24.86	67.01	0.0	0.0
67	Kathmandu	671846	822930	27.71	85.31	27.72	85.32	0.0	0.0
68	Kiev	2676789	2491404	50.43	30.52	50.45	30.52	0.0	0.0
69	Kingston	579137	585300	17.99	-76.80	18.02	-76.81	0.0	0.0
70	Kinshasa	7785965	8096254	-4.31	15.32	-4.44	15.27	0.0	0.0
71	Kolkata	4572876	4638350	22.57	88.36	22.57	88.36	0.0	0.0
72	Kuala Lumpur	1551306	1482359	3.16	101.71	3.14	101.69	0.0	0.0
73	Kuwait City	32403	63596	29.38	47.99	29.38	47.98	0.0	0.0
74	Lagos	5195247	9020089	6.45	3.47	6.52	3.38	0.0	0.0
75	Lima	8445211	7857121	-12.07	-77.05	-12.27	-76.27	0.0	0.0
76	Lisbon	504726	508209	38.72	-9.14	38.72	-9.14	0.0	0.0
77	Ljubljana	250953	254188	46.06	14.51	46.06	14.51	0.0	0.0
78	London	7421209	7489022	51.52	-0.10	51.51	-0.13	0.0	0.3
79	Los Angeles, CA	3834340	3911500	34.11	-118.41	34.05	-118.24	0.0	0.0
80	Madrid	3128600	3146804	40.42	-3.71	40.42	-3.70	0.0	0.0
81	Manama	176909	147894	26.21	50.58	26.23	50.59	0.0	0.0
82	Manila	1581082	10546511	14.62	120.97	14.60	120.98	0.0	0.0
83	Melbourne	3806092	3780871	-37.81	144.96	-37.81	144.96	0.0	0.0
84	Mexico City	11285654	8659409	19.43	-99.14	19.43	-99.13	0.0	0.0
85	Milan	1306086	1316218	45.48	9.19	45.47	9.19	0.0	0.0
86	Minsk	1789098	1747482	53.91	27.55	53.90	27.56	0.0	0.0
87	Montevideo	1345010	1271664	-34.87	-56.17	-34.90	-56.16	0.0	0.0
88	Montreal	3268513	3280123	45.52	-73.57	45.50	-73.57	0.0	0.0
89	Moscow	10456490	10472629	55.75	37.62	55.76	37.62	0.0	0.0
90	Mumbai	11978450	12883645	18.96	72.82	19.08	72.88	0.0	0.0
91	Nagoya	2215062	2194748	35.15	136.91	35.18	136.91	0.0	0.0
92	Naihati	215303	263628	22.91	88.43	22.89	88.42	0.0	0.0
93	Nairobi	2948109	2864667	-1.29	36.82	-1.29	36.82	0.0	0.0
94	New York, NY	8274527	8124427	40.67	-73.94	40.71	-74.01	0.0	0.0
95	Osaka	2628811	2590815	34.68	135.50	34.69	135.50	0.0	0.0
96	Ottawa	812129	885542	45.42	-75.71	45.42	-75.70	0.0	0.0
97	Panama City	484261	406070	8.97	-79.53	8.98	-79.52	0.0	0.0
98	Paris	2125017	2141839	48.86	2.34	48.86	2.35	0.0	0.0
99	Phnom Penh	703963	1673131	11.57	104.92	11.54	104.89	0.0	0.0
100	Phoenix, AZ	1552259	1450884	33.54	-112.07	33.45	-112.07	0.0	0.0

Continued on next page

Table C.3: *Continued*

	UrbMet		world.cities			WebWIMP			
c_id	**City**	**N_0**	**N_1**	Φ_1	Λ_1	Φ_2	Λ_2	δ_Φ	δ_Λ
101	Port-of-Spain	43396	49764	10.66	-61.51	10.65	-61.50	0.0	0.0
102	Prague	1188126	1168374	50.08	14.43	50.08	14.44	0.0	0.0
103	Quezon City	2173831	11800	10.43	123.28	14.68	121.04	0.4	0.0
104	Quito	1559295	1399814	-0.19	-78.50	-0.18	-78.47	-0.1	0.0
105	Rabat	642000	1688738	34.02	-6.84	33.97	-6.85	0.0	0.0
106	Riga	719928	738386	56.97	24.13	56.95	24.11	0.0	0.0
107	Riyadh	4087152	4328067	24.65	46.77	24.71	46.68	0.0	0.0
108	Rome	2626640	2561181	41.89	12.50	41.90	12.50	0.0	0.0
109	San Salvador	507665	534409	13.69	-89.19	13.69	-89.22	0.0	0.0
110	Sana'a	954448	1921589	15.38	44.21	15.37	44.19	0.0	0.0
111	Santiago	4960815	4893495	-33.46	-70.64	-33.45	-70.67	0.0	0.0
112	Santo Domingo	913540	2253437	18.48	-69.91	18.49	-69.93	0.0	0.0
113	Sao Paulo	11016703	10059502	-23.53	-46.63	-23.55	-46.63	0.0	0.0
114	Sarajevo	527049	737350	43.85	18.38	43.86	18.41	0.0	0.0
115	Seattle, WA	594210	570430	47.62	-122.35	47.61	-122.33	0.0	0.0
116	Seoul	10020123	10409345	37.56	126.99	37.57	126.98	0.0	0.0
117	Shanghai	14348535	15017783	31.23	121.47	31.23	121.47	0.0	0.0
118	Shenzhen	7008831	1072093	22.53	114.13	22.54	114.06	0.0	0.0
119	Singapore	4588600	3601745	1.30	103.85	1.35	103.82	0.0	0.0
120	Sofia	1155403	1166143	42.69	23.31	42.70	23.32	0.0	0.0
121	St. John's	99182	109555	47.58	-52.69	47.56	-52.71	0.0	0.0
122	St. Petersburg	4569616	4014710	59.93	30.32	59.93	30.34	0.0	0.0
123	Stockholm	789024	1260712	59.33	18.07	59.33	18.07	0.0	0.0
124	Surabaya	2611506	2366850	-7.24	112.74	-7.26	112.75	0.0	0.0
125	Sydney	4336374	4444513	-33.87	151.21	-33.87	151.21	0.0	0.0
126	Tashkent	2137218	1967879	41.31	69.30	41.30	69.24	0.0	0.0
127	Tbilisi	1108600	1038343	41.72	44.79	41.72	44.83	0.0	0.0
128	Tehran	7088287	7160094	35.67	51.43	35.69	51.39	0.0	0.0
129	Tel Aviv	387234	384276	32.07	34.77	32.09	34.78	0.0	0.0
130	Tokyo	8489653	8372440	35.67	139.77	35.69	139.69	0.0	0.0
131	Toronto	4612191	4670783	43.65	-79.38	43.65	-79.38	0.0	0.0
132	Tripoli	1500000	1164634	32.87	13.18	36.74	3.10	0.1	-0.8
133	Tunis	702330	693294	36.84	10.22	36.81	10.18	0.0	0.0
134	Ulaanbaatar	1012733	862842	47.93	106.91	47.89	106.91	0.0	0.0
135	Vancouver, BC	1837969	1839314	49.28	-123.13	49.28	-123.12	0.0	0.0
136	Victoria, BC	289625	289837	48.43	-123.37	48.43	-123.37	0.0	0.0
137	Vienna	1664146	1570976	48.22	16.37	48.21	16.37	0.0	0.0
138	Vilnius	543494	542014	54.70	25.27	54.69	25.28	0.0	0.0
139	Vladivostok	579811	584496	43.13	131.90	43.12	131.89	0.0	0.0
140	Warsaw	1704717	1634441	52.26	21.02	52.23	21.01	0.0	0.0
141	Yangon	4477638	4572948	16.79	96.15	16.87	96.20	0.0	0.0
142	Yokohama	3579628	3603710	35.47	139.62	35.44	139.64	0.0	0.0

Table C.4: Latitude and longitude for UrbMet cities, adjusted for WebWIMP: comparison of original values.

cs_id	City	Φ_1	Λ_1	$\Phi_1\prime$	$\Lambda_1\prime$	$\Delta\Phi$	$\Delta\Lambda$	$d(\Phi, \Lambda)$	$\Phi\prime$	$\Lambda\prime$
1	Abu Dhabi	24.48	54.37	24.5	54.5	0.0	0.0	0.0	24.5	54.5
2	Abuja	9.18	7.17	9.0	7.0	0.0	0.0	0.0	9.0	7.0
3	Accra	5.50	-1.50	5.5	-1.5	0.0	0.0	0.0	5.5	-1.5
4	Addis Ababa	9.03	38.74	9.0	38.5	0.0	0.0	0.0	9.0	38.5
5	Amman	31.95	35.93	32.0	36.0	0.0	0.0	0.0	32.0	36.0
6	Amsterdam	52.37	4.89	52.5	5.0	0.0	0.0	0.0	52.5	5.0
7	Anchorage, AK	61.18	-149.19	61.0	-149.0	0.0	0.0	0.0	61.0	-149.0
8	Ankara	39.93	32.85	40.0	33.0	0.0	0.0	0.0	40.0	33.0
9	Asuncion	-25.30	-57.63	-25.5	-57.5	0.0	0.0	0.0	-25.5	-57.5
10	Athens	37.98	23.73	38.0	23.5	0.5	0.0	0.5	38.5	23.5
11	Bandar seri begawan	4.93	114.95	5.0	115.0	0.0	0.0	0.0	5.0	115.0
12	Bangalore	12.97	77.56	13.0	77.5	0.0	0.0	0.0	13.0	77.5
13	Bangkok	13.73	100.50	13.5	100.5	0.0	0.5	0.5	13.5	101.0
14	Barcelona	41.40	2.17	41.5	2.0	0.0	-0.5	0.5	41.5	1.5
15	Beijing	39.93	116.40	40.0	116.5	0.0	0.0	0.0	40.0	116.5
16	Beirut	33.88	35.50	34.0	35.5	0.0	0.0	0.0	34.0	35.5
17	Belo Horizonte	-19.92	-43.94	-20.0	-44.0	0.0	0.0	0.0	-20.0	-44.0
18	Berlin	52.52	13.38	52.5	13.5	0.0	0.0	0.0	52.5	13.5
19	Bern	46.95	7.44	47.0	7.5	0.0	0.0	0.0	47.0	7.5
20	Bishkek	42.87	74.57	43.0	74.5	0.0	0.0	0.0	43.0	74.5
21	Bogota	4.63	-74.09	4.5	-74.0	0.0	0.0	0.0	4.5	-74.0
22	Boston, MA	42.34	-71.02	42.5	-71.0	-0.5	0.0	0.5	42.0	-71.0
23	Brasilia	-15.78	-47.91	-16.0	-48.0	0.0	0.0	0.0	-16.0	-48.0
24	Brussels	50.83	4.33	51.0	4.5	0.0	0.0	0.0	51.0	4.5
25	Bucharest	44.44	26.10	44.5	26.0	0.0	0.0	0.0	44.5	26.0
26	Budapest	47.51	19.08	47.5	19.0	0.0	0.0	0.0	47.5	19.0
27	Buenos Aires	-34.61	-58.37	-34.5	-58.5	0.0	0.0	0.0	-34.5	-58.5
28	Cairo	30.06	31.25	30.0	31.5	0.0	0.0	0.0	30.0	31.5
29	Cali	3.44	-76.52	3.5	-76.5	0.0	0.0	0.0	3.5	-76.5
30	Cape Town	-33.93	18.46	-34.0	18.5	0.0	0.5	0.5	-34.0	19.0
31	Caracas	10.54	-66.93	10.5	-67.0	0.0	0.0	0.0	10.5	-67.0
32	Casablanca	33.60	-7.62	33.5	-7.5	0.0	0.0	0.0	33.5	-7.5
33	Cebu	10.32	123.90	10.5	124.0	-0.5	0.0	0.5	10.0	124.0
34	Chennai	13.09	80.27	13.0	80.5	0.5	-0.5	0.7	13.5	80.0
35	Chicago, IL	41.84	-87.68	42.0	-87.5	0.0	0.0	0.0	42.0	-87.5
36	Chisinau	47.03	28.83	47.0	29.0	0.0	0.0	0.0	47.0	29.0
37	Colombo	6.93	79.85	7.0	80.0	0.0	0.0	0.0	7.0	80.0
38	Copenhagen	55.68	12.57	55.5	12.5	0.0	-0.5	0.5	55.5	12.0
39	Curitiba	-25.42	-49.29	-25.5	-49.5	0.0	0.0	0.0	-25.5	-49.5
40	Dakar	14.72	-17.48	14.5	-17.5	0.0	0.5	0.5	14.5	-17.0
41	Damascus	33.50	36.32	33.5	36.5	0.0	0.0	0.0	33.5	36.5
42	Dar es Salaam	-6.82	39.28	-7.0	39.5	0.0	-0.5	0.5	-7.0	39.0

Continued on the next page

Table C.4: *Cont.*

cs_id	City	Φ_1	Λ_1	$\Phi_1\prime$	$\Lambda_1\prime$	$\Delta\Phi$	$\Delta\Lambda$	$d(\Phi,\Lambda)$	$\Phi\prime$	$\Lambda\prime$
43	Delhi	28.67	77.21	28.5	77.0	0.0	0.0	0.0	28.5	77.0
44	Denver, CO	39.77	-104.87	40.0	-105.0	0.0	0.0	0.0	40.0	-105.0
45	Detroit, MI	42.38	-83.10	42.5	-83.0	0.0	0.0	0.0	42.5	-83.0
46	Dhaka	23.70	90.39	23.5	90.5	0.0	0.0	0.0	23.5	90.5
47	Doha	25.30	51.51	25.5	51.5	0.0	-0.5	0.5	25.5	51.0
48	Dubai	25.27	55.33	25.5	55.5	0.0	0.0	0.0	25.5	55.5
49	Dublin	53.33	-6.25	53.5	-6.5	0.0	0.0	0.0	53.5	-6.5
50	Durban	-29.87	30.99	-30.0	31.0	0.0	-0.5	0.5	-30.0	30.5
51	Florianopolis	-27.60	-48.54	-27.5	-48.5	0.0	-0.5	0.5	-27.5	-49.0
52	Geneva	46.21	6.14	46.0	6.0	0.0	0.0	0.0	46.0	6.0
53	Guadalajara	20.67	-103.35	20.5	-103.5	0.0	0.0	0.0	20.5	-103.5
54	Guangzhou	23.12	113.25	23.0	113.5	0.0	-0.5	0.5	23.0	113.0
55	Guatemala City	14.63	-90.55	14.5	-90.5	0.0	0.0	0.0	14.5	-90.5
56	Hamburg	53.55	10.00	53.5	10.0	0.0	0.0	0.0	53.5	10.0
57	Hanover	52.50	10.00	52.5	10.0	0.0	0.0	0.0	52.5	10.0
58	Helsinki	60.17	24.94	60.0	25.0	-0.5	0.0	0.5	59.5	25.0
59	Ho Chi Minh City	10.78	106.69	11.0	106.5	0.0	0.0	0.0	11.0	106.5
60	Hyderabad	17.40	78.48	17.5	78.5	0.0	0.0	0.0	17.5	78.5
61	Islamabad	33.72	73.06	33.5	73.0	0.0	0.0	0.0	33.5	73.0
62	Istanbul	41.10	29.00	41.0	29.0	-0.5	0.0	0.5	40.5	29.0
63	Jakarta	-6.18	106.83	-6.0	107.0	0.0	0.0	0.0	-6.0	107.0
64	Jerusalem	31.78	35.22	32.0	35.0	0.0	0.0	0.0	32.0	35.0
65	Johannesburg	-26.19	28.04	-26.0	28.0	0.0	0.0	0.0	-26.0	28.0
66	Karachi	24.86	67.01	25.0	67.0	0.0	0.5	0.5	25.0	67.5
67	Kathmandu	27.71	85.31	27.5	85.5	0.0	0.0	0.0	27.5	85.5
68	Kiev	50.43	30.52	50.5	30.5	0.0	0.0	0.0	50.5	30.5
69	Kingston	17.99	-76.80	18.0	-77.0	0.5	0.0	0.5	18.5	-77.0
70	Kinshasa	-4.31	15.32	-4.5	15.5	0.0	0.0	0.0	-4.5	15.5
71	Kolkata	22.57	88.36	22.5	88.5	0.0	0.0	0.0	22.5	88.5
72	Kuala Lumpur	3.16	101.71	3.0	101.5	0.0	0.0	0.0	3.0	101.5
73	Kuwait City	29.38	47.99	29.5	48.0	-0.5	0.0	0.5	29.0	48.0
74	Lagos	6.45	3.47	6.5	3.5	0.5	0.0	0.5	7.0	3.5
75	Lima	-12.07	-77.05	-12.0	-77.0	0.0	0.0	0.0	-12.0	-77.0
76	Lisbon	38.72	-9.14	38.5	-9.0	0.0	0.0	0.0	38.5	-9.0
77	Ljubljana	46.06	14.51	46.0	14.5	0.0	0.0	0.0	46.0	14.5
78	London	51.52	-0.10	51.5	-0.0	0.0	0.0	0.0	51.5	-0.0
79	Los Angeles, CA	34.11	-118.41	34.0	-118.5	0.0	0.0	0.0	34.0	-118.5
80	Madrid	40.42	-3.71	40.5	-3.5	0.0	0.0	0.0	40.5	-3.5
81	Manama	26.21	50.58	26.0	50.5	0.0	0.5	0.5	26.0	51.0
82	Manila	14.62	120.97	14.5	121.0	0.0	0.0	0.0	14.5	121.0
83	Melbourne	-37.81	144.96	-38.0	145.0	0.0	0.5	0.5	-38.0	145.5
84	Mexico City	19.43	-99.14	19.5	-99.0	0.0	0.0	0.0	19.5	-99.0
85	Milan	45.48	9.19	45.5	9.0	0.0	0.0	0.0	45.5	9.0

Continued on the next page

Table C.4: *Cont.*

cs_id	City	Φ_1	Λ_1	$\Phi_1\prime$	$\Lambda_1\prime$	$\Delta\Phi$	$\Delta\Lambda$	$d(\Phi,\Lambda)$	$\Phi\prime$	$\Lambda\prime$
86	Minsk	53.91	27.55	54.0	27.5	0.0	0.0	0.0	54.0	27.5
87	Montevideo	-34.87	-56.17	-35.0	-56.0	0.5	0.0	0.5	-34.5	-56.0
88	Montreal	45.52	-73.57	45.5	-73.5	0.0	0.0	0.0	45.5	-73.5
89	Moscow	55.75	37.62	56.0	37.5	0.0	0.0	0.0	56.0	37.5
90	Mumbai	18.96	72.82	19.0	73.0	0.0	0.0	0.0	19.0	73.0
91	Nagoya	35.15	136.91	35.0	137.0	0.0	0.5	0.5	35.0	137.5
92	Naihati	22.91	88.43	23.0	88.5	0.0	0.0	0.0	23.0	88.5
93	Nairobi	-1.29	36.82	-1.5	37.0	0.0	0.0	0.0	-1.5	37.0
94	New York, NY	40.67	-73.94	40.5	-74.0	0.0	-0.5	0.5	40.5	-74.5
95	Osaka	34.68	135.50	34.5	135.5	0.0	0.0	0.0	34.5	135.5
96	Ottawa	45.42	-75.71	45.5	-75.5	0.0	0.0	0.0	45.5	-75.5
97	Panama City	8.97	-79.53	9.0	-79.5	0.5	0.0	0.5	9.5	-79.5
98	Paris	48.86	2.34	49.0	2.5	0.0	0.0	0.0	49.0	2.5
99	Phnom Penh	11.57	104.92	11.5	105.0	0.0	0.0	0.0	11.5	105.0
100	Phoenix, AZ	33.54	-112.07	33.5	-112.0	0.0	0.0	0.0	33.5	-112.0
101	Port-of-Spain	10.66	-61.51	10.5	-61.5	0.0	0.0	0.0	10.5	-61.5
102	Portland, OR	45.50	-122.80	45.5	-123.0	0.0	0.0	0.0	45.5	-123.0
103	Prague	50.08	14.43	50.0	14.5	0.0	0.0	0.0	50.0	14.5
104	Quezon City	10.43	123.28	10.5	123.5	0.0	0.0	0.0	10.5	123.5
105	Quito	-0.19	-78.50	-0.0	-78.5	0.0	0.0	0.0	-0.0	-78.5
106	Rabat	34.02	-6.84	34.0	-7.0	0.0	0.0	0.0	34.0	-7.0
107	Riga	56.97	24.13	57.0	24.0	0.0	0.0	0.0	57.0	24.0
108	Riyadh	24.65	46.77	24.5	47.0	0.0	0.0	0.0	24.5	47.0
109	Rome	41.89	12.50	42.0	12.5	0.0	0.0	0.0	42.0	12.5
110	San Salvador	13.69	-89.19	13.5	-89.0	0.0	0.0	0.0	13.5	-89.0
111	Sana'a	15.38	44.21	15.5	44.0	0.0	0.0	0.0	15.5	44.0
112	Santiago	-33.46	-70.64	-33.5	-70.5	0.0	0.0	0.0	-33.5	-70.5
113	Santo Domingo	18.48	-69.91	18.5	-70.0	0.0	-0.5	0.5	18.5	-70.5
114	Sao Paulo	-23.53	-46.63	-23.5	-46.5	0.0	0.0	0.0	-23.5	-46.5
115	Sarajevo	43.85	18.38	44.0	18.5	0.0	0.0	0.0	44.0	18.5
116	Seattle, WA	47.62	-122.35	47.5	-122.5	0.0	0.0	0.0	47.5	-122.5
117	Seoul	37.56	126.99	37.5	127.0	0.0	0.0	0.0	37.5	127.0
118	Shanghai	31.23	121.47	31.0	121.5	0.0	-0.5	0.5	31.0	121.0
119	Shenzhen	22.53	114.13	22.5	114.0	0.5	0.0	0.5	23.0	114.0
120	Singapore	1.30	103.85	1.5	104.0	-0.5	0.0	0.5	1.0	104.0
121	Sofia	42.69	23.31	42.5	23.5	0.0	0.0	0.0	42.5	23.5
122	St. John's	47.58	-52.69	47.5	-52.5	0.5	-0.5	0.7	48.0	-53.0
123	St. Petersburg	59.93	30.32	60.0	30.5	0.0	0.0	0.0	60.0	30.5
124	Stockholm	59.33	18.07	59.5	18.0	0.0	0.0	0.0	59.5	18.0
125	Surabaya	-7.24	112.74	-7.0	112.5	0.0	0.0	0.0	-7.0	112.5
126	Sydney	-33.87	151.21	-34.0	151.0	0.0	-0.5	0.5	-34.0	150.5
127	Tashkent	41.31	69.30	41.5	69.5	0.0	0.0	0.0	41.5	69.5
128	Tbilisi	41.72	44.79	41.5	45.0	0.0	0.0	0.0	41.5	45.0

Continued on the next page

Table C.4: *Cont.*

cs_id	City	Φ_1	Λ_1	$\Phi_1\prime$	$\Lambda_1\prime$	$\Delta\Phi$	$\Delta\Lambda$	$d(\Phi,\Lambda)$	$\Phi\prime$	$\Lambda\prime$
129	Tehran	35.67	51.43	35.5	51.5	0.0	0.0	0.0	35.5	51.5
130	Tel Aviv	32.07	34.77	32.0	35.0	0.0	0.0	0.0	32.0	35.0
131	Tokyo	35.67	139.77	35.5	140.0	0.0	0.0	0.0	35.5	140.0
132	Toronto	43.65	-79.38	43.5	-79.5	0.0	0.0	0.0	43.5	-79.5
133	Tripoli	32.87	13.18	33.0	13.0	0.0	0.0	0.0	33.0	13.0
134	Tunis	36.84	10.22	37.0	10.0	0.0	0.0	0.0	37.0	10.0
135	Ulaanbaatar	47.93	106.91	48.0	107.0	0.0	0.0	0.0	48.0	107.0
136	Vancouver, BC	49.28	-123.13	49.5	-123.0	0.0	0.0	0.0	49.5	-123.0
137	Victoria, BC	48.43	-123.37	48.5	-123.5	-0.5	0.0	0.5	48.0	-123.5
138	Vienna	48.22	16.37	48.0	16.5	0.0	0.0	0.0	48.0	16.5
139	Vilnius	54.70	25.27	54.5	25.5	0.0	0.0	0.0	54.5	25.5
140	Vladivostok	43.13	131.90	43.0	132.0	0.5	0.0	0.5	43.5	132.0
141	Warsaw	52.26	21.02	52.5	21.0	0.0	0.0	0.0	52.5	21.0
142	Yangon	16.79	96.15	17.0	96.0	0.0	0.0	0.0	17.0	96.0
143	Yokohama	35.47	139.62	35.5	139.5	0.0	0.5	0.5	35.5	140.0

C.1.4 Water Balance Data

Table C.5: WebWIMP monthly water balance data for UrbMet cities. Variable TEMP is in °C and variables UPE—SST are in units of *mm/month*. Variable definitions and descriptions provided in Table A.2.

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Abu Dhabi	Jan	19.8	38	35	35	0	2	2	33	2	0	0	0
Abu Dhabi	Feb	20.1	40	35	37	2	7	5	32	3	0	0	0
Abu Dhabi	Mar	23.2	74	76	21	-55	4	-3	24	52	0	0	0
Abu Dhabi	Apr	27.7	145	154	4	-150	1	-3	7	147	0	0	0
Abu Dhabi	May	31.5	171	196	1	-195	0	-1	2	194	0	0	0
Abu Dhabi	Jun	33.2	178	202	0	-202	0	0	0	202	0	0	0
Abu Dhabi	Jul	34.5	182	211	1	-210	0	0	1	210	0	0	0
Abu Dhabi	Aug	33.9	180	201	2	-199	0	0	1	200	0	0	0
Abu Dhabi	Sep	32.3	174	178	1	-177	0	0	1	177	0	0	0
Abu Dhabi	Oct	29.3	158	156	0	-156	0	0	0	156	0	0	0
Abu Dhabi	Nov	24.4	91	83	5	-78	0	0	5	78	0	0	0
Abu Dhabi	Dec	21.0	48	44	18	-26	0	0	18	26	0	0	0
Abuja	Jan	26.1	125	125	3	-122	4	-10	12	113	0	0	0
Abuja	Feb	28.1	149	137	7	-130	1	-3	10	127	0	0	0
Abuja	Mar	29.6	160	166	27	-139	0	-1	27	139	0	0	0
Abuja	Apr	29.6	160	164	76	-88	0	0	75	89	0	0	0
Abuja	May	27.7	145	157	147	-10	3	3	144	13	0	0	0
Abuja	Jun	25.8	120	126	166	40	42	39	126	0	0	0	0
Abuja	Jul	25.2	109	119	218	99	140	98	119	0	0	0	0
Abuja	Aug	24.9	105	112	241	129	150	10	112	0	119	0	0
Abuja	Sep	25.2	109	111	248	137	150	0	111	0	137	0	0
Abuja	Oct	26.0	123	126	122	-4	130	-20	126	0	16	0	0
Abuja	Nov	26.3	129	125	8	-117	43	-87	96	29	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Abuja	Dec	25.6	116	116	3	-113	14	-29	31	85	0	0	0
Accra	Jan	27.0	139	142	21	-121	2	-5	25	117	0	0	0
Accra	Feb	27.5	144	134	27	-107	1	-1	27	107	0	0	0
Accra	Mar	27.8	146	152	89	-63	0	-1	89	63	0	0	0
Accra	Apr	27.8	146	149	111	-38	0	0	110	39	0	0	0
Accra	May	27.4	143	152	198	46	45	45	152	0	0	0	0
Accra	Jun	25.9	122	126	318	192	150	105	126	0	87	0	0
Accra	Jul	24.9	105	112	108	-4	133	-17	112	0	13	0	0
Accra	Aug	24.6	101	106	39	-67	72	-61	100	6	0	0	0
Accra	Sep	25.4	113	115	58	-57	42	-30	88	27	0	0	0
Accra	Oct	26.2	127	131	126	-5	42	0	126	5	0	0	0
Accra	Nov	27.0	139	138	67	-71	21	-21	88	50	0	0	0
Accra	Dec	26.8	138	140	18	-122	7	-14	33	107	0	0	0
Addis Ababa	Jan	16.4	58	58	10	-48	31	-19	29	29	0	0	0
Addis Ababa	Feb	17.5	65	60	26	-34	23	-8	34	26	0	0	0
Addis Ababa	Mar	18.7	73	76	53	-23	18	-5	57	19	0	0	0
Addis Ababa	Apr	18.9	74	76	67	-9	17	-1	68	8	0	0	0
Addis Ababa	May	19.2	76	82	63	-19	14	-3	66	16	0	0	0
Addis Ababa	Jun	18.0	68	72	103	31	44	30	72	0	0	0	0
Addis Ababa	Jul	16.8	60	66	212	146	150	106	66	0	40	0	0
Addis Ababa	Aug	16.8	60	65	238	173	150	0	65	0	173	0	0
Addis Ababa	Sep	16.8	60	61	101	40	150	0	61	0	40	0	0
Addis Ababa	Oct	16.6	59	61	33	-28	123	-27	61	0	0	0	0
Addis Ababa	Nov	16.0	55	54	6	-48	79	-44	50	4	0	0	0
Addis Ababa	Dec	15.6	53	53	5	-48	50	-29	34	19	0	0	0
Amman	Jan	8.1	15	13	37	24	49	24	13	0	0	0	0
Amman	Feb	9.1	19	16	35	19	67	18	16	0	0	0	0
Amman	Mar	12.4	33	34	37	3	72	5	33	1	0	0	0
Amman	Apr	16.1	54	58	9	-49	45	-27	36	22	0	0	0
Amman	May	20.6	84	100	2	-98	18	-27	30	70	0	0	0
Amman	Jun	23.6	108	128	0	-128	5	-13	13	115	0	0	0
Amman	Jul	25.2	122	148	0	-148	1	-4	4	144	0	0	0
Amman	Aug	25.4	124	142	0	-142	0	-1	1	141	0	0	0
Amman	Sep	23.6	108	111	0	-111	0	0	0	111	0	0	0
Amman	Oct	20.4	83	81	5	-76	0	0	5	76	0	0	0
Amman	Nov	14.7	45	39	18	-21	0	0	18	21	0	0	0
Amman	Dec	9.7	21	18	44	26	25	25	18	0	0	0	0
Amsterdam	Jan	2.5	11	8	73	65	150	0	8	0	65	0	0
Amsterdam	Feb	2.0	8	7	52	45	150	0	7	0	45	0	0
Amsterdam	Mar	5.5	25	25	53	28	150	0	25	0	28	0	0
Amsterdam	Apr	8.3	39	44	48	4	149	-1	44	0	5	0	0
Amsterdam	May	12.6	60	80	54	-26	122	-27	80	0	0	0	0
Amsterdam	Jun	15.3	74	101	65	-36	87	-35	100	1	0	0	0
Amsterdam	Jul	17.1	84	115	78	-37	61	-26	104	11	0	0	0
Amsterdam	Aug	16.9	83	103	90	-13	54	-7	97	6	0	0	0
Amsterdam	Sep	14.4	70	72	77	5	59	5	72	0	0	0	0
Amsterdam	Oct	10.9	52	47	73	26	85	26	47	0	0	0	0
Amsterdam	Nov	6.6	30	22	80	58	142	57	22	0	0	0	0
Amsterdam	Dec	4.3	19	13	75	62	150	8	13	0	54	0	0
Anchorage, AK	Jan	-6.4	0	0	387	-3	150	0	0	0	0	0	1593

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Anchorage, AK	Feb	-6.1	0	0	267	-1	150	0	0	0	0	0	1861
Anchorage, AK	Mar	-4.9	0	0	207	0	150	0	0	0	0	0	2068
Anchorage, AK	Apr	-0.4	0	0	256	251	150	0	0	0	251	83	2073
Anchorage, AK	May	4.0	40	53	355	829	150	0	53	0	829	527	1546
Anchorage, AK	Jun	8.5	65	89	230	1041	150	0	89	0	1041	901	646
Anchorage, AK	Jul	10.7	76	104	186	728	150	0	104	0	728	646	0
Anchorage, AK	Aug	10.0	73	90	380	290	150	0	90	0	290	0	0
Anchorage, AK	Sep	6.0	52	54	442	388	150	0	54	0	388	0	0
Anchorage, AK	Oct	-0.3	0	0	519	310	150	0	0	0	310	0	209
Anchorage, AK	Nov	-4.6	0	0	467	1	150	0	0	0	0	0	675
Anchorage, AK	Dec	-5.6	0	0	527	-1	150	0	0	0	0	0	1203
Ankara	Jan	-0.6	0	0	43	43	102	43	0	0	0	0	0
Ankara	Feb	0.8	2	2	38	36	139	37	2	0	0	0	0
Ankara	Mar	4.7	16	17	38	21	150	11	17	0	10	0	0
Ankara	Apr	10.2	42	46	42	-4	145	-5	46	0	1	0	0
Ankara	May	14.7	65	81	54	-27	118	-27	81	0	0	0	0
Ankara	Jun	18.5	87	108	32	-76	57	-61	93	15	0	0	0
Ankara	Jul	21.6	105	133	14	-119	18	-39	53	80	0	0	0
Ankara	Aug	21.6	105	124	9	-115	6	-12	21	103	0	0	0
Ankara	Sep	17.5	81	83	14	-69	3	-3	17	66	0	0	0
Ankara	Oct	11.9	51	48	27	-21	2	-1	27	21	0	0	0
Ankara	Nov	6.1	22	18	33	15	17	15	18	0	0	0	0
Ankara	Dec	1.6	4	3	46	43	59	42	3	0	0	0	0
Asuncion	Jan	27.3	142	166	172	6	150	3	166	0	3	0	0
Asuncion	Feb	26.7	137	138	150	12	150	0	138	0	12	0	0
Asuncion	Mar	25.3	119	126	154	28	150	0	126	0	28	0	0
Asuncion	Apr	22.3	85	82	178	96	150	0	82	0	96	0	0
Asuncion	May	19.6	60	56	116	60	150	0	56	0	60	0	0
Asuncion	Jun	17.8	46	41	84	43	150	0	41	0	43	0	0
Asuncion	Jul	17.5	44	41	50	9	150	0	41	0	9	0	0
Asuncion	Aug	19.2	57	55	72	17	150	0	55	0	17	0	0
Asuncion	Sep	20.7	70	70	97	27	150	0	70	0	27	0	0
Asuncion	Oct	23.5	98	108	148	40	150	0	108	0	40	0	0
Asuncion	Nov	25.3	119	134	161	27	150	0	134	0	27	0	0
Asuncion	Dec	27.1	140	166	165	-1	147	-3	166	0	2	0	0
Athens	Jan	8.9	19	16	75	59	150	41	16	0	18	0	0
Athens	Feb	9.4	21	18	57	39	150	0	18	0	39	0	0
Athens	Mar	10.7	27	28	61	33	150	0	28	0	33	0	0
Athens	Apr	14.7	47	52	33	-19	131	-19	52	0	0	0	0
Athens	May	19.1	75	92	24	-68	71	-60	85	7	0	0	0
Athens	Jun	23.4	108	133	11	-122	22	-49	60	73	0	0	0
Athens	Jul	25.9	129	161	6	-155	5	-17	23	138	0	0	0
Athens	Aug	26.0	130	152	7	-145	1	-4	11	141	0	0	0
Athens	Sep	22.2	98	101	18	-83	1	0	18	83	0	0	0
Athens	Oct	17.7	66	63	61	-2	7	6	54	9	0	0	0
Athens	Nov	13.7	42	35	74	39	45	38	35	0	0	0	0
Athens	Dec	10.8	27	22	86	64	109	64	22	0	0	0	0
Bandar seri begawan	Jan	26.7	137	140	326	186	150	0	140	0	186	0	0
Bandar seri begawan	Feb	26.5	135	125	254	129	150	0	125	0	129	0	0
Bandar seri begawan	Mar	27.3	142	148	245	97	150	0	148	0	97	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Bandar seri begawan	Apr	27.6	145	147	319	172	150	0	147	0	172	0	0
Bandar seri begawan	May	27.6	145	154	352	198	150	0	154	0	198	0	0
Bandar seri begawan	Jun	27.5	144	149	257	108	150	0	149	0	108	0	0
Bandar seri begawan	Jul	27.2	141	151	282	131	150	0	151	0	131	0	0
Bandar seri begawan	Aug	27.2	141	149	295	146	150	0	149	0	146	0	0
Bandar seri begawan	Sep	27.1	140	142	365	223	150	0	142	0	223	0	0
Bandar seri begawan	Oct	26.9	139	143	393	250	150	0	143	0	250	0	0
Bandar seri begawan	Nov	26.8	138	136	404	268	150	0	136	0	268	0	0
Bandar seri begawan	Dec	26.8	138	140	382	242	150	0	140	0	242	0	0
Bangalore	Jan	21.2	70	69	3	-66	49	-43	46	23	0	0	0
Bangalore	Feb	23.3	92	84	5	-79	23	-26	31	53	0	0	0
Bangalore	Mar	25.9	126	131	9	-122	7	-16	24	107	0	0	0
Bangalore	Apr	27.5	144	149	38	-111	2	-5	42	107	0	0	0
Bangalore	May	27.2	141	155	108	-47	2	0	109	46	0	0	0
Bangalore	Jun	24.9	112	121	70	-51	1	-1	71	50	0	0	0
Bangalore	Jul	23.9	100	110	90	-20	1	0	89	21	0	0	0
Bangalore	Aug	23.7	97	105	108	3	7	6	101	4	0	0	0
Bangalore	Sep	23.9	100	101	176	75	81	74	101	0	0	0	0
Bangalore	Oct	23.4	93	95	163	68	150	69	95	0	0	0	0
Bangalore	Nov	22.1	79	76	61	-15	133	-17	76	0	2	0	0
Bangalore	Dec	20.8	66	64	22	-42	92	-41	64	0	0	0	0
Bangkok	Jan	25.9	118	116	10	-106	12	-21	31	85	0	0	0
Bangkok	Feb	27.4	143	130	22	-108	4	-8	29	101	0	0	0
Bangkok	Mar	28.6	153	158	43	-115	1	-3	45	113	0	0	0
Bangkok	Apr	29.2	157	163	74	-89	1	0	73	90	0	0	0
Bangkok	May	29.0	155	171	184	13	18	17	166	5	0	0	0
Bangkok	Jun	28.8	154	166	128	-38	12	-6	134	32	0	0	0
Bangkok	Jul	28.2	149	165	163	-2	14	2	161	4	0	0	0
Bangkok	Aug	28.0	148	160	171	11	26	12	158	2	0	0	0
Bangkok	Sep	27.6	145	147	291	144	150	124	147	0	20	0	0
Bangkok	Oct	27.1	140	143	236	93	150	0	143	0	93	0	0
Bangkok	Nov	26.5	135	129	64	-65	88	-62	128	1	0	0	0
Bangkok	Dec	25.5	110	108	6	-102	33	-55	61	47	0	0	0
Barcelona	Jan	9.2	24	20	38	18	92	18	20	0	0	0	0
Barcelona	Feb	10.0	28	23	29	6	98	6	23	0	0	0	0
Barcelona	Mar	11.6	35	36	46	10	108	10	36	0	0	0	0
Barcelona	Apr	13.3	44	49	53	4	112	4	49	0	0	0	0
Barcelona	May	16.7	64	79	54	-25	89	-23	78	1	0	0	0
Barcelona	Jun	20.4	88	111	42	-69	46	-43	85	26	0	0	0
Barcelona	Jul	23.5	110	141	23	-118	15	-31	54	87	0	0	0
Barcelona	Aug	23.7	112	133	50	-83	6	-9	58	75	0	0	0
Barcelona	Sep	21.0	92	95	83	-12	6	0	83	12	0	0	0
Barcelona	Oct	17.3	67	63	77	14	19	13	63	0	0	0	0
Barcelona	Nov	12.7	41	33	57	24	43	24	33	0	0	0	0
Barcelona	Dec	10.0	28	22	53	31	74	31	22	0	0	0	0
Beijing	Jan	-4.6	0	0	2	0	34	0	0	0	0	0	4
Beijing	Feb	-1.4	0	0	6	10	44	10	0	0	0	7	0
Beijing	Mar	5.2	13	13	6	-7	42	-2	8	5	0	0	0
Beijing	Apr	13.8	53	58	20	-38	29	-13	32	26	0	0	0
Beijing	May	19.8	88	110	31	-79	14	-15	46	64	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Beijing	Jun	23.9	116	145	77	-68	7	-7	82	63	0	0	0
Beijing	Jul	25.7	128	163	188	25	33	26	161	2	0	0	0
Beijing	Aug	24.6	121	142	170	28	62	29	142	0	0	0	0
Beijing	Sep	19.6	87	90	48	-42	42	-20	69	21	0	0	0
Beijing	Oct	12.7	47	44	20	-24	33	-9	29	15	0	0	0
Beijing	Nov	4.3	10	8	8	0	33	0	8	0	0	0	0
Beijing	Dec	-2.2	0	0	3	1	34	1	0	0	0	0	2
Beirut	Jan	0.8	3	3	266	263	150	0	3	0	263	0	0
Beirut	Feb	1.5	7	6	193	187	150	0	6	0	187	0	0
Beirut	Mar	3.7	17	17	202	185	150	0	17	0	185	0	0
Beirut	Apr	7.5	36	39	75	36	150	0	39	0	36	0	0
Beirut	May	11.2	54	65	33	-32	119	-31	65	0	0	0	0
Beirut	Jun	14.5	71	86	1	-85	53	-66	67	19	0	0	0
Beirut	Jul	16.4	81	99	1	-98	20	-33	33	66	0	0	0
Beirut	Aug	17.1	85	98	1	-97	8	-12	13	85	0	0	0
Beirut	Sep	15.1	74	76	2	-74	4	-4	6	70	0	0	0
Beirut	Oct	11.9	58	56	50	-6	12	8	41	15	0	0	0
Beirut	Nov	7.3	35	30	102	72	82	70	30	0	0	0	0
Beirut	Dec	3.1	14	12	209	197	150	68	12	0	129	0	0
Belo Horizonte	Jan	22.2	91	104	290	186	150	0	104	0	186	0	0
Belo Horizonte	Feb	22.3	92	91	211	120	150	0	91	0	120	0	0
Belo Horizonte	Mar	22.0	89	94	181	87	150	0	94	0	87	0	0
Belo Horizonte	Apr	20.8	79	77	86	9	148	-2	77	0	11	0	0
Belo Horizonte	May	19.1	66	63	36	-27	121	-27	63	0	0	0	0
Belo Horizonte	Jun	17.8	56	51	19	-32	90	-31	50	1	0	0	0
Belo Horizonte	Jul	17.2	52	50	16	-34	65	-25	41	9	0	0	0
Belo Horizonte	Aug	18.7	63	62	21	-41	44	-21	42	20	0	0	0
Belo Horizonte	Sep	20.0	73	73	57	-16	38	-6	62	11	0	0	0
Belo Horizonte	Oct	20.8	79	86	115	29	66	28	86	0	0	0	0
Belo Horizonte	Nov	21.1	82	89	225	136	150	84	89	0	52	0	0
Belo Horizonte	Dec	21.5	85	98	320	222	150	0	98	0	222	0	0
Berlin	Jan	-0.3	0	0	38	38	125	38	0	0	0	0	0
Berlin	Feb	0.4	2	1	33	32	150	25	1	0	7	0	0
Berlin	Mar	3.9	17	18	32	14	150	0	18	0	14	0	0
Berlin	Apr	8.5	40	46	39	-7	143	-7	46	0	0	0	0
Berlin	May	13.7	66	88	51	-37	107	-36	88	0	0	0	0
Berlin	Jun	16.9	83	113	60	-53	64	-43	102	11	0	0	0
Berlin	Jul	18.6	92	126	66	-60	36	-28	94	32	0	0	0
Berlin	Aug	17.9	88	110	59	-51	22	-14	73	37	0	0	0
Berlin	Sep	14.0	68	71	39	-32	16	-6	45	26	0	0	0
Berlin	Oct	9.2	43	39	37	-2	18	2	35	4	0	0	0
Berlin	Nov	4.1	18	14	43	29	47	29	14	0	0	0	0
Berlin	Dec	0.9	4	3	43	40	87	40	3	0	0	0	0
Bern	Jan	-1.5	0	0	103	0	150	0	0	0	0	0	117
Bern	Feb	-1.3	0	0	88	47	150	0	0	0	47	14	158
Bern	Mar	1.5	12	12	102	248	150	0	12	0	248	158	0
Bern	Apr	3.9	26	30	118	88	150	0	30	0	88	0	0
Bern	May	8.9	53	69	136	67	150	0	69	0	67	0	0
Bern	Jun	11.8	67	89	172	83	150	0	89	0	83	0	0
Bern	Jul	14.3	79	106	164	58	150	0	106	0	58	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Bern	Aug	13.8	77	94	147	53	150	0	94	0	53	0	0
Bern	Sep	10.9	63	65	114	49	150	0	65	0	49	0	0
Bern	Oct	6.9	43	40	101	61	150	0	40	0	61	0	0
Bern	Nov	2.1	16	12	95	83	150	0	12	0	83	0	0
Bern	Dec	-0.5	0	0	103	89	150	0	0	0	89	0	14
Bishkek	Jan	-6.6	0	0	21	0	38	0	0	0	0	0	59
Bishkek	Feb	-5.5	0	0	30	0	38	0	0	0	0	0	89
Bishkek	Mar	0.2	1	1	44	132	150	112	1	0	20	101	0
Bishkek	Apr	7.7	37	42	69	27	150	0	42	0	27	0	0
Bishkek	May	12.7	63	79	70	-9	140	-10	79	0	1	0	0
Bishkek	Jun	17.2	86	110	41	-69	76	-64	105	5	0	0	0
Bishkek	Jul	19.9	100	129	24	-105	28	-48	73	56	0	0	0
Bishkek	Aug	18.9	95	113	12	-101	10	-18	29	84	0	0	0
Bishkek	Sep	13.4	66	68	16	-52	6	-4	20	48	0	0	0
Bishkek	Oct	6.3	30	28	37	9	18	12	25	3	0	0	0
Bishkek	Nov	-0.8	0	0	36	20	38	20	0	0	0	0	16
Bishkek	Dec	-5.0	0	0	22	0	38	0	0	0	0	0	38
Bogota	Jan	17.4	66	67	39	-28	121	-29	67	0	1	0	0
Bogota	Feb	18.1	70	65	80	15	134	13	65	0	0	0	0
Bogota	Mar	18.0	70	72	149	77	150	16	72	0	61	0	0
Bogota	Apr	17.4	66	67	346	279	150	0	67	0	279	0	0
Bogota	May	16.8	62	66	422	356	150	0	66	0	356	0	0
Bogota	Jun	16.2	58	60	403	343	150	0	60	0	343	0	0
Bogota	Jul	15.9	56	60	351	291	150	0	60	0	291	0	0
Bogota	Aug	16.1	58	61	302	241	150	0	61	0	241	0	0
Bogota	Sep	16.4	59	60	264	204	150	0	60	0	204	0	0
Bogota	Oct	16.7	61	63	310	247	150	0	63	0	247	0	0
Bogota	Nov	16.5	60	60	266	206	150	0	60	0	206	0	0
Bogota	Dec	16.6	61	62	125	63	150	0	62	0	63	0	0
Boston, MA	Jan	-1.4	0	0	105	18	150	0	0	0	18	0	87
Boston, MA	Feb	-0.8	0	0	98	99	150	0	0	0	99	40	86
Boston, MA	Mar	3.1	10	10	114	190	150	0	10	0	190	86	0
Boston, MA	Apr	8.2	32	36	107	71	150	0	36	0	71	0	0
Boston, MA	May	13.8	61	76	96	20	150	0	76	0	20	0	0
Boston, MA	Jun	19.0	90	114	78	-36	115	-35	114	0	0	0	0
Boston, MA	Jul	22.5	110	142	78	-64	62	-53	130	12	0	0	0
Boston, MA	Aug	21.9	107	127	108	-19	52	-10	119	8	0	0	0
Boston, MA	Sep	17.7	82	85	95	10	62	10	85	0	0	0	0
Boston, MA	Oct	12.1	52	49	93	44	105	43	49	0	0	0	0
Boston, MA	Nov	6.9	26	21	118	97	150	45	21	0	52	0	0
Boston, MA	Dec	1.0	2	2	112	110	150	0	2	0	110	0	0
Brasilia	Jan	21.9	86	97	258	161	150	0	97	0	161	0	0
Brasilia	Feb	21.9	86	85	222	137	150	0	85	0	137	0	0
Brasilia	Mar	21.8	86	90	235	145	150	0	90	0	145	0	0
Brasilia	Apr	21.6	84	82	131	49	150	0	82	0	49	0	0
Brasilia	May	20.3	72	71	28	-43	108	-42	71	0	0	0	0
Brasilia	Jun	19.0	62	58	8	-50	67	-41	49	9	0	0	0
Brasilia	Jul	18.6	59	57	8	-49	42	-25	33	24	0	0	0
Brasilia	Aug	20.1	71	71	14	-57	24	-18	31	40	0	0	0
Brasilia	Sep	22.1	88	89	38	-51	15	-9	46	43	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Brasilia	Oct	22.3	90	97	159	62	75	60	97	0	0	0	0
Brasilia	Nov	21.8	86	92	244	152	150	75	92	0	77	0	0
Brasilia	Dec	21.4	82	92	318	226	150	0	92	0	226	0	0
Brussels	Jan	2.4	10	8	63	55	150	0	8	0	55	0	0
Brussels	Feb	2.8	12	10	51	41	150	0	10	0	41	0	0
Brussels	Mar	5.8	26	27	60	33	150	0	27	0	33	0	0
Brussels	Apr	8.6	40	46	52	6	149	-1	46	0	7	0	0
Brussels	May	12.8	62	82	62	-20	130	-19	82	0	0	0	0
Brussels	Jun	15.5	76	103	72	-31	99	-31	103	0	0	0	0
Brussels	Jul	17.3	85	117	76	-41	66	-33	108	9	0	0	0
Brussels	Aug	17.0	84	104	71	-33	49	-17	89	15	0	0	0
Brussels	Sep	14.0	68	71	61	-10	44	-5	65	6	0	0	0
Brussels	Oct	10.1	48	44	64	20	64	20	44	0	0	0	0
Brussels	Nov	5.4	25	18	71	53	117	53	18	0	0	0	0
Brussels	Dec	3.4	15	10	72	62	150	33	10	0	29	0	0
Bucharest	Jan	-1.7	0	0	33	0	72	0	0	0	0	0	34
Bucharest	Feb	0.4	1	1	30	63	135	63	1	0	0	35	0
Bucharest	Mar	5.0	16	16	35	19	150	15	16	0	4	0	0
Bucharest	Apr	11.5	46	52	44	-8	142	-8	52	0	0	0	0
Bucharest	May	16.8	75	96	61	-35	107	-35	96	0	0	0	0
Bucharest	Jun	20.7	98	127	74	-53	65	-42	117	10	0	0	0
Bucharest	Jul	22.5	109	143	54	-89	28	-37	92	51	0	0	0
Bucharest	Aug	22.0	106	128	50	-78	13	-15	65	63	0	0	0
Bucharest	Sep	17.6	80	82	43	-39	9	-4	47	35	0	0	0
Bucharest	Oct	11.6	47	44	37	-7	9	0	36	8	0	0	0
Bucharest	Nov	5.3	17	14	41	27	36	27	14	0	0	0	0
Bucharest	Dec	0.6	1	1	37	35	72	36	1	0	0	0	1
Budapest	Jan	-0.9	0	0	31	31	129	32	0	0	0	0	0
Budapest	Feb	0.6	1	1	30	29	150	21	1	0	8	0	0
Budapest	Mar	5.6	20	20	33	13	150	0	20	0	13	0	0
Budapest	Apr	11.4	48	54	46	-8	142	-8	54	0	0	0	0
Budapest	May	16.5	75	98	64	-34	109	-33	98	0	0	0	0
Budapest	Jun	19.5	92	123	61	-62	61	-48	110	13	0	0	0
Budapest	Jul	21.9	106	143	49	-94	25	-36	85	58	0	0	0
Budapest	Aug	21.2	102	125	48	-77	12	-13	61	64	0	0	0
Budapest	Sep	16.6	76	78	47	-31	9	-3	50	28	0	0	0
Budapest	Oct	10.8	44	41	46	5	15	6	39	2	0	0	0
Budapest	Nov	4.5	15	12	54	42	57	42	12	0	0	0	0
Budapest	Dec	0.6	1	1	41	40	97	40	1	0	0	0	0
Buenos Aires	Jan	23.4	108	132	105	-27	90	-26	131	1	0	0	0
Buenos Aires	Feb	22.8	103	107	95	-12	79	-11	105	2	0	0	0
Buenos Aires	Mar	20.3	84	90	112	22	101	22	90	0	0	0	0
Buenos Aires	Apr	17.2	63	59	93	34	134	33	59	0	0	0	0
Buenos Aires	May	14.0	44	39	75	36	150	16	39	0	20	0	0
Buenos Aires	Jun	11.2	30	25	56	31	150	0	25	0	31	0	0
Buenos Aires	Jul	10.3	26	23	56	33	150	0	23	0	33	0	0
Buenos Aires	Aug	11.7	32	30	62	32	150	0	30	0	32	0	0
Buenos Aires	Sep	13.7	43	43	68	25	150	0	43	0	25	0	0
Buenos Aires	Oct	16.6	59	67	100	33	150	0	67	0	33	0	0
Buenos Aires	Nov	19.4	78	91	99	8	149	-1	91	0	9	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Buenos Aires	Dec	22.1	98	122	88	-34	116	-33	122	0	0	0	0
Cairo	Jan	10.6	21	19	6	-13	0	0	6	13	0	0	0
Cairo	Feb	11.8	26	23	3	-20	0	0	3	20	0	0	0
Cairo	Mar	14.5	40	41	3	-38	0	0	3	38	0	0	0
Cairo	Apr	18.4	64	69	1	-68	0	0	1	68	0	0	0
Cairo	May	21.8	91	107	1	-106	0	0	1	106	0	0	0
Cairo	Jun	24.6	116	135	0	-135	0	0	0	135	0	0	0
Cairo	Jul	25.5	124	148	0	-148	0	0	0	148	0	0	0
Cairo	Aug	25.5	124	141	0	-141	0	0	0	141	0	0	0
Cairo	Sep	23.5	105	108	0	-108	0	0	0	108	0	0	0
Cairo	Oct	20.9	83	81	1	-80	0	0	1	80	0	0	0
Cairo	Nov	16.2	50	44	5	-39	0	0	5	39	0	0	0
Cairo	Dec	12.3	29	25	5	-20	0	0	5	20	0	0	0
Cali	Jan	26.0	125	129	91	-38	53	-23	114	15	0	0	0
Cali	Feb	26.3	130	122	127	5	58	5	121	1	0	0	0
Cali	Mar	26.2	129	134	172	38	95	37	134	0	0	0	0
Cali	Apr	26.0	125	127	139	12	107	12	127	0	0	0	0
Cali	May	25.8	122	129	122	-7	101	-6	129	0	0	0	0
Cali	Jun	25.6	119	122	50	-72	51	-50	101	21	0	0	0
Cali	Jul	25.7	120	128	32	-96	20	-31	63	65	0	0	0
Cali	Aug	26.1	127	134	35	-99	8	-12	47	87	0	0	0
Cali	Sep	25.8	122	123	80	-43	5	-3	81	42	0	0	0
Cali	Oct	25.5	117	121	189	68	72	67	121	0	0	0	0
Cali	Nov	25.3	114	114	163	49	123	51	114	0	0	0	0
Cali	Dec	25.6	119	122	69	-53	76	-47	117	5	0	0	0
Cape Town	Jan	21.2	91	111	17	-94	11	-15	33	78	0	0	0
Cape Town	Feb	21.0	89	93	21	-72	5	-6	26	67	0	0	0
Cape Town	Mar	19.9	81	87	28	-59	3	-2	30	57	0	0	0
Cape Town	Apr	17.9	68	63	50	-13	3	0	50	13	0	0	0
Cape Town	May	15.5	53	47	66	19	21	18	47	0	0	0	0
Cape Town	Jun	13.8	43	35	75	40	61	40	35	0	0	0	0
Cape Town	Jul	12.9	38	33	71	38	98	37	33	0	0	0	0
Cape Town	Aug	13.4	41	38	70	32	131	33	38	0	0	0	0
Cape Town	Sep	14.6	47	47	44	-3	127	-4	47	0	0	0	0
Cape Town	Oct	16.5	59	66	39	-27	101	-26	66	0	0	0	0
Cape Town	Nov	18.5	72	83	28	-55	59	-42	70	13	0	0	0
Cape Town	Dec	20.1	83	103	18	-85	26	-33	51	52	0	0	0
Caracas	Jan	23.2	87	86	15	-71	5	-5	20	66	0	0	0
Caracas	Feb	23.9	96	88	12	-76	2	-3	14	74	0	0	0
Caracas	Mar	24.7	107	111	10	-101	1	-1	11	100	0	0	0
Caracas	Apr	25.9	125	128	46	-82	0	-1	46	82	0	0	0
Caracas	May	26.3	131	143	92	-51	0	0	91	52	0	0	0
Caracas	Jun	25.6	120	127	156	29	29	29	127	0	0	0	0
Caracas	Jul	25.5	118	129	147	18	47	18	129	0	0	0	0
Caracas	Aug	26.0	126	136	117	-19	40	-7	125	11	0	0	0
Caracas	Sep	26.6	136	138	100	-38	28	-12	112	26	0	0	0
Caracas	Oct	26.4	133	136	102	-34	20	-8	110	26	0	0	0
Caracas	Nov	25.1	112	109	78	-31	15	-5	84	25	0	0	0
Caracas	Dec	23.9	96	95	50	-45	10	-5	56	39	0	0	0
Casablanca	Jan	10.6	27	24	60	36	92	36	24	0	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Casablanca	Feb	11.6	32	27	53	26	118	26	27	0	0	0	0
Casablanca	Mar	13.3	40	41	44	3	121	3	41	0	0	0	0
Casablanca	Apr	14.8	48	52	38	-14	107	-14	52	0	0	0	0
Casablanca	May	17.8	67	80	16	-64	58	-49	65	15	0	0	0
Casablanca	Jun	20.8	88	105	3	-102	22	-36	39	66	0	0	0
Casablanca	Jul	23.8	111	135	1	-134	6	-16	17	118	0	0	0
Casablanca	Aug	24.2	114	131	1	-130	2	-4	5	126	0	0	0
Casablanca	Sep	22.3	99	101	8	-93	1	-1	9	92	0	0	0
Casablanca	Oct	18.6	72	70	33	-37	0	-1	32	38	0	0	0
Casablanca	Nov	14.4	46	40	54	14	15	15	39	1	0	0	0
Casablanca	Dec	11.5	31	27	68	41	56	41	27	0	0	0	0
Cebu	Jan	23.0	88	88	205	117	150	0	88	0	117	0	0
Cebu	Feb	23.1	89	82	110	28	150	0	82	0	28	0	0
Cebu	Mar	23.8	98	101	103	2	146	-4	101	0	6	0	0
Cebu	Apr	24.8	111	114	60	-54	93	-53	113	1	0	0	0
Cebu	May	25.4	119	129	118	-11	84	-9	126	3	0	0	0
Cebu	Jun	25.0	113	120	201	81	150	66	120	0	15	0	0
Cebu	Jul	24.7	109	119	211	92	150	0	119	0	92	0	0
Cebu	Aug	24.8	111	118	198	80	150	0	118	0	80	0	0
Cebu	Sep	24.7	109	111	201	90	150	0	111	0	90	0	0
Cebu	Oct	24.4	105	108	224	116	150	0	108	0	116	0	0
Cebu	Nov	24.1	101	98	258	160	150	0	98	0	160	0	0
Cebu	Dec	23.6	95	94	208	114	150	0	94	0	114	0	0
Chennai	Jan	24.8	92	90	24	-66	87	-63	89	1	0	0	0
Chennai	Feb	26.3	120	110	6	-104	32	-55	61	49	0	0	0
Chennai	Mar	28.6	153	158	4	-154	7	-25	29	129	0	0	0
Chennai	Apr	30.7	166	172	12	-160	1	-6	17	155	0	0	0
Chennai	May	32.5	175	192	41	-151	0	-1	42	150	0	0	0
Chennai	Jun	32.1	174	187	65	-122	0	0	65	122	0	0	0
Chennai	Jul	30.5	165	183	102	-81	0	0	101	82	0	0	0
Chennai	Aug	29.9	162	175	141	-34	0	0	140	35	0	0	0
Chennai	Sep	29.5	159	161	127	-34	2	2	124	37	0	0	0
Chennai	Oct	28.2	149	152	247	95	95	93	152	0	0	0	0
Chennai	Nov	26.2	118	113	350	237	150	55	113	0	182	0	0
Chennai	Dec	25.0	95	93	147	54	150	0	93	0	54	0	0
Chicago, IL	Jan	-4.1	0	0	49	-1	150	0	0	0	0	0	81
Chicago, IL	Feb	-2.4	0	0	45	18	150	0	0	0	18	7	108
Chicago, IL	Mar	2.9	8	8	62	162	150	0	8	0	162	108	0
Chicago, IL	Apr	9.9	38	42	97	55	150	0	42	0	55	0	0
Chicago, IL	May	15.6	68	85	100	15	150	0	85	0	15	0	0
Chicago, IL	Jun	21.0	100	127	111	-16	135	-15	127	0	0	0	0
Chicago, IL	Jul	23.5	115	148	98	-50	87	-48	146	2	0	0	0
Chicago, IL	Aug	22.9	111	133	94	-39	60	-27	121	12	0	0	0
Chicago, IL	Sep	18.5	85	87	76	-11	54	-6	82	5	0	0	0
Chicago, IL	Oct	12.8	53	50	73	23	77	23	50	0	0	0	0
Chicago, IL	Nov	5.3	17	14	68	54	130	53	14	0	0	0	0
Chicago, IL	Dec	-1.2	0	0	57	26	150	20	0	0	6	0	31
Chisinau	Jan	-3.0	0	0	34	0	67	0	0	0	0	0	46
Chisinau	Feb	-1.8	0	0	34	15	82	15	0	0	0	8	65
Chisinau	Mar	2.3	8	8	29	86	150	68	8	0	18	65	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Chisinau	Apr	9.7	41	47	40	-7	140	-10	47	0	3	0	0
Chisinau	May	15.6	72	94	50	-44	97	-43	94	0	0	0	0
Chisinau	Jun	19.5	94	124	68	-56	57	-40	108	16	0	0	0
Chisinau	Jul	21.3	104	139	58	-81	26	-31	89	50	0	0	0
Chisinau	Aug	20.6	100	122	49	-73	13	-13	62	60	0	0	0
Chisinau	Sep	16.0	74	77	46	-31	9	-4	49	28	0	0	0
Chisinau	Oct	10.0	43	40	32	-8	10	1	31	9	0	0	0
Chisinau	Nov	3.7	13	10	42	32	41	31	10	0	0	0	0
Chisinau	Dec	-0.4	0	0	38	26	67	26	0	0	0	0	12
Colombo	Jan	26.7	137	138	229	91	150	0	138	0	91	0	0
Colombo	Feb	27.0	139	129	159	30	150	0	129	0	30	0	0
Colombo	Mar	27.7	145	151	189	38	150	0	151	0	38	0	0
Colombo	Apr	28.1	149	152	295	143	150	0	152	0	143	0	0
Colombo	May	28.0	148	158	356	198	150	0	158	0	198	0	0
Colombo	Jun	27.4	143	149	274	125	150	0	149	0	125	0	0
Colombo	Jul	27.1	140	151	212	61	150	0	151	0	61	0	0
Colombo	Aug	27.2	141	150	185	35	150	0	150	0	35	0	0
Colombo	Sep	27.1	140	142	286	144	150	0	142	0	144	0	0
Colombo	Oct	26.8	138	142	396	254	150	0	142	0	254	0	0
Colombo	Nov	26.5	135	133	380	247	150	0	133	0	247	0	0
Colombo	Dec	26.6	136	137	318	181	150	0	137	0	181	0	0
Copenhagen	Jan	0.1	1	0	49	49	150	17	0	0	32	0	0
Copenhagen	Feb	0.1	1	0	35	35	150	0	0	0	35	0	0
Copenhagen	Mar	2.3	12	12	43	31	150	0	12	0	31	0	0
Copenhagen	Apr	6.4	33	38	40	2	147	-3	38	0	5	0	0
Copenhagen	May	11.3	58	77	43	-34	114	-33	77	0	0	0	0
Copenhagen	Jun	15.0	77	104	51	-53	68	-46	96	8	0	0	0
Copenhagen	Jul	16.8	86	117	65	-52	41	-27	92	25	0	0	0
Copenhagen	Aug	16.5	84	104	68	-36	29	-12	80	24	0	0	0
Copenhagen	Sep	12.9	66	69	59	-10	26	-3	61	8	0	0	0
Copenhagen	Oct	8.8	45	41	60	19	45	19	41	0	0	0	0
Copenhagen	Nov	4.5	23	17	53	36	80	35	17	0	0	0	0
Copenhagen	Dec	1.7	9	6	60	54	133	53	6	0	0	0	0
Curitiba	Jan	20.8	88	103	213	110	150	0	103	0	110	0	0
Curitiba	Feb	20.7	87	88	158	70	150	0	88	0	70	0	0
Curitiba	Mar	20.1	83	88	148	60	150	0	88	0	60	0	0
Curitiba	Apr	18.3	71	68	90	22	150	0	68	0	22	0	0
Curitiba	May	14.6	48	45	124	79	150	0	45	0	79	0	0
Curitiba	Jun	13.2	40	35	124	89	150	0	35	0	89	0	0
Curitiba	Jul	12.6	37	34	85	51	150	0	34	0	51	0	0
Curitiba	Aug	14.0	44	43	82	39	150	0	43	0	39	0	0
Curitiba	Sep	15.1	51	51	130	79	150	0	51	0	79	0	0
Curitiba	Oct	16.5	59	65	142	77	150	0	65	0	77	0	0
Curitiba	Nov	18.3	71	79	97	18	150	0	79	0	18	0	0
Curitiba	Dec	19.9	82	97	130	33	150	0	97	0	33	0	0
Dakar	Jan	23.5	86	84	1	-83	4	-5	6	78	0	0	0
Dakar	Feb	24.6	101	92	2	-90	2	-2	5	87	0	0	0
Dakar	Mar	25.5	116	120	0	-120	1	-1	1	119	0	0	0
Dakar	Apr	25.9	122	127	0	-127	0	-1	0	127	0	0	0
Dakar	May	26.6	136	150	1	-149	0	0	1	149	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Dakar	Jun	28.0	148	160	30	-130	0	0	29	131	0	0	0
Dakar	Jul	27.9	147	163	168	5	15	15	151	12	0	0	0
Dakar	Aug	27.3	142	154	269	115	130	115	154	0	0	0	0
Dakar	Sep	27.6	145	147	232	85	150	20	147	0	65	0	0
Dakar	Oct	28.1	149	151	78	-73	82	-68	148	3	0	0	0
Dakar	Nov	26.9	139	132	1	-131	23	-59	60	72	0	0	0
Dakar	Dec	24.3	97	94	2	-92	9	-14	16	78	0	0	0
Damascus	Jan	7.0	12	10	42	32	66	32	10	0	0	0	0
Damascus	Feb	8.4	16	14	37	23	89	23	14	0	0	0	0
Damascus	Mar	11.5	29	30	32	2	91	2	30	0	0	0	0
Damascus	Apr	16.2	54	59	13	-46	59	-32	45	14	0	0	0
Damascus	May	20.8	86	103	6	-97	23	-36	42	61	0	0	0
Damascus	Jun	24.5	116	139	1	-138	6	-17	18	121	0	0	0
Damascus	Jul	26.4	133	162	0	-162	1	-5	5	157	0	0	0
Damascus	Aug	26.4	133	153	0	-153	0	-1	1	152	0	0	0
Damascus	Sep	23.8	110	113	0	-113	0	0	0	113	0	0	0
Damascus	Oct	19.3	75	73	12	-61	0	0	12	61	0	0	0
Damascus	Nov	13.3	38	33	30	-3	4	4	26	7	0	0	0
Damascus	Dec	8.5	17	14	45	31	34	30	14	0	0	0	0
Dar es Salaam	Jan	27.7	145	156	93	-63	2	-1	95	61	0	0	0
Dar es Salaam	Feb	27.9	147	141	90	-51	1	-1	90	51	0	0	0
Dar es Salaam	Mar	27.7	145	152	148	-4	9	8	138	14	0	0	0
Dar es Salaam	Apr	26.8	138	137	279	142	150	141	137	0	0	0	0
Dar es Salaam	May	25.8	121	123	166	43	150	0	123	0	43	0	0
Dar es Salaam	Jun	24.6	102	100	35	-65	88	-62	99	1	0	0	0
Dar es Salaam	Jul	24.0	94	95	24	-71	44	-44	68	27	0	0	0
Dar es Salaam	Aug	24.1	95	97	20	-77	21	-23	43	54	0	0	0
Dar es Salaam	Sep	24.7	104	105	20	-85	9	-12	31	74	0	0	0
Dar es Salaam	Oct	25.7	120	127	56	-71	5	-4	59	68	0	0	0
Dar es Salaam	Nov	26.5	135	140	116	-24	4	-1	116	24	0	0	0
Dar es Salaam	Dec	27.4	143	154	124	-30	3	-1	125	29	0	0	0
Delhi	Jan	14.3	16	15	17	2	11	3	14	1	0	0	0
Delhi	Feb	17.3	31	27	14	-13	9	-2	16	11	0	0	0
Delhi	Mar	22.7	79	81	12	-69	5	-4	16	65	0	0	0
Delhi	Apr	28.7	153	164	7	-157	1	-4	11	153	0	0	0
Delhi	May	32.9	177	207	16	-191	0	-1	16	191	0	0	0
Delhi	Jun	33.9	180	209	57	-152	0	0	56	153	0	0	0
Delhi	Jul	30.9	168	199	193	-6	8	8	184	15	0	0	0
Delhi	Aug	29.7	160	181	224	43	51	43	181	0	0	0	0
Delhi	Sep	29.3	158	161	122	-39	35	-16	139	22	0	0	0
Delhi	Oct	26.0	126	123	24	-99	14	-21	47	76	0	0	0
Delhi	Nov	20.5	56	50	3	-47	9	-5	8	42	0	0	0
Delhi	Dec	15.7	22	20	8	-12	8	-1	9	11	0	0	0
Denver, CO	Jan	-1.3	0	0	12	8	36	9	0	0	0	5	5
Denver, CO	Feb	0.7	2	1	14	18	52	16	1	0	0	5	0
Denver, CO	Mar	3.5	12	12	30	18	70	18	12	0	0	0	0
Denver, CO	Apr	8.2	33	37	45	8	79	9	37	0	0	0	0
Denver, CO	May	13.7	61	76	69	-7	74	-5	74	2	0	0	0
Denver, CO	Jun	19.3	92	115	47	-68	39	-35	83	32	0	0	0
Denver, CO	Jul	22.6	111	141	57	-84	17	-22	79	62	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Denver, CO	Aug	21.7	106	125	49	-76	8	-9	58	67	0	0	0
Denver, CO	Sep	16.9	79	81	29	-52	5	-3	33	48	0	0	0
Denver, CO	Oct	10.7	46	43	26	-17	4	-1	27	16	0	0	0
Denver, CO	Nov	3.6	13	10	21	11	15	11	10	0	0	0	0
Denver, CO	Dec	-0.2	0	0	14	13	27	12	0	0	0	0	1
Detroit, MI	Jan	-4.4	0	0	54	0	122	0	0	0	0	0	94
Detroit, MI	Feb	-2.8	0	0	48	0	122	0	0	0	0	0	142
Detroit, MI	Mar	1.3	4	4	66	204	150	28	4	0	176	142	0
Detroit, MI	Apr	8.2	33	37	87	50	150	0	37	0	50	0	0
Detroit, MI	May	14.3	64	81	64	-17	132	-18	81	0	1	0	0
Detroit, MI	Jun	19.7	94	120	83	-37	96	-36	120	0	0	0	0
Detroit, MI	Jul	22.5	111	143	69	-74	47	-49	117	26	0	0	0
Detroit, MI	Aug	21.5	105	125	75	-50	29	-18	93	32	0	0	0
Detroit, MI	Sep	17.3	81	83	79	-4	29	0	79	4	0	0	0
Detroit, MI	Oct	11.3	48	46	60	14	43	14	46	0	0	0	0
Detroit, MI	Nov	4.5	16	13	65	52	95	52	13	0	0	0	0
Detroit, MI	Dec	-1.5	0	0	67	27	122	27	0	0	0	0	40
Dhaka	Jan	19.4	44	41	8	-33	44	-17	25	16	0	0	0
Dhaka	Feb	21.9	68	60	18	-42	30	-14	32	28	0	0	0
Dhaka	Mar	25.9	123	127	50	-77	14	-16	64	63	0	0	0
Dhaka	Apr	28.3	150	159	122	-37	11	-3	123	36	0	0	0
Dhaka	May	28.8	154	176	259	83	91	80	176	0	0	0	0
Dhaka	Jun	28.5	152	172	430	258	150	59	172	0	199	0	0
Dhaka	Jul	28.3	150	174	460	286	150	0	174	0	286	0	0
Dhaka	Aug	28.3	150	167	405	238	150	0	167	0	238	0	0
Dhaka	Sep	28.6	153	155	265	110	150	0	155	0	110	0	0
Dhaka	Oct	27.7	145	145	159	14	147	-3	145	0	17	0	0
Dhaka	Nov	24.4	100	91	37	-54	94	-53	91	0	0	0	0
Dhaka	Dec	20.6	55	50	6	-44	61	-33	39	11	0	0	0
Doha	Jan	17.1	24	22	16	-6	0	0	16	6	0	0	0
Doha	Feb	18.0	29	25	17	-8	0	0	17	8	0	0	0
Doha	Mar	21.2	55	57	16	-41	0	0	16	41	0	0	0
Doha	Apr	25.7	117	124	7	-117	0	0	7	117	0	0	0
Doha	May	30.8	167	193	2	-191	0	0	2	191	0	0	0
Doha	Jun	33.8	180	206	0	-206	0	0	0	206	0	0	0
Doha	Jul	34.7	182	213	0	-213	0	0	0	213	0	0	0
Doha	Aug	34.3	181	203	0	-203	0	0	0	203	0	0	0
Doha	Sep	32.2	174	177	0	-177	0	0	0	177	0	0	0
Doha	Oct	28.9	155	153	1	-152	0	0	1	152	0	0	0
Doha	Nov	24.1	91	82	2	-80	0	0	2	80	0	0	0
Doha	Dec	19.1	37	33	10	-23	0	0	10	23	0	0	0
Dubai	Jan	19.7	42	39	95	56	56	56	39	0	0	0	0
Dubai	Feb	16.3	20	18	14	-4	56	0	14	4	0	0	0
Dubai	Mar	22.1	65	67	9	-58	33	-23	33	34	0	0	0
Dubai	Apr	26.1	124	132	4	-128	9	-24	27	105	0	0	0
Dubai	May	30.8	167	193	1	-192	1	-8	9	184	0	0	0
Dubai	Jun	32.8	176	202	0	-202	0	-1	1	201	0	0	0
Dubai	Jul	33.4	179	209	1	-208	0	0	1	208	0	0	0
Dubai	Aug	33.6	179	201	4	-197	0	0	3	198	0	0	0
Dubai	Sep	31.5	171	174	5	-169	0	0	5	169	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Dubai	Oct	28.3	150	149	0	-149	0	0	0	149	0	0	0
Dubai	Nov	24.0	90	81	5	-76	0	0	4	77	0	0	0
Dubai	Dec	21.0	54	49	16	-33	0	0	15	34	0	0	0
Dublin	Jan	3.1	20	14	79	65	150	0	14	0	65	0	0
Dublin	Feb	3.3	21	16	59	43	150	0	16	0	43	0	0
Dublin	Mar	4.4	27	27	61	34	150	0	27	0	34	0	0
Dublin	Apr	6.1	36	41	50	9	150	0	41	0	9	0	0
Dublin	May	8.7	49	65	61	-4	145	-5	65	0	0	0	0
Dublin	Jun	11.6	64	87	61	-26	119	-26	87	0	0	0	0
Dublin	Jul	13.5	73	100	60	-40	81	-38	97	3	0	0	0
Dublin	Aug	13.1	71	88	73	-15	70	-11	84	4	0	0	0
Dublin	Sep	10.8	60	62	71	9	78	8	62	0	0	0	0
Dublin	Oct	8.1	46	42	78	36	114	36	42	0	0	0	0
Dublin	Nov	5.1	31	23	73	50	150	36	23	0	14	0	0
Dublin	Dec	3.8	23	16	88	72	150	0	16	0	72	0	0
Durban	Jan	24.1	108	129	145	16	142	15	129	0	0	0	0
Durban	Feb	24.4	111	114	114	0	143	1	114	0	0	0	0
Durban	Mar	23.7	104	110	124	14	150	7	110	0	7	0	0
Durban	Apr	21.8	86	81	61	-20	130	-20	81	0	0	0	0
Durban	May	19.5	66	61	46	-15	115	-15	61	0	0	0	0
Durban	Jun	17.1	49	42	43	1	116	1	42	0	0	0	0
Durban	Jul	17.2	50	45	43	-2	115	-1	45	0	0	0	0
Durban	Aug	18.1	56	53	47	-6	108	-7	53	0	0	0	0
Durban	Sep	19.4	66	66	57	-9	100	-8	66	0	0	0	0
Durban	Oct	20.4	74	82	90	8	108	8	82	0	0	0	0
Durban	Nov	21.7	85	97	108	11	118	10	97	0	0	0	0
Durban	Dec	23.1	98	119	128	9	127	9	119	0	0	0	0
Florianopolis	Jan	24.4	112	132	276	144	150	0	132	0	144	0	0
Florianopolis	Feb	24.5	113	115	227	112	150	0	115	0	112	0	0
Florianopolis	Mar	23.0	99	104	184	80	150	0	104	0	80	0	0
Florianopolis	Apr	21.2	83	79	91	12	150	0	79	0	12	0	0
Florianopolis	May	17.6	55	51	82	31	150	0	51	0	31	0	0
Florianopolis	Jun	15.2	40	35	100	65	150	0	35	0	65	0	0
Florianopolis	Jul	15.3	41	37	94	57	150	0	37	0	57	0	0
Florianopolis	Aug	16.5	48	46	100	54	150	0	46	0	54	0	0
Florianopolis	Sep	18.1	59	59	157	98	150	0	59	0	98	0	0
Florianopolis	Oct	19.7	70	78	140	62	150	0	78	0	62	0	0
Florianopolis	Nov	22.1	90	102	143	41	150	0	102	0	41	0	0
Florianopolis	Dec	23.8	106	127	143	16	150	0	127	0	16	0	0
Geneva	Jan	-0.9	0	0	98	98	150	0	0	0	98	0	0
Geneva	Feb	0.7	3	3	97	94	150	0	3	0	94	0	0
Geneva	Mar	4.6	22	23	95	72	150	0	23	0	72	0	0
Geneva	Apr	7.7	38	43	88	45	150	0	43	0	45	0	0
Geneva	May	11.9	59	76	94	18	150	0	76	0	18	0	0
Geneva	Jun	15.2	76	100	106	6	149	-1	100	0	7	0	0
Geneva	Jul	18.0	90	120	89	-31	119	-30	120	0	0	0	0
Geneva	Aug	17.1	86	104	102	-2	116	-3	104	0	0	0	0
Geneva	Sep	13.8	69	71	107	36	150	34	71	0	2	0	0
Geneva	Oct	9.2	45	42	108	66	150	0	42	0	66	0	0
Geneva	Nov	3.9	19	15	114	99	150	0	15	0	99	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Geneva	Dec	0.6	3	2	116	114	150	0	2	0	114	0	0
Guadalajara	Jan	16.6	48	45	13	-32	45	-16	29	16	0	0	0
Guadalajara	Feb	18.2	59	52	6	-46	29	-16	22	30	0	0	0
Guadalajara	Mar	20.1	73	75	6	-69	15	-14	19	56	0	0	0
Guadalajara	Apr	22.1	90	95	8	-87	6	-9	16	79	0	0	0
Guadalajara	May	23.5	103	116	26	-90	3	-3	28	88	0	0	0
Guadalajara	Jun	23.0	98	110	183	73	75	72	110	0	0	0	0
Guadalajara	Jul	21.6	86	98	223	125	150	75	98	0	50	0	0
Guadalajara	Aug	21.6	86	94	191	97	150	0	94	0	97	0	0
Guadalajara	Sep	21.5	85	86	163	77	150	0	86	0	77	0	0
Guadalajara	Oct	20.6	77	77	57	-20	128	-22	77	0	2	0	0
Guadalajara	Nov	18.8	63	58	15	-43	86	-42	57	1	0	0	0
Guadalajara	Dec	17.1	51	48	11	-37	61	-25	37	11	0	0	0
Guangzhou	Jan	14.0	25	23	34	11	106	11	23	0	0	0	0
Guangzhou	Feb	14.8	29	25	63	38	144	38	25	0	0	0	0
Guangzhou	Mar	18.1	49	51	88	37	150	6	51	0	31	0	0
Guangzhou	Apr	22.2	84	89	180	91	150	0	89	0	91	0	0
Guangzhou	May	26.0	129	147	279	132	150	0	147	0	132	0	0
Guangzhou	Jun	27.8	146	165	296	131	150	0	165	0	131	0	0
Guangzhou	Jul	28.9	155	179	252	73	150	0	179	0	73	0	0
Guangzhou	Aug	28.7	153	170	277	107	150	0	170	0	107	0	0
Guangzhou	Sep	27.6	145	147	189	42	150	0	147	0	42	0	0
Guangzhou	Oct	24.4	109	108	75	-33	117	-33	108	0	0	0	0
Guangzhou	Nov	20.1	65	59	39	-20	97	-20	59	0	0	0	0
Guangzhou	Dec	15.9	35	32	31	-1	95	-2	32	0	0	0	0
Guatemala City	Jan	21.4	77	75	20	-55	60	-42	62	13	0	0	0
Guatemala City	Feb	20.9	72	66	17	-49	38	-22	39	27	0	0	0
Guatemala City	Mar	22.5	88	91	19	-72	19	-19	37	54	0	0	0
Guatemala City	Apr	23.7	101	105	57	-48	19	0	54	51	0	0	0
Guatemala City	May	23.7	101	111	368	257	150	131	111	0	126	0	0
Guatemala City	Jun	23.9	103	111	423	312	150	0	111	0	312	0	0
Guatemala City	Jul	23.7	101	112	462	350	150	0	112	0	350	0	0
Guatemala City	Aug	23.2	95	103	422	319	150	0	103	0	319	0	0
Guatemala City	Sep	23.0	93	94	571	477	150	0	94	0	477	0	0
Guatemala City	Oct	22.6	89	90	265	175	150	0	90	0	175	0	0
Guatemala City	Nov	21.7	80	76	100	24	148	-2	76	0	26	0	0
Guatemala City	Dec	21.0	73	71	25	-46	102	-46	71	0	0	0	0
Hamburg	Jan	0.4	2	1	55	54	150	10	1	0	44	0	0
Hamburg	Feb	0.8	4	3	41	38	150	0	3	0	38	0	0
Hamburg	Mar	3.4	17	17	48	31	150	0	17	0	31	0	0
Hamburg	Apr	7.5	38	43	45	2	148	-2	43	0	4	0	0
Hamburg	May	12.3	62	82	53	-29	119	-29	82	0	0	0	0
Hamburg	Jun	15.5	78	106	66	-40	81	-38	103	3	0	0	0
Hamburg	Jul	17.3	87	120	73	-47	52	-29	103	17	0	0	0
Hamburg	Aug	16.6	84	104	72	-32	38	-14	86	18	0	0	0
Hamburg	Sep	13.1	66	68	56	-12	34	-4	61	7	0	0	0
Hamburg	Oct	8.8	44	40	52	12	46	12	40	0	0	0	0
Hamburg	Nov	3.9	19	14	55	41	85	39	14	0	0	0	0
Hamburg	Dec	1.5	7	5	60	55	140	55	5	0	0	0	0
Hanover	Jan	0.7	3	2	59	57	150	0	2	0	57	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Hanover	Feb	1.1	5	4	45	41	150	0	4	0	41	0	0
Hanover	Mar	4.1	19	20	53	33	150	0	20	0	33	0	0
Hanover	Apr	8.0	39	44	50	6	149	-1	44	0	7	0	0
Hanover	May	12.8	63	84	63	-21	129	-20	84	0	0	0	0
Hanover	Jun	15.7	78	106	72	-34	95	-34	106	0	0	0	0
Hanover	Jul	17.7	88	121	83	-38	66	-29	112	9	0	0	0
Hanover	Aug	17.1	85	106	70	-36	47	-19	89	17	0	0	0
Hanover	Sep	13.7	68	70	60	-10	42	-5	64	6	0	0	0
Hanover	Oct	9.3	45	41	59	18	60	18	41	0	0	0	0
Hanover	Nov	4.6	22	16	56	40	99	39	16	0	0	0	0
Hanover	Dec	2.1	10	7	65	58	150	51	7	0	7	0	0
Helsinki	Jan	-5.2	0	0	47	0	150	0	0	0	0	0	111
Helsinki	Feb	-5.7	0	0	34	-1	150	0	0	0	0	0	146
Helsinki	Mar	-2.4	0	0	32	27	150	0	0	0	27	17	151
Helsinki	Apr	3.8	23	27	44	168	150	0	27	0	168	151	0
Helsinki	May	9.8	55	73	47	-26	124	-26	73	0	0	0	0
Helsinki	Jun	14.7	79	107	58	-49	79	-45	103	4	0	0	0
Helsinki	Jul	16.8	89	122	80	-42	53	-26	106	16	0	0	0
Helsinki	Aug	15.5	83	103	85	-18	44	-9	93	10	0	0	0
Helsinki	Sep	10.8	60	62	83	21	64	20	62	0	0	0	0
Helsinki	Oct	5.9	35	32	75	43	107	43	32	0	0	0	0
Helsinki	Nov	0.5	4	3	71	59	150	43	3	0	16	0	9
Helsinki	Dec	-2.8	0	0	56	1	150	0	0	0	0	0	64
Ho Chi Minh City	Jan	26.1	123	122	10	-112	18	-37	46	76	0	0	0
Ho Chi Minh City	Feb	26.8	138	126	4	-122	6	-12	16	110	0	0	0
Ho Chi Minh City	Mar	28.0	148	153	12	-141	1	-5	16	137	0	0	0
Ho Chi Minh City	Apr	29.2	157	162	51	-111	0	-1	50	112	0	0	0
Ho Chi Minh City	May	28.6	153	166	209	43	44	44	163	3	0	0	0
Ho Chi Minh City	Jun	27.7	145	155	288	133	150	106	155	0	27	0	0
Ho Chi Minh City	Jul	27.4	143	156	276	120	150	0	156	0	120	0	0
Ho Chi Minh City	Aug	27.3	142	153	274	121	150	0	153	0	121	0	0
Ho Chi Minh City	Sep	27.1	140	142	315	173	150	0	142	0	173	0	0
Ho Chi Minh City	Oct	26.8	138	141	264	123	150	0	141	0	123	0	0
Ho Chi Minh City	Nov	26.5	135	130	107	-23	122	-28	130	0	5	0	0
Ho Chi Minh City	Dec	25.9	119	118	33	-85	55	-67	101	17	0	0	0
Hyderabad	Jan	22.2	68	65	5	-60	6	-5	9	56	0	0	0
Hyderabad	Feb	25.1	108	97	6	-91	3	-3	9	88	0	0	0
Hyderabad	Mar	28.5	152	157	10	-147	1	-2	12	145	0	0	0
Hyderabad	Apr	31.2	169	177	20	-157	0	-1	21	156	0	0	0
Hyderabad	May	32.9	177	197	31	-166	0	0	30	167	0	0	0
Hyderabad	Jun	29.4	158	174	96	-78	0	0	95	79	0	0	0
Hyderabad	Jul	27.0	139	157	141	-16	0	0	140	17	0	0	0
Hyderabad	Aug	26.4	130	143	155	12	13	13	143	0	0	0	0
Hyderabad	Sep	26.4	130	132	157	25	37	24	132	0	0	0	0
Hyderabad	Oct	25.7	118	119	93	-26	29	-8	102	17	0	0	0
Hyderabad	Nov	23.4	82	77	26	-51	18	-11	38	39	0	0	0
Hyderabad	Dec	21.4	59	56	7	-49	11	-7	14	42	0	0	0
Islamabad	Jan	10.4	12	11	47	36	58	37	11	0	0	0	0
Islamabad	Feb	12.7	20	17	57	40	97	39	17	0	0	0	0
Islamabad	Mar	17.5	47	48	53	5	103	6	48	0	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Islamabad	Apr	22.8	92	100	47	-53	63	-40	87	13	0	0	0
Islamabad	May	27.8	146	175	33	-142	16	-47	80	95	0	0	0
Islamabad	Jun	31.6	171	205	43	-162	3	-13	54	151	0	0	0
Islamabad	Jul	29.9	162	197	174	-23	3	0	173	24	0	0	0
Islamabad	Aug	28.5	152	175	213	38	41	38	175	0	0	0	0
Islamabad	Sep	27.2	141	145	77	-68	22	-19	98	47	0	0	0
Islamabad	Oct	22.8	92	89	23	-66	11	-11	33	56	0	0	0
Islamabad	Nov	16.8	42	36	14	-22	9	-2	16	20	0	0	0
Islamabad	Dec	12.1	18	15	28	13	21	12	15	0	0	0	0
Istanbul	Jan	5.8	13	11	100	89	150	0	11	0	89	0	0
Istanbul	Feb	6.7	16	13	85	72	150	0	13	0	72	0	0
Istanbul	Mar	8.6	23	24	74	50	150	0	24	0	50	0	0
Istanbul	Apr	13.3	46	51	67	16	149	-1	51	0	17	0	0
Istanbul	May	17.8	72	90	55	-35	115	-34	90	0	0	0	0
Istanbul	Jun	21.8	99	124	39	-85	51	-64	103	21	0	0	0
Istanbul	Jul	24.1	116	147	25	-122	16	-35	60	87	0	0	0
Istanbul	Aug	24.1	116	137	20	-117	5	-11	30	107	0	0	0
Istanbul	Sep	20.6	91	93	36	-57	3	-2	37	56	0	0	0
Istanbul	Oct	16.1	62	59	73	14	19	16	56	3	0	0	0
Istanbul	Nov	11.7	38	31	87	56	74	55	31	0	0	0	0
Istanbul	Dec	7.9	21	16	103	87	150	76	16	0	11	0	0
Jakarta	Jan	26.0	124	132	313	181	150	32	132	0	149	0	0
Jakarta	Feb	26.2	127	122	239	117	150	0	122	0	117	0	0
Jakarta	Mar	26.5	135	141	202	61	150	0	141	0	61	0	0
Jakarta	Apr	26.8	138	137	139	2	146	-4	137	0	6	0	0
Jakarta	May	27.1	140	143	117	-26	120	-26	143	0	0	0	0
Jakarta	Jun	26.6	136	133	89	-44	80	-40	130	3	0	0	0
Jakarta	Jul	26.3	129	131	63	-68	42	-38	102	29	0	0	0
Jakarta	Aug	26.3	129	132	51	-81	19	-23	74	58	0	0	0
Jakarta	Sep	26.7	137	138	48	-90	8	-11	59	79	0	0	0
Jakarta	Oct	26.9	139	146	117	-29	6	-2	118	28	0	0	0
Jakarta	Nov	26.6	136	140	192	52	57	51	140	0	0	0	0
Jakarta	Dec	26.4	131	141	203	62	118	61	141	0	0	0	0
Jerusalem	Jan	9.7	20	18	88	70	146	70	18	0	0	0	0
Jerusalem	Feb	10.6	24	20	81	61	150	4	20	0	57	0	0
Jerusalem	Mar	12.8	34	35	73	38	150	0	35	0	38	0	0
Jerusalem	Apr	16.9	57	62	17	-45	107	-43	62	0	0	0	0
Jerusalem	May	20.8	85	101	3	-98	42	-65	68	33	0	0	0
Jerusalem	Jun	23.3	105	125	1	-124	13	-29	30	95	0	0	0
Jerusalem	Jul	24.4	115	138	0	-138	3	-10	9	129	0	0	0
Jerusalem	Aug	24.8	118	135	0	-135	1	-2	2	133	0	0	0
Jerusalem	Sep	23.4	106	108	0	-108	0	-1	1	107	0	0	0
Jerusalem	Oct	21.3	89	86	13	-73	0	0	13	73	0	0	0
Jerusalem	Nov	16.3	53	46	51	5	10	10	40	6	0	0	0
Jerusalem	Dec	11.6	28	24	90	66	76	66	24	0	0	0	0
Johannesburg	Jan	20.5	88	103	120	17	82	17	103	0	0	0	0
Johannesburg	Feb	19.9	84	85	92	7	89	7	85	0	0	0	0
Johannesburg	Mar	18.7	76	80	94	14	103	14	80	0	0	0	0
Johannesburg	Apr	16.0	59	57	45	-12	93	-10	56	1	0	0	0
Johannesburg	May	12.6	40	37	19	-18	78	-15	34	3	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Johannesburg	Jun	9.7	26	23	7	-16	67	-11	18	5	0	0	0
Johannesburg	Jul	9.8	27	25	5	-20	55	-12	16	9	0	0	0
Johannesburg	Aug	12.2	38	37	6	-31	41	-14	20	17	0	0	0
Johannesburg	Sep	15.9	59	59	24	-35	30	-11	35	24	0	0	0
Johannesburg	Oct	18.1	72	79	69	-10	27	-3	71	8	0	0	0
Johannesburg	Nov	18.9	77	87	111	24	51	24	87	0	0	0	0
Johannesburg	Dec	19.9	84	100	114	14	65	14	100	0	0	0	0
Karachi	Jan	18.3	32	30	7	-23	0	0	7	23	0	0	0
Karachi	Feb	20.6	50	44	7	-37	0	0	6	38	0	0	0
Karachi	Mar	25.2	109	112	8	-104	0	0	8	104	0	0	0
Karachi	Apr	28.9	155	164	3	-161	0	0	3	161	0	0	0
Karachi	May	31.5	171	196	1	-195	0	0	1	195	0	0	0
Karachi	Jun	32.3	174	199	10	-189	0	0	9	190	0	0	0
Karachi	Jul	31.0	168	196	87	-109	0	0	86	110	0	0	0
Karachi	Aug	29.6	160	178	56	-122	0	0	57	121	0	0	0
Karachi	Sep	29.4	158	161	27	-134	0	0	27	134	0	0	0
Karachi	Oct	28.4	151	150	3	-147	0	0	3	147	0	0	0
Karachi	Nov	24.2	93	85	2	-83	0	0	2	83	0	0	0
Karachi	Dec	19.9	44	40	5	-35	0	0	5	35	0	0	0
Kathmandu	Jan	15.9	30	28	17	-11	51	-6	22	6	0	0	0
Kathmandu	Feb	18.1	44	39	8	-31	38	-13	21	18	0	0	0
Kathmandu	Mar	21.7	75	77	25	-52	23	-15	40	37	0	0	0
Kathmandu	Apr	25.4	119	128	43	-85	10	-13	55	73	0	0	0
Kathmandu	May	27.5	144	167	106	-61	6	-4	109	58	0	0	0
Kathmandu	Jun	28.6	153	176	254	78	80	74	176	0	0	0	0
Kathmandu	Jul	28.2	149	176	415	239	150	70	176	0	169	0	0
Kathmandu	Aug	28.3	150	169	355	186	150	0	169	0	186	0	0
Kathmandu	Sep	27.5	144	147	222	75	150	0	147	0	75	0	0
Kathmandu	Oct	24.6	109	107	75	-32	117	-33	107	0	1	0	0
Kathmandu	Nov	20.6	65	58	10	-48	74	-43	53	5	0	0	0
Kathmandu	Dec	17.5	40	36	8	-28	57	-17	25	11	0	0	0
Kiev	Jan	-5.1	0	0	42	0	71	0	0	0	0	0	83
Kiev	Feb	-4.0	0	0	37	-1	71	0	0	0	0	0	121
Kiev	Mar	0.5	2	2	36	155	150	79	2	0	76	129	0
Kiev	Apr	8.5	39	45	45	0	143	-7	45	0	7	0	0
Kiev	May	15.0	72	96	54	-42	101	-42	96	0	0	0	0
Kiev	Jun	18.6	92	125	69	-56	59	-42	111	14	0	0	0
Kiev	Jul	19.9	98	135	76	-59	34	-25	102	33	0	0	0
Kiev	Aug	19.0	94	116	65	-51	21	-13	79	37	0	0	0
Kiev	Sep	13.9	67	69	50	-19	17	-4	54	15	0	0	0
Kiev	Oct	7.9	36	33	39	6	24	7	32	1	0	0	0
Kiev	Nov	1.4	6	4	48	44	68	44	4	0	0	0	0
Kiev	Dec	-2.4	0	0	45	4	71	3	0	0	0	0	41
Kingston	Jan	24.6	102	98	165	67	150	0	98	0	67	0	0
Kingston	Feb	24.8	105	94	138	44	150	0	94	0	44	0	0
Kingston	Mar	25.3	113	117	77	-40	111	-39	117	0	0	0	0
Kingston	Apr	25.9	123	128	163	35	144	33	128	0	0	0	0
Kingston	May	26.4	132	147	204	57	150	6	147	0	51	0	0
Kingston	Jun	27.2	141	156	127	-29	121	-29	156	0	0	0	0
Kingston	Jul	27.4	143	162	111	-51	75	-46	157	5	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Kingston	Aug	27.9	147	161	163	2	77	2	160	1	0	0	0
Kingston	Sep	27.1	140	143	188	45	122	45	143	0	0	0	0
Kingston	Oct	26.7	137	137	270	133	150	28	137	0	105	0	0
Kingston	Nov	26.2	128	120	286	166	150	0	120	0	166	0	0
Kingston	Dec	25.2	111	105	237	132	150	0	105	0	132	0	0
Kinshasa	Jan	23.7	100	106	148	42	150	0	106	0	42	0	0
Kinshasa	Feb	23.9	102	97	124	27	150	0	97	0	27	0	0
Kinshasa	Mar	24.3	107	111	170	59	150	0	111	0	59	0	0
Kinshasa	Apr	24.1	104	104	223	119	150	0	104	0	119	0	0
Kinshasa	May	23.7	100	102	130	28	144	-6	102	0	34	0	0
Kinshasa	Jun	21.7	78	77	7	-70	78	-66	74	3	0	0	0
Kinshasa	Jul	20.6	68	69	2	-67	41	-37	39	30	0	0	0
Kinshasa	Aug	21.5	76	79	6	-73	20	-21	26	53	0	0	0
Kinshasa	Sep	23.0	92	92	35	-57	12	-8	42	50	0	0	0
Kinshasa	Oct	23.5	97	102	122	20	32	20	100	2	0	0	0
Kinshasa	Nov	23.5	97	100	244	144	150	118	100	0	26	0	0
Kinshasa	Dec	23.4	96	103	181	78	150	0	103	0	78	0	0
Kolkata	Jan	20.1	46	43	12	-31	54	-18	30	13	0	0	0
Kolkata	Feb	23.0	77	68	20	-48	34	-20	40	28	0	0	0
Kolkata	Mar	27.1	140	145	27	-118	11	-23	50	95	0	0	0
Kolkata	Apr	29.9	162	171	47	-124	3	-8	54	117	0	0	0
Kolkata	May	30.6	166	189	106	-83	1	-2	106	83	0	0	0
Kolkata	Jun	29.8	161	181	258	77	76	75	181	0	0	0	0
Kolkata	Jul	29.1	156	180	345	165	150	74	180	0	91	0	0
Kolkata	Aug	29.0	155	172	310	138	150	0	172	0	138	0	0
Kolkata	Sep	29.0	155	158	287	129	150	0	158	0	129	0	0
Kolkata	Oct	27.8	146	146	162	16	148	-2	146	0	18	0	0
Kolkata	Nov	24.4	96	88	44	-44	104	-44	88	0	0	0	0
Kolkata	Dec	20.7	51	47	9	-38	72	-32	42	5	0	0	0
Kuala Lumpur	Jan	26.9	139	143	124	-19	129	-21	143	0	2	0	0
Kuala Lumpur	Feb	27.5	144	134	123	-11	118	-11	134	0	0	0	0
Kuala Lumpur	Mar	27.8	146	152	189	37	150	32	152	0	5	0	0
Kuala Lumpur	Apr	27.9	147	149	210	61	150	0	149	0	61	0	0
Kuala Lumpur	May	27.9	147	155	185	30	150	0	155	0	30	0	0
Kuala Lumpur	Jun	27.7	145	149	136	-13	136	-14	149	0	1	0	0
Kuala Lumpur	Jul	27.2	141	149	160	11	147	11	149	0	0	0	0
Kuala Lumpur	Aug	27.2	141	148	192	44	150	3	148	0	41	0	0
Kuala Lumpur	Sep	27.2	141	143	197	54	150	0	143	0	54	0	0
Kuala Lumpur	Oct	27.1	140	146	244	98	150	0	146	0	98	0	0
Kuala Lumpur	Nov	27.2	141	141	271	130	150	0	141	0	130	0	0
Kuala Lumpur	Dec	26.8	138	142	211	69	150	0	142	0	69	0	0
Kuwait City	Jan	13.7	12	11	19	8	15	9	11	0	0	0	0
Kuwait City	Feb	15.4	18	16	12	-4	15	0	12	4	0	0	0
Kuwait City	Mar	19.0	39	40	11	-29	11	-4	14	26	0	0	0
Kuwait City	Apr	24.4	98	105	11	-94	5	-6	18	87	0	0	0
Kuwait City	May	30.2	163	192	3	-189	1	-4	7	185	0	0	0
Kuwait City	Jun	34.9	183	213	0	-213	0	-1	1	212	0	0	0
Kuwait City	Jul	36.6	185	220	0	-220	0	0	0	220	0	0	0
Kuwait City	Aug	36.2	185	209	0	-209	0	0	0	209	0	0	0
Kuwait City	Sep	33.0	177	181	0	-181	0	0	0	181	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Kuwait City	Oct	27.7	145	143	1	-142	0	0	1	142	0	0	0
Kuwait City	Nov	21.4	61	54	18	-36	0	0	17	37	0	0	0
Kuwait City	Dec	15.6	19	17	21	4	6	6	14	3	0	0	0
Lagos	Jan	27.0	139	141	20	-121	7	-16	36	105	0	0	0
Lagos	Feb	28.1	149	138	39	-99	3	-4	43	95	0	0	0
Lagos	Mar	28.6	153	159	80	-79	1	-2	80	79	0	0	0
Lagos	Apr	28.3	150	154	141	-13	5	4	136	18	0	0	0
Lagos	May	27.4	143	153	223	70	73	68	153	0	0	0	0
Lagos	Jun	26.2	127	132	332	200	150	77	132	0	123	0	0
Lagos	Jul	25.1	107	115	271	156	150	0	115	0	156	0	0
Lagos	Aug	24.9	104	111	118	7	145	-5	111	0	12	0	0
Lagos	Sep	25.5	114	115	202	87	150	5	115	0	82	0	0
Lagos	Oct	26.1	125	128	193	65	150	0	128	0	65	0	0
Lagos	Nov	27.2	141	139	63	-76	79	-71	135	4	0	0	0
Lagos	Dec	27.1	140	142	15	-127	23	-56	71	71	0	0	0
Lima	Jan	21.3	90	98	12	-86	0	0	11	87	0	0	0
Lima	Feb	21.7	93	90	17	-73	0	0	17	73	0	0	0
Lima	Mar	21.7	93	97	17	-80	0	0	17	80	0	0	0
Lima	Apr	19.9	79	78	7	-71	0	0	7	71	0	0	0
Lima	May	17.5	62	62	7	-55	0	0	7	55	0	0	0
Lima	Jun	15.4	49	47	9	-38	0	0	8	39	0	0	0
Lima	Jul	14.5	44	44	14	-30	0	0	14	30	0	0	0
Lima	Aug	14.7	45	46	12	-34	0	0	12	34	0	0	0
Lima	Sep	15.0	47	47	13	-34	0	0	13	34	0	0	0
Lima	Oct	16.1	54	57	8	-49	0	0	8	49	0	0	0
Lima	Nov	17.5	62	66	11	-55	0	0	11	55	0	0	0
Lima	Dec	19.5	76	84	11	-73	0	0	11	73	0	0	0
Lisbon	Jan	10.9	29	24	98	74	150	24	24	0	50	0	0
Lisbon	Feb	11.8	33	27	74	47	150	0	27	0	47	0	0
Lisbon	Mar	14.0	44	45	75	30	150	0	45	0	30	0	0
Lisbon	Apr	15.8	54	60	55	-5	143	-7	60	0	2	0	0
Lisbon	May	17.4	64	79	48	-31	112	-31	79	0	0	0	0
Lisbon	Jun	20.9	89	110	15	-95	45	-67	82	28	0	0	0
Lisbon	Jul	22.8	103	129	3	-126	13	-32	35	94	0	0	0
Lisbon	Aug	22.9	104	122	2	-120	4	-9	11	111	0	0	0
Lisbon	Sep	21.7	95	97	21	-76	2	-2	23	74	0	0	0
Lisbon	Oct	18.5	72	68	61	-7	5	3	58	10	0	0	0
Lisbon	Nov	15.2	51	42	78	36	39	34	42	0	0	0	0
Lisbon	Dec	11.6	32	26	113	87	126	87	26	0	0	0	0
Ljubljana	Jan	-1.6	0	0	94	0	150	0	0	0	0	0	148
Ljubljana	Feb	-0.8	0	0	84	91	150	0	0	0	91	41	141
Ljubljana	Mar	3.1	15	15	115	241	150	0	15	0	241	141	0
Ljubljana	Apr	7.7	38	42	122	80	150	0	42	0	80	0	0
Ljubljana	May	12.3	61	79	162	83	150	0	79	0	83	0	0
Ljubljana	Jun	16.0	80	105	139	34	150	0	105	0	34	0	0
Ljubljana	Jul	18.2	91	121	98	-23	127	-23	121	0	0	0	0
Ljubljana	Aug	17.3	87	105	147	42	150	23	105	0	19	0	0
Ljubljana	Sep	13.5	67	69	164	95	150	0	69	0	95	0	0
Ljubljana	Oct	9.0	44	41	200	159	150	0	41	0	159	0	0
Ljubljana	Nov	4.5	22	17	164	147	150	0	17	0	147	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Ljubljana	Dec	-0.8	0	0	124	70	150	0	0	0	70	0	54
London	Jan	3.7	17	12	68	56	150	0	12	0	56	0	0
London	Feb	3.7	17	13	49	36	150	0	13	0	36	0	0
London	Mar	6.0	28	28	49	21	150	0	28	0	21	0	0
London	Apr	8.3	39	45	51	6	149	-1	45	0	7	0	0
London	May	11.5	55	74	46	-28	122	-27	74	0	0	0	0
London	Jun	14.5	71	97	49	-48	78	-44	92	5	0	0	0
London	Jul	16.7	82	113	52	-61	43	-35	86	27	0	0	0
London	Aug	16.3	80	99	57	-42	29	-14	72	27	0	0	0
London	Sep	13.8	67	70	60	-10	27	-2	62	8	0	0	0
London	Oct	10.5	50	46	70	24	50	23	46	0	0	0	0
London	Nov	6.7	31	23	78	55	105	55	23	0	0	0	0
London	Dec	4.7	21	15	72	57	150	45	15	0	12	0	0
Los Angeles, CA	Jan	13.8	40	35	54	19	23	19	35	0	0	0	0
Los Angeles, CA	Feb	14.6	44	38	56	18	41	18	38	0	0	0	0
Los Angeles, CA	Mar	15.1	47	48	53	5	48	7	47	1	0	0	0
Los Angeles, CA	Apr	16.5	55	60	18	-42	32	-16	34	26	0	0	0
Los Angeles, CA	May	18.0	65	78	6	-72	16	-16	22	56	0	0	0
Los Angeles, CA	Jun	19.9	79	94	1	-93	6	-10	11	83	0	0	0
Los Angeles, CA	Jul	22.2	96	118	0	-118	2	-4	5	113	0	0	0
Los Angeles, CA	Aug	22.9	102	118	2	-116	1	-1	3	115	0	0	0
Los Angeles, CA	Sep	22.2	96	99	5	-94	0	-1	5	94	0	0	0
Los Angeles, CA	Oct	20.0	79	77	8	-69	0	0	8	69	0	0	0
Los Angeles, CA	Nov	16.8	57	49	32	-17	0	0	31	18	0	0	0
Los Angeles, CA	Dec	14.2	42	36	38	2	4	4	34	2	0	0	0
Madrid	Jan	5.5	14	11	48	37	133	37	11	0	0	0	0
Madrid	Feb	6.8	19	16	47	31	150	17	16	0	14	0	0
Madrid	Mar	9.5	31	31	44	13	150	0	31	0	13	0	0
Madrid	Apr	11.4	40	44	48	4	149	-1	44	0	5	0	0
Madrid	May	15.7	63	78	54	-24	125	-24	78	0	0	0	0
Madrid	Jun	20.1	90	113	31	-82	58	-67	98	15	0	0	0
Madrid	Jul	24.1	117	149	13	-136	16	-42	56	93	0	0	0
Madrid	Aug	23.7	114	135	13	-122	5	-11	24	111	0	0	0
Madrid	Sep	19.8	88	91	34	-57	3	-2	36	55	0	0	0
Madrid	Oct	14.2	55	52	55	3	10	7	48	4	0	0	0
Madrid	Nov	8.9	28	23	55	32	42	32	23	0	0	0	0
Madrid	Dec	6.2	16	13	67	54	96	54	13	0	0	0	0
Manama	Jan	17.3	26	24	17	-7	0	0	17	7	0	0	0
Manama	Feb	18.1	30	27	17	-10	0	0	17	10	0	0	0
Manama	Mar	21.1	55	57	17	-40	0	0	17	40	0	0	0
Manama	Apr	25.3	110	117	7	-110	0	0	7	110	0	0	0
Manama	May	30.1	163	188	2	-186	0	0	2	186	0	0	0
Manama	Jun	33.2	178	204	0	-204	0	0	0	204	0	0	0
Manama	Jul	34.3	181	212	0	-212	0	0	0	212	0	0	0
Manama	Aug	34.1	181	203	0	-203	0	0	0	203	0	0	0
Manama	Sep	32.2	174	177	0	-177	0	0	0	177	0	0	0
Manama	Oct	29.0	155	154	1	-153	0	0	1	153	0	0	0
Manama	Nov	24.2	93	84	3	-81	0	0	2	82	0	0	0
Manama	Dec	19.3	39	35	11	-24	0	0	11	24	0	0	0
Manila	Jan	26.1	122	119	73	-46	104	-46	119	0	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Manila	Feb	26.4	128	116	34	-82	48	-56	91	25	0	0	0
Manila	Mar	27.6	145	150	32	-118	15	-33	65	85	0	0	0
Manila	Apr	28.9	155	161	39	-122	5	-10	49	112	0	0	0
Manila	May	29.3	158	174	161	-13	9	4	156	18	0	0	0
Manila	Jun	28.6	153	165	239	74	82	73	165	0	0	0	0
Manila	Jul	27.9	147	163	245	82	150	68	163	0	14	0	0
Manila	Aug	27.8	146	159	359	200	150	0	159	0	200	0	0
Manila	Sep	27.6	145	147	262	115	150	0	147	0	115	0	0
Manila	Oct	27.6	145	147	279	132	150	0	147	0	132	0	0
Manila	Nov	27.2	141	134	215	81	150	0	134	0	81	0	0
Manila	Dec	26.5	135	131	177	46	150	0	131	0	46	0	0
Melbourne	Jan	18.6	82	103	56	-47	92	-46	102	1	0	0	0
Melbourne	Feb	18.9	84	89	48	-41	62	-30	78	11	0	0	0
Melbourne	Mar	17.4	75	80	65	-15	54	-8	73	7	0	0	0
Melbourne	Apr	14.4	58	53	78	25	77	23	53	0	0	0	0
Melbourne	May	11.8	44	38	102	64	140	63	38	0	0	0	0
Melbourne	Jun	9.5	32	26	95	69	150	10	26	0	59	0	0
Melbourne	Jul	9.0	30	25	102	77	150	0	25	0	77	0	0
Melbourne	Aug	9.7	33	31	105	74	150	0	31	0	74	0	0
Melbourne	Sep	11.1	40	40	107	67	150	0	40	0	67	0	0
Melbourne	Oct	13.3	52	59	100	41	150	0	59	0	41	0	0
Melbourne	Nov	14.7	59	71	88	17	150	0	71	0	17	0	0
Melbourne	Dec	16.9	72	92	80	-12	138	-12	92	0	0	0	0
Mexico City	Jan	8.4	37	36	11	-25	89	-23	34	2	0	0	0
Mexico City	Feb	9.6	43	39	11	-28	68	-21	31	8	0	0	0
Mexico City	Mar	11.5	53	55	15	-40	46	-22	36	19	0	0	0
Mexico City	Apr	12.7	59	62	41	-21	38	-8	49	13	0	0	0
Mexico City	May	12.9	60	68	87	19	56	18	68	0	0	0	0
Mexico City	Jun	12.4	58	64	187	123	150	94	64	0	29	0	0
Mexico City	Jul	11.2	51	58	236	178	150	0	58	0	178	0	0
Mexico City	Aug	11.4	52	58	208	150	150	0	58	0	150	0	0
Mexico City	Sep	11.3	52	53	186	133	150	0	53	0	133	0	0
Mexico City	Oct	11.0	50	50	76	26	150	0	50	0	26	0	0
Mexico City	Nov	9.9	45	42	25	-17	133	-17	42	0	0	0	0
Mexico City	Dec	8.8	39	37	16	-21	112	-21	37	0	0	0	0
Milan	Jan	0.7	1	1	62	61	150	0	1	0	61	0	0
Milan	Feb	3.4	9	7	47	40	150	0	7	0	40	0	0
Milan	Mar	8.5	29	30	51	21	150	0	30	0	21	0	0
Milan	Apr	12.2	47	53	64	11	149	-1	53	0	12	0	0
Milan	May	16.6	72	92	66	-26	124	-25	92	0	0	0	0
Milan	Jun	20.5	95	124	60	-64	69	-55	116	8	0	0	0
Milan	Jul	23.1	112	147	44	-103	25	-44	87	60	0	0	0
Milan	Aug	22.2	106	128	53	-75	12	-13	66	62	0	0	0
Milan	Sep	18.5	83	86	68	-18	11	-1	69	17	0	0	0
Milan	Oct	13.1	52	48	87	39	49	38	48	0	0	0	0
Milan	Nov	6.4	20	16	71	55	104	55	16	0	0	0	0
Milan	Dec	2.3	5	4	58	54	150	46	4	0	8	0	0
Minsk	Jan	-6.2	0	0	39	0	92	0	0	0	0	0	99
Minsk	Feb	-5.4	0	0	33	1	92	0	0	0	0	0	131
Minsk	Mar	-1.4	0	0	39	75	150	58	0	0	17	57	95

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Minsk	Apr	6.1	32	37	42	100	145	-5	37	0	105	95	0
Minsk	May	12.7	66	87	62	-25	120	-25	87	0	0	0	0
Minsk	Jun	16.4	84	115	77	-38	85	-35	112	3	0	0	0
Minsk	Jul	17.7	91	124	82	-42	57	-28	110	14	0	0	0
Minsk	Aug	16.7	86	106	71	-35	40	-17	88	18	0	0	0
Minsk	Sep	11.7	61	63	59	-4	40	0	60	3	0	0	0
Minsk	Oct	6.2	33	30	47	17	56	16	30	0	0	0	0
Minsk	Nov	0.2	1	1	49	35	92	36	1	0	0	0	13
Minsk	Dec	-3.9	0	0	47	0	92	0	0	0	0	0	60
Montevideo	Jan	22.9	104	128	87	-41	78	-36	124	4	0	0	0
Montevideo	Feb	22.4	101	104	99	-5	73	-5	103	1	0	0	0
Montevideo	Mar	20.9	89	95	92	-3	72	-1	94	1	0	0	0
Montevideo	Apr	17.3	65	60	104	44	115	43	60	0	0	0	0
Montevideo	May	14.1	46	40	91	51	150	35	40	0	16	0	0
Montevideo	Jun	11.2	31	25	94	69	150	0	25	0	69	0	0
Montevideo	Jul	10.8	29	25	80	55	150	0	25	0	55	0	0
Montevideo	Aug	11.6	33	31	88	57	150	0	31	0	57	0	0
Montevideo	Sep	13.2	41	41	86	45	150	0	41	0	45	0	0
Montevideo	Oct	15.9	56	63	92	29	150	0	63	0	29	0	0
Montevideo	Nov	18.7	74	86	92	6	148	-2	86	0	8	0	0
Montevideo	Dec	21.4	93	116	81	-35	114	-34	116	0	0	0	0
Montreal	Jan	-9.5	0	0	70	-1	150	0	0	0	0	0	159
Montreal	Feb	-8.4	0	0	57	1	150	0	0	0	0	0	215
Montreal	Mar	-2.3	0	0	70	68	150	0	0	0	68	41	217
Montreal	Apr	5.9	26	30	81	268	150	0	30	0	268	217	0
Montreal	May	13.1	63	80	74	-6	141	-9	80	0	3	0	0
Montreal	Jun	18.5	91	119	94	-25	116	-25	119	0	0	0	0
Montreal	Jul	21.1	105	138	84	-54	69	-47	131	7	0	0	0
Montreal	Aug	19.6	97	117	99	-18	58	-11	110	7	0	0	0
Montreal	Sep	14.8	71	74	97	23	81	23	74	0	0	0	0
Montreal	Oct	8.6	40	37	81	44	125	44	37	0	0	0	0
Montreal	Nov	2.2	9	7	84	66	150	25	7	0	41	0	11
Montreal	Dec	-6.0	0	0	77	0	150	0	0	0	0	0	88
Moscow	Jan	-8.7	0	0	45	0	95	0	0	0	0	0	128
Moscow	Feb	-7.8	0	0	37	0	95	0	0	0	0	0	165
Moscow	Mar	-2.6	0	0	34	42	138	43	0	0	0	32	157
Moscow	Apr	5.8	31	35	39	161	144	6	35	0	155	157	0
Moscow	May	12.7	66	88	53	-35	110	-34	88	0	0	0	0
Moscow	Jun	16.9	87	118	76	-42	74	-36	112	6	0	0	0
Moscow	Jul	18.4	94	129	83	-46	47	-27	109	20	0	0	0
Moscow	Aug	16.6	85	106	80	-26	36	-11	90	16	0	0	0
Moscow	Sep	11.0	57	60	62	2	41	5	58	2	0	0	0
Moscow	Oct	5.0	27	24	61	37	77	36	24	0	0	0	0
Moscow	Nov	-1.9	0	0	53	19	95	18	0	0	0	0	34
Moscow	Dec	-6.1	0	0	48	-1	95	0	0	0	0	0	83
Mumbai	Jan	23.4	82	79	1	-78	6	-8	8	71	0	0	0
Mumbai	Feb	24.2	94	84	1	-83	3	-3	4	80	0	0	0
Mumbai	Mar	26.6	136	140	1	-139	1	-2	3	137	0	0	0
Mumbai	Apr	28.6	153	160	4	-156	0	-1	4	156	0	0	0
Mumbai	May	29.9	162	181	18	-163	0	0	18	163	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Mumbai	Jun	28.6	153	169	541	372	150	150	159	10	232	0	0
Mumbai	Jul	27.1	140	159	1065	906	150	0	159	0	906	0	0
Mumbai	Aug	26.7	137	150	722	572	150	0	150	0	572	0	0
Mumbai	Sep	26.9	139	141	386	245	150	0	141	0	245	0	0
Mumbai	Oct	27.6	145	145	95	-50	90	-60	144	1	11	0	0
Mumbai	Nov	26.4	130	122	17	-105	33	-57	75	47	0	0	0
Mumbai	Dec	24.6	100	94	4	-90	14	-19	23	71	0	0	0
Nagoya	Jan	5.8	10	9	58	49	150	0	9	0	49	0	0
Nagoya	Feb	6.3	12	10	78	68	150	0	10	0	68	0	0
Nagoya	Mar	9.4	23	24	146	122	150	0	24	0	122	0	0
Nagoya	Apr	14.4	48	52	184	132	150	0	52	0	132	0	0
Nagoya	May	18.3	72	86	191	105	150	0	86	0	105	0	0
Nagoya	Jun	21.7	96	115	267	152	150	0	115	0	152	0	0
Nagoya	Jul	25.4	125	153	212	59	150	0	153	0	59	0	0
Nagoya	Aug	26.8	138	159	195	36	150	0	159	0	36	0	0
Nagoya	Sep	23.8	112	115	284	169	150	0	115	0	169	0	0
Nagoya	Oct	18.5	73	70	169	99	150	0	70	0	99	0	0
Nagoya	Nov	13.5	43	37	105	68	150	0	37	0	68	0	0
Nagoya	Dec	8.4	19	16	60	44	150	0	16	0	44	0	0
Naihati	Jan	19.4	41	39	12	-27	54	-16	28	11	0	0	0
Naihati	Feb	22.3	69	61	20	-41	37	-17	38	23	0	0	0
Naihati	Mar	26.8	138	142	29	-113	12	-25	53	89	0	0	0
Naihati	Apr	30.2	163	173	56	-117	4	-8	63	110	0	0	0
Naihati	May	30.6	166	189	116	-73	2	-2	116	73	0	0	0
Naihati	Jun	29.7	160	181	267	86	86	84	181	0	0	0	0
Naihati	Jul	29.1	156	180	341	161	150	64	180	0	97	0	0
Naihati	Aug	28.9	155	172	311	139	150	0	172	0	139	0	0
Naihati	Sep	28.9	155	158	276	118	150	0	158	0	118	0	0
Naihati	Oct	27.7	145	145	144	-1	140	-10	145	0	9	0	0
Naihati	Nov	23.9	90	82	39	-43	97	-43	82	0	0	0	0
Naihati	Dec	20.0	46	43	8	-35	70	-27	36	7	0	0	0
Nairobi	Jan	18.7	70	73	51	-22	86	-20	71	2	0	0	0
Nairobi	Feb	19.5	75	71	37	-34	62	-24	61	10	0	0	0
Nairobi	Mar	19.9	78	82	79	-3	62	0	78	4	0	0	0
Nairobi	Apr	19.5	75	76	163	87	150	88	76	0	0	0	0
Nairobi	May	18.3	67	69	73	4	146	-4	69	0	8	0	0
Nairobi	Jun	17.0	58	58	13	-45	101	-45	58	0	0	0	0
Nairobi	Jul	16.0	52	54	7	-47	64	-37	43	11	0	0	0
Nairobi	Aug	16.5	55	57	5	-52	39	-25	30	27	0	0	0
Nairobi	Sep	17.8	63	64	7	-57	23	-16	23	41	0	0	0
Nairobi	Oct	19.1	72	76	47	-29	18	-5	50	26	0	0	0
Nairobi	Nov	18.9	71	72	147	75	93	75	72	0	0	0	0
Nairobi	Dec	18.4	67	71	83	12	106	13	71	0	0	0	0
New York, NY	Jan	0.2	0	0	96	96	150	0	0	0	96	0	0
New York, NY	Feb	1.4	2	2	86	84	150	0	2	0	84	0	0
New York, NY	Mar	4.7	13	13	122	109	150	0	13	0	109	0	0
New York, NY	Apr	11.0	41	45	95	50	150	0	45	0	50	0	0
New York, NY	May	16.5	71	88	107	19	149	-1	88	0	20	0	0
New York, NY	Jun	21.6	102	128	84	-44	107	-42	128	0	0	0	0
New York, NY	Jul	24.3	119	152	115	-37	75	-32	147	5	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
New York, NY	Aug	23.4	113	134	113	-21	61	-14	127	7	0	0	0
New York, NY	Sep	19.1	86	89	113	24	85	24	89	0	0	0	0
New York, NY	Oct	13.5	54	51	91	40	124	39	51	0	0	0	0
New York, NY	Nov	8.3	28	23	113	90	150	26	23	0	64	0	0
New York, NY	Dec	2.2	5	4	97	93	150	0	4	0	93	0	0
Osaka	Jan	1.8	3	2	80	78	150	0	2	0	78	0	0
Osaka	Feb	2.3	4	3	87	84	150	0	3	0	84	0	0
Osaka	Mar	5.8	15	16	119	103	150	0	16	0	103	0	0
Osaka	Apr	11.8	42	46	152	106	150	0	46	0	106	0	0
Osaka	May	16.6	69	83	159	76	150	0	83	0	76	0	0
Osaka	Jun	20.3	92	110	259	149	150	0	110	0	149	0	0
Osaka	Jul	24.3	119	145	250	105	150	0	145	0	105	0	0
Osaka	Aug	25.0	124	143	176	33	150	0	143	0	33	0	0
Osaka	Sep	21.1	97	99	244	145	150	0	99	0	145	0	0
Osaka	Oct	15.1	60	58	141	83	150	0	58	0	83	0	0
Osaka	Nov	9.5	31	26	98	72	150	0	26	0	72	0	0
Osaka	Dec	4.6	11	9	79	70	150	0	9	0	70	0	0
Ottawa	Jan	-9.9	0	0	67	0	149	0	0	0	0	0	153
Ottawa	Feb	-8.6	0	0	53	0	149	0	0	0	0	0	206
Ottawa	Mar	-2.5	0	0	61	59	150	1	0	0	58	38	208
Ottawa	Apr	6.0	27	30	72	250	150	0	30	0	250	208	0
Ottawa	May	13.2	63	81	68	-13	135	-15	81	0	2	0	0
Ottawa	Jun	18.2	90	117	80	-37	99	-36	117	0	0	0	0
Ottawa	Jul	20.8	103	136	85	-51	60	-39	123	13	0	0	0
Ottawa	Aug	19.6	97	117	88	-29	45	-15	103	14	0	0	0
Ottawa	Sep	14.7	71	74	91	17	62	17	74	0	0	0	0
Ottawa	Oct	8.5	39	37	73	36	98	36	37	0	0	0	0
Ottawa	Nov	1.9	8	6	83	64	149	51	6	0	13	0	13
Ottawa	Dec	-6.3	0	0	73	0	149	0	0	0	0	0	86
Panama City	Jan	25.0	110	110	85	-25	121	-29	110	0	4	0	0
Panama City	Feb	25.4	116	107	47	-60	69	-52	99	8	0	0	0
Panama City	Mar	26.0	126	131	39	-92	28	-41	79	52	0	0	0
Panama City	Apr	26.4	133	136	102	-34	26	-2	101	35	0	0	0
Panama City	May	26.0	126	136	349	213	150	124	136	0	89	0	0
Panama City	Jun	25.5	118	125	343	218	150	0	125	0	218	0	0
Panama City	Jul	25.7	121	132	329	197	150	0	132	0	197	0	0
Panama City	Aug	25.6	119	128	367	239	150	0	128	0	239	0	0
Panama City	Sep	25.4	116	118	387	269	150	0	118	0	269	0	0
Panama City	Oct	25.1	112	114	466	352	150	0	114	0	352	0	0
Panama City	Nov	25.0	110	107	333	226	150	0	107	0	226	0	0
Panama City	Dec	25.1	112	111	225	114	150	0	111	0	114	0	0
Paris	Jan	3.5	13	10	58	48	150	21	10	0	27	0	0
Paris	Feb	4.2	16	13	48	35	150	0	13	0	35	0	0
Paris	Mar	6.9	28	29	52	23	150	0	29	0	23	0	0
Paris	Apr	9.5	41	47	49	2	149	-1	47	0	3	0	0
Paris	May	13.8	63	83	63	-20	129	-20	83	0	0	0	0
Paris	Jun	16.5	78	105	56	-49	83	-46	102	3	0	0	0
Paris	Jul	18.9	91	123	57	-66	44	-39	96	27	0	0	0
Paris	Aug	18.5	89	109	51	-58	25	-19	70	39	0	0	0
Paris	Sep	15.5	72	75	62	-13	22	-3	65	10	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Paris	Oct	11.2	50	45	60	15	36	14	45	0	0	0	0
Paris	Nov	6.6	27	20	62	42	78	42	20	0	0	0	0
Paris	Dec	4.1	15	11	63	52	129	51	11	0	0	0	0
Phnom Penh	Jan	26.4	127	126	8	-118	18	-40	47	79	0	0	0
Phnom Penh	Feb	27.4	143	131	6	-125	5	-13	19	112	0	0	0
Phnom Penh	Mar	28.9	155	161	26	-135	1	-4	30	131	0	0	0
Phnom Penh	Apr	29.7	160	165	73	-92	1	0	73	92	0	0	0
Phnom Penh	May	28.9	155	169	167	-2	6	5	161	8	0	0	0
Phnom Penh	Jun	28.4	151	161	148	-13	6	0	149	12	0	0	0
Phnom Penh	Jul	27.9	147	161	150	-11	5	-1	151	10	0	0	0
Phnom Penh	Aug	27.8	146	157	154	-3	9	4	150	7	0	0	0
Phnom Penh	Sep	27.5	144	146	210	64	72	63	146	0	0	0	0
Phnom Penh	Oct	27.3	142	145	263	118	150	78	145	0	40	0	0
Phnom Penh	Nov	26.9	139	134	124	-10	133	-17	134	0	7	0	0
Phnom Penh	Dec	26.1	121	119	29	-90	58	-75	105	14	0	0	0
Phoenix, AZ	Jan	11.2	15	13	26	13	22	13	13	0	0	0	0
Phoenix, AZ	Feb	13.1	23	19	21	2	24	2	18	1	0	0	0
Phoenix, AZ	Mar	15.5	35	35	22	-13	22	-2	25	10	0	0	0
Phoenix, AZ	Apr	19.8	64	70	7	-63	12	-10	17	53	0	0	0
Phoenix, AZ	May	24.6	112	134	3	-131	3	-9	12	122	0	0	0
Phoenix, AZ	Jun	29.8	161	193	1	-192	1	-2	4	189	0	0	0
Phoenix, AZ	Jul	32.8	176	215	17	-198	0	-1	17	198	0	0	0
Phoenix, AZ	Aug	31.7	172	197	33	-164	0	0	32	165	0	0	0
Phoenix, AZ	Sep	28.9	155	159	20	-139	0	0	20	139	0	0	0
Phoenix, AZ	Oct	22.9	93	90	17	-73	0	0	17	73	0	0	0
Phoenix, AZ	Nov	15.9	37	32	16	-16	0	0	16	16	0	0	0
Phoenix, AZ	Dec	11.5	16	14	23	9	9	9	14	0	0	0	0
Port-of-Spain	Jan	25.2	111	111	87	-24	126	-24	111	0	0	0	0
Port-of-Spain	Feb	25.4	114	105	59	-46	83	-43	102	3	0	0	0
Port-of-Spain	Mar	26.1	126	131	46	-85	37	-46	92	39	0	0	0
Port-of-Spain	Apr	26.7	137	141	56	-85	16	-21	76	65	0	0	0
Port-of-Spain	May	27.1	140	152	104	-48	11	-5	108	44	0	0	0
Port-of-Spain	Jun	26.6	136	144	244	100	110	99	144	0	0	0	0
Port-of-Spain	Jul	26.5	135	147	239	92	150	40	147	0	52	0	0
Port-of-Spain	Aug	26.6	136	146	232	86	150	0	146	0	86	0	0
Port-of-Spain	Sep	26.8	138	139	178	39	150	0	139	0	39	0	0
Port-of-Spain	Oct	26.6	136	139	191	52	150	0	139	0	52	0	0
Port-of-Spain	Nov	26.2	128	124	220	96	150	0	124	0	96	0	0
Port-of-Spain	Dec	25.5	116	115	172	57	150	0	115	0	57	0	0
Portland, OR	Jan	4.6	16	13	188	175	150	0	13	0	175	0	0
Portland, OR	Feb	6.0	23	18	130	112	150	0	18	0	112	0	0
Portland, OR	Mar	7.7	30	31	126	95	150	0	31	0	95	0	0
Portland, OR	Apr	10.1	42	47	88	41	150	0	47	0	41	0	0
Portland, OR	May	13.2	58	75	67	-8	140	-10	75	0	2	0	0
Portland, OR	Jun	16.1	74	96	47	-49	92	-48	95	1	0	0	0
Portland, OR	Jul	19.1	91	120	17	-103	34	-58	75	45	0	0	0
Portland, OR	Aug	19.2	91	110	22	-88	14	-20	41	69	0	0	0
Portland, OR	Sep	16.5	76	79	55	-24	12	-2	57	22	0	0	0
Portland, OR	Oct	12.0	52	48	92	44	54	42	48	0	0	0	0
Portland, OR	Nov	7.4	29	23	198	175	150	96	23	0	79	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Portland, OR	Dec	4.7	17	13	180	167	150	0	13	0	167	0	0
Prague	Jan	-2.1	0	0	27	1	91	0	0	0	0	0	39
Prague	Feb	-1.2	0	0	27	37	129	38	0	0	0	25	29
Prague	Mar	2.6	14	14	32	47	150	21	14	0	26	29	0
Prague	Apr	7.2	37	42	42	0	148	-2	42	0	2	0	0
Prague	May	12.4	63	84	70	-14	134	-14	84	0	0	0	0
Prague	Jun	15.6	79	108	76	-32	101	-33	108	0	0	0	0
Prague	Jul	17.4	89	121	82	-39	69	-32	114	7	0	0	0
Prague	Aug	16.6	84	105	77	-28	53	-16	94	11	0	0	0
Prague	Sep	12.6	64	67	50	-17	45	-8	58	9	0	0	0
Prague	Oct	7.7	40	36	40	4	49	4	36	0	0	0	0
Prague	Nov	2.7	14	10	33	23	71	22	10	0	0	0	0
Prague	Dec	-0.7	0	0	33	20	91	20	0	0	0	0	13
Quezon City	Jan	25.5	115	114	103	-11	137	-13	114	0	2	0	0
Quezon City	Feb	25.6	117	107	77	-30	108	-29	107	0	0	0	0
Quezon City	Mar	26.4	131	136	69	-67	57	-51	120	16	0	0	0
Quezon City	Apr	27.4	143	147	81	-66	30	-27	108	39	0	0	0
Quezon City	May	27.8	146	159	131	-28	23	-7	137	22	0	0	0
Quezon City	Jun	27.3	142	151	193	42	65	42	151	0	0	0	0
Quezon City	Jul	26.8	138	150	225	75	139	74	150	0	0	0	0
Quezon City	Aug	26.9	139	149	195	46	150	11	149	0	35	0	0
Quezon City	Sep	26.8	138	139	207	68	150	0	139	0	68	0	0
Quezon City	Oct	26.5	135	138	222	84	150	0	138	0	84	0	0
Quezon City	Nov	26.3	129	125	195	70	150	0	125	0	70	0	0
Quezon City	Dec	25.8	120	119	161	42	150	0	119	0	42	0	0
Quito	Jan	6.8	45	47	86	39	150	0	47	0	39	0	0
Quito	Feb	6.6	44	42	107	65	150	0	42	0	65	0	0
Quito	Mar	6.7	45	47	127	80	150	0	47	0	80	0	0
Quito	Apr	6.9	46	46	142	96	150	0	46	0	96	0	0
Quito	May	6.9	46	48	107	59	150	0	48	0	59	0	0
Quito	Jun	6.3	43	43	77	34	150	0	43	0	34	0	0
Quito	Jul	6.2	42	44	67	23	150	0	44	0	23	0	0
Quito	Aug	6.6	44	46	59	13	150	0	46	0	13	0	0
Quito	Sep	6.7	45	45	79	34	150	0	45	0	34	0	0
Quito	Oct	6.5	44	46	103	57	150	0	46	0	57	0	0
Quito	Nov	6.6	44	45	103	58	150	0	45	0	58	0	0
Quito	Dec	6.7	45	47	86	39	150	0	47	0	39	0	0
Rabat	Jan	11.6	29	25	71	46	125	46	25	0	0	0	0
Rabat	Feb	12.6	34	29	61	32	150	25	29	0	7	0	0
Rabat	Mar	14.4	43	44	55	11	150	0	44	0	11	0	0
Rabat	Apr	15.6	50	54	48	-6	143	-7	54	0	1	0	0
Rabat	May	18.7	70	84	19	-65	81	-62	81	3	0	0	0
Rabat	Jun	21.6	92	110	4	-106	29	-52	56	54	0	0	0
Rabat	Jul	24.1	112	137	0	-137	8	-21	22	115	0	0	0
Rabat	Aug	24.9	119	137	1	-136	2	-6	7	130	0	0	0
Rabat	Sep	22.9	102	105	8	-97	1	-1	9	96	0	0	0
Rabat	Oct	19.8	78	75	34	-41	1	0	34	41	0	0	0
Rabat	Nov	15.4	49	42	66	24	25	24	42	0	0	0	0
Rabat	Dec	12.9	35	30	84	54	79	54	30	0	0	0	0
Riga	Jan	-4.6	0	0	38	-1	130	0	0	0	0	0	80

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Riga	Feb	-4.5	0	0	31	0	130	0	0	0	0	0	111
Riga	Mar	-0.9	0	0	33	84	150	20	0	0	64	65	60
Riga	Apr	5.4	30	34	44	70	148	-2	34	0	72	60	0
Riga	May	11.5	61	81	48	-33	115	-33	81	0	0	0	0
Riga	Jun	15.4	80	109	63	-46	75	-40	103	6	0	0	0
Riga	Jul	17.2	89	122	76	-46	48	-27	103	19	0	0	0
Riga	Aug	16.4	85	105	71	-34	35	-13	85	20	0	0	0
Riga	Sep	11.9	63	65	74	9	44	9	65	0	0	0	0
Riga	Oct	7.0	38	34	63	29	72	28	34	0	0	0	0
Riga	Nov	1.7	10	7	59	52	123	51	7	0	0	0	0
Riga	Dec	-2.2	0	0	48	7	130	7	0	0	0	0	41
Riyadh	Jan	15.4	15	14	16	2	3	3	13	1	0	0	0
Riyadh	Feb	17.6	26	23	8	-15	2	-1	8	15	0	0	0
Riyadh	Mar	21.9	61	63	30	-33	2	0	31	32	0	0	0
Riyadh	Apr	26.5	135	143	26	-117	1	-1	28	115	0	0	0
Riyadh	May	32.2	174	200	9	-191	0	-1	10	190	0	0	0
Riyadh	Jun	35.1	183	208	0	-208	0	0	0	208	0	0	0
Riyadh	Jul	36.1	185	215	0	-215	0	0	0	215	0	0	0
Riyadh	Aug	35.9	184	206	0	-206	0	0	0	206	0	0	0
Riyadh	Sep	32.9	177	180	0	-180	0	0	0	180	0	0	0
Riyadh	Oct	27.8	146	145	1	-144	0	0	1	144	0	0	0
Riyadh	Nov	21.9	61	56	5	-51	0	0	5	51	0	0	0
Riyadh	Dec	16.7	21	19	10	-9	0	0	10	9	0	0	0
Rome	Jan	5.0	15	12	102	90	150	0	12	0	90	0	0
Rome	Feb	5.3	16	13	106	93	150	0	13	0	93	0	0
Rome	Mar	7.5	25	26	75	49	150	0	26	0	49	0	0
Rome	Apr	10.5	39	44	81	37	150	0	44	0	37	0	0
Rome	May	14.2	59	74	61	-13	136	-14	74	0	1	0	0
Rome	Jun	18.4	83	105	42	-63	77	-59	101	4	0	0	0
Rome	Jul	21.4	101	130	23	-107	27	-50	73	57	0	0	0
Rome	Aug	21.6	103	122	34	-88	12	-15	49	73	0	0	0
Rome	Sep	18.3	82	85	90	5	22	10	79	6	0	0	0
Rome	Oct	13.8	57	53	118	65	86	64	53	0	0	0	0
Rome	Nov	9.3	34	27	132	105	150	64	27	0	41	0	0
Rome	Dec	6.4	21	16	115	99	150	0	16	0	99	0	0
San Salvador	Jan	26.7	137	134	11	-123	8	-20	30	104	0	0	0
San Salvador	Feb	27.0	139	127	8	-119	3	-5	14	113	0	0	0
San Salvador	Mar	27.5	144	149	8	-141	1	-2	10	139	0	0	0
San Salvador	Apr	28.5	152	157	37	-120	0	-1	36	121	0	0	0
San Salvador	May	28.3	150	165	163	-2	14	14	147	18	0	0	0
San Salvador	Jun	27.7	145	157	319	162	150	136	157	0	26	0	0
San Salvador	Jul	27.4	143	158	305	147	150	0	158	0	147	0	0
San Salvador	Aug	27.5	144	155	304	149	150	0	155	0	149	0	0
San Salvador	Sep	27.0	139	141	378	237	150	0	141	0	237	0	0
San Salvador	Oct	27.0	139	142	305	163	150	0	142	0	163	0	0
San Salvador	Nov	26.8	138	132	69	-63	87	-63	130	2	2	0	0
San Salvador	Dec	26.4	129	126	9	-117	28	-59	68	58	0	0	0
Sana'a	Jan	13.6	43	42	4	-38	21	-9	13	29	0	0	0
Sana'a	Feb	15.0	51	46	5	-41	14	-7	11	35	0	0	0
Sana'a	Mar	16.2	58	60	41	-19	11	-3	42	18	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Sana'a	Apr	17.5	66	69	96	27	39	28	69	0	0	0	0
Sana'a	May	18.9	75	84	68	-16	34	-5	73	11	0	0	0
Sana'a	Jun	20.2	84	92	32	-60	19	-15	46	46	0	0	0
Sana'a	Jul	20.2	84	94	100	6	27	8	91	3	0	0	0
Sana'a	Aug	20.3	85	93	144	51	79	52	93	0	0	0	0
Sana'a	Sep	17.7	67	69	29	-40	54	-25	55	14	0	0	0
Sana'a	Oct	14.7	49	50	18	-32	40	-14	32	18	0	0	0
Sana'a	Nov	15.0	51	48	20	-28	31	-9	30	18	0	0	0
Sana'a	Dec	12.2	36	35	30	-5	30	-1	31	4	0	0	0
Santiago	Jan	13.5	73	89	5	-84	17	-22	27	62	0	0	0
Santiago	Feb	12.6	69	72	4	-68	9	-8	12	60	0	0	0
Santiago	Mar	10.6	59	63	9	-54	5	-4	12	51	0	0	0
Santiago	Apr	7.5	44	41	38	-3	11	6	32	9	0	0	0
Santiago	May	4.6	28	25	113	88	98	87	25	0	0	0	0
Santiago	Jun	1.9	13	11	112	101	150	52	11	0	49	0	0
Santiago	Jul	1.6	11	10	122	112	150	0	10	0	112	0	0
Santiago	Aug	2.2	15	14	76	62	150	0	14	0	62	0	0
Santiago	Sep	3.6	23	23	38	15	150	0	23	0	15	0	0
Santiago	Oct	7.3	43	48	23	-25	126	-24	48	0	0	0	0
Santiago	Nov	9.8	55	64	17	-47	82	-44	61	3	0	0	0
Santiago	Dec	12.6	69	86	8	-78	39	-43	51	35	0	0	0
Santo Domingo	Jan	25.2	108	103	43	-60	2	-1	44	59	0	0	0
Santo Domingo	Feb	25.4	111	100	40	-60	1	-1	41	59	0	0	0
Santo Domingo	Mar	26.3	128	132	38	-94	0	-1	39	93	0	0	0
Santo Domingo	Apr	27.3	142	149	56	-93	0	0	55	94	0	0	0
Santo Domingo	May	27.6	145	162	158	-4	5	5	153	9	0	0	0
Santo Domingo	Jun	27.6	145	160	136	-24	4	-1	137	23	0	0	0
Santo Domingo	Jul	28.2	149	169	95	-74	2	-2	97	72	0	0	0
Santo Domingo	Aug	28.7	153	168	134	-34	1	-1	134	34	0	0	0
Santo Domingo	Sep	28.2	149	152	144	-8	1	0	144	8	0	0	0
Santo Domingo	Oct	27.7	145	146	153	7	10	9	145	1	0	0	0
Santo Domingo	Nov	26.7	137	128	81	-47	6	-4	85	43	0	0	0
Santo Domingo	Dec	26.0	122	116	35	-81	3	-3	38	78	0	0	0
Sao Paulo	Jan	19.8	82	94	436	342	150	0	94	0	342	0	0
Sao Paulo	Feb	20.0	83	83	437	354	150	0	83	0	354	0	0
Sao Paulo	Mar	19.1	77	81	393	312	150	0	81	0	312	0	0
Sao Paulo	Apr	17.5	66	64	320	256	150	0	64	0	256	0	0
Sao Paulo	May	15.5	54	51	222	171	150	0	51	0	171	0	0
Sao Paulo	Jun	14.1	46	41	194	153	150	0	41	0	153	0	0
Sao Paulo	Jul	14.4	48	44	183	139	150	0	44	0	139	0	0
Sao Paulo	Aug	14.4	48	46	203	157	150	0	46	0	157	0	0
Sao Paulo	Sep	14.3	47	47	269	222	150	0	47	0	222	0	0
Sao Paulo	Oct	15.8	56	61	326	265	150	0	61	0	265	0	0
Sao Paulo	Nov	17.1	64	71	348	277	150	0	71	0	277	0	0
Sao Paulo	Dec	19.9	82	96	386	290	150	0	96	0	290	0	0
Sarajevo	Jan	-2.1	0	0	62	0	150	0	0	0	0	0	93
Sarajevo	Feb	-0.5	0	0	62	102	150	0	0	0	102	58	53
Sarajevo	Mar	3.2	15	15	67	105	150	0	15	0	105	53	0
Sarajevo	Apr	7.8	38	42	73	31	150	0	42	0	31	0	0
Sarajevo	May	12.7	63	80	78	-2	146	-4	80	0	2	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Sarajevo	Jun	16.0	80	103	89	-14	133	-13	103	0	0	0	0
Sarajevo	Jul	17.8	89	116	77	-39	94	-39	115	1	0	0	0
Sarajevo	Aug	17.7	88	106	68	-38	65	-29	97	9	0	0	0
Sarajevo	Sep	13.8	68	70	87	17	81	16	70	0	0	0	0
Sarajevo	Oct	9.4	46	43	85	42	123	42	43	0	0	0	0
Sarajevo	Nov	4.4	21	17	91	74	150	27	17	0	47	0	0
Sarajevo	Dec	-0.6	0	0	81	50	150	0	0	0	50	0	31
Seattle, WA	Jan	4.1	16	12	163	151	150	0	12	0	151	0	0
Seattle, WA	Feb	5.9	24	19	123	104	150	0	19	0	104	0	0
Seattle, WA	Mar	7.3	31	31	109	78	150	0	31	0	78	0	0
Seattle, WA	Apr	9.6	42	48	83	35	150	0	48	0	35	0	0
Seattle, WA	May	13.0	59	77	57	-20	129	-21	77	0	0	0	0
Seattle, WA	Jun	15.8	74	99	50	-49	83	-46	96	3	0	0	0
Seattle, WA	Jul	17.9	86	115	25	-90	35	-48	74	41	0	0	0
Seattle, WA	Aug	18.0	86	105	33	-72	17	-18	51	54	0	0	0
Seattle, WA	Sep	15.3	72	74	52	-22	15	-2	54	20	0	0	0
Seattle, WA	Oct	11.0	49	45	99	54	67	52	45	0	0	0	0
Seattle, WA	Nov	6.8	28	22	170	148	150	83	22	0	65	0	0
Seattle, WA	Dec	4.5	18	13	166	153	150	0	13	0	153	0	0
Seoul	Jan	-4.4	0	0	22	0	150	0	0	0	0	0	35
Seoul	Feb	-1.8	0	0	27	36	150	0	0	0	36	25	26
Seoul	Mar	3.9	10	10	48	64	150	0	10	0	64	26	0
Seoul	Apr	10.9	40	44	85	41	150	0	44	0	41	0	0
Seoul	May	16.3	70	85	89	4	150	0	85	0	4	0	0
Seoul	Jun	21.0	98	121	135	14	150	0	121	0	14	0	0
Seoul	Jul	24.7	122	152	337	185	150	0	152	0	185	0	0
Seoul	Aug	25.3	126	147	261	114	150	0	147	0	114	0	0
Seoul	Sep	19.7	90	92	143	51	150	0	92	0	51	0	0
Seoul	Oct	12.7	50	48	49	1	148	-2	48	0	3	0	0
Seoul	Nov	5.6	16	14	47	33	150	2	14	0	31	0	0
Seoul	Dec	-1.3	0	0	26	13	150	0	0	0	13	0	13
Shanghai	Jan	4.3	5	5	50	45	150	18	5	0	27	0	0
Shanghai	Feb	5.2	7	6	63	57	150	0	6	0	57	0	0
Shanghai	Mar	8.5	17	18	85	67	150	0	18	0	67	0	0
Shanghai	Apr	14.7	47	50	92	42	150	0	50	0	42	0	0
Shanghai	May	19.8	79	94	130	36	150	0	94	0	36	0	0
Shanghai	Jun	24.2	114	134	153	19	150	0	134	0	19	0	0
Shanghai	Jul	28.6	153	183	153	-30	120	-30	183	0	0	0	0
Shanghai	Aug	28.5	152	173	129	-44	79	-41	170	3	0	0	0
Shanghai	Sep	24.0	112	115	134	19	98	19	115	0	0	0	0
Shanghai	Oct	19.2	75	73	47	-26	77	-21	70	3	0	0	0
Shanghai	Nov	13.3	39	34	52	18	94	17	34	0	0	0	0
Shanghai	Dec	7.9	15	13	51	38	132	38	13	0	0	0	0
Shenzhen	Jan	14.6	28	26	30	4	98	3	26	0	0	0	0
Shenzhen	Feb	15.3	32	28	67	39	136	38	28	0	0	0	0
Shenzhen	Mar	18.2	50	52	84	32	150	14	52	0	18	0	0
Shenzhen	Apr	22.1	84	89	164	75	150	0	89	0	75	0	0
Shenzhen	May	25.7	125	142	269	127	150	0	142	0	127	0	0
Shenzhen	Jun	27.3	142	160	331	171	150	0	160	0	171	0	0
Shenzhen	Jul	28.3	150	173	321	148	150	0	173	0	148	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Shenzhen	Aug	28.1	149	165	337	172	150	0	165	0	172	0	0
Shenzhen	Sep	27.1	140	143	237	94	150	0	143	0	94	0	0
Shenzhen	Oct	24.3	108	107	83	-24	123	-27	107	0	3	0	0
Shenzhen	Nov	20.4	68	62	39	-23	100	-23	62	0	0	0	0
Shenzhen	Dec	16.4	38	35	29	-6	95	-5	35	0	0	0	0
Singapore	Jan	25.8	121	126	230	104	150	0	126	0	104	0	0
Singapore	Feb	26.1	126	118	155	37	150	0	118	0	37	0	0
Singapore	Mar	26.4	131	137	187	50	150	0	137	0	50	0	0
Singapore	Apr	26.3	130	131	236	105	150	0	131	0	105	0	0
Singapore	May	27.1	140	147	218	71	150	0	147	0	71	0	0
Singapore	Jun	26.6	136	138	188	50	150	0	138	0	50	0	0
Singapore	Jul	26.4	131	138	201	63	150	0	138	0	63	0	0
Singapore	Aug	26.5	135	141	168	27	150	0	141	0	27	0	0
Singapore	Sep	26.3	130	131	206	75	150	0	131	0	75	0	0
Singapore	Oct	26.2	128	133	220	87	150	0	133	0	87	0	0
Singapore	Nov	26.0	124	125	286	161	150	0	125	0	161	0	0
Singapore	Dec	25.8	121	126	326	200	150	0	126	0	200	0	0
Sofia	Jan	-1.7	0	0	63	0	150	0	0	0	0	0	63
Sofia	Feb	-1.3	0	0	51	15	150	0	0	0	15	2	99
Sofia	Mar	0.3	3	3	71	167	150	0	3	0	167	99	0
Sofia	Apr	4.2	26	29	65	36	150	0	29	0	36	0	0
Sofia	May	9.0	52	65	87	22	150	0	65	0	22	0	0
Sofia	Jun	12.4	69	87	89	2	147	-3	87	0	5	0	0
Sofia	Jul	14.7	80	103	65	-38	110	-37	103	0	0	0	0
Sofia	Aug	15.2	82	98	55	-43	73	-37	92	6	0	0	0
Sofia	Sep	11.5	64	66	42	-24	58	-15	57	9	0	0	0
Sofia	Oct	7.9	46	43	53	10	67	9	43	0	0	0	0
Sofia	Nov	3.3	21	17	69	52	118	51	17	0	0	0	0
Sofia	Dec	0.2	2	1	62	61	150	32	1	0	29	0	0
St. John’s	Jan	-4.6	0	0	150	0	150	0	0	0	0	0	257
St. John’s	Feb	-5.1	0	0	145	-1	150	0	0	0	0	0	403
St. John’s	Mar	-2.5	0	0	129	19	150	0	0	0	19	2	513
St. John’s	Apr	1.5	12	14	117	317	150	0	14	0	317	214	299
St. John’s	May	5.6	37	48	100	351	150	0	48	0	351	299	0
St. John’s	Jun	10.6	62	83	88	5	148	-2	83	0	7	0	0
St. John’s	Jul	15.0	83	112	81	-31	117	-31	112	0	0	0	0
St. John’s	Aug	15.2	84	103	109	6	122	5	103	0	0	0	0
St. John’s	Sep	11.6	67	70	129	59	150	28	70	0	31	0	0
St. John’s	Oct	6.9	44	40	143	103	150	0	40	0	103	0	0
St. John’s	Nov	2.9	21	16	149	133	150	0	16	0	133	0	0
St. John’s	Dec	-1.9	0	0	146	39	150	0	0	0	39	0	107
St. Petersburg	Jan	-7.1	0	0	40	0	104	0	0	0	0	0	109
St. Petersburg	Feb	-7.1	0	0	31	0	104	0	0	0	0	0	140
St. Petersburg	Mar	-2.9	0	0	32	18	122	18	0	0	0	12	154
St. Petersburg	Apr	3.7	23	26	33	161	147	25	26	0	136	154	0
St. Petersburg	May	10.0	56	74	39	-35	112	-35	74	0	0	0	0
St. Petersburg	Jun	15.1	81	110	61	-49	71	-41	103	7	0	0	0
St. Petersburg	Jul	17.2	91	124	73	-51	43	-28	100	24	0	0	0
St. Petersburg	Aug	15.6	83	103	76	-27	33	-10	86	17	0	0	0
St. Petersburg	Sep	10.7	59	61	66	5	39	6	60	1	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
St. Petersburg	Oct	5.1	30	28	61	33	71	32	28	0	0	0	0
St. Petersburg	Nov	-0.5	0	0	53	33	104	33	0	0	0	0	20
St. Petersburg	Dec	-4.5	0	0	48	-1	104	0	0	0	0	0	69
Stockholm	Jan	-3.0	0	0	41	0	97	0	0	0	0	0	76
Stockholm	Feb	-3.3	0	0	29	0	97	0	0	0	0	0	105
Stockholm	Mar	-0.6	0	0	27	75	150	53	0	0	22	58	57
Stockholm	Apr	3.6	23	26	34	65	147	-3	26	0	68	57	0
Stockholm	May	9.1	52	69	33	-36	111	-36	69	0	0	0	0
Stockholm	Jun	13.8	75	102	45	-57	64	-47	92	10	0	0	0
Stockholm	Jul	16.3	87	119	59	-60	36	-28	87	32	0	0	0
Stockholm	Aug	15.3	82	102	68	-34	26	-10	78	24	0	0	0
Stockholm	Sep	10.8	61	63	53	-10	24	-2	56	7	0	0	0
Stockholm	Oct	6.4	38	35	52	17	40	16	35	0	0	0	0
Stockholm	Nov	1.6	11	8	50	42	81	41	8	0	0	0	0
Stockholm	Dec	-1.4	0	0	50	15	97	16	0	0	0	0	35
Surabaya	Jan	27.2	141	152	318	166	150	76	152	0	90	0	0
Surabaya	Feb	27.1	140	135	267	132	150	0	135	0	132	0	0
Surabaya	Mar	27.4	143	150	251	101	150	0	150	0	101	0	0
Surabaya	Apr	27.8	146	146	161	15	148	-2	146	0	17	0	0
Surabaya	May	27.6	145	147	104	-43	106	-42	147	0	0	0	0
Surabaya	Jun	27.1	140	137	59	-78	50	-56	115	22	0	0	0
Surabaya	Jul	26.6	136	137	26	-111	17	-33	59	78	0	0	0
Surabaya	Aug	26.8	138	141	24	-117	5	-12	35	106	0	0	0
Surabaya	Sep	27.6	145	146	23	-123	2	-3	26	120	0	0	0
Surabaya	Oct	28.5	152	161	45	-116	1	-1	45	116	0	0	0
Surabaya	Nov	28.6	153	158	130	-28	2	1	127	31	0	0	0
Surabaya	Dec	27.5	144	155	229	74	74	72	155	0	0	0	0
Sydney	Jan	21.0	92	113	108	-5	117	-5	113	0	0	0	0
Sydney	Feb	21.0	92	96	127	31	148	31	96	0	0	0	0
Sydney	Mar	19.6	83	88	118	30	150	2	88	0	28	0	0
Sydney	Apr	16.6	63	59	95	36	150	0	59	0	36	0	0
Sydney	May	13.4	45	40	84	44	150	0	40	0	44	0	0
Sydney	Jun	10.9	32	27	127	100	150	0	27	0	100	0	0
Sydney	Jul	9.8	27	24	59	35	150	0	24	0	35	0	0
Sydney	Aug	10.8	32	30	86	56	150	0	30	0	56	0	0
Sydney	Sep	13.2	44	44	52	8	150	0	44	0	8	0	0
Sydney	Oct	15.8	59	66	92	26	150	0	66	0	26	0	0
Sydney	Nov	17.8	71	83	92	9	150	0	83	0	9	0	0
Sydney	Dec	19.8	84	104	77	-27	122	-28	104	0	0	0	0
Tashkent	Jan	-0.9	0	0	63	63	150	16	0	0	47	0	0
Tashkent	Feb	0.0	0	0	75	75	150	0	0	0	75	0	0
Tashkent	Mar	4.8	15	15	105	90	150	0	15	0	90	0	0
Tashkent	Apr	11.9	48	53	95	42	150	0	53	0	42	0	0
Tashkent	May	16.7	74	93	56	-37	115	-35	93	0	0	0	0
Tashkent	Jun	21.7	104	131	20	-111	39	-76	95	36	0	0	0
Tashkent	Jul	24.0	118	152	7	-145	10	-29	37	115	0	0	0
Tashkent	Aug	22.6	110	130	2	-128	3	-7	8	122	0	0	0
Tashkent	Sep	16.7	74	76	7	-69	1	-2	8	68	0	0	0
Tashkent	Oct	10.6	41	39	59	20	25	24	34	5	0	0	0
Tashkent	Nov	4.9	15	12	60	48	72	47	12	0	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Tashkent	Dec	0.2	0	0	62	62	134	62	0	0	0	0	0
Tbilisi	Jan	4.8	8	7	19	12	32	13	7	0	0	0	0
Tbilisi	Feb	5.5	10	9	26	17	49	17	9	0	0	0	0
Tbilisi	Mar	9.3	24	25	35	10	59	10	25	0	0	0	0
Tbilisi	Apr	15.0	53	59	50	-9	55	-4	54	5	0	0	0
Tbilisi	May	19.5	81	102	75	-27	42	-13	87	15	0	0	0
Tbilisi	Jun	23.0	106	135	79	-56	25	-17	96	39	0	0	0
Tbilisi	Jul	26.1	131	167	46	-121	8	-17	63	104	0	0	0
Tbilisi	Aug	25.8	128	152	37	-115	2	-6	42	110	0	0	0
Tbilisi	Sep	21.8	97	100	35	-65	1	-1	36	64	0	0	0
Tbilisi	Oct	16.3	61	57	39	-18	1	0	39	18	0	0	0
Tbilisi	Nov	10.6	30	24	32	8	9	8	24	0	0	0	0
Tbilisi	Dec	6.4	13	10	21	11	19	10	10	0	0	0	0
Tehran	Jan	3.6	3	2	26	24	37	24	2	0	0	0	0
Tehran	Feb	6.1	8	6	18	12	48	11	6	0	0	0	0
Tehran	Mar	10.4	21	22	22	0	51	3	20	2	0	0	0
Tehran	Apr	16.7	54	59	9	-50	32	-19	28	31	0	0	0
Tehran	May	23.3	104	126	11	-115	10	-22	32	94	0	0	0
Tehran	Jun	27.8	146	177	3	-174	2	-8	12	165	0	0	0
Tehran	Jul	31.2	169	208	1	-207	0	-2	3	205	0	0	0
Tehran	Aug	30.0	162	188	2	-186	0	0	2	186	0	0	0
Tehran	Sep	24.9	119	122	1	-121	0	0	1	121	0	0	0
Tehran	Oct	18.5	66	64	10	-54	0	0	10	54	0	0	0
Tehran	Nov	11.1	24	21	24	3	7	7	17	4	0	0	0
Tehran	Dec	7.2	10	9	16	7	13	6	9	0	0	0	0
Tel Aviv	Jan	9.7	20	18	88	70	146	70	18	0	0	0	0
Tel Aviv	Feb	10.6	24	20	81	61	150	4	20	0	57	0	0
Tel Aviv	Mar	12.8	34	35	73	38	150	0	35	0	38	0	0
Tel Aviv	Apr	16.9	57	62	17	-45	107	-43	62	0	0	0	0
Tel Aviv	May	20.8	85	101	3	-98	42	-65	68	33	0	0	0
Tel Aviv	Jun	23.3	105	125	1	-124	13	-29	30	95	0	0	0
Tel Aviv	Jul	24.4	115	138	0	-138	3	-10	9	129	0	0	0
Tel Aviv	Aug	24.8	118	135	0	-135	1	-2	2	133	0	0	0
Tel Aviv	Sep	23.4	106	108	0	-108	0	-1	1	107	0	0	0
Tel Aviv	Oct	21.3	89	86	13	-73	0	0	13	73	0	0	0
Tel Aviv	Nov	16.3	53	46	51	5	10	10	40	6	0	0	0
Tel Aviv	Dec	11.6	28	24	90	66	76	66	24	0	0	0	0
Tokyo	Jan	4.7	8	7	85	78	150	0	7	0	78	0	0
Tokyo	Feb	5.3	10	8	122	114	150	0	8	0	114	0	0
Tokyo	Mar	8.0	19	20	177	157	150	0	20	0	157	0	0
Tokyo	Apr	13.7	46	50	213	163	150	0	50	0	163	0	0
Tokyo	May	18.1	72	88	162	74	150	0	88	0	74	0	0
Tokyo	Jun	20.9	91	111	205	94	150	0	111	0	94	0	0
Tokyo	Jul	25.1	123	151	124	-27	120	-30	151	0	3	0	0
Tokyo	Aug	26.4	133	154	152	-2	116	-4	154	0	0	0	0
Tokyo	Sep	23.3	109	112	273	161	150	34	112	0	127	0	0
Tokyo	Oct	17.4	68	65	304	239	150	0	65	0	239	0	0
Tokyo	Nov	12.6	40	34	164	130	150	0	34	0	130	0	0
Tokyo	Dec	7.5	17	15	113	98	150	0	15	0	98	0	0
Toronto	Jan	-3.4	0	0	65	0	134	0	0	0	0	0	95

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Toronto	Feb	-2.9	0	0	52	0	134	0	0	0	0	0	147
Toronto	Mar	1.2	3	4	66	209	150	16	4	0	193	149	0
Toronto	Apr	7.6	31	34	73	39	150	0	34	0	39	0	0
Toronto	May	13.7	62	78	64	-14	135	-15	78	0	1	0	0
Toronto	Jun	19.3	92	119	71	-48	89	-46	117	2	0	0	0
Toronto	Jul	22.2	109	142	66	-76	43	-46	112	30	0	0	0
Toronto	Aug	21.2	103	124	79	-45	28	-15	94	30	0	0	0
Toronto	Sep	17.4	82	84	78	-6	26	-2	79	5	0	0	0
Toronto	Oct	11.2	49	45	61	16	41	15	45	0	0	0	0
Toronto	Nov	5.3	20	16	74	58	99	58	16	0	0	0	0
Toronto	Dec	-0.8	0	0	65	35	134	35	0	0	0	0	30
Tripoli	Jan	12.2	23	20	55	35	90	35	20	0	0	0	0
Tripoli	Feb	13.3	28	24	27	3	94	4	24	0	0	0	0
Tripoli	Mar	15.9	42	43	27	-16	81	-13	40	3	0	0	0
Tripoli	Apr	18.7	61	66	19	-47	52	-29	48	18	0	0	0
Tripoli	May	22.5	93	111	6	-105	19	-33	39	72	0	0	0
Tripoli	Jun	26.4	133	159	1	-158	4	-15	16	143	0	0	0
Tripoli	Jul	27.7	145	177	0	-177	1	-3	3	174	0	0	0
Tripoli	Aug	28.2	149	172	1	-171	0	-1	1	171	0	0	0
Tripoli	Sep	26.5	135	138	12	-126	0	0	11	127	0	0	0
Tripoli	Oct	22.8	96	93	40	-53	0	0	40	53	0	0	0
Tripoli	Nov	17.6	53	46	54	8	11	11	43	3	0	0	0
Tripoli	Dec	13.4	28	24	69	45	55	44	24	0	0	0	0
Tunis	Jan	11.1	24	21	65	44	106	45	21	0	0	0	0
Tunis	Feb	11.6	26	22	52	30	136	30	22	0	0	0	0
Tunis	Mar	13.4	35	36	45	9	146	10	36	0	0	0	0
Tunis	Apr	15.9	49	54	42	-12	134	-12	54	0	0	0	0
Tunis	May	19.6	74	90	25	-65	75	-59	85	5	0	0	0
Tunis	Jun	23.7	108	132	12	-120	24	-51	63	69	0	0	0
Tunis	Jul	26.5	135	168	3	-165	5	-19	22	146	0	0	0
Tunis	Aug	27.0	139	162	9	-153	1	-4	12	150	0	0	0
Tunis	Sep	24.8	118	121	37	-84	0	-1	37	84	0	0	0
Tunis	Oct	20.4	80	77	68	-9	2	2	66	11	0	0	0
Tunis	Nov	15.6	47	40	61	21	23	21	40	0	0	0	0
Tunis	Dec	12.0	28	23	62	39	61	38	23	0	0	0	0
Ulaanbaatar	Jan	-22.8	0	0	1	0	13	0	0	0	0	0	10
Ulaanbaatar	Feb	-18.9	0	0	2	0	13	0	0	0	0	0	12
Ulaanbaatar	Mar	-9.7	0	0	3	-1	13	0	0	0	0	0	16
Ulaanbaatar	Apr	0.2	2	3	9	22	36	23	2	1	0	18	0
Ulaanbaatar	May	8.8	54	70	14	-56	21	-15	29	41	0	0	0
Ulaanbaatar	Jun	14.4	81	108	49	-59	12	-9	58	50	0	0	0
Ulaanbaatar	Jul	16.5	90	122	75	-47	8	-4	79	43	0	0	0
Ulaanbaatar	Aug	14.6	82	100	67	-33	6	-2	69	31	0	0	0
Ulaanbaatar	Sep	7.7	48	50	30	-20	7	1	29	21	0	0	0
Ulaanbaatar	Oct	-0.9	0	0	8	6	13	6	0	0	0	0	2
Ulaanbaatar	Nov	-12.7	0	0	4	-1	13	0	0	0	0	0	7
Ulaanbaatar	Dec	-20.6	0	0	2	0	13	0	0	0	0	0	9
Vancouver, BC	Jan	2.7	11	8	204	196	150	0	8	0	196	0	0
Vancouver, BC	Feb	4.4	19	15	153	138	150	0	15	0	138	0	0
Vancouver, BC	Mar	6.3	28	28	132	104	150	0	28	0	104	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Vancouver, BC	Apr	9.2	42	48	100	52	150	0	48	0	52	0	0
Vancouver, BC	May	12.8	60	80	68	-12	136	-14	80	0	2	0	0
Vancouver, BC	Jun	15.4	74	100	68	-32	104	-32	100	0	0	0	0
Vancouver, BC	Jul	17.8	87	118	44	-74	51	-53	98	20	0	0	0
Vancouver, BC	Aug	17.6	86	106	49	-57	29	-22	70	36	0	0	0
Vancouver, BC	Sep	15.0	72	75	80	5	39	10	70	5	0	0	0
Vancouver, BC	Oct	10.3	47	43	148	105	142	103	43	0	0	0	0
Vancouver, BC	Nov	5.9	26	19	208	189	150	8	19	0	181	0	0
Vancouver, BC	Dec	3.3	14	10	214	204	150	0	10	0	204	0	0
Victoria, BC	Jan	-3.6	0	0	206	0	150	0	0	0	0	0	532
Victoria, BC	Feb	-2.0	0	0	172	0	150	0	0	0	0	0	704
Victoria, BC	Mar	-0.6	0	0	140	143	150	0	0	0	143	39	701
Victoria, BC	Apr	1.9	23	26	86	306	150	0	26	0	306	246	455
Victoria, BC	May	5.2	45	59	54	450	149	-1	59	0	451	455	0
Victoria, BC	Jun	7.7	59	79	45	-34	116	-33	79	0	0	0	0
Victoria, BC	Jul	10.1	71	96	27	-69	60	-56	82	14	0	0	0
Victoria, BC	Aug	10.1	71	87	29	-58	34	-26	55	32	0	0	0
Victoria, BC	Sep	7.8	59	62	47	-15	33	-1	48	14	0	0	0
Victoria, BC	Oct	3.5	34	31	119	88	118	85	31	0	0	0	0
Victoria, BC	Nov	-0.7	0	0	209	113	150	32	0	0	81	0	96
Victoria, BC	Dec	-3.0	0	0	230	0	150	0	0	0	0	0	326
Vienna	Jan	-0.4	0	0	31	31	139	32	0	0	0	0	0
Vienna	Feb	0.9	2	2	38	36	150	11	2	0	25	0	0
Vienna	Mar	5.6	21	22	39	17	150	0	22	0	17	0	0
Vienna	Apr	10.5	45	51	46	-5	144	-6	51	0	1	0	0
Vienna	May	15.6	72	94	61	-33	111	-33	94	0	0	0	0
Vienna	Jun	18.7	89	119	74	-45	73	-38	112	7	0	0	0
Vienna	Jul	20.9	101	137	81	-56	42	-31	111	26	0	0	0
Vienna	Aug	20.0	96	118	65	-53	25	-17	82	36	0	0	0
Vienna	Sep	16.0	74	77	53	-24	20	-5	58	19	0	0	0
Vienna	Oct	10.4	44	41	46	5	26	6	40	1	0	0	0
Vienna	Nov	5.2	19	15	54	39	65	39	15	0	0	0	0
Vienna	Dec	1.2	3	2	44	42	107	42	2	0	0	0	0
Vilnius	Jan	-5.9	0	0	35	0	105	0	0	0	0	0	89
Vilnius	Feb	-6.3	0	0	32	0	105	0	0	0	0	0	121
Vilnius	Mar	-1.9	0	0	34	51	150	45	0	0	6	38	104
Vilnius	Apr	5.2	29	33	45	116	149	-1	33	0	117	104	0
Vilnius	May	12.2	65	86	61	-25	124	-25	86	0	0	0	0
Vilnius	Jun	15.1	79	108	77	-31	94	-30	107	1	0	0	0
Vilnius	Jul	17.3	90	123	73	-50	58	-36	109	14	0	0	0
Vilnius	Aug	16.1	84	104	80	-24	46	-12	92	12	0	0	0
Vilnius	Sep	11.4	61	63	63	0	47	1	62	1	0	0	0
Vilnius	Oct	6.3	35	32	54	22	69	22	32	0	0	0	0
Vilnius	Nov	0.6	4	3	50	36	105	36	3	0	0	0	11
Vilnius	Dec	-4.0	0	0	43	0	105	0	0	0	0	0	54
Vladivostok	Jan	-13.2	0	0	13	1	150	0	0	0	0	0	45
Vladivostok	Feb	-10.0	0	0	14	0	150	0	0	0	0	0	59
Vladivostok	Mar	-2.3	0	0	23	52	150	0	0	0	52	41	30
Vladivostok	Apr	5.4	26	29	43	44	150	0	29	0	44	30	0
Vladivostok	May	10.8	53	68	65	-3	147	-3	68	0	0	0	0

Continued on the next page

Table C.5: *Cont.*

City	MON	TEMP	UPE	APE	PREC	DIFF	ST	DST	AE	DEF	SURP	SMT	SST
Vladivostok	Jun	14.5	72	93	93	0	147	0	93	0	0	0	0
Vladivostok	Jul	18.8	94	122	111	-11	136	-11	122	0	0	0	0
Vladivostok	Aug	20.8	105	125	135	10	145	9	125	0	0	0	0
Vladivostok	Sep	16.0	80	82	125	43	150	5	82	0	38	0	0
Vladivostok	Oct	8.7	43	40	61	21	150	0	40	0	21	0	0
Vladivostok	Nov	-1.1	0	0	34	19	150	0	0	0	19	0	15
Vladivostok	Dec	-9.5	0	0	17	-1	150	0	0	0	0	0	33
Warsaw	Jan	-3.6	0	0	23	0	62	0	0	0	0	0	42
Warsaw	Feb	-2.5	0	0	23	2	65	3	0	0	0	1	63
Warsaw	Mar	1.8	8	8	26	81	145	80	8	0	0	63	0
Warsaw	Apr	8.3	40	46	36	-10	135	-10	46	0	0	0	0
Warsaw	May	13.4	66	88	52	-36	100	-35	88	0	0	0	0
Warsaw	Jun	17.0	84	115	69	-46	65	-35	105	10	0	0	0
Warsaw	Jul	18.6	93	127	72	-55	38	-27	98	29	0	0	0
Warsaw	Aug	17.9	89	111	58	-53	23	-15	73	38	0	0	0
Warsaw	Sep	13.5	66	69	44	-25	18	-5	49	20	0	0	0
Warsaw	Oct	8.4	40	37	38	1	21	3	35	2	0	0	0
Warsaw	Nov	2.9	13	10	38	28	48	27	10	0	0	0	0
Warsaw	Dec	-1.3	0	0	33	14	62	14	0	0	0	0	19
Yangon	Jan	25.0	103	99	4	-95	11	-17	20	79	0	0	0
Yangon	Feb	26.4	128	116	4	-112	4	-7	11	105	0	0	0
Yangon	Mar	28.5	152	157	8	-149	1	-3	11	146	0	0	0
Yangon	Apr	30.4	165	172	16	-156	0	-1	16	156	0	0	0
Yangon	May	29.5	159	177	282	105	111	111	167	10	0	0	0
Yangon	Jun	27.5	144	157	482	325	150	39	157	0	286	0	0
Yangon	Jul	27.0	139	157	501	344	150	0	157	0	344	0	0
Yangon	Aug	27.0	139	152	564	412	150	0	152	0	412	0	0
Yangon	Sep	27.4	143	145	339	194	150	0	145	0	194	0	0
Yangon	Oct	27.9	147	148	193	45	149	-1	148	0	46	0	0
Yangon	Nov	27.4	143	135	51	-84	73	-76	129	6	0	0	0
Yangon	Dec	25.3	108	103	5	-98	28	-45	50	53	0	0	0
Yokohama	Jan	4.7	8	7	85	78	150	0	7	0	78	0	0
Yokohama	Feb	5.3	10	8	122	114	150	0	8	0	114	0	0
Yokohama	Mar	8.0	19	20	177	157	150	0	20	0	157	0	0
Yokohama	Apr	13.7	46	50	213	163	150	0	50	0	163	0	0
Yokohama	May	18.1	72	88	162	74	150	0	88	0	74	0	0
Yokohama	Jun	20.9	91	111	205	94	150	0	111	0	94	0	0
Yokohama	Jul	25.1	123	151	124	-27	120	-30	151	0	3	0	0
Yokohama	Aug	26.4	133	154	152	-2	116	-4	154	0	0	0	0
Yokohama	Sep	23.3	109	112	273	161	150	34	112	0	127	0	0
Yokohama	Oct	17.4	68	65	304	239	150	0	65	0	239	0	0
Yokohama	Nov	12.6	40	34	164	130	150	0	34	0	130	0	0
Yokohama	Dec	7.5	17	15	113	98	150	0	15	0	98	0	0

Table C.6: WebWIMP annual water balance data for UrbMet cities. Variables APE—SMT are in units of *mm/month*. Variable definitions and descriptions provided in Table A.2.

cs_id	**City**	**APE**	**PREC**	**DIFF**	**AE**	**DEF**	**SURP**	**SMT**
1	Abu Dhabi	1571	125	-1446	124	1447	0	0

Continued on the next page

Table C.6: *Cont.*

cs_id	City	APE	PREC	DIFF	AE	DEF	SURP	SMT
2	Abuja	1584	1266	-318	989	595	272	0
3	Accra	1597	1180	-417	1076	521	100	0
4	Addis Ababa	784	917	133	663	121	253	0
5	Amman	888	187	-701	187	701	0	0
6	Amsterdam	637	818	181	619	18	197	0
7	Anchorage, AK	390	4223	3833	390	0	3837	2157
8	Ankara	663	390	-273	378	285	11	0
9	Asuncion	1183	1547	364	1183	0	364	0
10	Athens	873	513	-360	422	451	90	0
11	Bandar seri begawan	1724	3874	2150	1724	0	2150	0
12	Bangalore	1260	853	-407	849	411	2	0
13	Bangkok	1756	1392	-364	1276	480	113	0
14	Barcelona	805	605	-200	604	201	0	0
15	Beijing	773	579	-194	577	196	0	7
16	Beirut	587	1135	548	332	255	800	0
17	Belo Horizonte	938	1577	639	897	41	678	0
18	Berlin	629	540	-89	519	110	21	0
19	Bern	517	1443	926	517	0	926	172
20	Bishkek	570	422	-148	374	196	48	101
21	Bogota	763	3057	2294	763	0	2292	0
22	Boston, MA	662	1202	540	642	20	560	126
23	Brasilia	981	1663	682	865	116	795	0
24	Brussels	640	775	135	610	30	165	0
25	Bucharest	704	539	-165	537	167	4	35
26	Budapest	696	550	-146	531	165	21	0
27	Buenos Aires	828	1009	181	825	3	183	0
28	Cairo	941	25	-916	25	916	0	0
29	Cali	1505	1269	-236	1269	236	0	0
30	Cape Town	806	527	-279	526	280	0	0
31	Caracas	1426	925	-501	925	501	0	0
32	Casablanca	833	379	-454	377	456	0	0
33	Cebu	1282	2097	815	1278	4	818	0
34	Chennai	1786	1266	-520	1026	760	236	0
35	Chicago, IL	694	930	236	675	19	256	115
36	Chisinau	661	520	-141	498	163	21	73
37	Colombo	1732	3279	1547	1732	0	1547	0
38	Copenhagen	585	626	41	520	65	103	0
39	Curitiba	796	1523	727	796	0	727	0
40	Dakar	1574	784	-790	718	856	65	0
41	Damascus	903	218	-685	217	686	0	0
42	Dar es Salaam	1527	1171	-356	1124	403	43	0
43	Delhi	1437	697	-740	697	740	0	0
44	Denver, CO	641	413	-228	414	227	0	10
45	Detroit, MI	652	817	165	590	62	227	142
46	Dhaka	1517	2219	702	1363	154	850	0
47	Doha	1488	71	-1417	71	1417	0	0
48	Dubai	1514	154	-1360	151	1363	0	0
49	Dublin	581	814	233	574	7	237	0
50	Durban	999	1006	7	999	0	7	0
51	Florianopolis	965	1737	772	965	0	772	0

Continued on the next page

Table C.6: *Cont.*

cs_id	City	APE	PREC	DIFF	AE	DEF	SURP	SMT
52	Geneva	599	1214	615	599	0	615	0
53	Guadalajara	954	902	-52	673	281	226	0
54	Guangzhou	1195	1803	608	1195	0	607	0
55	Guatemala City	1105	2749	1644	960	145	1785	0
56	Hamburg	603	676	73	558	45	117	0
57	Hanover	621	735	114	589	32	145	0
58	Helsinki	529	712	183	499	30	211	168
59	Ho Chi Minh City	1724	1843	119	1269	455	569	0
60	Hyderabad	1551	747	-804	745	806	0	0
61	Islamabad	1213	809	-404	807	406	0	0
62	Istanbul	796	764	-32	522	274	239	0
63	Jakarta	1636	1773	137	1439	197	333	0
64	Jerusalem	898	417	-481	322	576	95	0
65	Johannesburg	772	706	-66	705	67	0	0
66	Karachi	1555	216	-1339	214	1341	0	0
67	Kathmandu	1308	1538	230	1100	208	431	0
68	Kiev	625	606	-19	525	100	83	129
69	Kingston	1568	2129	561	1562	6	565	0
70	Kinshasa	1142	1392	250	1004	138	385	0
71	Kolkata	1588	1627	39	1247	341	376	0
72	Kuala Lumpur	1751	2242	491	1751	0	491	0
73	Kuwait City	1401	96	-1305	95	1306	0	0
74	Lagos	1627	1697	70	1255	372	438	0
75	Lima	816	138	-678	136	680	0	0
76	Lisbon	829	643	-186	512	317	129	0
77	Ljubljana	594	1613	1019	594	0	1019	182
78	London	635	701	66	568	67	132	0
79	Los Angeles, CA	850	273	-577	273	577	0	0
80	Madrid	756	509	-247	478	278	32	0
81	Manama	1482	75	-1407	74	1408	0	0
82	Manila	1766	2115	349	1526	240	588	0
83	Melbourne	707	1026	319	688	19	335	0
84	Mexico City	622	1099	477	580	42	516	0
85	Milan	736	731	-5	589	147	142	0
86	Minsk	563	647	84	525	38	122	152
87	Montevideo	814	1086	272	808	6	279	0
88	Montreal	602	968	366	588	14	380	258
89	Moscow	560	671	111	516	44	155	189
90	Mumbai	1624	2855	1231	888	736	1966	0
91	Nagoya	846	1949	1103	846	0	1103	0
92	Naihati	1565	1619	54	1252	313	363	0
93	Nairobi	823	712	-111	702	121	8	0
94	New York, NY	729	1232	503	717	12	516	0
95	Osaka	740	1844	1104	740	0	1104	0
96	Ottawa	598	894	296	571	27	323	246
97	Panama City	1455	3072	1617	1360	95	1708	0
98	Paris	670	681	11	591	79	88	0
99	Phnom Penh	1775	1358	-417	1310	465	47	0
100	Phoenix, AZ	1171	206	-965	205	966	0	0
101	Port-of-Spain	1594	1828	234	1443	151	382	0

Continued on the next page

Table C.6: *Cont.*

cs_id	City	APE	PREC	DIFF	AE	DEF	SURP	SMT
102	Portland, OR	673	1210	537	536	137	671	0
103	Prague	587	589	2	560	27	28	54
104	Quezon City	1634	1859	225	1557	77	301	0
105	Quito	546	1143	597	546	0	597	0
106	Rabat	872	451	-421	433	439	19	0
107	Riga	557	648	91	512	45	136	125
108	Riyadh	1472	105	-1367	106	1366	0	0
109	Rome	707	979	272	567	140	410	0
110	San Salvador	1743	1916	173	1188	555	724	0
111	Sana'a	782	587	-195	586	196	0	0
112	Santiago	546	565	19	326	220	238	0
113	Santo Domingo	1685	1113	-572	1112	573	0	0
114	Sao Paulo	779	3717	2938	779	0	2938	0
115	Sarajevo	592	920	328	582	10	337	111
116	Seattle, WA	660	1130	470	542	118	586	0
117	Seoul	713	1269	556	713	0	556	51
118	Shanghai	898	1139	241	892	6	248	0
119	Shenzhen	1182	1991	809	1182	0	808	0
120	Singapore	1591	2621	1030	1591	0	1030	0
121	Sofia	512	772	260	497	15	274	101
122	St. John's	486	1486	1000	486	0	1000	515
123	St. Petersburg	526	613	87	477	49	136	166
124	Stockholm	524	541	17	451	73	90	115
125	Surabaya	1765	1637	-128	1292	473	340	0
126	Sydney	774	1117	343	774	0	342	0
127	Tashkent	701	611	-90	355	346	254	0
128	Tbilisi	847	494	-353	492	355	0	0
129	Tehran	1004	143	-861	142	862	0	0
130	Tel Aviv	898	417	-481	322	576	95	0
131	Tokyo	815	2094	1279	815	0	1277	0
132	Toronto	646	814	168	579	67	233	149
133	Tripoli	1073	311	-762	309	764	0	0
134	Tunis	946	481	-465	481	465	0	0
135	Ulaanbaatar	453	264	-189	266	187	0	18
136	Vancouver, BC	650	1468	818	589	61	877	0
137	Victoria, BC	440	1364	924	380	60	981	740
138	Vienna	678	632	-46	589	89	43	0
139	Vilnius	552	647	95	524	28	123	142
140	Vladivostok	559	734	175	559	0	174	71
141	Warsaw	611	512	-99	512	99	0	64
142	Yangon	1718	2449	731	1163	555	1282	0
143	Yokohama	815	2094	1279	815	0	1277	0

C.2 Data from Chapters 4 and 5

C.2.1 Demographics

Los Angeles

Table C.7: Table of historical Los Angeles demographics, water use, and climate data. Units were: population (N) in capita; total water use (W_N) in 10^6 $m^3 \cdot yr^{-1}$; water use intensity (w_N) in $m^3 \cdot yr^{-1} \cdot cap^{-1}$; , city area ($A_N$) in km^2.

	Various Sources					WebWIMP[190] ($m \cdot yr^{-1}$)			
Year	N	W_N	w_N	A_N	q_P [230]	q_P	q_S	q_{def}	q_{Net}
1960	2479015	—	—	1177.9	0.2	0.307	0.000	0.623	0.078
1961	—	—	—	1177.9	0.1	0.211	0.000	0.723	0.113
1962	—	—	—	1177.9	0.5	0.513	0.084	0.516	0.242
1963	—	—	—	—	0.2	0.445	0.000	0.578	0.155
1964	—	—	—	—	0.2	0.346	0.000	0.616	0.198
1965	—	—	—	—	0.3	0.974	0.064	0.521	0.674
1966	—	—	—	—	0.5	0.409	0.000	0.582	0.103
1967	—	—	—	—	0.6	0.709	0.005	0.461	0.289
1968	—	—	—	—	0.4	0.206	0.000	0.616	-0.022
1969	—	—	—	—	0.7	0.968	0.246	0.556	0.672
1970	2816111	722.9	256.7	1203.6	0.2	0.539	0.000	0.664	0.344
1971	—	715.7	—	—	0.3	0.295	0.000	0.650	0.113
1972	2853322	745.1	261.1	—	0.2	0.155	0.000	0.652	-0.047
1973	—	704.6	—	—	0.5	0.575	0.017	0.500	0.264
1974	—	694.4	—	—	0.4	0.518	0.000	0.588	0.264
1975	—	701.3	—	—	0.4	0.337	0.000	0.518	0.056
1976	—	763.5	—	—	0.2	0.403	0.000	0.702	0.213
1977	—	708.4	—	—	0.3	0.521	0.000	0.615	0.293
1978	—	616.7	—	—	0.8	1.061	0.283	0.428	0.647
1979	—	729.7	—	—	0.5	0.508	0.050	0.574	0.224
1980	2966850	730.4	246.2	1215.5	0.7	0.751	0.184	0.513	0.426
1981	—	776.2	—	—	0.2	0.349	0.000	0.630	0.067
1982	3075040	756.3	245.9	—	0.3	0.614	0.000	0.529	0.320
1983	—	769.5	—	—	0.8	1.129	0.136	0.365	0.602
1984	—	834.8	—	—	0.3	0.243	0.000	0.764	0.113
1985	—	796.1	—	—	0.3	0.226	0.000	0.592	-0.018
1986	—	838.3	—	—	0.5	0.437	0.000	0.529	0.114
1987	—	867.4	—	—	0.2	0.325	0.000	0.680	0.167
1988	—	846.0	—	—	0.3	0.411	0.000	0.687	0.256
1989	—	855.7	—	—	0.2	0.130	0.000	0.734	-0.002
1990	3490590	845.1	242.1	—	0.2	0.162	0.000	0.764	0.045
1991	3506553	780.6	222.6	—	0.3	0.536	0.000	0.597	0.321
1992	3545560	689.8	194.5	—	0.5	0.680	0.000	0.611	0.371
1993	3557602	716.5	201.4	—	0.7	0.791	0.141	0.539	0.465
1994	3543970	735.0	207.4	—	0.2	0.278	0.000	0.704	0.091
1995	3535798	708.7	200.4	—	0.6	0.790	0.128	0.505	0.424
1996	3543182	744.6	210.2	—	0.3	0.610	0.000	0.588	0.311
1997	3565297	777.8	218.2	—	0.3	0.280	0.000	0.642	-0.009
1998	3597920	730.8	203.1	—	0.8	0.841	0.274	0.404	0.423
1999	3635576	763.1	209.9	—	0.2	0.229	0.000	0.636	0.060
2000	3702574	812.5	219.4	—	0.3	0.430	0.000	0.613	0.197

Continued on next page

Table C.7: *Continued*

Year	Various Sources					WebWIMP[190] ($m \cdot yr^{-1}$)			
	N	W_N	w_N	A_N	q_P [230]	q_P	q_S	q_{def}	q_{Net}
2001	3736059	815.1	218.2	—	0.5	0.560	0.010	0.488	0.243
2002	3767555	818.2	217.2	—	0.1	0.228	0.000	0.745	0.152
2003	3788396	807.9	213.2	—	0.4	0.449	0.000	0.525	0.124
2004	3794784	846.8	223.2	—	0.2	0.645	0.000	0.601	0.387
2005	3792836	761.8	200.8	—	1.0	0.771	0.174	0.438	0.414
2006	3774517	776.8	205.8	—	0.3	0.415	0.000	0.613	0.183
2007	3774880	824.8	218.5	—	0.1	0.214	0.000	0.692	0.112
2008	3797421	800.9	210.9	—	0.3	0.509	0.000	0.536	0.191
2009	3824479	755.8	197.6	—	0.2	0.277	0.000	0.557	0.053
2010	3839642	668.7	174.2	1302.8	0.4	0.723	0.023	0.370	0.330
2011	3806499	655.3	172.2	—	0.5	0.383	—	—	—
2012	3827240	671.4	175.4	—	0.2	0.323	—	—	—
2013	3866133	698.9	180.8	—	0.1	0.169	—	—	—
2014	3904657	730.7	187.1	1302.8	0.2	0.324	—	—	—
2015	3971883	672.6	169.3	1302.8	0.2	—	—	—	—

Singapore

Table C.8: Table of historical Singapore demographics, water use, and climate data. Units not shown in the table were: population (N) in 10^6 $m^3 \cdot yr^{-1}$ and water use intensity (w_N) in $m^3 \cdot yr^{-1} \cdot cap^{-1}$.

Year	Various Sources					Water Balance[190] ($m \cdot yr^{-1}$)			
	N	W_N	w_N	A_N	q_P [355]	q_P	q_S	q_{def}	q_{Net}
				km^2	$m \cdot yr^{-1}$	$m \cdot yr^{-1}$	$m \cdot yr^{-1}$	$m \cdot yr^{-1}$	$m \cdot yr^{-1}$
1960	1646400	88	53	581.5	2.1	2.556	0.886	0.000	0.930
1961	1702400	102	60	581.5	1.5	2.366	0.715	0.028	0.779
1962	1750200	106	61	581.5	2.2	2.685	1.021	0.000	1.105
1963	1795000	94	52	581.5	1.6	2.069	0.426	0.055	0.508
1964	1841600	114	62	581.5	2.9	3.176	1.519	0.000	1.586
1965	1886900	124	66	581.5	1.9	2.370	0.747	0.005	0.794
1966	1934400	136	70	581.5	2.0	2.574	0.874	0.001	0.932
1967	1977600	137	69	583.0	2.5	3.082	1.438	0.004	1.495
1968	2012000	150	75	584.3	2.1	2.426	0.775	0.003	0.834
1969	2042500	157	77	585.3	2.3	2.728	1.020	0.000	1.078
1970	2074500	166	80	586.4	2.1	2.622	0.959	0.000	1.016
1971	2112900	161	76	586.4	1.5	2.170	0.548	0.001	0.606
1972	2152400	176	82	586.4	1.7	2.089	0.447	0.029	0.507
1973	2193000	175	80	586.4	2.2	2.812	1.214	0.000	1.221
1974	2229800	175	79	587.6	2.1	2.112	0.556	0.002	0.564
1975	2262600	190	84	596.8	1.9	2.272	0.671	0.001	0.723

Continued on next page

Table C.8: *Continued*

Year	N	W_N	w_N	A_N km^2	q_P [355] $\mathrm{m \cdot yr^{-1}}$	q_P $\mathrm{m \cdot yr^{-1}}$	q_S $\mathrm{m \cdot yr^{-1}}$	q_{def} $\mathrm{m \cdot yr^{-1}}$	q_{Net} $\mathrm{m \cdot yr^{-1}}$
	Various Sources					Water Balance[190] ($\mathrm{m \cdot yr^{-1}}$)			
1976	2293300	198	86	602.0	2.0	2.530	1.005	0.031	1.027
1977	2325300	210	90	616.3	1.8	2.187	0.693	0.002	0.625
1978	2353600	224	95	616.3	2.6	2.467	1.217	0.000	0.885
1979	2383500	238	100	617.8	2.1	2.458	1.018	0.001	0.873
1980	2413900	256	106	617.8	2.6	2.735	0.875	0.000	1.143
1981	2443300	276	113	617.9	2.0	2.179	0.445	0.013	0.577
1982	2471800	289	117	618.1	2.2	2.031	0.396	0.071	0.475
1983	2502000	307	123	618.1	1.8	2.301	0.599	0.053	0.677
1984	2529100	318	126	620.2	2.9	3.033	1.458	0.000	1.500
1985	2558000	323	126	620.5	2.2	2.341	0.677	0.006	0.779
1986	2586200	325	126	621.7	2.5	2.598	0.977	0.002	1.029
1987	2612800	335	128	622.6	2.3	2.406	0.762	0.001	0.783
1988	2647100	344	130	625.6	2.6	2.708	1.098	0.001	1.106
1989	2647600	356	134	626.4	2.5	2.805	1.046	0.000	1.256
1990	3047100	374	123	633.0	1.7	2.228	0.587	0.003	0.597
1991	3135700	399	127	639.1	2.0	2.458	0.864	0.015	0.847
1992	3230700	420	130	641.0	2.7	2.522	0.915	0.000	0.908
1993	3313500	434	131	641.4	2.4	2.408	0.768	0.016	0.799
1994	3419000	456	133	646.1	2.1	2.263	0.615	0.031	0.643
1995	3524500	469	133	647.5	2.8	2.819	1.142	0.000	1.172
1996	3670400	481	131	647.5	3.0	2.426	0.773	0.000	0.779
1997	3793700	495	130	647.8	1.3	1.606	0.094	0.220	0.103
1998	3922000	491	125	648.1	2.7	2.614	0.874	0.002	0.874
1999	3958700	488	123	659.9	2.3	2.349	0.736	0.000	0.687
2000	4027900	506	126	682.7	2.6	2.613	0.954	0.001	0.933
2001	4138000	508	123	682.3	2.9	2.766	1.111	0.002	1.095
2002	4163700	513	123	687.1	1.9	1.951	0.314	0.043	0.292
2003	4114800	499	121	693.4	2.6	2.332	0.757	0.059	0.712
2004	4166700	491	118	696.2	2.3	2.157	0.590	0.020	0.492
2005	4265800	492	115	697.9	2.1	2.111	0.439	0.075	0.464
2006	4401400	502	114	699.5	2.8	2.644	0.882	0.001	0.948
2007	4588600	508	111	705.1	3.0	2.730	1.022	0.000	1.063
2008	4839400	463	96	710.2	2.6	2.545	0.849	0.000	0.887
2009	4987600	468	94	710.3	2.2	2.209	0.458	0.010	0.500
2010	5076732	476	94	712.4	—	2.458	0.696	0.015	0.733
2011	5183688	478	92	714.3	—	2.630	—	—	—
2012	5312437	491	92	715.8	—	2.408	—	—	—
2013	5399162	499	92	716.5	—	2.820	—	—	—
2014	5469724	506	93	718.9	—	1.982	—	—	—
2015	5535002	430	78	719.1	—	—	—	—	—

Appendix D

Appendix: Code

D.1 Code for Chapter 3

D.1.1 Python script

```
%%%% water_balance_scraper(filename='urbmet_cities.csv', failed=True, prev=False)
\begin{verbatim}
from mechanize import Browser
from bs4 import BeautifulSoup
import time
import math
import pandas as pd
import xlrd


#### FUNCTION FOR DEALING WITH MISSING DATAFRAMES
def error_script(dataFile, dataSeriesName, functionName):
	read_error_msg0 		= 'Create new ' + dataSeriesName + ' database? (y/n):'
	read_error_input0 		 = raw_input(read_error_msg0)
	if str.lower(read_error_input0) == 'n':
		read_error_msg1 	= 'Enter new filename for ' + dataSeriesName + ' (y/n):'
		read_error_input1 	 = raw_input(read_error_msg1)
		if str.lower(read_error_input1) == 'n':
			print 'Exiting function ' , functionName , '()...'
			return None
		else:
			data_readfile 	 = raw_input('Enter new file name (e.g. filename.csv):')
			try:
				df_cities 	 = pd.read_csv(data_readfile, header=0, index_col=0)
			except:
				print 'Error in reading previous file for ' , print_str
				print 'Exiting...'
				return None
	else:
		return none
	return df_cities
#### FUNCTION TO GENERATE NEW PANDAS DATAFRAMES
def new_pdf(df, type=None):
	#### HEADERS ####
	#### Header for city status
```

```
        cstatus_header          = ['cs_id' , 'city', 'latitude', 'longitude',
                   'round_lat' , 'round_lon', 'status', 'delta_lat',
                   'delta_long', 'dist' , 'new_lat', 'new_long']
  #### Header for Temp/Precip
        tp_header                 = ['tp_id' , 'cs_id' , 'city', 'unit',
                   'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun',
                   'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec', 'Ann']
        #### Header for Water Balance
        wb_header                 = ['wb_id' , 'cs_id' , 'city', 'MON', 'TEMP',
                   'UPE', 'APE', 'PREC', 'DIFF', 'ST', 'DST',
                   'AE', 'DEF', 'SURP', 'SMT', 'SST']
        if type         == 'status':
                header                    = cstatus_header
                print_str         = 'City Status'
        elif type         == 'tp':
                header                    = tp_header
                print_str         = 'Temp/Precip'
                data_range         = 2
        else:
                header                    = wb_header
                print_str         = 'Water Balance'
                data_range         = 12

        pdf_columns         = header[1:]
        print 'Creating new, empty dataframe for ' , print_str , ' data...'
        if type == 'status':
                pdf_ids                  = df.index.values
        else:
                len_new_id                  = (int(len(df.city)))*data_range + 1
                pdf_ids                  = range(1, len_new_id)
        #### PANDAS DATAFRAME
        pdf                          = pd.DataFrame(index=pdf_ids, columns = pdf_columns)
        return pdf
#### FUNCTION TO SUBMIT QUERIES TO WEBWIMP ####
def query_webwimp(latitude, longitude, index):
    br = Browser()
    br.open("http://climate.geog.udel.edu/~wimp/wimp_map.php")
    #### Select the first form on the page
    br.select_form(nr=0)
    #### Fill in and submit the form
    lat = float(latitude)
    lon = float(longitude)
    br.form['lati'] = str(lat)
    br.form['long'] = str(lon)
    #### Get and read the form
    response = br.submit()
    html = response.read()
    if index == 1:
        br.select_form(nr=1)
        response = br.submit()
        html = response.read()
    #### Parse html
    soup = BeautifulSoup(html, "html5lib")
    if index == 0:
        #### Get first and only table
        table = soup.find('table')
    else:
```

```
        #### Get 3rd table
        table = soup.findAll('table')[2]

    try:
        rows = table.findAll('tr')
    except:
        rows = 'Fail'
    return rows
#### FUNCTION TO ACQUIRE STATUS FOR CITIES ####
def get_status(latitude, longitude):
    #### This function applies the query_webwimp function for a specific city. If the query was
    #### successful, an html table is returned. If the query was successful, a status of 'true'
    #### is saved to the list 'is_there', along with lat/long coordinates associated with the
    #### success.
    #### However, the query will fail if the location appears to fall on a body of water. This
    #### may happen since the resolution of the map is only 0.5 degrees. If this is the case, or
    #### if no entry exists, then the function will attempt to search local locations.
    web_index = 0

    html = query_webwimp(latitude, longitude, web_index)
    if html == "Fail":
        cstatus = False
    else:
        cstatus = True
    return cstatus
#### FUNCTION TO PRODUCE A DICTIONARY OF DISTANCES FOR FAILED OBSERVATIONS ####
def distance_dict(locationBuffer=1):
    distance_dict_header = ['delta_lat' , 'delta_long' , 'dist' , 'status']
    LB = locationBuffer

    latlong_delta       = LB*2
    latlong_range       = range(-(latlong_delta), latlong_delta+1)
    len_latlong_tuples  = pow(len(latlong_range), 2)

    test_range = len_latlong_tuples
    ll_id = range(0, test_range)
    dict_dist = pd.DataFrame(index=ll_id, columns = distance_dict_header)

    # #### Generate a matrix or set of coordinates to try
    latlong_floats = [round(float(i)/2, 1) for i in latlong_range]
    latlong_tuples = []
    for i in latlong_floats:
        for j in latlong_floats:
            latlong_tuples.append([i, j])

    #### Add these to the dataframe:
    latlong_tuple_x = [i[0] for i in latlong_tuples]
    latlong_tuple_y = [j[1] for j in latlong_tuples]
    float_x         = [float(i) for i in latlong_tuple_x]
    float_y         = [float(j) for j in latlong_tuple_y]

    dict_dist['delta_lat']  = float_x
    dict_dist['delta_long'] = float_y

    for k in dict_dist.index.values:
        k_x     = dict_dist.loc[k, 'delta_lat']
        k_y     = dict_dist.loc[k, 'delta_long']
```

```
        k_xy    = round(math.sqrt(pow(k_x, 2) + pow(k_y, 2)), 3)
        dict_dist.loc[k, 'dist'] = k_xy
    return dict_dist
#### FUNCTION TO FIND A NEW LOCATION USING DISTANCE DICTIONARY ####
def find_new_location(city_lat, city_lon, df_dist):
    for k in df_dist.index.values:
        #### Adjust location
        lat_adj         = df_dist.loc[k, 'delta_lat' ]
        long_adj        = df_dist.loc[k, 'delta_long']
        new_latitude    = city_lat   + lat_adj
        new_longitude   = city_lon   + long_adj
        #### Submit query to webwimp
        html            = query_webwimp(new_latitude, new_longitude, 0)
        #### Update dict_adj with status
        if html == "Fail":
            df_dist.loc[k, 'status'] = False
            # print 'False'
        else:
            df_dist.loc[k, 'status'] = True
            # print 'True'
    #### Find the observations that are true and closest to the original location
    #### Find all of the observations for which status=='true':
    trues       = df_dist.ix[df_dist['status']==True]
    #### Find the best distance
    best_trues  = trues.loc[trues['dist']==min(trues['dist'])]
    bt_index    = best_trues.index.values[0]
    best_true   = best_trues[best_trues.index.values==bt_index]
    return best_true
#### FUNCTION TO SCRAPE WEBWIMP FOR A GIVEN TABLE ####
def get_index(g_index, g_range=2):
    data_range          = int(g_range)
    #### Calculate base value:
    city_index_base     = (int(g_index)-1)*data_range + 1
    city_index_bases    = range(city_index_base, city_index_base+data_range)
    return city_index_bases

def scrape_wimp(city, type ='tp'):
    #### This scrapes the table data for water balance for a single city
    wb_header           = ['wb_id' , 'cs_id' , 'city', 'MON', 'TEMP', 'UPE', 'APE', 'PREC',
                               'DIFF', 'ST', 'DST', 'AE', 'DEF', 'SURP', 'SMT', 'SST']

    tp_header           = ['tp_id' , 'cs_id' , 'city', 'unit', 'Jan', 'Feb', 'Mar', 'Apr',
                               'May', 'Jun', 'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec', 'Ann']

    city_name           = city['city'].values
    #### Get city indicies
    city_index          = city.index.values
    city_index_base     = (int(city_index)-1)*2 + 1
    city_index_bases    = [city_index_base, city_index_base+1]
    if type == 'tp':
        header          = tp_header
        query_index     = 0
        data_range      = 2
        cindex          = get_index(city_index, data_range)
    else:
        header          = wb_header
        query_index     = 1
```

```
        data_range      = 12
        cindex          = get_index(city_index, data_range)
    #### Create dataframe
    data_df             = pd.DataFrame(index=cindex, columns = header[1:])
    data_df[['cs_id']]  = city_index
    data_df[['city']]   = city_name

    html                = query_webwimp(city['new_lat'], city['new_long'], query_index)
    table_data          = []
    if  html == 'Fail':
        data_df         = 'Fail'
    else:
        for row in html:
            cols        = row.findAll('td')
            cols        = [i.text.strip() for i in cols]
            table_data.append( [str(i) for i in cols] )
        for m in range(0, data_range):
            data_df.ix[cindex[m], header[3:]] = table_data[m+1]
    return data_df
#### FUNCTION TO GET CITY STATUS FOR ALL CITIES ####
#### A FUNCTION THAT GETS THE CITY STATUS FROM WEBWIMP OR UPDATES
#### FAILED VALUES WITH NEW, NEARBY LOCATIONS
#### This function takes as input a .csv file with the city status.
def get_city_status(df, date = None, drange=None , failed=False, silent=False,
        prev=False, prevDate=None):

        type                         = 'status'
        print_str                 = 'status'

        #### CITY DATA ####
        try:
                city_range         = range(drange[0], drange[1]+1)
        except:
                city_range         = df.index.values

        #### GET DATE ####
        if date is None:
                date_str         = time.strftime('\%Y\%m\%d')
        else:
                date_str         = date
        #### Previous date
        if prevDate is None:
                prevDate_str = date_str
        else:
                prevDate_str = str(prevDate)

        #### CITY STATUS FILE ####
        cs_file                 = '_city_status.csv'
        cstatus_file           = date_str                 + cs_file
        prev_cs_file         = prevDate_str         + cs_file
        #### Header for city status
        cstatus_header         = ['cs_id' , 'city', 'latitude', 'longitude', 'round_lat' ,
                                        'round_lon', 'status', 'delta_lat', 'delta_long', 'dist' ,
                                        'new_lat', 'new_long']

        #### CITY STATUS DATA ####
        # try:
```

```
if prev == True:
        try:
                cities_status = pd.read_csv(prev_cs_file, header=0,
                     index_col=0, names=cstatus_header)
                cstatus_ids        = cities_status.index.values
        except:
                print 'Could not find city status file.'
                return None
else:
    print 'Creating new cities_status dataframe....'
    cities_status        = new_pdf(df, type='status')
    cstatus_ids                 = cities_status.index.values
    cities_status[['city', 'latitude', 'longitude']] = df[['city', 'latitude', 'longitude']]

print cities_status.head()

#### New cities for which status is needed:
# new_cities0                        = cities_status.loc[cstatus_ids[drange[0], drange[1]]]
new_cities0                               = cities_status.ix[city_range, cstatus_header[1:]]
new_cities                          = new_cities0.ix[new_cities0['status'].isnull()==True]

#### Cities with status not == NA:
prev_cities                          = new_cities0.ix[new_cities0['status'].isnull()==False]
prev_cities_ids                 = prev_cities.index.values
if silent == False:
        print len(prev_cities) , ' cities already in status file.'
        print 'First five cities in status file: '
        print prev_cities.city[prev_cities_ids[0:4]]
else:
        pass

#### If all of the cities in df list appear in prev_cities list
#### then len(new_cities) will be zero. If this is the case, get the lat/long of
#### cities in the failed city list from the df.
if len(new_cities) > 0:
        new_cities_ids                          = new_cities.index.values
        #### Print progress statements on city status
        if silent == False:
                print 'Getting city status for new cities...'
                print len(new_cities)  , ' cities not in status file.'
                print 'First five cities not in status file:'
                print new_cities.city[new_cities_ids[0:4]]
        else:
                pass
        #### Query webwimp and parse results to see if location works
        for i in cities_status.city[new_cities_ids]:
                if silent == False:
                        print 'Getting city status for ' , i , '...'
                else:
                        pass
                #### New city status information
                new_city         = cities_status.ix[cities_status['city'] ==i]
                new_city_index   = new_city.index.values
                #### Update cities_status dataframe with lat/long from df:
                new_city_lat     = float(new_city['latitude'])
                new_city_lon     = float(new_city['longitude'])
                #### Round latitude and longitude to the nearest half a degree:
```

```
            round_lat                          = round(round(new_city_lat*2, 0)/2, 1)
            round_lon                          = round(round(new_city_lon*2, 0)/2, 1)
            round_names                         = ['round_lat' , 'round_lon']
            cities_status.ix[new_city_index, round_names]        = [round_lat, round_lon]
            #### Get status
            cities_status.ix[new_city_index, 'status']        = get_status(round_lat, round_lon)
            if silent == False:
                print i, 'status: ' , cities_status.ix[new_city_index, 'status']
            else:
                    pass
            #### Update columns with defaults
            cstatus_cols2update = ['delta_lat', 'delta_long', 'dist', 'new_lat', 'new_long']
            cstatus_vals2update = [0, 0, 0, round_lat, round_lon]
            cities_status.ix[new_city_index,
              cstatus_cols2update]                = cstatus_vals2update
            #### Save file to .csv
            cities_status.to_csv(cstatus_file, index=True, index_label        ='cs_id')
            if silent == False:
                    print 'Saving updated city status file to .csv...'
                    print 'Moving on...'
            else:
                    pass
else:
        pass

#### Failed cities:
failed_cities0                    = cities_status.ix[city_range, cstatus_header[1:]]
failed_cities                         = failed_cities0.ix[failed_cities0['status']==False]
#### Option for failed cities to find new locations:
if failed==True:
#### Print progress statements for failed cities
        if len(failed_cities) == 0:
                if silent==False:
                        print 'All failed cities are already in the ' , print_str , ' file.'
                else:
                        pass
        else:
                pass
        #### Update failed cities: look for new locations
        if len(failed_cities) == 0:
                pass
        else:
                failed_ids                                = failed_cities.index.values
                if silent==False:
                        print 'Updating the status of cities for which status == False...'
                        print len(failed_cities) , ' cities in status file with status == False'
                        print 'First five cities with status == False:'
                        print failed_cities.city[failed_ids[0:4]]
                #### Print progress statements
                if silent == False:
                        print 'Querying WebWIMP on failed cities to try new locations...'
                        print 'Generating distance matrix...'
                else:
                        pass
                #### Generate distance matrix
                dict_adj = distance_dict(1)
                for i in cities_status.city[failed_ids]:
```

```
                                #### Print progress statements on web query for failed cities
                                if silent == False:
                                        print 'Looking for new location for ' , i
                                else:
                                        pass
                                #### Web query for failed cities:
                                #### Failed cities
                                failed_city       = cities_status.ix[cities_status['city']==i]
                                failed_city_index = failed_city.index.values
                                failed_city_lat   = failed_city.ix[failed_city_index, 'round_lat']
                                failed_city_lon   = failed_city.ix[failed_city_index, 'round_lon']
                                #### Try new locations
                                new_city_data   = find_new_location(failed_city_lat, failed_city_lon, dict_adj)
                                new_city_lat                 = failed_city_lat + new_city_data['delta_lat' ].values
                                new_city_long         = failed_city_lon + new_city_data['delta_long'].values
                                cities_status.ix[failed_city_index, 'new_lat' ] = new_city_lat
                                cities_status.ix[failed_city_index, 'new_long'] = new_city_long
                                #### Update location values
                                cstatus_cols2update = new_city_data.columns
                                cstatus_vals2update = new_city_data.values
                                cities_status.ix[failed_city_index,
                                  cstatus_cols2update] = cstatus_vals2update
                                #### Print progress statements on web query for failed cities
                                if silent == False:
                                        print 'New location information for ' , i
                                        print cities_status.ix[cities_status.index.values == failed_city_index]
                                else:
                                        pass
                                cities_status.to_csv(cstatus_file, index=True)
                                if silent == False:
                                        print 'Saving updated status of failed cities to .csv...'
                                else:
                                        pass

        #### Print final status update:
        print 'All cities in the city status file have been updated.'

        return cities_status
#### FUNCTION TO SCRAPE WEBWIMP AND RETURN DATA FOR TEMP, PRECIP, AND WATER BALANCE ####
#### This function takes as input a .csv file with the city status.
def scrape_wimp_df(df, date= None, type=None, drange=None, silent=False,
        prev=False, prevDate=None, cstatus=None):
        if type == None:
                dtype = 'tp'
        else:
                dtype = type

        #### GET DATA ####
        if date is None:
                date_str         =  time.strftime('\%Y\%m\%d')
        else:
                date_str         = date
        #### Previous Date
        if prevDate is None:
                prevDate_str = date_str
        else:
                prevDate_str = str(prevDate)
```

```
    #### File names
    cstatus_file           = date_str + '_city_status.csv'
    tp_file                = '_temp_precip.csv'
    wb_file                = '_water_balance.csv'

    #### HEADERS ####
    #### Header for city status
    cstatus_header        = ['cs_id' , 'city', 'latitude', 'longitude',
              'round_lat' , 'round_lon', 'status', 'delta_lat', 'delta_long',
              'dist' , 'new_lat', 'new_long']
#### Header for Temp/Precip
    tp_header                = ['tp_id' , 'cs_id' , 'city', 'unit',
              'Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun',
              'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec', 'Ann']
    #### Header for Water Balance
    wb_header                = ['wb_id' , 'cs_id' , 'city', 'MON', 'TEMP',
              'UPE', 'APE', 'PREC', 'DIFF', 'ST', 'DST',
              'AE', 'DEF', 'SURP', 'SMT', 'SST']

    #### Dataset type
    if dtype=='tp':
        print_fstr        = 'temp_precip'
        print_str        = 'Temp/Precip'
        data_range        = 2
        header                  = tp_header
        datafile          = tp_file
        save_file        = date_str + datafile
        id_label          = 'tp' + '_id'

    else:
        print_fstr        = 'water_balance'
        print_str        = 'Water Balance'
        data_range        = 12
        header                  = wb_header
        datafile          = wb_file
        save_file        = date_str + datafile
        id_label          = 'wb' + '_id'

    #### CITY STATUS DATA ####
    #### Read in the city status data from the previous step
    if cstatus is None:
        try:
            cities_status = pd.read_csv(cstatus_file, header=0,
                index_col=0, names=cstatus_header)
            cstatus_ids        = cities_status.index.values
        except:
            print 'Could not find city status file.'
            print 'Try running water_balance_scraper(data_types=[\'status\'])'
            return None
    else:
        cities_status        = cstatus

    #### Set the data range to given input; otherwise to the whole data range
    try:
        # city_range = range(drange[0], drange[1])
        city_range          = range(drange[0], drange[1]+1)
```

```
except:
        city_range          = df.index.values

#### Set of cities for which status == True
new_cities0                = cities_status.ix[city_range]
trues_city                 = new_cities0.ix[new_cities0['status']==True]
trues_ids                 = trues_city.index.values

#### WATER BALANCE OR TEMP/PRECIP DATA ####
#### Should I look for an existing file for the scraped data?
if prev==False:
        df_cities         = new_pdf(cities_status, type)
else:
        #### File to try to read in
        data_readfile                 = prevDate_str + datafile
        #### Try to read in the data
        try:
                df_cities = pd.read_csv(data_readfile, header=0,
                   index_col=0, names=header)
        except:
                #### If reading in the data was unsuccessful, create a new dataframe.
                df_ReadError         = error_script(data_readfile, print_str, 'scrape_wimp_df')
                if df_ReadError is None:
                        return None
                else:
                        df_cities = new_pdf(cities_status, type)

#### Otherwise, try to read in the file
#### Cities for which Temp/Precip or Water Balance data are needed
new_cities = trues_city.ix[~trues_city['city'].isin(df_cities.city)]
new_cities_ids = new_cities.index.values
if len(new_cities.city) == 0:
        if silent == False:
                print 'Data for all cities in the ' , print_str , ' file have already been scraped.'
        else:
                pass
else:
        if silent == False:
            print print_str , ' data will be scraped for ' , len(new_cities.city), ' cities...'
            print 'First five cities: \n' , new_cities.city[new_cities_ids[0:4]]
        else:
                pass
    #### Scrape WebWIMP data tables ####
    for j in cities_status.city[new_cities_ids]:
            if silent == False:
                    print 'Getting ' , print_str , ' for ' , j , '...'
            else:
                    pass
            #### Query WebWIMP for data #### Generate list of indices
            j_city         = cities_status.ix[cities_status['city']==j]
        #### Scrape WebWIMP
        j_table = scrape_wimp(j_city, type=dtype)
        #### Update table
        df_cities.ix[j_table.index.values, header[1:]]         = j_table
        #### Let the computer (and server) rest
        time.sleep(0.1)
        #### Save file to .csv ####
```

```
		df_cities.to_csv(save_file, index=True, index_label=id_label)
	else:
		pass

	print 'Finished scraping ' , print_str , ' data.'

	return df_cities
#### MAIN FUNCTION ####
def water_balance_scraper(filename, date=None, drange=None,
	failed=False, silent=False, data_types=['status', 'tp', 'wb'],
	prev=False, prevDate=None):
	#### GET DATA ####
	if date == None:
		today			=  time.strftime('\%Y\%m\%d')
	else:
		today			= date

	cities_file		= filename

	#### CSV HEADERS ####
	c2scrape_header		= ['c_id' , 'city', 'latitude', 'longitude']

	### UrbMet city list and lat/long data:
	try:
		c2scrape			= pd.read_csv(cities_file, header=0,
		   index_col=0, names=c2scrape_header)
		c2scrape_ids			= c2scrape.index.values
	except:
		print('Could not find file of cities to scrape. ')
		c2s_error			= raw_input('Enter new filename? (y/n): ')
		if str.lower(c2s_error) == 'n':
			return None
		else:
			new_c2scrape		= raw_input('Enter new filename: ')
			try:
				c2scrape		= pd.read_csv(str(new_c2scrape), header=0,
				  index_col=0, names=c2scrape_header)
				c2scrape_ids = c2scrape.index.values
			except:
				print('Could not find the new file. Exiting...')
				return None
		c2scrape_ids = c2scrape.index.values
		print c2scrape_ids

	if drange is None:
		drange = [c2scrape_ids[0] , c2scrape_ids[-1]]
	else:
		drange = drange

	if silent == False:
		print 'Scraping data for ' , len(c2scrape_ids) , ' cities...'
		print 'First five cities: '
		print c2scrape.city[c2scrape_ids[0:4]]
	else:
		pass

	#### CITY STATUS
```

```
        if 'status' in data_types:
                if failed == True:
                        cities_status = get_city_status(c2scrape, date=today, drange=drange,
                                failed=True, silent=silent, prev=prev, prevDate=prevDate)
                else:
                        cities_status = get_city_status(c2scrape, date=today, drange=drange,
                                failed=False, silent=silent, prev=prev, prevDate=prevDate)
        else:
                cities_status = None
        if silent == False:
                if cities_status is not None:
                        print 'cities_status.head() : ' , cities_status.head()
                else:
                        pass
        else:
                pass

        #### TEMP/PRECIP
        if 'tp' in data_types:
                tp = scrape_wimp_df(c2scrape, drange=drange, type='tp', silent=silent,
                        cstatus=cities_status, prev=prev, prevDate=prevDate)
        else:
                tp = None
        if silent == False:
                if tp is not None:
                        print 'tp.head() : ' , tp.head()
                else:
                        pass
        else:
                pass

        #### WATER BALANCE
        if 'wb' in data_types:
                wb = scrape_wimp_df(c2scrape, drange=drange, type='wb', silent=silent,
                        cstatus=cities_status, prev=prev, prevDate=prevDate)
        else:
                wb = None
        if silent == False:
                if wb is not None:
                        print 'wb.head() : ' , wb.head()
                else:
                        pass
        else:
                pass
        return
\end{verbatim}
```

D.2 Code for Chapters 4 and 5

D.2.1 Statistics

stats_list

```
stats_list <- function(data_col, list_probs=seq(0, 1, 0.25)){
  require(moments)
```

```
  stats_object <- list()
  na_status    <- ifelse(any(is.na(data_col)) > 0, T, F)
  quant_matrix              <- t(as.matrix(quantile(data_col, probs=list_probs, na.rm=na_status)))
  colnames(quant_matrix)    <- paste0(list_probs*100, "%")

  stats_object["n"]         <- length(na.exclude(data_col))
  stats_object["mean"]      <- mean(data_col, na.rm=na_status)
  stats_object["sd"]        <- sd(data_col, na.rm=na_status)
  stats_object["skewness"]  <- skewness(data_col, na.rm=na_status)
  stats_object["kurtosis"]  <- kurtosis(data_col, na.rm=na_status)
  stats_object["quantiles"] <- list(quant_matrix)

  stats_object
}
```

stats_matrix

```
#### TAKE A stats_col OBJECT AND CREATE A MATRIX ####
stats_matrix <- function(stats_list_obj, type="row", name=NULL, latex=F, dset="both", n_obs=F){
  stats_matrix_obj <- as.matrix(cbind(c(stats_list_obj["n"]),
                               c(stats_list_obj["mean"]),
                               c(stats_list_obj["sd"]),
                               c(stats_list_obj["skewness"]),
                               c(stats_list_obj["kurtosis"]),
                               rbind(lapply(stats_list_obj["quantiles"][[1]], function(x) c(x)))), nrow=1)
  #### COLUMN NAMES ####
  names_R <- c("n", "mean", "sd", "skewness", "kurtosis", colnames(stats_list_obj["quantiles"][[1]]))
  names_latex <- c("$n$", "$\\mu$", "$\\sigma$", "$\\gamma$", "\\kappa",
                 paste0(sub("%", "", colnames(stats_list_obj["quantiles"][[1]])), "\\%"))
  if(latex==F){colnames(stats_matrix_obj) <- names_R}
  else{colnames(stats_matrix_obj) <- names_latex}
  #### ROW NAMES ####
  if(!is.null(name)){rownames(stats_matrix_obj) <- name}
  else{rownames(stats_matrix_obj) <- "dataset"}
  #### STATS, QUANTILES, OR BOTH? ####
  if(dset=="stats"){
    if(n_obs==F){
      stats_matrix_obj <- t(as.matrix(stats_matrix_obj[,2:5]))
      }
    else{stats_matrix_obj <- t(as.matrix(stats_matrix_obj[,1:5]))}}
  if(dset=="quants"){
    if(n_obs==F){stats_matrix_obj <- t(as.matrix(c(stats_matrix_obj[,6:length(stats_matrix_obj)])))}
    else{stats_matrix_obj <- t(as.matrix(c(stats_matrix_obj[,1], stats_matrix_obj[,6:length(stats_matrix_obj)])))}}
  else{
    if(n_obs==F){stats_matrix_obj <- t(as.matrix(stats_matrix_obj[,2:length(stats_matrix_obj)]))}
    else{stats_matrix_obj <- stats_matrix_obj}
  }
  #### ROW OR COLUMN? ####
  if(type=="row"){stats_matrix_obj}
  else{t(stats_matrix_obj)}
}
```

stats_ts

```
 #### STATISTICS ON MONTE CARLO TIME SERIES ####
stats_ts <- function(dfStatsList, StatsProbs){
  stats_ts_probs    <- seq(0, 1, 0.25)
```

```
  labStats_ts_probs <- paste0(stats_ts_probs*100, "%")

  dfNew     <- dfStatsList[, -1]
  dfNewTime <- dfStatsList[, 1]

  #### Make matrix ####
  dimnames_dfNewMatrix <- list(Year = dfStatsList$Year[1:length(dfNewTime)],
                               sim = names(dfStatsList[1:length(dfNewTime), -1]))
  dfNewMatrix <- as.matrix(as.data.frame(lapply(dfNew, as.numeric)))
  dimnames(dfNewMatrix) <- dimnames_dfNewMatrix

  #### Transpose matrix ####
  transNewMatrix <- t(dfNewMatrix)

  #### Get Quantiles ####
  tsQuantsMatrix <- t(apply(transNewMatrix, 2, function(x) quantile(x, stats_ts_probs)))
  tsQuantsDF <- as.data.frame(tsQuantsMatrix)
  df_quants  <- cbind(dfNewTime, tsQuantsDF)
  names(df_quants) <- c("Year", paste0(stats_ts_probs*100, "%"))

  row.names(df_quants) <- seq(1, length(dfNewTime))

  df_quants #### Return results ####
}
```

D.2.2 Simulation

case_scenarios

```
case_scenarios <- function(caseQuantiles, simTime){
  #### QUANTILE INFO ####
  num_quants <- length(caseQuantiles)
  df_scen  <- data.frame(simTime)
  for(i in 1:num_quants){
    scen    <- data.frame(rep(caseQuantiles[i], length(simTime)))
    df_scen <- cbind(df_scen, scen)
  }
  names(df_scen) <- c("Year", colnames(caseQuantiles))
  df_scen
}
```

caseSims

```
caseSims <- function(x, simTime, numTrials, randomType="normal"){
  #### Generate Simulations ####
  simRandomSeeds <- runif(n=numTrials, min=1, max=3000) #### Set random seeds
  df_sims        <- as.data.frame(simTime)              #### Dataframe of simulation results
  #### Distribution Parameters ####
  xMean <- x$mean
  xSd   <- x$sd
  for(i in 1:numTrials){            #### Iterate over number of simulations
    set.seed(simRandomSeeds[i])     #### Set a new seed
    #### Generate Random Values ####
    if( randomType=="Gamma" ){      #### Gamma ####
      rArg1     <- xMean^2 / xSd^2  #### Shape
      rArg2     <- xSd^2   / xMean  #### Scale
      simRandom <- c(rgamma(n=length(simTime), shape=rArg1, scale=rArg2))
```

```
    } else{                         #### Normal ####
      rArg1     <- xMean
      rArg2     <- xSd
      simRandom <- c(rnorm( n=length(simTime), mean =rArg1,    sd=rArg2))
    }                               #### Combine simulations
    df_sims     <- as.data.frame(cbind(df_sims, simRandom))
  }                                 #### Rename simulations:
  names(df_sims) <- c("Year", paste0("simrun", seq(1, numTrials, 1)))
  df_sims                           #### Return simulations
}
```

cCombineCases

```
#### Combine the simulation data for cases ####
cCombineCases <- function(dfCaseObject, newProbs = seq(0, 1, 0.25)){
  cCombineCasesVars <- c( "prec", "wui", "kpop", "area", "pop", "demand",
                          "supply", "wuci", "totWuci", "Rss" )
  newVarList  <- list( "prec", "wui", "kpop", "area", "pop", "demand",
                       "supply", "wuci", "totWuci", "Rss" )
  df_allCases <- list(simTypes = c("historical", "projection"),
                      historical = newVarList,
                      projection = newVarList)

  simTypes = c("historical", "projection")

  cCombineCasesList <- c("la", "sg")
  for( j in 1:length(simTypes) ){
    simType = simTypes[j]

    #### Iterate over variables ####
    for( i in 1:length(cCombineCasesVars) ){
      caseVar <- cCombineCasesVars[i]
      inputTypeSims <- data.frame()
      #### Object to hold the case data ####
      CasesData <- NULL

      #### Iterate over cases ####
      for( k in 1:length(cCombineCasesList) ){
        case <- cCombineCasesList[k]
        caseLabel <- dfCaseObject[[ case ]]$label

        if( caseVar != "area"){
          #### Melted Simulation Data ####
          caseDataSims <-
            as.data.frame(dfCaseObject[[ case ]]$simulations$simulation[[
              simType ]][[ caseVar ]]$sims$base$melt$sims)
          caseDataSims$Case <- caseLabel #### Add columns for cases ####
          caseDataSims$sim_type <- "Simulation"
          caseDataSims$quant_type <- "Simulation"
          #### Melted Quantile Data ####
          caseDataQuants <- as.data.frame(dfCaseObject[[ case ]]$simulations$simulation[[
            simType ]][[ caseVar ]]$sims$base$melt$quants)
          caseDataQuants$Case <- paste(caseLabel, "Quantiles")
          caseDataQuants$sim_type <- "Quantile"
          caseDataQuants$quant_type <- factor(caseDataQuants$trial)

          caseData  <- rbind(caseDataSims, caseDataQuants)
```

```
        CasesData <- rbind(CasesData, caseData)
      } else{ CasesData <- NULL }
    }
    #### Relevel the variables for plotting ####
    CasesDataCasesLevels <- c("Los Angeles", "Los Angeles Quantiles", "Los Angeles Historical",
                              "Singapore"  , "Singapore Quantiles"  , "Singapore Historical")
    CasesData$Case       <- factor(CasesData$Case, levels = factor(CasesDataCasesLevels))
    CasesData$sim_type   <- factor(CasesData$sim_type, levels = factor(c("Simulation", "Quantile")))
    CasesData$quant_type   <- factor(CasesData$quant_type,
                                     levels = factor(c("Simulation", "0%", "25%", "50%", "75%", "100%")))
    df_allCases[[ simType ]][[ caseVar ]]$sims    <- CasesData

    df_allCases[[ simType ]][[ caseVar ]]$subQuants <- subset(CasesData,
                                                              (!(as.character(CasesData$trial)=="0%") &
                                                                 !(as.character(CasesData$trial)=="100%")))
   }
  }
  df_allCases
}
```

cSimExp

```
#### GENERATE POPULATION FROM AVERAGE GROWTH CONSTANT KPOP
cSimExp <- function(x, baseValue=NULL, series1=NULL, seriesName="newSeries"){
  series1  <- ifelse(is.null(series1), "base", series1)
  df1      <- x[[series1]]$sims      #### Grab series from x (i.e. kpop):

  df1.Time   <- df1[,1]              #### Take the first column as the time
  lengthSeries <- length(df1.Time)
  #### Initial and Final Times       ####
  t_0 <- df1.Time[1]                 #### Initial Time
  t_i <- t_0 + 1                     #### Initial simulation time
  t_f <- df1.Time[lengthSeries]      #### Final Time
  #### Subset simulations  ####
  df1.Trials  <- df1[ which(df1$Year > t_0) , ]
  df1.simTime <- df1.Trials[ ,  1]   #### Simulation Time: first column
  df1.simRuns <- df1.Trials[, - 1]   #### #### Simulation Runs: Everything but the first column
  df1.numSims <- length(df1.simRuns) #### Number of simulations equals the number of columns
  #### Simulation dataframe ####
  newSims <- data.frame(df1.Time)    #### First column is time

  #### Find population for each column ####
  popSimsFunction <- function(X, baseValue=NULL){
    newSimX   <- c(baseValue)           #### Add the base value as the initial population in the series
    for( j in 1:length(X) ){            #### Iterate over the simulations
      pop1    <- newSimX[j]             #### Population at time j - 1
      popAdj  <- pop1*X[j]              #### X[j] is really the observation for the preceding time
      pop2    <- pop1 + popAdj          #### Population at time j
      newSimX <- append(newSimX, pop2)} #### Add population at time j to the series
    data.frame(newSimX)                 #### This should be of length equal to df1.Time
  }

  for(i in 1:length(df1.simRuns)){               #### Iterate over all of the simulations
    kpopX   <- df1.simRuns[,i]                   #### kpop to use for the simulation run
    popSim  <- popSimsFunction(kpopX, baseValue) #### Get the population for the simulations
    newSims <- cbind(newSims, popSim)
  }
```

```
  #### Base values ####
  names(newSims) <- names(df1)     #### Get the names from the original dataframe
  row.names(newSims) <- seq(1, length(df1.Time))
  newSims                          #### Return population
}
```

cSimProduct

```
#### Take the product of two cSimObjects
#### Modifications to Make: specify simulation runs
cSimProduct <- function(x, y, ...){ UseMethod("cSimProduct", y) }

cSimProduct.numeric <- function(x, y, baseValue=NULL, series1 = "base"){
  df1         <- x[[series1]]$sims #### Grab series from x ####
  df1.Time    <- df1[,1]           #### Take first column in x as the time
  lengthSeries <- length(df1.Time)
  #### Initial and Final Times ####
  t_0 <- df1.Time[1]             #### Initial time
  t_i <- t_0 + 1                 #### Initial simulation time
  t_f <- df1.Time[lengthSeries]
  #### Simulation Setup ####
  df1.Trials  <- df1[ which(df1$Year > t_0) , ] #### Subset simulations greater than baseTime
  df1.simTime <- df1.Trials[ ,  1]             #### Simulation Time: first column
  df1.simRuns <- df1.Trials[ , -1]             #### Simulations: everything but the first column
  df1.numSims <- length(df1.simRuns)           #### Number of simulations equals the number of columns
  #### Simulations ####
  newSims     <- data.frame(df1.Time)  #### New dataframe to hold the simulations: first column is time
  BaseValues  <- c(t_0, rep(baseValue, df1.numSims)) #### Base values, if provided
  #### Product of x and y ####
  if((length(y)==1) | (length(y)==length(df1.simTime))){
    newSims    <- df1.simRuns*y
    df.newSims <- data.frame(rbind(BaseValues, cbind(df1.simTime, newSims)))
    names(df.newSims) <- names(df1)
  } else{
  print(paste0("Error: ", sQuote("y"), " must have the same length as the simulation time."))
  print(paste0("length(x)=", length(x), ", length(y)=", length(y), " (where y is numeric)"))
  }
  df.newSims
}

cSimProduct.cSimObject <- function(x, y, #### First and second cSimObjects
                                   #### Which series to use in the cSimObjects
                                   series1 = "base", series2 = "base",
                                   baseValue=NULL, inverse = F){
  #### x is a cSimObject
  df1         <- x[[series1]]$sims        #### Grab data from x ####
  df1.Time    <- df1[,1]                  #### Take first column in x as the time
  lengthSeries <- length(df1.Time)
  #### y is a cSimObject
  df2         <- y[[series2]]$sims        #### Grab data from y ####
  df2.Time    <- df2[,1]                  #### Take first column in x as the time
  #### Initial and Final Times ####
  t_0         <- max(df1.Time[1], df2.Time[1])              #### Initial time
  t_i         <- t_0 + 1                  #### Initial simulation time
  t_f         <- max(df1.Time, df2.Time)  #### Final   simulation time
  # t_f          <- df1.Time[lengthSeries] #### Final   simulation time
  #### Simulation Setup ####
```

```
  #### df1 ####
  df1.Trials  <- df1[ which(df1$Year > t_0) , ] #### Subset simulations greater than baseTime
  df1.simTime <- df1.Trials[ ,  1]             #### Simulation Time: first column
  df1.simRuns <- df1.Trials[ , -1]             #### Simulations: everything but the first column
  df1.numSims <- length(df1.simRuns)           #### Number of simulations equals the number of columns
  #### df2 ####
  df2.Trials  <- df2[ which(df2$Year > t_0) , ] #### Subset simulations greater than baseTime
  df2.simTime <- df2.Trials[ ,  1]             #### Simulation Time: first column
  df2.simRuns <- df2.Trials[ , -1]             #### Simulations: everything but the first column
  df2.numSims <- length(df2.simRuns)           #### Number of simulations equals the number of columns

  #### Check Compatibility ####
  if((length(df1.simRuns) != length(df2.simRuns))){ #### Number of simulation runs
    print("Error: datasets must have the same number of simulation runs.")
  }
  if((length(df1.simTime) != length(df2.simTime))){ #### Time
    print("Error: datasets must have the same simulation time.")
    print("x has time: ")
    print(df1.simTime)
    print("y has time: ")
    print(df2.simTime)
  }
  #### Product ####
  BaseValues <- c(t_0, rep(baseValue, df1.numSims))
  if(inverse==T){
    newSims <- df1.simRuns/df2.simRuns
  } else{
    newSims <- df1.simRuns*df2.simRuns
      }

  df.newSims        <- data.frame(cbind(df1.simTime, newSims))
  names(df.newSims) <- names(df1)
  df.newSims
}

cSimProduct.default    <- function(x, y, #### First and second cSimObjects
                                 series1 = "base", #### Which series to use in the cSimObjects
                                 baseValue=NULL, inverse = F){
  #### x is a cSimObject
  df1 <- x[[series1]]$sims         #### Grab series from x ####
  df1.Time   <- df1[,1]            #### Take first column in x as the time
  lengthSeries <- length(df1.Time)
  #### y is a Data frame or matrix
  df2        <- y                  #### Grab series from y ####
  df2.Time   <- df2[,1]
  #### Initial and Final Times ####
  t_0          <- max(df1.Time[1], df2.Time[1])             #### Initial time
  t_i          <- t_0 + 1                  #### Initial simulation time
  t_f          <- max(df1.Time, df2.Time)  #### Final   simulation time

  #### Simulation Setup ####
  #### df1 ####
  df1.Trials  <- df1[ which(df1$Year > t_0) , ] #### Subset simulations greater than baseTime
  df1.simTime <- df1.Trials[ ,  1]             #### Simulation Time: first column
  df1.simRuns <- df1.Trials[ , -1]             #### Simulations: everything but the first column
  df1.numSims <- length(df1.simRuns)           #### Number of simulations equals the number of columns
  #### df2 ####
```

```
  df2.Trials  <- df2[ which(df2$Year > t_0) , ] #### Subset simulations greater than baseTime
  df2.simTime <- df2.Trials[ ,  1]              #### Simulation Time: first column
  df2.simRuns <- df2.Trials[ , -1]              #### Simulations: everything but the first column
  df2.numSims <- length(df2.simRuns)            #### Number of simulations equals the number of columns

  #### Check Compatibility ####
  if((length(df1.simRuns) != length(df2.simRuns))){ #### Number of simulation runs
    print("Error: datasets must have the same number of simulation runs.")
  }
  if((length(df1.simTime) != length(df2.simTime))){ #### Time
    print("Error: datasets must have the same simulation time.")
    print("x has time: ")
    print(df1.simTime)
    print("y has time: ")
    print(df2.simTime)
  }
  #### Product ####
  BaseValues <- c(t_0, rep(baseValue, df1.numSims))
  if(inverse==T){
    newSims <- df1.simRuns/df2.simRuns
  } else{
    newSims <- df1.simRuns*df2.simRuns
  }
  df.newSims        <- data.frame(cbind(df1.simTime, newSims))
  names(df.newSims) <- names(df1)
  df.newSims
}
```

cSimulate

```
#### c_simulate ####
#### THIS FUNCTION TAKES A DATA SERIES, DATAFRAME, OR CASE_LIST OBJECT AND PERFORMS SIMULATION.
c_simulate <- function(dataInput, ...){ UseMethod("c_simulate", dataInput) }

c_simulate.cSimObject <- function(dataInput, FUN, FUN.arguments = list(),
                                  seriesName = NULL, probs = NULL){
  #### Create a new series in the cSimObject
  seriesName              <- ifelse(is.null(seriesName), "base", seriesName)
  dataOutput              <- dataInput                       #### Save a copy of the data

  statProbs = ifelse(is.null(probs), seq(0, 1, 0.25), probs) #### Quantiles
  FUN.arguments$x <- dataInput                               #### Function Input
  dataOutputSims   <- do.call(FUN, FUN.arguments)            #### Sims ####
  #### Quants
  dataOutputQuants <- stats_ts(dataOutputSims, statProbs)    #### Quants ####

  #### Combine Sims and Quants
  dataOutputList   <- list(sims = dataOutputSims, quants = dataOutputQuants)
  dataOutput[[seriesName]] <- dataOutputList

  #### Melt ####
  dataOutputMelt <- cMeltSims(dataOutputList)
  dataOutputList <- append(dataOutputList, list(melt = dataOutputMelt))
  dataOutput[[seriesName]]    <- dataOutputList
  #### Return cSim Object ####
  dataOutput
}
```

```
c_simulate.default <- function(dataInput, t_i=1 , t_f=NULL, numTrials = 100  , randomType = "normal",
                               probs=seq(0, 1, 0.25), valType = "prec"){
  #### Simulation Data Input ####
  simDataSummary <- dataInput$summary[[valType]]
  #### Simulation Time ####
  simTrialTime      <- seq(t_i, t_f, 1)
  simLength    <- length(simTrialTime)

  #### Define cSimObject ####
  cSimObject        <- list(simOptions = list(initialTime = t_i, finalTime = t_f,
                                              numSims = numTrials, randomType = randomType))

  #### Base Simulation ####
  baseSims   <- caseSims(simDataSummary,
                         simTime    = simTrialTime,
                         numTrials  = numTrials,
                         randomType = randomType)
  print("Finished simulations...")
  #### Simulation Quantiles ####
  baseSimsQuants  <- stats_ts(baseSims, probs)
  #### Format simulations ####
  print("Formatting simulations...")
  baseSimsList <- list(sims = baseSims, quants= baseSimsQuants)
  baseSimsMelt <- cMeltSims(baseSimsList)
  #### Return cSimObject ####
  baseSimsObject <- list(sims = baseSims, quants = baseSimsQuants, melt = baseSimsMelt)
  cSimObject$base <- baseSimsObject

  class(cSimObject) <- append(class(cSimObject), "cSimObject")
  cSimObject
}
```

cSimBaseValues

```
lookupInput <- function(data, year, colName){data[data$Year==year, colName]}
#### This function looks up values in a data series
cSimBaseValues <- function(data, xCol=NULL, xTime="last", timeCol=NULL, na.omit=T){
  #### Specify data column ####
  if( is.null(xCol) ){ #### What to do if no xCol is specified
    #### Default first column if data has only one column, otherwise second column
    if(length(data==1)){ #### Default first column
      xCol=colnames(data[1])
    } else{ #### Default second column
      xCol=colnames(data[2])
    }
  } else{ #### What to do if a column was specified
    if(is.numeric(xCol)){ #### If 'xCol' is numeric, get column name
      xCol = colnames(data[xCol])
    } #### Otherwise, use column specified, i.e. xCol = xCol
  }
  #### Specify time column ####
  if( is.null(timeCol) ){ #### What to do if 'timeCol' is NULL
    if( length(data) == 1 ){ #### If dataframe has only one column
      data$Time <- row.names(data) #### Set 'Time' equal to row names
    } else{
      data$Time <- data[ , 1] #### Set 'Time' to first column
```

```
    }
  } else{ #### What to do if 'timeCol' is not NULL
    data$Time <- data[, timeCol]
  }

  #### Na.omit = T ####
  newData <- data[ which(is.na(data[xCol]) == F ) , ]
  #### Get Base Value and Time ####
  lengthSeries <- length(newData[, xCol]) #### Get length of series
  #### Set index to look up
  lookupIndex  <- ifelse(xTime == "last", lengthSeries,
                         ifelse(xTime == "first", 1, which(newData$Time == xTime)))
  #### Return Values
  returnValue <- newData[lookupIndex, xCol]
  returnTime  <- newData$Time[lookupIndex]
  list(baseValue = returnValue, baseTime = returnTime)
}
```

cSimCaseObject

```
#### Create a case object based on a dataframe
cSimCaseObject <- function(dataInput, colNames = NULL, probs = NULL) {
  #### Input the dataframe and column names
  probs      <- ifelse(is.null(probs), seq(0, 1, 0.25), probs)

  #### Subset data based on column names
  inputTypes <- c("prec", "wui", "kpop", "area", "pop", "demand", "supply", "wuci", "totWuci", "Rss")
  if( is.null(colNames) ){
    simCols <- inputTypes
  } else{
    simCols    <- c()
    for(i in 1:length(inputTypes)){
      if( is.numeric( colNames[[ inputTypes[i] ]][[1]] ) ){
        simCols[i] <- colnames( dataInput[ colNames[[ inputTypes[i] ]][[1]] ] )
      } else{
        simCols[i] <- colNames[[ inputTypes[i] ]][[1]]
      }
    }
  }

  caseData   <- dataInput[ , c("Year", simCols)] #### Subset data

  CaseObject <- list(name = deparse(substitute(dataInput)),
                     data = dataInput,
                     dataColumns = simCols,
                     summary=list())

  for(i in inputTypes){
    colI <- simColList[[i]]
    CaseObject$summary[[i]] <- stats_list(dataInput[, colI ], list_probs = simProbs)
    CaseObject$summary[[i]]$column <- colI
  }

  CaseObject
}
```

cSimDataSummary

```
#### cSimDataSummary ####
cSimDataSummary         <- function(dataInput, ...) {UseMethod("cSimDataSummary", dataInput)}

cSimDataSummary.list    <- function(dataInput){ #### For list object
  if(is.null(dataInput$mean) | is.null(dataInput$sd)){
    print("Error: Must enter data series or a list object with slots ",
          sQuote("mean"), "and ", sQuote("sd"))
  } else{
    dataInput
  }
}

cSimDataSummary.default <- function(dataInput, probs){#### For numeric vectors
  stats_list(dataInput, probs)
}
```

cMeltSims

```
#### cMeltSims ####
cMeltSims <- function(cSimObjectSims){
  cSimObjectSimsMelt <- list()
  if( !is.null(cSimObjectSims$sims)){
    cSimObjectSimsMelt$sims        <- melt(cSimObjectSims$sims, id.vars = c("Year"))
    names(cSimObjectSimsMelt$sims) <- c("Year", "trial", "value")
  } else{
    cSimObjectSimsMelt$sims <- NULL
  }
  if( !is.null(cSimObjectSims$quants)){
    cSimObjectSimsMelt$quants        <- melt(cSimObjectSims$quants, id.vars = c("Year"))
    names(cSimObjectSimsMelt$quants) <- c("Year", "trial", "value")
    print("finished quants")
  } else{
    cSimObjectSimsMelt$quants <- NULL
  }
  print("Returning simulations...")
  cSimObjectSimsMelt
}
```

cSim

```
cSim <- function( x, simVars  = NULL, simTypes = NULL, numTrials = 100,
                  baseValues = T, t_F=NULL, t_f=NULL, t_i = NULL, t_I = NULL, case = NULL, probs=NULL){
  #### Variables to be simulated ####
  probs <- ifelse(is.null(probs), seq(0, 1, 0.25), probs)
  defaultVars <- c( "prec", "wui", "kpop" , "area", "pop", "demand", "supply", "wuci", "totWuci", "Rss")
  CinputTypes <- defaultVars

  print(CinputTypes)
  colList      <- c("ann.prec.webwimp.m", "Water.Use.Intensity.m3.cap.year", "kpop", "City.Area.m2",
                  "Population", "Total.Water.Use.m3.year", "rainfall.volume.m3.per.yr", "wuci.m2.cap",
                  "wuci.m2", "Rss")

  simTypes <- c("historical", "projection")
  n_trials <- numTrials
```

```
#### Input Times ####
cCaseObject <- list() #### Define cCaseObject ####

#### Simulation Type ####
for(simType in 1:length(simTypes)){   #### Historical or projection ####
  j <- simTypes[simType]
  print(paste0("simulation type: ",  j))

  cCaseObjectJ           <- list( prec = list(),
                                  wui = list(),
                                  kpop = list(),
                                  pop = list(),
                                  demand = list(),
                                  supply = list(),
                                  wuci = list(),
                                  totWuci = list(),
                                  Rss = list())
  #### Initial Values Lookup ####
  lookupTime                   <- ifelse(j == "historical", "first", "last")

  for( i in 1:length(CinputTypes) ){
    #### Iterate over input types, e.g. prec, wui, kpop, etc.
    CinputType                  <- CinputTypes[i]
    print(paste0("Simulating for variable ", CinputType))

    cCaseObjectJ_I <- list()
    cCaseObjectJ_I$initial     <- cSimBaseValues(x$data, xCol = colList[i], xTime=lookupTime)

    #### Simulation Times ####
    T_0       <- ifelse( j == "historical", t_i, t_I)
    T_i       <- ifelse(baseValues == F, T_0, T_0 + 1)
    T_f       <- ifelse( j == "historical", t_f, t_F)

    print(c(paste0("Input: ", CinputType), paste0("initial time: ", T_i),
            paste0("final time: ", T_f)))

    simTimes  <- seq(T_i, T_f, 1)
    simPeriod <- length(simTimes)
    #### Base Value ####
    # x_i       <- ifelse(cCaseObjectJ_I$initial$baseTime <= T_0,
                        # cCaseObjectJ_I$initial$baseValue, NA)
    x_i       <- cCaseObjectJ_I$initial$baseValue
    print(CinputType)

    if( i <= 3 ){
      randomType <- ifelse(CinputType == "kpop", "normal", "Gamma")
      cCaseObjectJ_I$sims <-  c_simulate(x, t_i = T_i, t_f = T_f, numTrials = n_trials,
                                         randomType=randomType, valType = CinputType)
     } else if(CinputType == "area"){
       4
     } else if(CinputType == "pop"){ #### Pop ####
        x1 <- cCaseObjectJ[[ "kpop" ]]$sims
        cCaseObjectJ_I$sims       <- c_simulate(x1 , FUN = "cSimExp",
                                                FUN.arguments = list(baseValue = x_i))
        cCaseObjectJ_I$baseSeries <- c("kpop")
      } else if(CinputType == "demand"){ #### Demand ####
        #### Multiply population by WUI
```

```
        x1 <- cCaseObjectJ[[ "pop" ]]$sims
        x2 <- cCaseObjectJ[[ "wui" ]]$sims
        cCaseObjectJ_I$sims <-  c_simulate(x1, FUN = "cSimProduct",
                                   FUN.arguments = list(x2, baseValue = x_i))
        cCaseObjectJ_I$baseSeries <- c("pop", "wui")
      } else if(CinputType == "supply"){ #### Supply ####
        #### Use data if j = historic
        if(j == "historic"){
          # areaColName <- ifelse(case == "la", "City.Area.interp.constant.m2", "City.Area.m2")
          areaColName <- "City.Area.m2"
          x2 <- x$data[, areaColName]
        } else{
          x2 <- c(cCaseObjectJ[[ "area"]]$initial$baseValue)
        }
        # print(x2)
        x1 <- cCaseObjectJ[[ "prec" ]]$sims
        print(x_i)
        cCaseObjectJ_I$sims <-  c_simulate(x1, FUN = "cSimProduct",
                                   FUN.arguments = list(x2, baseValue = x_i))
      } else if(CinputType == "wuci"){ #### WUCI ####
        x1 <- cCaseObjectJ[[ "wui" ]]$sims
        x2 <- cCaseObjectJ[[ "prec" ]]$sims
        cCaseObjectJ_I$sims <-  c_simulate(x1, FUN = "cSimProduct",
                                   FUN.arguments = list(x2, baseValue = x_i, inverse = T))
      } else if(CinputType == "totWuci"){ #### Total WUCI ####
        x1 <- cCaseObjectJ[[ "wuci" ]]$sims
        x2 <- cCaseObjectJ[[ "pop" ]]$sims
        cCaseObjectJ_I$sims <-  c_simulate(x1, FUN = "cSimProduct",
                                   FUN.arguments = list(x2, baseValue = x_i))
      } else{ #### Rss ####
        x1 <- cCaseObjectJ$supply$sims
        x2 <- cCaseObjectJ$demand$sims
        cCaseObjectJ_I$sims <-  c_simulate(x1, FUN = "cSimProduct",
                                   FUN.arguments = list(x2, baseValue = x_i, inverse =T))
      }

    cCaseObjectJ[[ CinputType ]] <- cCaseObjectJ_I
   }
   if( j == "historical"){
    x$simulation$historical <- cCaseObjectJ
   } else{
    x$simulation$projection <- cCaseObjectJ
   }
  }
  x
}
```