

Winter 4-4-2014

Household Water Demand and Land Use Context: A Multilevel Approach

Elizabeth Yancey Breyer
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/open_access_etds



Part of the [Water Resource Management Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Breyer, Elizabeth Yancey, "Household Water Demand and Land Use Context: A Multilevel Approach" (2014). *Dissertations and Theses*. Paper 1670.
<https://doi.org/10.15760/etd.1669>

This Thesis is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Household Water Demand and Land Use Context:

A Multilevel Approach

by

Elizabeth Yancey Breyer

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science
in
Geography

Thesis Committee:
Heejun Chang, Chair
Teresa Bulman
Geoffrey Duh

Portland State University
2014

Abstract

Urban water use arises from a mix of scale-dependent biophysical and socioeconomic factors. In Portland, Oregon, single-family residential water use exhibits a tightly coupled relationship with summertime weather, although this relationship varies with land use patterns across households and neighborhoods. This thesis developed a multilevel regression model to evaluate the relative importance of weather variability, parcel land use characteristics, and neighborhood geographic context in explaining single-family residential water demand patterns in the Portland metropolitan area. The model drew on a high-resolution panel dataset of weekly mean summer water use over five years (2001-2005) for a sample of 460 single-family households spanning an urban-to-suburban gradient. Water use was found to be most elastic with respect to parcel-scale building size. Building age was negatively related to water use at both the parcel and neighborhood scale. Half the variation in water use can be attributed to between-household factors. Between-neighborhood variation exerted a modest but statistically significant effect. The analysis decomposed household temperature sensitivity into four components: a fixed effect common to all households, a household-specific deviation from the fixed effect, a separate extreme heat effect, and a land use effect, where lot size exaggerated the effect of temperature on water use. Results suggested that land use planning may be an effective non-price mechanism for long-range management of peak demand, as land use decisions have water use implications. The combined effects of population growth, urbanization, and climate change expose water providers to risk of water stress. Modeling fine-grain relationships among heat, land use, and water use across scales plays a role in long-range climate change planning and adaptation.

Dedication

This thesis is dedicated to my mother and father, who raised me in a dense urban area, taught me to value water at a very young age, and have never been Lawn People.

Acknowledgements

Many thanks are due to my advisor, Dr. Heejun Chang, whose ongoing encouragement and support have been invaluable to my development as a graduate student. I would also like to thank committee members Dr. Teresa Bulman, and Dr. Geoffrey Duh for their time and careful feedback throughout this thesis process. Dr. Jason Newsom taught multilevel modeling class that served as the knowledge base for the methodology I used. Members of Dr. Chang's water research group also provided helpful feedback as I developed the groundwork for this thesis.

This research also owes a debt of gratitude for background information, data provision, and feedback from Portland Water Bureau staff members Lorna Stickel and Dr. Hossien Parandvash, Tualatin Valley Water District staff member Steven Carper, and Metro Regional Government staff members Dennis Yee and Jim Csar.

Special thanks are due to my mom and dad, my brother Bobby, and to Auntie Amy and Uncle Jack. I also want to thank my friends, particularly the French family as well as Nili Yosha and Annie Johnson, who generously hosted me as I completed my thesis. Thanks also go out to the people who kept me sane, especially Jennifer Knipling, Niki Harrison, Sara Tretter, Julia Kuhns, Ben O'Donnell and Adonia Lugo, Fiona Gladstone and Austin Brown, Annie Scriven, Sonia Singh, Neil, Josh, and Jake.

This research was funded by a grant from the National Oceanic and Atmospheric Administration Sectoral Applications Research Program. I am grateful for additional funding from the Courtney Scholarship, Oregon Geographical Alliance, and the Rockey Scholarship.

Contents

Abstract.....	i
Dedication.....	ii
Acknowledgements.....	iii
List of Tables	vii
List of Figures.....	viii
Chapter 1: Introduction.....	1
1.1. Climate change and urban water resource management	1
1.2. Multiple scales, multiple populations	2
1.3. Problem to be investigated.....	4
1.4. Research questions	5
1.5. Structure of the thesis.....	6
Chapter 2: Background and Literature Review	9
2.1. Urban water use as coupled human-natural system	9
2.1.1. Water use and scale	10
2.1.2. Seasonality of water use	10
2.1.3. Drivers of single-family residential (SFR) water use.....	12
2.1.4. The ‘nature’ of residential landscapes	13
2.2. Nested hierarchies	14
2.2.1. Multilevel data structures	15
2.2.2. Ecological and atomistic fallacies	15
2.2.3. Geographic hierarchies	16
2.2.4. Problems of neighborhood context.....	18
2.2.5. Neighborhood effects	19
2.3. The study area: Portland, Oregon.....	21
2.3.1. Water supply infrastructure	21
2.3.2. Characterizing regional water use	23
2.3.3. Climate change	24
2.3.4. Land use change	25
Chapter 3: Theory	27
3.1. Mixed effects.....	27
3.2. Conceptualizing a multilevel approach to residential water use	28
3.3. Developing a multilevel model of residential water use	30
3.3.1. An unconditional means model	30
3.3.2. A three-level random slopes model with cross-scale interactions.....	33
3.4. Diagnostics	36

3.4.1. Statistical significance	37
3.4.2. Goodness of fit.....	37
Chapter 4: Data and Methods	39
4.1. Data	39
4.1.1. Response variable: Single-family residential (SFR) water use	39
4.1.2. Explanatory variables	42
4.1.3. Variable transformations	46
4.2. Model development.....	47
4.2.1. Initial random intercepts structure.....	47
4.2.2. Specifying random effects.....	49
4.2.3. Back fitting fixed effects	49
4.3. Model diagnostics	51
Chapter 5: Results.....	52
5.1. Results of unconditional means model	52
5.2. Results of model development	56
5.2.1. Fitting random effects.....	56
5.2.2. Fitting fixed effects.....	57
5.3. Full model	57
5.3.1. Estimated fixed effects	58
5.3.2. Estimated variance components	60
5.4. Diagnostics.....	62
5.4.1. Normality and independence of residuals	62
5.4.2. Goodness of fit	62
5.4.3. Confidence intervals and hypothesis tests.....	64
5.4.4. Predictive power of the multilevel model.....	65
Chapter 6: Discussion	67
6.1. Decomposition of weather sensitivity	67
6.2. Linkages between land use and water use.....	69
6.3. Why was land cover not more predictive?.....	73
6.4. Variance decomposition.....	74
6.5. Policy implications.....	75
6.5.1. Triple exposure	76
6.5.2. Vegetation feedbacks and landscape regulations	79
6.5.3. Land use planning as equitable water demand management.....	80
Chapter 7: Conclusions.....	83
7.1. Summary of main findings.....	83
7.2. Addressing issues of scale and aggregation	85
7.3. Climate change, urban form, and resilience.....	85

7.4. Future work	86
References.....	88

List of Tables

Table 4.1: Candidate explanatory variables.....	43
Table 4.2: Functions used for model development with R Statistical Software.....	48
Table 4.3: Candidate cross-scale interactions.....	50
Table 5.1: Results from unconditional means models.....	53
Table 5.2: Design effects from unconditional means models.....	55
Table 5.3: Regression coefficients.....	59
Table 5.4: Variance components.....	59
Table 5.5: Relative goodness of fit across models.....	64

List of Figures

Figure 2.1: Urban water supply infrastructure	22
Figure 2.2: Urban density in the study area	25
Figure 3.1: Conceptual diagram of household water use as a nested data structure.....	29
Figure 4.1: Seasonality of daily household water use by month, 2001-2005	41
Figure 4.2: Spatial variation of daily household water use by month, 2001-2005	41
Figure 4.3: Distribution of temperature and precipitation patterns, 2001-2005.	41
Figure 4.4: Spatial distribution of households and neighborhood boundaries.....	45
Figure 5.1: Distribution of random effects from unconditional means models.....	54
Figure 5.2: Distribution the raw data and response variable	56
Figure 5.3: Correlation of household-specific intercepts and slopes	61
Figure 5.4: Normality and independence of level 1 residuals	63
Figure 5.5: Correlation between lot size and random effect for temperature	63
Figure 5.6: Confidence intervals for fixed effects in full model.....	65
Figure 5.7: Predicted water use from the multilevel model.....	66
Figure 6.1: Decomposition of temperature response	69
Figure 6.2: Linkages between land use and water use across scales	70
Figure 6.3: The contingent effect of lot size	72
Figure 6.4: Future climate projections in the study area.....	77
Figure 6.5: Projected change in mean single-family residential lot area.....	78

Chapter 1: Introduction

1.1. Climate change and urban water resource management

Clean, potable water is an increasingly scarce, valuable resource. Although U.S. per-capita water use has declined relative to the 1970s, population growth has stressed urban water supply infrastructure, leading to management of demand through higher prices, technological change, and conservation incentives (Gleick 2003; Coomes et al. 2011). Climate change threatens to add a new source of stress to urban water provisioning systems (Miller and Yates 2006). In the Pacific Northwest, a wide range of climate change scenarios have projected hotter, drier summers, which could reduce snowpack and modify the timing and magnitude of streamflow, potentially limiting water supply during peak-season demand in late summer (Mote and Salathé 2010). At the same time, peak demand may itself increase, as air temperature drives summer water consumption (Praskievicz and Chang 2009). The prospect of climate-induced reductions in supply and increases in demand, compounded by the effects of population growth, paints a potentially grim picture of intensified water stress associated with warming climate patterns.

The reliability of urban water provisioning systems in the next century will depend on stabilizing peak demand (Vano et al. 2010). A growing body of literature on the geography of urban water use suggests that building structural and land cover variables mediate the relationship between climate and peak demand (Domene and Saurí 2006; March and Saurí 2009). In summer, single-family residential (SFR) water use tends to be less weather sensitive in denser neighborhoods with smaller lots or more

multifamily housing, as there is less space for irrigated vegetation (Balling et al. 2008; Breyer et al. 2012). These findings point to a role for urban densification in softening the impact of higher summer temperatures on water use as part of a broader plan for adaptation to climate change in cities.

Clear articulation of the relationships among climate, urban form, and water use could also inform long-range forecasting. Several scenario-based analyses have incorporated the effects of both climate and land use change on projected water demand, yielding surprising results. Low-density residential development dominated climate change in driving future neighborhood-level water use increases in Hillsboro, Oregon, a suburb of Portland (House-Peters et al. 2010). Conversely, a broader study that jointly modeled urban water supply and demand in the Seattle metropolitan area found that continued urban densification was key to stabilizing peak demand and averting shortages (Polebitski et al. 2011). These findings support the notion of a conservation effect associated with urban density that may help alleviate the severity of regional water stress. Long-range water demand models that assume a static urban form or otherwise neglect the place-specific effects of land use on water use may misgauge the trajectory demand, overestimating in densifying areas and underestimating in areas experiencing rapid low-density development.

1.2. Multiple scales, multiple populations

Land use shapes the relationship between climate and water use. Evidence for this is compelling, although the bulk of analyses focused on this relationship have relied on spatially and temporally aggregated data. Because fine-scale water consumption data

have historically been unavailable, it is largely unknown how much the scale of analysis influences the strength of this relationship. Does the purported conservation effect of urban density aggregate up from daily household processes or operate independently as a contextual effect at the neighborhood scale? Alternatively, does density interact with finer-scale household processes in ways that may not be evident from a single-scale perspective? To model the effect of land use on household water use without inadvertently introducing ecological or atomistic fallacies (Robinson 1950; Alker 1969), it is necessary to examine water use at multiple scales.

This analysis drew on a rare, high-resolution panel dataset of daily SFR water use to examine the micro-foundations of established meso-scale relationships between weather, water use, and land use. Even at this fine scale, however, it should be recognized that a certain level of abstraction from direct experience is present in the data. The finest resolution available is the household, an aggregation of one or more individuals in space, on a specific day, which represents an aggregation of water-consuming moments. Each of these moments is intricately bound up with a complex mix of factors, including "local cultures, everyday routines, institutional practices and structures" (Jones and Duncan 1995, 29). As a consequence of this abstraction, the signal connecting the consumptive moment with its context has been encoded, albeit imperfectly, in the data.

When attributed with parcel-scale and neighborhood-scale land use characteristics, multiple related populations are being represented in these data: a population of water use observations, a population of SFR households, and a (geographically uncertain) population of neighborhoods (Kwan 2012). The study period

can also be thought of as a sample from a super-population of summer days in the Portland (Crowder and South 2011). Broader socio-economic, political, and climactic factors play a role in shaping these parameters within the Pacific Northwest or between regions. This analysis takes a quantitative approach to grappling with a small piece of this complexity, SFR water consumption across households and neighborhoods in the Portland metropolitan area.

1.3. Problem to be investigated

Outdoor water use figures prominently in management of residential landscapes, even in the maritime Pacific Northwest, where the climate is marked by distinct rainy and dry seasons. In Portland, Oregon, water use exhibits a tightly coupled relationship with summertime weather and climate, although this relationship varies with patterns of land cover and building structure across households and neighborhoods. This thesis developed a multilevel regression model to evaluate the relative importance of weather variability, parcel characteristics, and neighborhood geographic context in explaining SFR water demand patterns. Multilevel models tease out variance associated with each level of a nested set of relationships (Goldstein 2003). In this case, a time series of daily water demand observations was nested within individual households. Households were, to an unknown degree, nested within neighborhoods; both spatial units were attributed with land use characteristics. In addition to partitioning variance across scales, the analysis tested for heterogeneity in finer-scale coefficients and interactions between drivers across scales. Results are presented against current understanding of relationships among weather, land use, and residential water use. The research concluded by relating key

findings back to the problem of double (or triple) exposure and potential adaptation strategies for urban water provisioning systems to the combined effects of climate, population, and land use change in American cities. This thesis was a component of a broader study on the effects of combined climate and land use change on the sustainability of urban water resources.

1.4. Research questions

The goals of this thesis were fourfold. First, the thesis sought to detect and measure the degree of heterogeneity in weather sensitivity of water use patterns across households and neighborhoods. Second, the elasticity of water use with respect to parcel-scale and neighborhood scale building structure and land cover characteristics was estimated. Third, variance in water use was partitioned across scales. Finally, the thesis sought to determine whether the drivers of water use interact across scales and measure their effects. The analysis will provide answers to the following research questions:

- 1. How sensitive is SFR water use to weather variation? To what extent does this sensitivity vary across households?*

- 2. To what extent do individual household characteristics explain variance in SFR water use patterns? How do parcel characteristics affect water use and does this effect vary across neighborhoods?*

3. *To what extent do neighborhood characteristics explain variance in household water use patterns?*

4. *To what extent to these variables interact across scales?*

1.5. Structure of the thesis

Chapter 2 reviews relevant literature on water use. It examines evidence linking weather variation with water use at multiple spatial and temporal scales, paying particular attention to previous studies of SFR water use and land use covariates. Qualitative evidence for neighborhood influences on household land management decisions is presented. The literature review then turns to data hierarchies, describing how these data structures relate to ecological and atomistic fallacies and the modifiable areal unit problem. Current problems in the detection and explanation of neighborhood effects are presented. The chapter concludes by presenting relevant aspects of the urban water provisioning system in the study area of Portland, Oregon.

Chapter 3 constructs a theoretical scaffolding for multilevel regression. It explains the theory behind fixed and random effects, variance decomposition, and cross-scale interactions. Unconditional means and random slopes models for water use are developed and parameters are interpreted. Problems related to assessing goodness of fit and statistical significance in multilevel models are discussed. Chapter 4 describes data sources and pre-processing of water use, weather, and land use variables. It also details the model development process and diagnostic procedures.

Chapter 5 presents the results. An increasingly complex series of models are developed to illustrate how variance in water use is partitioned across scales. The final model includes random intercepts for household and neighborhood as well as random slopes for household response to maximum air temperature. Fixed effects are specified to account for persistent effects of heat and rain, parcel-scale effects of lot size, building size, and building age, as well as neighborhood scale effects of building age and a cross-level interaction involving maximum air temperature and lot size. The final model explains 67% of variance in SFR water use.

Chapter 6 discusses these results in light of existing research and explores implications for urban water resource sustainability, tying these implications back to central questions of climate change adaptation and the conservation effect of urban density. It suggests that land use planning represents an overlooked strategy for long-range water demand management and draws connections between results and the need for integrated land use planning and water resource management.

Chapter 7 concludes by summarizing key analytical results. On average, maximum air temperature tends to be the primary driver of temporal variation in summertime water use, while building size is the primary driver of spatial variation in water use. The effect of temperature was decomposed into an average or fixed effect, a household-specific deviation from that average, an extreme heat effect, and a land use effect, encapsulated by a cross-level interaction that was detected between temperature and lot size. Building age varies inversely with water use and is significant at both the parcel and neighborhood scales, indicating that elevated level of summertime water use were associated with new construction, primarily construction of low-density housing.

Results provide evidence that land use patterns have water use implications and underscored the utility of modeling fine-grain relationships among heat, land use, and water use across scales as part of a broader program of planning and adaptation to combined climate and land use change.

Chapter 2: Background and Literature Review

This chapter reviews the literature on urban water use through a framework of coupled human-natural systems. Section 2.1 draws on a diverse range of quantitative and qualitative studies to discuss the roles of space, scale, and seasonality in the patterning of SFR water use. Section 2.2 frames SFR water use as a nested hierarchy and discusses methodological issues related to statistical dependence and heterogeneity. Section 2.3 concludes by presenting relevant aspects of the water supply in the study area of Portland, Oregon, along with expected shifts associated with combined climate and land use change.

2.1. Urban water use as coupled human-natural system

A rich body of academic literature has developed around urban water demand modeling and forecasting. Throughout the 20th century, however, this research focused primarily on modeling temporal rather than spatial variation (House-Peters and Chang 2011a; Tanverakul and Lee 2012). Prices, technology, and climate were seen the key drivers of demand; econometric and time series analyses dominated the research agenda (Maidmont et al. 1985; Zhou et al. 2001; Arbués et al. 2003). However, water use research has carved out a more prominent role for geography in the first decade of the 21st century. Increasingly approached through the framework of a coupled human-natural system (Liu et al. 2007), urban water use has been understood as an explicitly spatial phenomenon arising from complex biophysical and social processes that operate at multiple spatial and temporal scales (March and Saurí 2009). The shift towards recognizing the roles of space and scale of water consumption parallels a second

quantitative revolution in geographic research (Kwan and Schwanen 2009) as well as a new emphasis on the place-based dimensions of consumption (Goodman et al. 2010) and effects of the build environment on behavior (Norman et al. 2006).

2.1.1. Water use and scale

Drivers of water use vary across scales. At the global scale, rising incomes dominate population growth in driving increased water use and subsequent water stress in large river basins (Alcamo et al. 2007). Weather and climate explain much of the variation in daily water use across U.S. cities (Maidmont and Miaou 1986), while water prices account for demand variation between cities at coarser temporal resolutions (Arbués et al. 2003; Grafton et al. 2011). As intraurban water use data have become available, socio-demographic factors, building structural characteristics, and land cover patterns have been found to account for spatial variation in water use within cities (Wentz and Gober 2007; Balling et al. 2008; Chang et al. 2010), filtering coarse-scale climate and economic signals into fine-scale water use outcomes (Saurí 2003). The relationship between water use and urban form is particularly prominent in lower-density suburban areas (March and Saurí 2010). Although the drivers of water use vary across scales, the role of scale itself in structuring water use is not well understood (Medd and Chappells 2007; Perveen and James 2010), particularly for urban residential water use.

2.1.2. Seasonality of water use

Weather and climate are the primary biophysical drivers of urban water use. Urban water use studies customarily subset total demand into two components, base and seasonal use (Howe and Linaweaver 1967; Maidmont et al. 1985; Miaou 1990; Zhou et

al. 2001). Generally taken as mean winter water demand, base use reflects daily indoor water consumption associated with cooking, drinking, cleaning, and sanitation. Base use is considered climate-invariant, fluctuating instead with incomes and technology, although this may not hold for warm-winter areas where households use water outdoors year-round (Gato et al. 2007). The difference between winter and summer water use is taken to represent seasonal (or outdoor) water use. Seasonal use responds to variations in weather and climate, reflecting water 'lost' to evapotranspiration (ET).

The temperature that elicits seasonal use varies across cities and regions. Akuoko-Asibey et al. (1993) found mean per-capita weekly water use became responsive to temperatures above 15°C in Calgary, Canada, while Maidmont and Miaou (1986) estimated the changepoint at 21°C across nine east coast U.S. cities. Protopapas et al. (2000) located the threshold temperature at 25.5°C in hyper-dense New York City, United States. Zhou et al. (2001) found a second, steeper nonlinear increase above 39°C in Melbourne, Australia. Urban water use varies inversely with precipitation patterns, although summer usage tends to respond more to rainfall occurrence than to depth (Miaou 1990; Adamowski and Karapataki 2010). The magnitude of seasonality varies widely within and across cities, while base use is relatively constant (Maidmont and Miaou 1986; Mayer and DeOreo 1999). Rockaway et al. (2011) compared residential use across 11 North American cities and found the lowest annual rates of water use in the rainy cities of Seattle, Washington, (annual average of 169 gallons per day, or GPD) and Cleveland, Ohio, (191 GPD), while the highest users were in the desert cities of Las Vegas, Nevada, (557 GPD) and Phoenix, Arizona, (441 GPD).

Since seasonal use varies with climate, it is the component of water use that exposes municipalities to risk of summer water scarcity, particularly in cases of prolonged drought or extreme heat. However, the very elasticity of seasonal use suggests that it is the component of water use with the greatest capacity for change. Outdoor watering restrictions are increasingly used to suppress seasonal use where drought has induced water scarcity (Kenny et al. 2004). Renwick and Green (2000) found outdoor use restrictions to be significantly more effective in reducing seasonal demand than voluntary measures. Water use restrictions and rationing have become routine in some desert cities (Hanak and Davis 2006) and are not without precedent in more humid climates, particularly where demand has outstripped the capacity of aging supply infrastructure (Lyon et al. 2005; Hill and Polsky 2007).

2.1.3. Drivers of single-family residential (SFR) water use

Seasonality is particularly pronounced in the SFR water use sector. Polebitski and Palmer (2010) found that a 10% increase in mean maximum temperature led to a 10% increase in SFR water in Seattle, Washington, while Balling and Gober (2007) estimated a 6% increase in Phoenix, Arizona, for a 10% increase in mean maximum temperature. Linkages with rainfall are smaller but statistically significant (Balling and Cubuque 2009; Polebitski and Palmer 2010). Spatially explicit studies have identified covariance of SFR water use, particularly the weather sensitivity of water use, with land use and socio-economic characteristics. In Phoenix, Arizona, for example, SFR water use was found to be most drought-sensitive in census tracts with higher incomes and more swimming pools per capita (Balling et al. 2008). In Portland, Oregon, SFR water use has tended to be

lower and less weather-sensitive in older neighborhoods with higher building densities (Chang et al. 2010; House-Peters et al. 2010; Breyer et al. 2012). Because of the connection between peak demand and land use, Shandas and Parandvash (2010) found that changes in urban design to limit opportunities for outdoor water use would aggregate to substantially reduce citywide peak water use in Portland. Taken together, these studies have suggested a prominent role for residential landscapes in patterning SFR water use and stabilizing peak demand.

2.1.4. The 'nature' of residential landscapes

Outdoor water-consuming behaviors are embedded in residential landscapes, where the coupling between human and natural systems becomes especially pronounced. Residential landscapes tend to reflect shared attitudes towards nature, functioning simultaneously as sites of neighborhood solidarity and social coercion (Cook et al. 2011). Both individuated spaces and communal places, residential landscapes link neighbors through common norms, shared streetscapes, and interdependent property values. Hydrological processes of rainfall, infiltration, and ET play out in tandem in social functions of recreation, emulation, display, and the reproduction of everyday life.

Social hierarchies become articulated in residential landscapes through water use. Highlighting the cultural and perceptual aspects of water use, Askew and McGuirk (2004) argued that irrigation for green, lush gardens in a newly suburbanized area of New South Wales, Australia, served to display status, while also denoting conformity to neighborhood norms. Similarly, Domene et al. (2005) revealed that higher income households in the water-stressed, suburban peripheries of Barcelona, Spain, tended to

maintain extensive gardens with irrigated turfgrass. Turfgrass functioned as a positional good, in these frameworks, intended to denote status and to distinguish the wealthy from lower income households, whose gardens made use of drought-tolerant native vegetation.

Robbins (2007) developed a theory of ‘lawn people’ to explain why well-educated individuals who profess to have more concern for the environment also consume vast amounts of resources, including toxic lawn chemicals, to maintain green lawns. The yard appears as a site of coercion, in this framework, where lawn people anxiously reproduce a turfgrass monoculture by simultaneously responding to neighborhood expectations and the biophysical demand of the lawn itself. As they labor for multiple ecologies, lawn people raise questions about individual agency and who has domesticated whom in turfgrass-dominated landscapes. In some newer residential developments, norms of lawn maintenance have become so coercive that they are codified and legally enforced through homeowners associations (Turner and Ibes 2011). Collectively, these points underscore a need for a synthetic understanding of SFR households as they are embedded in their ecologies, both biophysical and social.

2.2. Nested hierarchies

Chowdhury et al. (2011) argued that the dense weave of social and ecological elements at work in residential landscapes requires a multiscale analysis. An act of outdoor water use (or the decision to abstain) reflects a range of contexts, from the daily weather to the specificity of the household, the property’s land cover, neighborhood landscaping norms, and broader municipal or bioregional factors. To the extent that these contextual aspects jointly shape water-consuming behaviors, SFR water use may take on

a nested or 'hierarchical' data structure. This section explains why multilevel regression would be appropriate for water use, as nested data. It addresses problems of correlation, scale, aggregation between individuals and groups. Spatial applications of multilevel models are presented and possible critiques are discussed.

2.2.1. Multilevel data structures

A wide variety of social data exhibit a nested or clustered structure. Clustered data cannot be regarded as statistically independent, so simple linear regression and other statistical techniques that require an independent random sample are not appropriate. Multilevel regression is designed to handle clustered data. Also called hierarchical linear regression, the technique respects the grouped structure of the data, allowing for dependence among observations in the same group and heterogeneity among groups. Application of multilevel models to nested data structures originated in education research, with students nested in classrooms and schools, and has expanded to a wide range of disciplines (Goldstein 2003). Panel data present a temporal form of nested hierarchy, where a sequence of observations is nested in a subject (Singer and Willett 2003).

2.2.2. Ecological and atomistic fallacies

In his influential paper laying out the ecological fallacy, Robinson (1950) demonstrated that results from group-level data cannot be ascribed to individuals because aggregation smooths over individual variation such that within-group and between-group variability are conflated. Alker's (1969) atomistic fallacy showed that applying properties of individuals to groups is equally invalid. Jones and Duncan (1995) argued that the

multilevel approach avoids both the ecological and atomistic fallacies, providing an opportunity for jointly modeling individuals and their neighborhood contexts without reducing one to the other. By decomposing variance across scales and examining cross-scale interactions, multilevel models harness the underlying mechanisms of the ecological and atomistic fallacies—differences in variance within and between clusters—to measure contextual effects (Jones 1991).

2.2.3. Geographic hierarchies

Spatial data are often clustered at multiple scales; water use is no exception. Treating these spatial clusters as geographical groups and scales as levels would seem to be an appealing way to account for place and could shed light on how drivers interact across scales. However, conceptual and methodological critiques have been raised against multilevel approaches to spatial data analysis. These critiques are discussed below and are addressed in the analysis that follows.

One of the premises behind multilevel modeling is that the underlying population parameters are variable, or nonstationary, in the parlance. For example, the effect of daily maximum air temperature on SFR water use could truly vary across space. However, Fotheringham and Brundson (1999) note that apparent spatial nonstationarity in coefficients could also reflect bias from omitted variables. Parameters would be stable across space in constant in a correctly specified model, in this view, raising questions about whether spatial nonstationarity arises from irreducible place-based differences or human cognitive limits in the face of complexity. Fotheringham and Brundson (1999) also note that some multilevel models have generated spatially clustered residuals,

particularly at coarser scales of the data structure, leading to biased results. In a nested structure, dependence is captured by unambiguous membership to a neighborhood group. No dependencies are assumed between neighborhoods. This assumption may be unwarranted in a geographic context because it neglects spatial relations of adjacency, proximity, and continuity. For spatial processes, the boundaries of group membership may be fuzzy or even non-existent. Failing to account for this variance may lead to spatial clustering of model residuals.

Several studies have compared multilevel models with geographically weighted regression (GWR) to investigate whether correlation decays continuously across space or operates in discrete packets in unambiguously delineated places. Chaix et al. (2005) found GWR superior to multilevel regression, in that the latter accounted for less variance and generated spatially clustered residuals. They concluded that assigning individuals to unambiguous and unrelated groups implied an unwarranted fragmentation of space that was inadequate to capturing spatial processes, which, in their case, more realistically operated along a continuum. However, López-Carr et al. (2012) suggested that the multilevel place-based approach complimented the space-based, distance-decay approach of GWR, particularly with respect to policy-related processes that operate in discrete geographic units. Spatial implementations of multilevel models have attempted to accommodate correlation among coarse-scale spatial units by including spatial lags (Caughy et al. 2007) or a local index of spatial autocorrelation (Chen et al. 2010) as explanatory variables.

2.2.4. Problems of neighborhood context

Even if spatial correlation structures could be plausibly captured with discrete neighborhood units, the delineation of those boundaries opens up another forum for uncertainty. López-Carr et al. (2012) showed that spatial applications of multilevel models are vulnerable to a set of issues collectively referred to as the modifiable areal unit problem (MAUP). The MAUP implies that spatial correlations may vary by both scale and unit of aggregation such that alternative configurations of neighborhoods boundary could generate different results (Openshaw 1984). Neighborhood geometry is often represented as a set of contiguous, mutually exclusive spatial units that tessellate the study area, mostly for the convenience of the researcher—census geographies are a prime example. With spatiotemporal data, the uncertainty of bias increases exponentially with the number of aggregative and scalar possibilities along both spatial and temporal dimensions (Cressie 1998).

Not only do results depend on configuration of neighborhood boundaries, but there is uncertainty in how much the selected zonation coincides with the true contextually relevant geographic area (Kwan 2012). The area exerting the contextual influence may be spatially disjoint, may shift over time, have fuzzy boundaries, or otherwise vary from convenient, historically stable, and unambiguous enumeration units. These considerations have immediate implications for detection and interpretation of neighborhood effects.

2.2.5. Neighborhood effects

With the spatial turn in academic literature, a variegated literature on neighborhood effects has emerged to explain a variety of situations whereby neighborhood context influences individual behaviors (Durlauf 2004). Seeking to isolate the effect of neighborhood context, researchers (many operating completely outside the discipline of geography) have had to grapple with issues of spatial dependence and heterogeneity, as well as the MAUP and cognate problems of geographic correlation. Some researchers have eschewed zonal neighborhoods altogether in favor of bespoke neighborhoods, where a unique neighborhood geography is delineated for each research subject (Bolster et al. 2006, Johnston et al. 2004). Others have constructed flexible geographies (variable buffer sizes) to examine changes in contextual effects across scales (Caughy et al. 2007).

In addition to grappling with MAUP, the specific causal mechanisms undergirding the observed pattern of neighborhood effects remain in question. Johnston et al. (2005) enumerated possible compositional and contextual processes that link individual behaviors and neighborhood characteristics, including self-selection, emulation of neighbors, active peer-to-peer persuasion, and lobbying for new rules to govern peer behavior. Manski (1993) offered a typology of neighborhood effects. Correlated effects occur when a prior sorting process leads to self-similarity across individuals in the same neighborhood. Exogenous effects occur when neighborhood-scale factors structure but are unrelated to finer-scale processes. Finally, endogenous effects occur when individual behaviors are interdependent. An observer would not be able to

distinguish between the latter two pathways, creating a fundamental analytical difficulty, which Manski termed the reflection problem. The causal pathway linking individual behavior and neighborhood context remains an active area of research.

Critiquing multilevel approaches to neighborhood effects, Entwistle (2007) argued the term 'hierarchical' biases the direction of multilevel research by privileging the structure side of what is actually a dialectic between structure and agency. By locating the structure, the group or neighborhood, at the 'top' of the hierarchy, the modeling approach seems to imply that causation is coming from that scale, when correlated effects, endogeneity, or some other process of inter-causation may be at work. Interestingly, Subramanian et al. (2009) pointed out that the reverse could, paradoxically, also be true. Robinson's (1950) idea of an ecological fallacy, they posited, had an ecology of his own, one of methodological individualism that reflected Cold War anxiety around the Soviet Union and prioritized individual rationality over the group-level dynamics. By applying a multilevel model to Robinson's original data, the authors provided evidence for contextual effects that were overlooked in the original analysis.

Neighborhood effects have been observed in residential water usage patterns (Aitken et al. 1991; Ramachandran and Johnson 2011) and a multilevel approach has been found to explain more variation in SFR water use than single-scale analysis (Polebitski and Palmer 2010). The recursive connections between context and behavior involved in residential landscapes, noted by Robbins (2007) and Askew and McGuirk (2004), suggest a neighborhood effect operating in water use through the social dimensions of residential landscapes. Similar effects have been noted in electricity

consumption. Households, particularly those who did not respond to appeals to altruism or cost-savings, lowered their electricity use when utility bills were modified to show household consumption to be higher than the neighborhood average, and marked the disparity with a sad face (Allcott 2011).

2.3. The study area: Portland, Oregon

This section presents the hydrological and social context of residential water use in Portland, Oregon. It describes regional water supply infrastructure and current trends in water usage for the two water providers in this study. The section concludes by discussing implications of climate and land use change on water use in the region.

2.3.1. Water supply infrastructure

Summer water supply in the Portland metropolitan area relies heavily on winter precipitation stored in the Bull Run watershed (Figure 2.1). Located in the Cascade Range, this surface water supply system is managed by Portland Water Bureau (PWB), the largest water supplier in the State of Oregon. The system comprises two open-air reservoirs and a lake with a combined usable storage capacity of 14.2 billion gallons (BG). Discharge from Bull Run watershed far exceeds this volume, but limited storage capacity means that the reservoirs must be recharged annually—multi-year storage is not currently possible. The drawdown period for Bull Run reservoirs begins when outflows exceed inflows, usually in June. Drawdown typically ends with the onset of fall rains in mid-October, although it occasionally persists through December in drier years. In addition to providing water for human consumption, streamflow from the Bull Run

watershed must be managed for endangered species, particularly to cool streams for returning salmonids in autumn (Portland Water Bureau, 2010).

PWB supplements Bull Run reservoirs with a smaller groundwater reserve, the Columbia South Shore Well Fields, with an initial 30-day pumping capacity of 102 MGD. A series of smaller wells were added to the system when PWB annexed the region previously served by Powell Valley Water District in 2005. In addition to serving a retail population of over 500,000, PWB sells water wholesale to 22 smaller water providers, serving a combined retail and wholesale population of 770,250 in fiscal year 2004-2005. Currently, PWB serves a total population of over 900,000.



Figure 2.1: Urban water supply infrastructure. The bulk of the Portland metropolitan region's water supply originates from Bull Run Reservoir #1 (left), located in the Cascade Range and operated by Portland Water Bureau (PWB). PWB provides water for the City of Portland and also sells water wholesale to smaller water providers; Tualatin Valley Water District (TVWD) is their largest wholesale customer. As a member of the Joint Water Commission, TVWD also withdraws water from Hagg Lake (right), located in the Coast Range. Images sources: http://upload.wikimedia.org/wikipedia/commons/3/30/Bull_Run_Reservoir_1.jpg (left) and http://www.usbr.gov/projects/Project.jsp?proj_Name=Tualatin%20Project (right).

The state's second largest water provider, Tualatin Valley Water District (TVWD) serves a suburban and largely unincorporated area in Washington County, west of City of Portland, providing water to a population of nearly 200,000. Wholesale

purchases from PWB are augmented by water obtained through the Joint Water Commission (JWC), which operates the reservoir at Hagg Lake in the Upper Tualatin River and the Barney Reservoir in the Coast Range (Figure 2.1). A modest aquifer storage and recovery system has been built to store winter precipitation for additional supplies during peak season, since daily purchases are limited to 42.3 MGD from PWB and 10 MGD from JWC (Tualatin Valley Water District 2005). TVWD serves a rapidly growing area and expects to add 82,000 residents by 2042. Demand projections have indicated that peak demand will exceed current system capacity by 2025. TVWD ultimately plans to draw water from the Willamette River to meet the additional demand (Tualatin Valley Water District 2013).

2.3.2. Characterizing regional water use

According to their most recent Water Management and Conservation Plan (Portland Water Bureau 2010), total annual water consumption from PWB-managed water supplies was 32.9 BG in fiscal year 2004-2005. Average flow increased from 83 million gallons per day (MGD) in winter to 123 MGD in summer, with a peak flow of 187 MGD. Considering only the retail service area, annual consumption was 20.1 BG for a population of 494,200. Of that, 8.2 BG (41%) was associated with small-meter residential meters (presumably, most of these are SFR households), leading to daily per-household consumption of 163 GPD, comparable to Seattle's per-household consumption of 169 GPD (Rockaway et al. 2011). Average per-capita water consumption was 66 GPD in the denser retail service area, compared to 127 GPD in the broader, more suburbanized wholesale service area. Retail per-capita water use has declined an average of 19 GPD

compared to the 1987-1992 period. PWB attributes this to a combination of factors, principally plumbing code changes, which mandated water-efficient appliances and were effective after 1992. For TVWD, the residential sector comprises 70% of annual usage. The provider distributed an average of 23 MGD in 2005, with peak flow of 53 MGD (Tualatin Valley Water District 2005). In 2012, TVWD purchased 70% of its water through a wholesale contract with PWB (Tualatin Valley Water District 2012). Both PWB and TVWD are members of the Regional Supply Consortium, which pools resources for a range of passive and active water conservation programs and allows for coordination across water providers across the Portland metropolitan area.

2.3.3. Climate change

Downscaled climate change and emissions scenarios have projected exaggerated seasonality of future precipitation patterns in the Pacific Northwest, likely trending towards wetter winters, drier summers, and uncertain seasonal transitions (Mote and Salathé 2010; Hamlet 2011). The effect of these anticipated climactic shifts on the timing and magnitude of reservoir inflows is also uncertain (Palmer and Hahn 2002). Because surface water storage depends on antecedent climate conditions, the system may be vulnerable to drought, although actual shortages have only occurred in 1992, when groundwater was unavailable to make up for shortfalls in Bull Run supply. Water yield models under climate change scenarios suggest that the drawdown period will begin earlier and end later both in the Bull Run and at Hagg Lake (Palmer and Hahn 2002; Palmer et al. 2004), although no shortfalls are currently projected under climate change scenarios (personal communication, Lorna Stickel, Portland Water Bureau, May 2013).

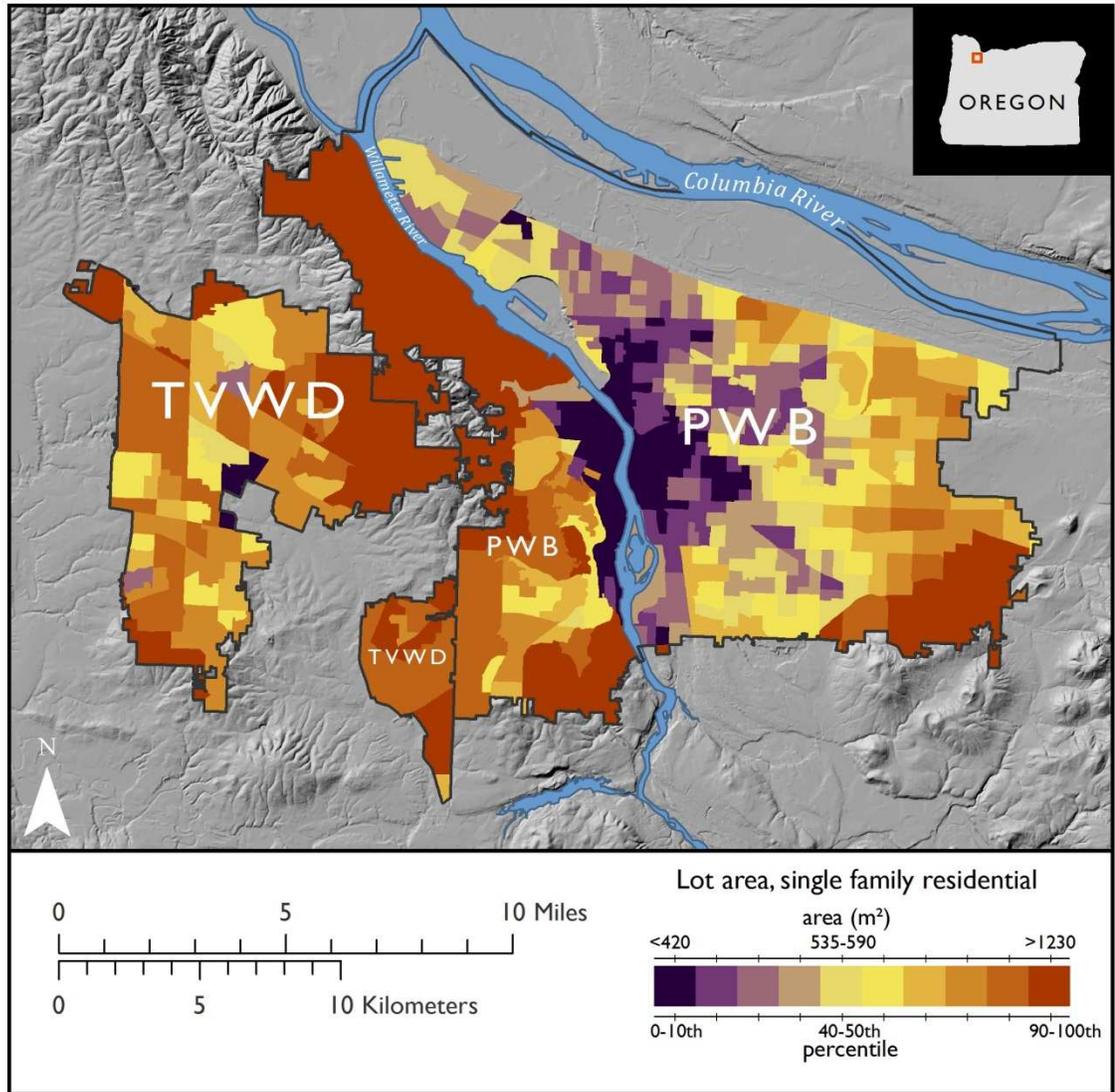


Figure 2.2: Urban density in the study area. The study area was defined by the service areas for Portland Water Bureau (PWB) and Tualatin Valley Water District (TVWD), the largest two water providers in the State of Oregon. To visualize the density along the urban-to-suburban gradient, mean SFR lot area was aggregated at the block-group level (excluding primarily industrial lands). Lot sizes are generally smallest and density is highest near the city center (shown in purple). Lot sizes increase and density decreases with distance from the city center (shown in orange).

2.3.4. Land use change

Two countervailing tendencies are at work with respect to seasonal water use in Portland. On the one hand, increased summer temperatures linked to climate change are

expected to drive up seasonal use. On the other hand, continued urban densification could mitigate the linkage between weather and water use (Shandas and Parandvash, 2010). Since 1973, suburban development in the Portland area has been contained by an urban growth boundary (UGB), which has encouraged an increasingly compact urban form with smaller lots, higher building densities, and thus fewer opportunities for outdoor watering. However, large-lot developments have also proliferated along the suburban peripheries of the UGB, leading to a pronounced urban density gradient (Figure 2.2). Older neighborhoods with smaller average lot sizes tend to be clustered near the city center and were associated with lower average water use (Chang et al. 2010) and less pronounced seasonality (Breyer et al. 2012). In short, the key driver of outdoor water use, temperature, can be expected to increase water use, but the temperature sensitivity of water use may vary spatially due to concurrent densification of the urban core and suburbanization of the periphery. The sum total of these offsetting trends is currently unknown.

Chapter 3: Theory

The previous chapter has reviewed evidence to suggest that SFR water use arises from a wide range of possible drivers arranged in a nested hierarchy of scales. This chapter takes the nested data structure as a point of departure to construct the theoretical scaffolding for the multilevel analysis that follows. Section 3.1 defines key conceptual aspects to multilevel models. Section 3.2 explains the utility of multilevel regression for exploring scalar aspects to SFR water use as a nested data structure. Section 3.3 develops equations for a three-level random slopes model with a cross-scale interaction. Section 3.4 discusses potential limitations related to assessing goodness of fit and statistical significance in multilevel models.

3.1. Mixed effects

A mixed-effects model can describe any regression model containing both fixed and random effects. Fixed effects refer to a single parameter estimated for all cases, as with pooled ordinary least squares (OLS) regression. Random effects imply estimation of a separate parameter for each group of cases, where those parameters are drawn from a probability distribution common to all groups (Kreft and de Leeuw 1998). The method tailors a regression model to each group of cases while stitching those group-specific regressions together through the shared distribution. In this way, mixed effects models are akin to partially pooled linear regression, in contrast to nonpooled OLS, where regression models for each group are unrelated (Gelman and Hill 2007). Not only does the shared distribution account for the broader context shared by all groups, but its mean

and variance can shed light on nuanced relationships in the data that would otherwise be missed (or, worse, misinterpreted) in OLS regression.

A mixed effects model is appropriate where the data are inherently a grouped or clustered. Clustered data cannot be treated as an independent random sample because of higher within-group dependence and between-group variance. By jointly fitting a set of related regression models to the sample of individuals and the corresponding sample of groups, mixed effects approach not only accounts for the structure of within-group correlation and between-group variance but actually places that structure at the center of the analysis. Mixed effects models have been used to honor a range of clustered data structures in both social and physical sciences, with applications ranging from education (Goldstein 2003) to two-stage sampling in ecological research (Zuur et al. 2009) and epidemiology (Gelman and Hill 2007), as well as a variety of repeated-measure data (Singer and Willett 2003).

3.2. Conceptualizing a multilevel approach to residential water use

A multilevel model is a specific type of mixed effects model developed for data arranged in a nested hierarchy. The central hypothesis for this thesis is that repeated-measure household water use follows a nested data structure that includes the neighborhood scale. If the data support this hypothesis, a time series of water use observations can be thought of as being grouped or nested within a household. Each household water use pattern, in turn, can be thought of as nested within a neighborhood, following Tobler's First Law of geography. As a consequence, there are multiple dimensions of statistical dependence in the data, across time (correlation among

sequential water use observations for a single household), across space (correlation among households in the same neighborhood), and across scales (Figure 3.1). Estimated variance in the distribution of random effects can be used to assess the relative importance of spatial and temporal drivers; this technique is known as variance decomposition. Neighborhood effects will be assessed by examining the component of this variance associated with the neighborhood scale. A multilevel approach can also be used to detect contextual effects or interactions between drivers at different scales. Cross-scale interactions involving weather and land use variables will be investigated below.

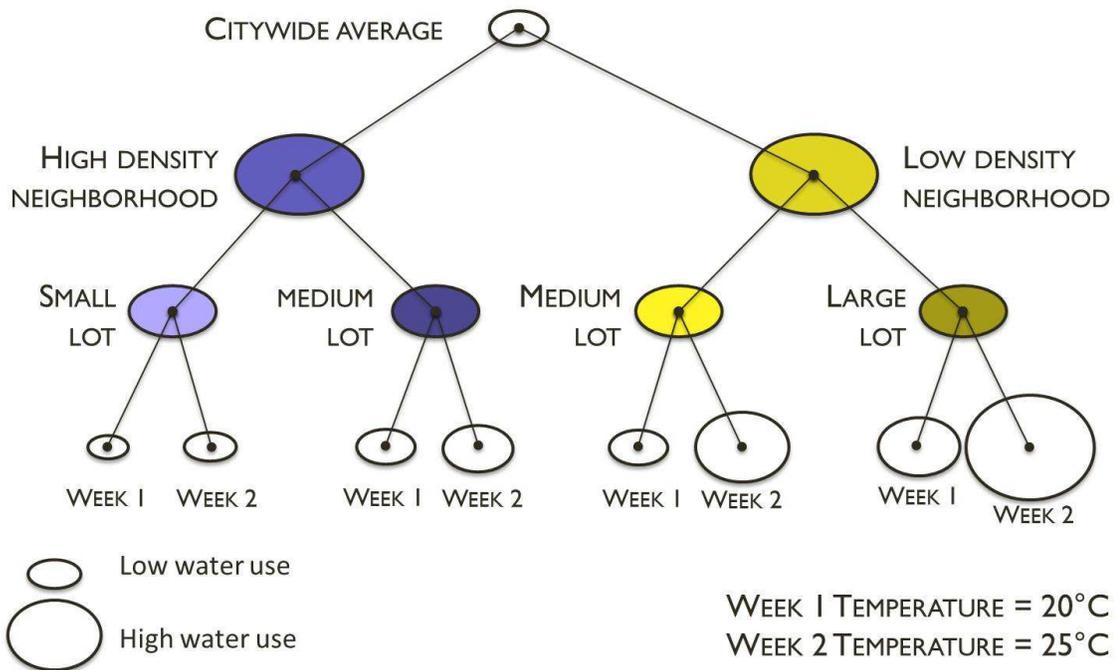


Figure 3.1: Conceptual diagram of household water use as a nested data structure. Weekly water use observations are nested in households, which are nested in neighborhoods. The structure is marked by high within-group correlation, as indicated by similarity in color—lots tend to be smaller in the denser neighborhood—and high between-group variance—heterogeneity in neighborhood density. Based on previous research, one would expect the lowest level of water use (shown by size of oval on Week 1 and Week 2) and the least temperature response (change in oval size between Week 1 and Week 2) to be associated with the small lot in the high density neighborhood.

3.3. *Developing a multilevel model of residential water use*

This analysis uses a panel dataset of SFR water use to investigate the extent to which land use contextualizes the relationship between weather and water use at the household and neighborhood scales. To underscore key methodological concepts, the following section steps through the development of a multilevel model of SFR water use. First, equations are presented for an unconditional means model, followed by a more complex, three-level model with random slopes and a cross-level interaction.

3.3.1. *An unconditional means model*

The first step to multilevel modeling is construction of an unconditional means model, where a random effects structure is specified but no predictors are included. Equivalent to a random effects analysis of variance, the unconditional means (or ‘empty’) model is useful for assessing the prominence of the grouping structure. An intercept is estimated for each group as the weighted mean of all data points falling within that group. Variance in this intercept serves to indicate the strength of the grouping structure—if groups are dissimilar, their intercepts are, too. The relative importance of scale can be assessed by partitioning total variance in water use into components that relate back to each scale in the hierarchy. To illustrate, an empty three-level random intercept model is developed below.

Over a given time period i , an individual household’s water use is given in Equation 1.

$$Y_{ijk} = \pi_{0jk} + \varepsilon_{ijk} \quad (1)$$

Equation 1 states that mean water use on time period i for household j located in neighborhood k is a function of two components: the mean water use for that household π_{0jk} and a random error term ε_{ijk} , which expresses the deviation of that specific week from the household mean. ε is an independent and random error term that adheres to the usual assumptions of linear regression ($\varepsilon \sim N(0, \sigma^2)$).

As a random coefficient, π_{0jk} appears fixed at the household scale but varies at the neighborhood scale because it arises from a distribution of coefficients shared by all households in a particular neighborhood. This relationship is given in Equation 2.

$$\pi_{0jk} = \beta_{0k} + R_{0jk} \quad (2)$$

In Equation 2, π_{0jk} is a function of the mean water use in neighborhood k , denoted as β_{0k} , and a random effect, R_{0jk} , reflecting the deviation of each household's mean water use and the neighborhood mean. The variance of R_{0jk} is given by τ_{00}^2 . Thus, the distribution of household coefficients in a particular neighborhood can be characterized in terms of the mean of β_{0k} and standard deviation of τ_{00} .

Also a random coefficient, β_{0k} appears fixed at the neighborhood scale but actually arises from a probability distribution shared by the set of neighborhoods. This relationship is stated in Equation 3.

$$\beta_{0k} = \gamma_{00} + U_{0k} \quad (3)$$

In Equation 3, γ_{00} is an intercept term that expresses the grand mean of water use across all neighborhoods. The neighborhood-specific deviation from this grand mean is given by random effect U_{0k} with variance denoted as φ_{00}^2 .

Equations 3 and 2 can be substituted into Equation 1 to yield the full expression for household water use:

$$Y_{ijk} = \gamma_{00} + U_{0k} + R_{0jk} + \varepsilon_{ijk} \quad (4)$$

Equation 4 states that an individual household's water use is a function of γ_{00} , a fixed intercept that represents the grand mean of water use, and a composite error term, $U_{0k} + R_{0jk} + \varepsilon_{ijk}$, that expresses how water use for a particular data point deviates from that grand mean. This deviation has been partitioned into three variance components: the component that can be attributed to the neighborhood (U_{0k}), to the household (R_{0jk}), and to the week (ε_{ijk}). Each component is assumed to follow a normal distribution centered on zero and with variance denoted as φ_{00}^2 , τ_{00}^2 , and σ^2 , respectively. The sum of these variance components is identical to the variance of the response variable. Total variance in household water use can be decomposed across scales based on that identity.

A key output of variance decomposition is intraclass correlation (ρ), which is calculated from variance components as shown in Equation 5:

$$\rho = \frac{\varphi_{00}^2 + \tau_{00}^2}{\varphi_{00}^2 + \tau_{00}^2 + \sigma^2} \quad (5)$$

Ranging from 0 to 1, ρ is a measure of within-group dependence. It reflects the portion of total variance that can be attributed to between-household and between-neighborhood factors. A larger value indicates that clustering by household and neighborhood is more prominent, suggesting that processes at these scales play a larger role in explaining variance in the response variable. Because there are three levels specified in this model, it is necessary to calculate a separate intraclass correlation ρ_N to capture the portion of ρ that can be attributable to between-neighborhood factors, as shown in Equation 6:

$$\rho_N = \frac{\Phi_{00}^2}{\Phi_{00}^2 + \tau_{00}^2} \quad (6)$$

Both ρ and ρ_N can be used to calculate design effects, which are important for assessing whether a multilevel approach is more efficient than OLS (Snijders 2005). Design effects provide a measure of redundancy in the data caused by within-group dependence. Simulation studies have found that data with design effects > 2 warrant multilevel analysis (Muthen and Satorra 1995). The formula for design effects is given in Equation 7.

$$\text{Design effect} = 1 + \rho(\theta - 1) \quad (7)$$

Here, θ denotes average cluster size, in this case, the number of observations per household and the number of households per neighborhood. Effective sample size is obtained by dividing the design effect by the sample size, and can be used to assess the statistical power of the analysis, since the effective sample size accounts for redundancy arising from dependence in the data. Statistical power increases with the effective sample size, which is particularly important at the ‘top’ of the hierarchy (Snijders 2005).

3.3.2. A three-level random slopes model with cross-scale interactions

This section develops a random slopes model as an extension of the empty model. Explanatory variables are specified at level 1 (weather), level 2 (parcel characteristics), and level 3 (neighborhood characteristics). A cross-scale interaction between weather and parcel characteristics is also specified. Model outputs can be used to assess the relative importance of drivers at each scale as well as the magnitude of any interactions across scales. Dependencies among observations are captured by a variance-covariance matrix that results from the structure of random effects. However, variance decomposition

becomes more complex because the variance of random slopes may be estimated in different units than the random intercepts.

On a given day, an individual household's water use is given in Equation 8, which extends Equation 1 to include weather variation.

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}W_i + \varepsilon_{ijk} \quad (8)$$

Equation 8 states that water use on a given day i for household j in neighborhood k is a function of three components. First, the intercept π_{0jk} reflects average household water use, the amount of water used on a given summer day regardless of weather variation. The value is constant over time for a given household but variable across households and neighborhoods. Second, the change in water use associated with a change in weather conditions is given by the slope term π_{1jk} , which modifies a time series of weather conditions W_i . Weather varies over time but, for a given time increment, is assumed to be constant over space, so lacks subscripts for spatial units j or k . Finally, a random error component is given by ε_{ijk} ($\varepsilon \sim N(0, \sigma^2)$).

Once again, π_{0jk} and π_{1jk} appear fixed at the household scale but actually arise from a neighborhood-specific distribution. Equations 9 and 10 expand on Equation 2 by specifying household slopes and intercepts is a function of parcel characteristics H_{jk} as well as the neighborhood mean and the random effect, as expressed in Equations 9 and 10.

$$\pi_{0jk} = \beta_{00k} + \beta_{01}H_{jk} + R_{0jk} \quad (9)$$

$$\pi_{1jk} = \beta_{10} + \beta_{11}H_{jk} + R_{1jk} \quad (10)$$

The term β_{00k} represents neighborhood mean water use—the subscript k indicates that this term varies across neighborhoods. The terms β_{01} , β_{10} , and β_{11} are purely fixed effects,

meaning that they do not vary across space or time, hence the absence of subscripts i , j , or k . β_{01} and β_{11} capture the effect of parcel characteristics on household intercepts and slopes, respectively, while β_{10} represents the baseline level of weather sensitivity for all households. One could also test for whether β_{01} , β_{10} , and β_{11} vary across neighborhoods, but this possibility is omitted here for simplicity. Random effects are, once again, given by terms R_{0jk} and R_{1jk} .

The neighborhood mean intercept β_{00k} arises from a probability distribution shared by all neighborhoods, as stated in Equation 11.

$$\beta_{00k} = \gamma_{00} + \gamma_{01N}N_k + U_{0k} \quad (11)$$

Here, β_{00k} varies across neighborhoods as a function of three components: the mean intercept γ_{00} , the effect of neighborhood characteristics γ_{01N} , which modifies a vector of neighborhood characteristics N_k , and the random effect U_{0k} . The variance of U_{0k} indicates the magnitude of contextual influence of neighborhood characteristics on neighborhood average water use.

To express the full multilevel equation for household water use, the fixed effects are first defined in Equations 12-14.

$$\beta_{10} = \gamma_{10} \quad (12)$$

$$\beta_{01} = \gamma_{01H} \quad (13)$$

$$\beta_{11} = \gamma_{11} \quad (14)$$

These equations simply restate that the effect of characteristics on household coefficients (β_{01} and β_{11}) and the baseline level of weather sensitivity (β_{10}) are truly fixed effects.

Equations 12-14 can be substituted into Equations 9 and 10, which can be substituted into

Equation 8 and re-arranged to obtain the long-form model for Y_{ijk} expressed in Equation 15.

$$Y_{ijk} = \gamma_{00} + \gamma_{10}W_i + \gamma_{10H}H_{jk} + \gamma_{10N}N_k + \gamma_{11}W_iH_{jk} + U_{0k} + R_{0jk} + R_{1jk}W_i + \varepsilon_{ijk} \quad (15)$$

Equation 15 states that weekly household water use arises from a complex mix of biophysical and social drivers. At level 1, the fixed effect of weather W_i is given by γ_{10} . At level 2, a fixed effect of parcel characteristics H_{jk} is given by γ_{10H} . At level 3, a fixed effect of neighborhood characteristics N_k is given by γ_{10N} . The magnitude of the cross-scale interaction between weather and parcel characteristics, given by γ_{11} , is itself a fixed effect. However, the overall effect on water use also depends on the values of W_i and H_{jk} , as well as the random term $R_{1jk}W_i$. The quantity $U_{0k} + R_{0jk} + R_{1jk}W_i + \varepsilon_{ijk}$ is a compound error term with neighborhood (U_{0k}), household ($R_{0jk} + R_{1jk}W_i$), and residual (ε_{ijk}) components. As in the unconditional means model, variance in R_{0jk} and U_{0k} is expressed as τ^2_{00} and φ^2_{00} , respectively. The estimated variance of household slopes π_{1jk} is given by τ^2_{10} . Slope-intercept correlation is given by τ_{01} . Note that intraclass correlation cannot be calculated for a random slopes model because the variance components are expressed in different units (Kreft & De Leeuw 1998).

3.4. Diagnostics

Although the relative magnitudes of variance components are useful for assessing the importance of drivers at each scale, obtaining multiple estimates of variance makes it difficult to evaluate the statistical significance of the estimate. For the same reasons, comparing overall goodness of fit across models is challenging. Approaches to these methodological problems are discussed below. Residual diagnostics at each level should

also be examined for normality and independence, including spatial independence (Chaix et al. 2005).

3.4.1. Statistical significance

Statistical significance is straightforward in OLS regression, but less so in multilevel regression (Baayen et al. 2008). There is not consensus about how to calculate p-values for fixed and random effects (Bates 2006). Some software applications—SPSS and HLM, for example—produce p-values as outputs for fixed and random effects, while others do not, namely lme4, a mixed-effects regression package available through R Statistical Software (Bates 2006). Parametric bootstrapping procedures and likelihood profiling have been proposed as alternative means to determine statistical significance of estimated coefficients (Bates 2010).

3.4.2. Goodness of fit

With OLS regression, model fit can be assessed with R^2 , a simple dimensionless measure that is readily calculated as the proportion of variance explained by model terms. Because mixed-effect models have separate variance estimates for each random effect and the residual variance, there are multiple possible ways to calculate a commensurate metric. A range of possible pseudo- R^2 calculations have been proposed to indicate overall model fit (Roberts et al. 2011), but each can be problematic. Those that assess model fit in terms of reduction in residual variance can obscure variance explained by higher-order random effects. In others, it is possible to obtain negative pseudo- R^2 values for more complex models. Metrics for assessing overall model fit remains an active area of research for mixed-effects models. Nakagawa and Schielzeth (2013) have proposed a pair

of measures, the marginal and conditional R^2 , to quantify overall fit. Respectively, they refer to the proportion of variance explained by fixed effects and by both fixed and random effects. Marginal and conditional R^2 will be reported—with caution—in the analysis that follows.

Since multilevel models use maximum likelihood estimation, a likelihood ratio test can be used to select the random effects structure that is best fitting (or, more accurately, most likely in relation to) the grouping structure in the data (Baayen et al. 2008). Model fit can also be compared using log-likelihood, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), or similar metrics obtained from likelihood-based estimation techniques—but, again, only with caution. These outputs provide no insight into overall model fit and are meaningless for comparing models that use different data. Their utility lies in examining relative fit between alternative model specifications based on the same dataset. AIC and BIC have the additional quality of penalizing model complexity, in contrast to the tendency of R^2 to increase with the number of covariates. However, any assessment of relative fit must be attentive to whether maximum likelihood (ML) or restricted maximum likelihood (REML) estimation is used. The former is preferred for assessing models with different fixed effects while the latter is preferred for assessing models with different random effects because it accounts for the number of parameters estimated (Bates 2010). AIC and deviance will be used to compare relative fit across models in the analysis that follows.

Chapter 4: Data and Methods

A multilevel model was developed to investigate the roles of weather and land use in driving variation in SFR water use. This chapter details data sources (Section 4.1), methods of model development (Section 4.2), and diagnostic procedures (Section 4.3) used in this implementation. Data were drawn from panel data on SFR water consumption distributed along an urban-to-suburban gradient in the Portland metropolitan area. Candidate explanatory variables included weather, building structure, and land cover variables for both sample taxlots and across neighborhoods. The model development process began by selecting an initial random intercepts structure through a series of unconditional means models. Random slopes, cross-scale interactions, and fixed effects were then fit to the data using iterative log-likelihood ratio tests. Bootstrapping procedures and likelihood profiling, respectively, were used to assess statistical significance and construct confidence intervals for model parameters.

4.1. Data

4.1.1. Response variable: Single-family residential (SFR) water use

The dependent variable, SFR water use, represents a time series of weekly mean water use observations in summer (June-August) over 2001-2005 for 460 households served by two water providers: PWB (n = 321) and TVWD (n = 139). In this case, a ‘household’ is defined by the water meter associated with a discrete billing unit at a unique address spatially referenced to a SFR tax parcel. In the analysis that follows, ‘household’ and ‘parcel’ refer to the same spatial unit. The term ‘household’ connotes behavior, while the term ‘parcel’ refers to aspects of building structure and land cover.

The number of residents per billing unit was unknown in this study. The service areas for these utilities span an urban gradient, ranging from established neighborhoods with relatively high building density near the urban core to newly constructed, lower-density subdivisions along the suburban periphery.

Selection of sample households was performed by PWB staff member Dr. Hossein Parandvash. Once household consent was obtained, each water meter was equipped with a flow monitoring device to transmit a time series of meter readings to PWB, although instrumentation for households in the TVWD service area differed from that used for PWB households. The resulting raw data were manipulated to show quantity of flow per day. There was some device error, particularly for TVWD households (more flow recorded than is possible to move through the pipe in that 24 hour period), with some very high values in a few cases; these values were removed by request of PWB staff before we obtained the data. Because the research question was aimed at weather sensitivity, data were limited to summer months (June-August), as the typical SFR water use pattern are higher and more variable in summer, with an average of 1271 liters per day (LPD) compared to 575 LPD in winter (Figure 4.1). However, sample households exhibited wide variation around the average, with some households consuming over 2000 LPD in summer while others maintained the winter average of around 600 LPD (Figure 4.2). Daily data were aggregated to weekly mean values because the signal connecting weather and water use is more pronounced at this scale (Akuoko-Asibey et al. 1990). Water use is substantially more variable at the daily scale, likely related to indoor water use drivers or weekend effects (Wong et al. 2010), and thus not related to either weather or land use.

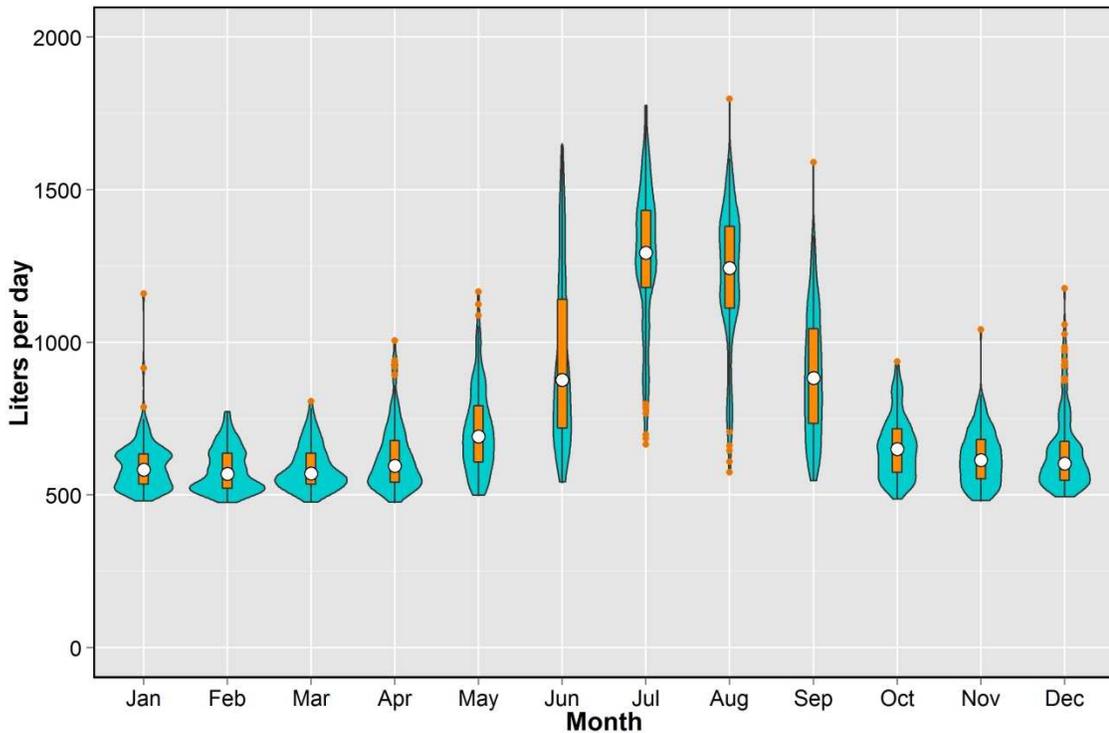


Figure 4.1: Seasonality of daily household water use by month, 2001-2005. For the typical Portland single-family residential household, median water use (white dot) increases from 575 liters per day in winter to 1271 liters per day in peak season (July-August). Boxplots (in orange) indicate that the variance in water use also becomes greater in summer. Violin plots (in blue) visualize the density of water use, underscoring the stability of winter use and the elasticity of summer use.

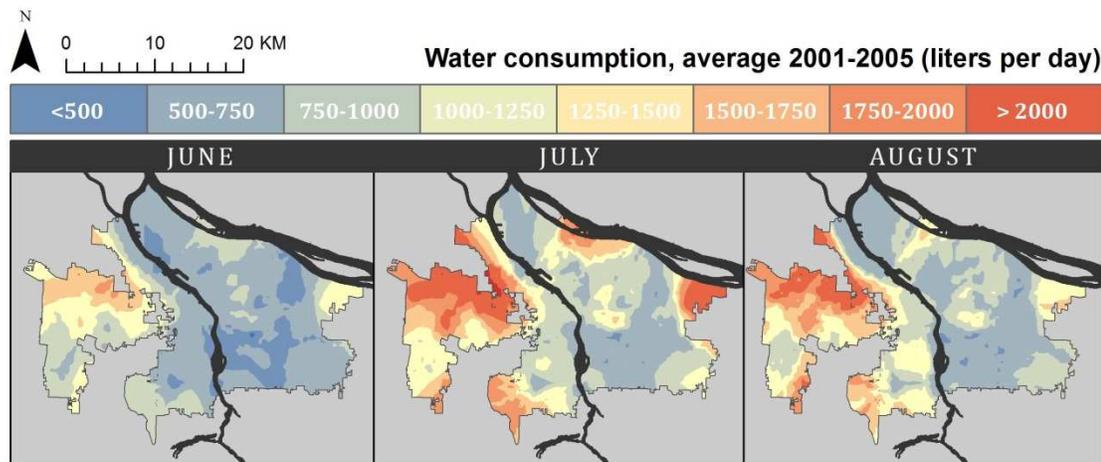


Figure 4.2: Spatial variation of daily household water use by month, 2001-2005. Single-family residential households exhibited wide variation around the average water use level of 1271 liters per day in summer. Consumption levels were particularly high for households in the northwesternmost part of the study area, in excess of 2000 liters per day, and were particularly low in the areas directly east of the Willamette River.

The source data contained observations for a larger sample of households ($n = 746$), but the data were highly imbalanced. Some households were monitored continuously from 1999-2008, while others were monitored for a month or less. Mixed-effects models can handle imbalanced data (Bates 2010), but a problematic spatial trend was evident in this missing-ness (Snijders and Bosker 2012). Nearly all TVWD households were missing substantial chunks of data due to issues with research design and data collection that preceded the present analysis. Thus, it was necessary to select a time period to temporally subset the data. After examining a set of six possible time periods, the years ranging from 2001 to 2005 were selected. This time period was associated with a fairly large ($n = 460$) and spatially unbiased distribution of households along an urban gradient, avoiding any spatial cluster to the missing data (Figure 4.4). This time period also has the advantage of including some interannual weather variation (hotter or cooler summers due to El Nino Southern Oscillation) while avoiding any marked decline in SFR water use. A ~6% decline was apparent for the households with most complete time series, 1999-2008.

4.1.2: Explanatory variables

Candidate explanatory variables were selected at three scales (Table 4.1): weather (level 1), SFR parcel (level 2), and neighborhood (level 3). These variables were selected because they have proven significant in single-level, fixed effect studies in the water demand literature (Balling et al. 2008; Praskievicz and Chang 2009; Chang et al. 2010; Breyer et al. 2012).

Table 4.1: Candidate explanatory variables. Data sources include Interactive Numeric and Spatial Information Data Engine (INSIDE) Idaho gridded weather data and Regional Land Information System (RLIS). At the neighborhood scale, three alternative boundaries were examined, census tracts (n = 125), RLIS neighborhood boundaries (n = 63), and regions (n = 10). For each, neighborhood-scale values were calculated from a spatial join to RLIS single-family residential taxlots.

Scale	Predictor	Source
Weather (level 1) n = 35,065	Minimum/maximum air temperature (°C)	INSIDE Idaho
	Minimum/maximum relative humidity (%)	INSIDE Idaho
	Precipitation (mm)	INSIDE Idaho
	Wind speed (km/hr)	INSIDE Idaho
	Five-day cumulative precipitation	Calculated value
	Extreme heat index	Calculated value
Households (level 2) n = 460	Lot area (m ²)	RLIS 2005
	Building area (m ²)	RLIS 2005
	Low vegetation (m ² , %)	RLIS 2007
	High vegetation (m ² , %)	RLIS 2007
	Impervious (m ² , %)	RLIS 2007
	Property value: land, building, total* (\$US)	RLIS 2005
Neighborhood (level 3)	Building age	RLIS 2005
	Mean SFR lot area (m ²)	Calculated value
	Mean SFR building area (m ²)	Calculated value
	Mean SFR property value (land, building, total)	Calculated value
	Mean SFR building age (years)	Calculated value
	Mean SFR low vegetation (m ² , %)	Calculated value
	Mean SFR high vegetation (m ² , %)	Calculated value
	Mean SFR impervious (m ² , %)	Calculated value

At level 1, daily air temperature (minimum and maximum, °C) and precipitation depth (mm) were derived from Interactive Numeric and Spatial Information Data Engine (INSIDE) Idaho gridded weather data, a spatially explicit (4 km² grid) and statistically downscaled dataset of daily weather variables generated from a regional climate model (Abatzoglou 2013). To capture spatial variation in water use over the study area, mean daily values for all grid cells with centroids falling inside a water provider service area

were calculated. The monthly distribution of temperature and precipitation in the study area is shown in Figure 4.3. Two additional indices were calculated to capture temporally lagged effects of heat or rain: the extreme heat index, defined by a count of the number of the previous five days with daily maximum air temperature $> 32^{\circ}\text{C}$, and the cumulative precipitation index, defined by the sum of precipitation depth for the previous five days. These values were then aggregated to their weekly means and subset to the summer months of 2001-2005 in order to correspond with weekly water use.

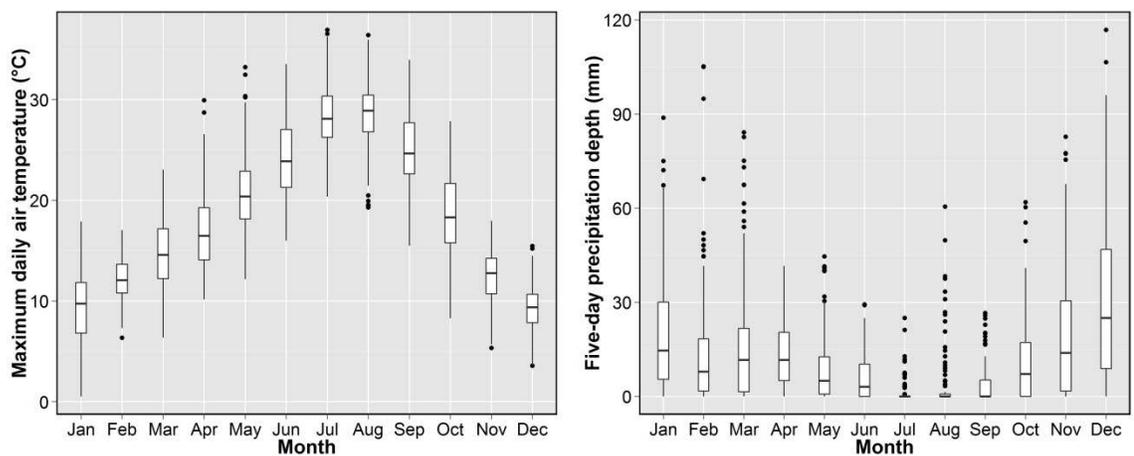


Figure 4.3: Distribution of temperature and precipitation patterns, 2001-2005. The climate in Portland, Oregon, is marked by distinct rainy and dry seasons. Temperature peaks during the dry summer months of July and August, when water demand reaches its peak.

At level 2, each SFR taxlot was attributed with the following building structural characteristics derived from a 2005 iteration the Regional Land Information System (RLIS): lot size (m^2), building size (m^2), building age (years built relative to 2012), and property value (land, building, and total value in 2005 \$US). Land cover data were derived from a 2007 land cover classification, provided by Metro Regional Government, with the following four classes: impervious surfaces, low vegetation, high vegetation, and water. Total pixel area (m^2) and % of total area assigned to each land cover type were

tabulated by taxlot in ArcGIS version 10.0. No water pixels were detected from the land cover data in sample taxlots.

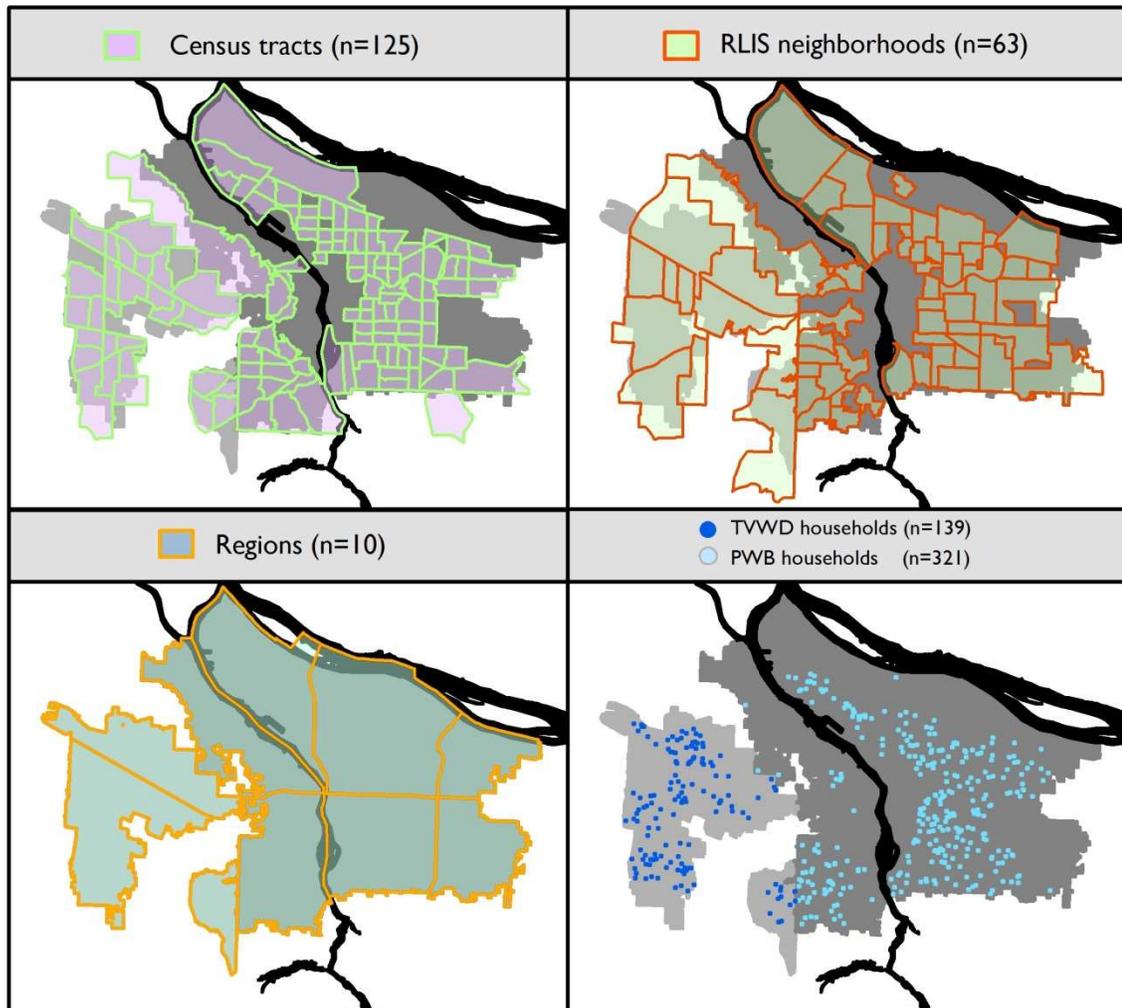


Figure 4.4: Spatial distribution of households and neighborhood boundaries. The analysis considered three candidate neighborhood structures: census tracts (top left), neighborhood association boundaries from the Regional Land Information System (RLIS, top right), and regions (bottom left). The spatial distribution of sample households (symbolized by water provider) is presented for comparison (bottom right). Each spatial dataset was overlaid on the study area (shown in grey).

At level 3, there were nearly infinite possibilities for zonation of neighborhood boundaries. The uncertain geographic context problem suggests that delineation of neighborhood boundaries may not coincide with the contextually relevant spatial area. In light of these concerns, three different neighborhoods were considered in this analysis:

census tracts (n = 125), neighborhood boundaries from the Regional Land Information System, or RLIS (n = 63), and broad regions defined for the purposes of this study (n = 10). The spatial configuration of these neighborhoods is shown in Figure 4.4. Mean values for all building structural and land cover variables for the population of SFR taxlots were obtained for each using ArcGIS 10.0.

Census tracts are an attractive zonal configuration because of the wealth of socioeconomic data available at that scale. By contrast, RLIS neighborhood boundaries have very little data, but may be more contextually relevant, as these boundaries are delineated locally, not federally, and are often used for land use planning or conducting municipal business. They also tend to coincide with the boundaries of neighborhood associations or community planning organizations. In addition, household location decisions may be influenced by these neighborhoods more than by census tracts, so RLIS neighborhoods capture neighborhood effects by self-selection. A total of 12 neighborhoods were merged so that at least three households fell within each RLIS neighborhood. The regions dataset was created for comparison with finer-scale neighborhoods by subsetting water provider study areas into the coarse-scale units using the Willamette River, Interstate 5, Interstate 205, and Highway 26.

4.1.3. Variable transformations

In this study, the response variable and all covariates were transformed to the natural log scale, then centered on their grand mean values. Log-transformation normalized variables and allowed estimated coefficients to be interpreted as elasticities, unitless measures with readily comparable effect sizes. Mean centering allows intercepts

to be interpreted as means, which is helpful for generating unbiased random effects. The decision for mean-centering weather and neighborhood-scale variables was straightforward because these data have only a grand mean. However, for variables at levels 2, it was possible to center on either a group mean or the grand mean for the sample. The grand mean was selected in all cases to highlight contextual effects (Enders and Tofghi 2005). After normalizing and centering, each variable represented the deviation from log-transformed global mean.

4.2. Model development

Models were fit in three stages using the `lmer` function for mixed-effects models from the `lme4` version 0.99999 (Bates, Maechler, Bolker and Walker 2013) deployed in R Statistical Software version 2.15.0 (R Core Team 2012). First, alternative unconditional means models were specified to compare plausible random intercept structures (Section 3.3.1) with iterative fitting algorithms available through `LMERConvenienceFunctions` package version 2.0 (Tremblay and Ransijn 2013). The selected random intercept structure was then used as input for an iterative process that fit random slopes and identify any higher-level predictors of lower-level random coefficients (Section 3.3.2). Variables with fixed effects were subsequently back-fit and cross-scale interactions were explored using the full random effects structure. Specific functions used at each step of model development are summarized in Table 4.2.

4.2.1. Initial random intercepts structure

To select an initial random intercepts structure, a series of unconditional means models were fit using REML estimation. This step determined whether intercepts varied

Table 4.2: Functions used for model development with R Statistical Software.

Model development step	Function	R package
Fit initial unconditional means model	<code>lmer</code>	<code>lme4</code> version 0.99999
Remove influential outliers	<code>romr.fnc</code>	<code>LMERConvenienceFunctions</code> version 2.0
Assess anticonservative p-values of cross-scale interactions	<code>pamer.fnc</code>	<code>LMERConvenienceFunctions</code> version 2.0
Forward-fit random effects	<code>ffRanefLMER.fnc</code>	<code>LMERConvenienceFunctions</code> version 2.0
Back-fit fixed effects	<code>bfFixefLMER_F.fnc</code>	<code>LMERConvenienceFunctions</code> version 2.0
Likelihood ratio testing	<code>anova</code>	<code>lme4</code> version 0.99999
Marginal and conditional R^2	<code>r.squaredGLMM</code>	<code>MuMIn</code> version 1.9.5
Parametric bootstrapping	<code>boot.mer</code>	<code>lme4</code> version 1.0.4
Multicollinearity	<code>vif.mer</code>	https://raw.githubusercontent.com/auf frank/R-hacks/master/mer-utils.R
Confidence intervals	<code>confint</code>	<code>lme4</code> version 1.0.4

by neighborhood and selected among alternative neighborhood boundaries. Water use observations were grouped by household, by neighborhood, and both households and neighborhoods. Three alternative neighborhood boundaries were considered: census tracts, neighborhood associations, and regions (Figure 4.4). Relative log-likelihood ratio tests were performed to compare pairs of models (significance level of 0.05). Three models with best fit (lowest deviance) were identified. Design effects and contextual relevance were examined to select a prevailing random effects structure from this subset.

Outliers were removed from the selected empty model using the `romr.fnc` function from `LMERConvenienceFunctions`, where an outlier was defined as any observation from the prevailing candidate model with standardized residuals greater than 2.5 standard deviations from the mean (Tremblay and Ransijn 2013). Note that `lme4` version 0.99999 was used to fit models input while `lme4` version 1.0.4 was used to bootstrap p-values and calculate confidence intervals.

4.2.2. Specifying random effects

Random slopes were forward-fit using the `ffRanefLMER.fnc` function from `LMERConvenienceFunctions`. This algorithm uses REML estimation to iteratively evaluate a list of possible random effects structures, applying a log-likelihood ratio test (significance level of 0.05) to compare relative fit with a baseline model. To avoid bias from omitted variables, the baseline model included all candidate explanatory variables (Zuur et al. 2009).

4.2.3. Back fitting fixed effects

Fixed effects were back-fit using the `bfFixefLMER_F.fnc` function from `LMERConvenienceFunctions`. The back-fitting process used by this script is outlined below. The input model contained all candidate explanatory variables (Table 4.1), along with all cross-scale interactions identified as significant using the anticonservative p-values generated through the `pamer.fnc` function from `LMERConvenienceFunctions` (Table 4.3). Using ML estimation, the following steps were iteratively performed. First, the least significant model parameter at the coarsest-scale was identified based on the t-statistic. The relative fit of models with and without this candidate variable were

compared using a log-likelihood ratio test (significance level of 0.05). If the test was significant, the variable was retained; otherwise, it was removed. The next iteration began by selecting the next-smallest t-statistic, in absolute value. Once all of the variables at the coarsest scale were tested, the process stepped down to the next-finest scale and repeated the same steps. However, in addition to checking model fit with a likelihood ratio test, these finer-scale variables were tested for cross-scale interactions with higher-level variables. To specify the most parsimonious model (and to avoid conflating nonstationarity with omitted variable bias), potential redundancies in the selected random effects structure were also examined at each iteration. If the variance could be accounted for with a fixed instead of a random effect, the latter was pruned. To most accurately express both the fixed and random effects, the final model was estimated in REML.

Table 4.3: Candidate cross-scale interactions. The variable pairs with interaction effects that were identified as significant using the `pamer.fnc` function from `LMERConvenienceFunctions` are marked with an ‘x’. Maximum air temperature was found to interact with all parcel-scale building structural variables as well as impervious land cover. Interactions between weather variables and lot size were most significant and most common. Each significant candidate interaction term was tested in back-fitting process.

Level 1 variable	Land use variables with significant cross-scale interaction				
	Lot size, parcel	Property value, parcel	Impervious area, parcel	Lot size, neighborhood	Building size, neighborhood
Maximum air temperature	x	x	x	x	x
Antecedent precipitation	x	-	-	-	-
Extreme heat index	x	-	-	-	-

4.3. Model diagnostics

Previous studies of water use have noted two key methodological problems with multilevel models: (1) a wide range of potentially collinear explanatory variables and (2) autocorrelated or nonstationary model error terms. Here, model diagnostics paid careful attention to possible multicollinearity between fixed effects. Residual diagnostics at each scale of analysis were undertaken, with particular attention paid to possible spatial correlation in ‘higher-order’ random effects, temporal correlation in ‘lower-order’ effects, or correlation between random effects and explanatory variables.

The lmer function does not generate p-values for fixed or random effects because of uncertainty in the degrees of freedom for the denominator of the F-test that assigns p-values (Bates 2006; Bates 2010). Bootstrapping procedures (1,000 replicates) were used to estimate p-values. Absolute and relative goodness of fit were compared for the unconditional means model, the models developed in Section 3.2, and the final model. Marginal and conditional R^2 measures were obtained to assess overall goodness of fit (Nakagawa and Schielzeth 2013). To conduct hypothesis tests about the fixed effects, a Wald approximation from the likelihood profile was used to calculate confidence intervals of fixed-effect parameters. For random effects, confidence intervals were identified from a likelihood profile (Bates 2010).

Chapter 5: Results

This chapter presents results of multilevel regression analysis of SFR water use. Section 5.1 reviews model outputs for candidate unconditional means models and provides justification for selecting the initial random intercepts structure. Section 5.2 covers the results from the model development process of forward-fitting random effects and back-fitting fixed effects. Section 5.3 examines the final model. Section 5.4 concludes by examining residual diagnostics and goodness of fit. It also compares the predictive power of a multilevel model to pooled OLS outputs.

5.1. Results of unconditional means model

Results from candidate unconditional means (or ‘empty’) models are reported in Table 5.1, with corresponding design effects in Table 5.2. Intraclass correlation (ρ) for each empty model was relatively large, suggesting that SFR water use exhibited a nested data structure. All empty models that included random intercepts by household (Models 1 and 5-7) had $\rho > 0.5$, indicating that roughly half of total variance in water use can be attributed to between-household or between-neighborhood variability. Between-week weather variability accounted for 47% of total water use variance in Models 1 and 5-7. Relatively less variation was attributed to between-neighborhood variability. Among three level models, ρ_N suggested that 22-30% of between-household variance was attributed to between-neighborhood factors (Table 5.2). Model fit (AIC, deviance) was best for three-level models (Models 5-7), with water use observations nested in households and households nested in neighborhoods. Log-likelihood testing confirmed that a model with both household and neighborhood structures were significantly more

likely, given the data ($p < 0.05$). Taken together, these findings suggest the possibility of neighborhood effects at work, although it is uncertain which (if any) of the three candidate neighborhood geographies adequately captured the area exerting the contextual effect (Kwan 2012).

Table 5.1: Results from unconditional means models. Models that grouped households into neighborhoods were a better fit to the data.

Model	Level 2	Level 3	AIC	deviance	Household variance	Neighborhood variance	Residual variance
1	Household	-	66415	66409	0.4206	-	0.3670
2	-	Tract	80227	80221	-	0.2701	0.5673
3	-	Region	86034	86028	-	0.0732	0.6952
4	-	RLIS	83275	83269	-	0.1487	0.6241
5	Household	Tract	66365	66357	0.2953	0.1244	0.3670
6	Household	Region	66339	66331	0.3365	0.1016	0.3670
7	Household	RLIS	66353	66345	0.3162	0.0883	0.3670

The neighborhoods with highest mean water use were located in the northwestern area of the TVWD service area, an area known as Wolf Creek (Figure 5.1). Households with lower water use areas were clustered in more established neighborhoods on the east side of the study area, although some areas of high water use were detected close to the city center at finer spatial scales—these areas coincide with more affluent close-in neighborhoods. Note that variance in intercepts, as well as the size of the confidence intervals, was highest for individual households and lowest for regions, consistent with MAUP. As more observations are grouped into a spatial unit, the variability in those observations was averaged away in the weighted average, illustrating the smoothing effects of aggregation at coarser scales.

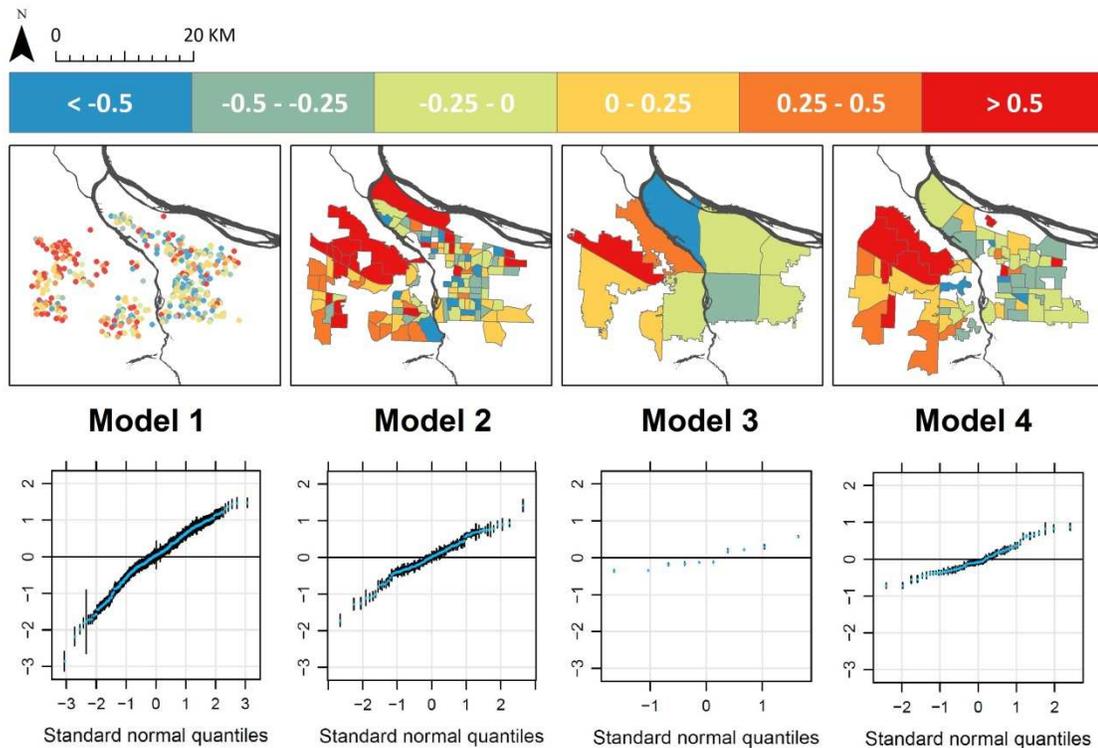


Figure 5.1: Distribution of random effects from unconditional means models. Maps and quantile plots of intercept variability in four two-level unconditional means models with random effects specified for addresses ($n = 460$, Model 1), tracts ($n = 125$, Model 2), neighborhood associations ($n = 63$, Model 3), and regions ($n = 10$, Model 4). Because random intercepts represent a weighted average of observations, the maps indicate where log-transformed water use observations tended to be high or low in the city relative to the grand mean. Corresponding quantile plots with confidence intervals for each grouping structure reveal the effect of MAUP in the results, in that intercept variability is most prominent in the most disaggregated data (Model 1).

Design effects in Table 5.2 indicated substantial dependence in the data, leading to reductions in effective sample size. This was particularly true at level 1 because of strong temporal correlation of water use observations for a given household. At level 2, all empty models exceeded the threshold design effect of 2 (Muthen and Satorra 1995), but results were mixed at level 3. Grouping households by RLIS neighborhood (Model 7) led to relatively large design effects of 2.38, while census tracts (Model 5) yielded a design effect of only 1.79. The small number of households per census tract (average of 3.68) may explain this result. Grouping households by region presented the reverse

problem. The calculated design effect of 11.43 led to a sharp reduction in effective household sample size, from 460 to 40, raising questions about the statistical power of the estimate (Scherbaum and Ferreter 2009). Given these considerations, as well as the plausible contextual relevance noted in Section 4.1.2, RLIS neighborhoods were selected as the most contextually relevant neighborhood boundary.

Table 5.2: Design effects from unconditional means models. Results suggested a three-level model with households grouped by neighborhood association boundaries from the Regional Land Information System were most appropriate for the data.

Model	Levels 1 and 2				Levels 2 and 3			
	ICC	Mean cluster size	Design effect	Effective sample size	ICC	Mean Cluster size	Design effect	Effective sample size
1	0.53	76.23	41.17	852	-	-	-	-
2	0.32	280.52	91.15	385	-	-	-	-
3	0.10	556.59	334.81	105	-	-	-	-
4	0.19	3506.50	107.90	325	-	-	-	-
5	0.53	76.23	41.13	852	0.30	3.68	1.79	256
6	0.54	76.23	41.94	836	0.23	46.00	11.43	40
7	0.52	76.23	40.44	867	0.22	7.30	2.38	194

A three-level random effects structure, with SFR water use observations grouped by households that were, in turn, nested in RLIS neighborhoods, was selected for further examination. A small number of data points (907 rows, or 2.59% of total) were identified as influential outliers in this empty model using the process as explained in Section 3.3.1. These outliers could have arisen from device error or from water leakage from pipes on the property. Once influential outliers were removed, model residuals approached the normal distribution (Figure 5.2). Parameter estimates for the selected empty model with

this modified dataset indicate that 50% of total variance in observations arose from between-household processes ($\tau_{00}^2 = 0.32$), with 15% attributed to between-neighborhood processes ($\phi_{00}^2 = 0.09$). The remaining 35% variation can be attributed to temporal or random variation, captured by the residual ($\sigma^2 = 0.37$).

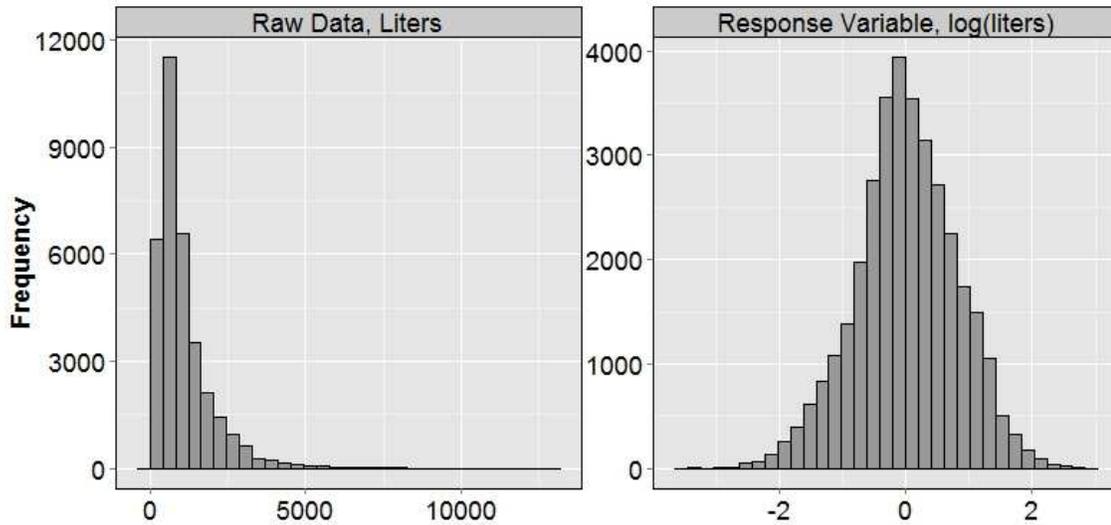


Figure 5.2: Distribution the raw data and response variable. The distribution of weekly mean single family residential water use data in summer was skewed by a small number of very large observations (left). For analytical purposes, water use data were transformed to the natural log scale and grand mean centered. Influential outliers were removed, resulting in the response variable (right).

5.2. Results of model development

5.2.1. Fitting random effects

The forward-fitting process identified significant heterogeneity in the slopes for daily maximum air temperature and cumulative five-day precipitation at the household scale, indicating that the weather-sensitivity of water use was household-specific.

However, because variation in the slope for cumulative five-day precipitation was small ($\tau_{10}^2 = 0.001$ compared to $\tau_{10}^2 = 0.15$ for temperature), this predictor was included only as

a fixed effect. No other random slopes were detected. Contrary to expectations, the effect of parcel-scale land use characteristics was not found to vary across neighborhoods.

5.2.2. Fitting fixed effects

The back-fitting process selected maximum air temperature (TMAX), five-day cumulative precipitation (CumPrp), and the extreme heat index (Extreme_Heat) as explanatory variables at level 1. Relative humidity and wind speed were also evaluated, but were not significantly related to water use, once temperature and precipitation were included in the model. Lot size (Lot_Size), building size (Building_Size), and building age (Building_Age) were selected at level 2, while property value was not. No land cover variables were selected at the household scale. At the level 3, models that included neighborhood mean building age (Building_Age_Nbo) and % of SFR area with low vegetation (LowVeg_%_Nbo) were identified as significantly more likely. Cross-scale interactions between TMAX and Lot_Size at both the household scale and neighborhood scale were selected. However, at the neighborhood scale, the coefficient was barely significant, approaching zero in value, so it was ultimately excluded from the final model. None of the other candidate cross-scale interactions in Table 4.3 were found to be significant.

5.3. Full model

Fixed effects, variance components, and model fit for the full model are reported in the rightmost columns of Tables 5.3 - 5.5, respectively. To illustrate how variance is explained across scales, these tables include estimates for a series of increasingly complex multilevel models leading to the full model. The tables also indicate parameters

named in Chapter 3 to link empirical results with statistical theory. The leftmost column presents the selected unconditional means model (Empty model), omitting influential outliers. To its left is an expanded model that includes TMAX, CumPrcp, and Extreme_Heat as fixed effects (Level 1 fixed). Next, the model was expanded to account for Building_Size, Lot_Size, and Building_Age (Level 2 fixed). Random slopes for TMAX were then specified (Random slopes). The full model expands the random slopes to include fixed effects for the neighborhood-scale predictors Building_Age_Nbo and LowVeg_%_Nbo, as well as the cross-scale interaction between TMAX and Lot_Size. Because all variables were natural log transformed, coefficients represent elasticities and can be interpreted as the percent change in water use associated with a 1% change in the explanatory variable. Interepts are interpreted as averages, with random effects denoting deviation from the average.

5.3.1. Estimated fixed effects

Water use increased with temperature and decreased with precipitation, as expected. A 10% increase in TMAX was associated with a 3.9% increase in water use. The connection between rainfall and water use was relatively weak (decrease of 0.2% for a 10% increase in CumPrcp). Extreme_Heat was also significant and positive, with a 10% increase associated with an additional 0.8% increase in water use, on top of the fixed effect for TMAX of 3.9%. Bootstrapping results indicated these coefficients were highly significant ($p < 0.001$).

At the household scale, water use was positively correlated with Building_size and Lot_size, while negatively correlated with Building_age. Building_size had the

largest effect of all explanatory variables—a 10% increase was associated with a 6.0% increase in water use. A 10% increase in Building_age was associated with a 1.6% decrease in water use. Estimates for these variables were significant across model runs ($p < 0.05$). Although Lot_size appeared to be a strong predictor of water use in the Level 2 fixed effects model, its effect weakened as variance was absorbed by other variables more complex models. In the full model, the effect of Lot_Size was less pronounced (1.6%) and less significant ($p = 0.06$) than either Building_size or Building_age. While Lot_Size was a poor predictor of water use on its own, the cross-scale interaction involving Lot_size and TMAX (γ_{11}) was relatively large and highly significant.

Table 5.3: Regression coefficients. Fixed effects results from increasingly complex model design are presented. In the empty model, only random intercepts were specified. Fixed effects were added for weather (Level 1 fixed) and parcel characteristics (Level 2 fixed), then temperature response was allowed to vary by household (Random slopes). Finally, neighborhood-scale variables and a cross-level interaction were specified (Full Model) Parametric bootstrapping procedures were used to obtain p-values.

Variable	Parameter	Empty model	Level 1 fixed	Level 2 fixed	Random slopes	Full model
Intercept	γ_{00}	-0.015	- 0.015	- 0.005	- 0.006	0.000
TMAX	γ_{10T}	-	0.392***	0.392**	0.396***	0.394***
CumPrcp	γ_{10P}	-	- 0.024***	- 0.024**	- 0.024***	- 0.024***
Extreme_Heat	γ_{10H}	-	0.083***	0.083**	0.082***	0.083***
Lot_Size	γ_{10L}	-	-	0.184***	0.144**	0.132*
Building_Size	γ_{10B}	-	-	0.607***	0.595***	0.600***
Building_Age	γ_{10A}	-	-	-0.236	- 0.249***	- 0.158**
LowVeg_%_Nbo	γ_{01N1}	-	-	-	-	0.049
Building_Age_Nbo	γ_{01N2}	-	-	-	-	- 0.218*
Interaction, TMAX and Lot_Size	γ_{11H}	-	-	-	-	0.187***

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Building_Age_Nbo also exhibited an inverse relationship with water use.

Compared to the household scale, the effect of Building_Age_Nbo was larger but was less significant ($p = 0.07$). Water use increased with LowVeg_%_Nbo, but the effect was not significant ($p = 0.63$). Multicollinearity was not an issue in the full model as variance inflation factors were less than 2.5 for all explanatory variables.

Table 5.4: Variance components. Random effects results from increasingly complex model design are presented. In the empty model, only random intercepts were specified. Fixed effects were added for weather (Level 1 fixed) and parcel characteristics (Level 2 fixed), then temperature response was allowed to vary by household (Random slopes). Finally, neighborhood-scale variables and a cross-level interaction were specified (Full Model) Parametric bootstrapping procedures were used to obtain p-values.

	Parameter	Empty model	Level 1 fixed	Level 2 fixed	Random slopes	Full model
Residual	σ^2	0.231***	0.223***	0.223***	0.219***	0.219***
Household intercept	τ_{00}^2	0.323***	0.323***	0.285***	0.287***	0.289***
TMAX slope	τ_{10}^2	-	-	-	0.147***	0.142***
Slope-intercept covariance	τ_{01}	-	-	-	0.300***	0.300***
Neighborhood intercept	ϕ_{00}^2	0.094***	0.095***	0.014	0.090*	0.006*
ICC _H	ρ_H	0.644	0.651	0.561	-	-
ICC _N	ρ_N	0.225	0.227	0.047	-	-

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

5.3.2. Estimated variance components

Variance components in Table 5.4 were largest in the unconditional means models and decreased as fixed and random effects account for progressively more variation in the response variable, as expected. Residual variance initially decreased from 0.231 to 0.223 as weather variables accounted for variance at level 1, then decreased to 0.219 with the

addition of random slopes for TMAX. Variance in the household level intercept declined from 0.323 to 0.285 as fixed effects at level 2 captured a portion of between-household variance. The slope-intercept correlation of 0.3 indicated that households with high water use, as reflected by larger intercepts, also tended to have more temperature-sensitive water use, indicated by steeper slopes (Figure 5.3). Variance in the neighborhood intercept dropped sharply from 0.09 to 0.006 once the model controlled for parcel characteristics. Although small, p-values generated through bootstrapping (1,000 replicates) indicated that variance in neighborhood intercepts was different from zero.

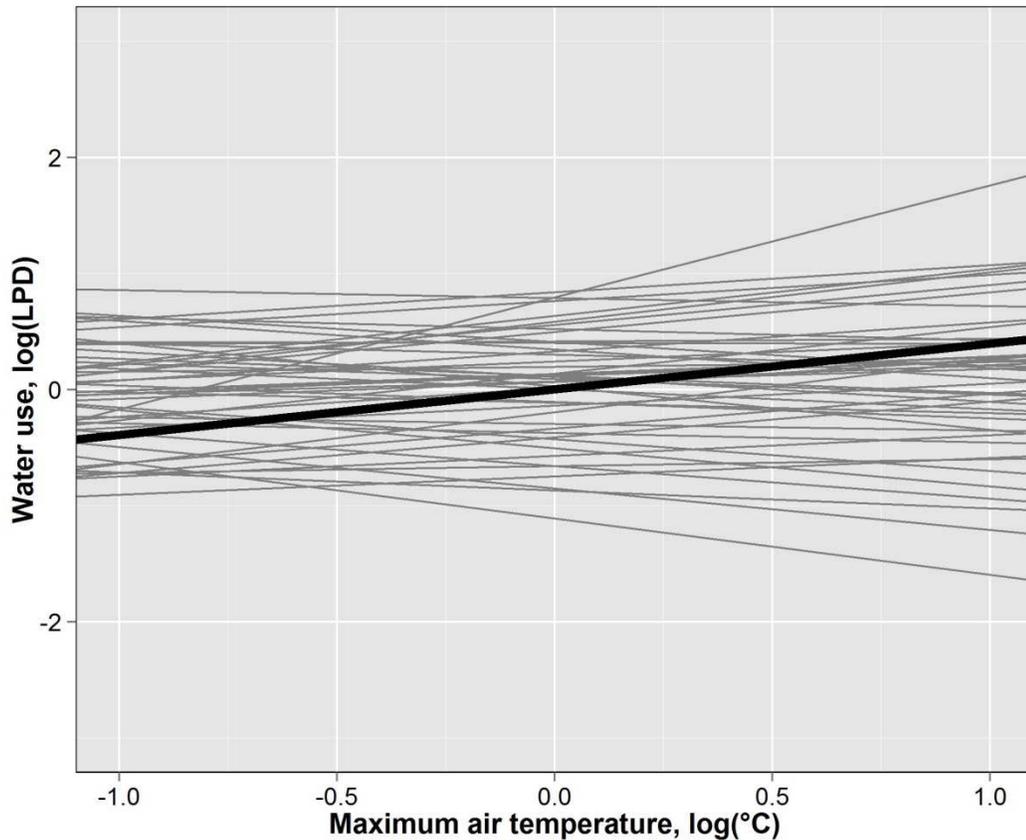


Figure 5.3: Correlation of household-specific slopes and intercepts. Slopes and intercepts were plotted for randomly selected subset of 10% of sample households for temperature (in grey) along with the fixed effect for temperature (in black). The resulting fan shape reflects the slope-intercept correlation of 0.3 and indicates that higher water-consuming households, with $\log(\text{liters per day}) > 0$, also tend to have more temperature-sensitive use patterns (slope > 0).

5.4. Diagnostics

5.4.1. Normality and independence of residuals

Level 1 residuals were not temporally correlated and appeared to follow a normal distribution centered on zero (Figure 5.4), satisfying requirement that $\varepsilon \sim N(0, \sigma^2)$.

Normality and independence in the residual suggested that the full model successfully captured dependencies the data at this scale. No spatial pattern was evident in the random slope for TMAX (Morans I p-value = 0.0991) or the household intercept (Morans I p-value = 0.40), indicating that nesting of households in neighborhoods led to a spatially random pattern. At the neighborhood-scale, residuals were also found to be spatially random in the full model (Morans I p-value = 0.77), although this result could also reflect the very limited variability in between-neighborhood variance, controlling for parcel characteristics. Cluster and outlier analysis in ArcGIS 10.0 indicated no significant hot or cold spots in random effects. The full model also had the effect of removing correlation between standardized residuals and predictors, particularly between Lot_Size and the random effect for TMAX (Figure 5.5).

5.4.2: Goodness of fit

AIC and deviance in Table 5.5 indicated that increasingly complex models accounted for progressively more variation in water use. Deviance decreased from 49085 in the unconditional model to 47631 in the full model, while the number of parameters increased from 3 to 15. Results from log-likelihood ratio tests, conducted with ML or REML estimation, as appropriate (see Section 3.4.2), indicated that each model explained significantly more variance than the previous, although the magnitude of this difference

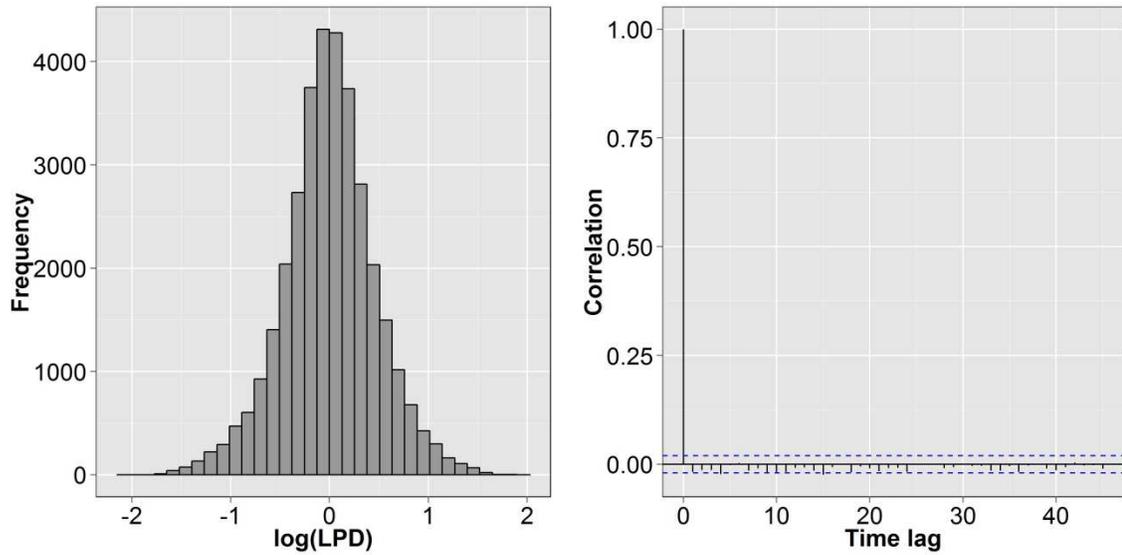


Figure 5.4: Normality and independence of level 1 residuals. Residuals from the full model were normally distributed, centered on zero (left), and were not not correlated across week-long time lags (right).

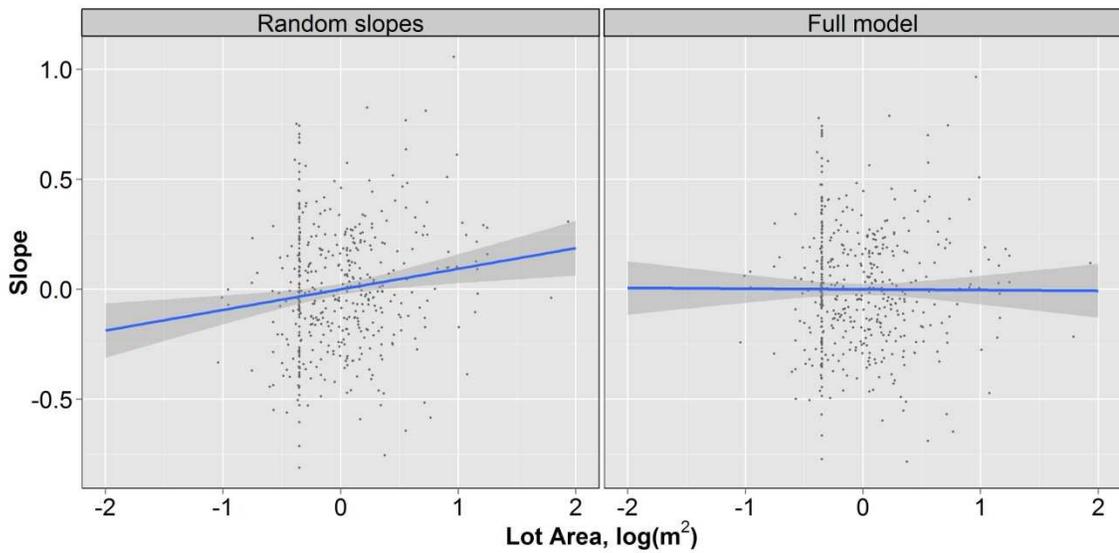


Figure 5.5: Correlation between lot size and random effect for temperature across models. Because it excluded neighborhood-scale variables, the Random slopes model resulted in correlation between explanatory variables and random effects, particularly between lot size and the random slope for temperature (left). The full model removed the correlation, resulting in superior performance as well as improved model fit (right). This result suggested that neighborhood effects were at work in the data, even if the estimated between-neighborhood intercept variability was itself small.

was relatively low between the random slopes model and the full model. Between the full and the empty model, residual variance decreased 5% at level 1, 11% at level 2, and 93% at level 3. The marginal and conditional R^2 were calculated for the full model as 0.22 and 0.67, respectively, indicating that fixed effects explained 22% of total variance, with that portion increasing to 67% of total variance when random effects were included.

Table 5.5: Relative goodness of fit across models. Results from increasingly complex model design are presented. In the empty model, only random intercepts were specified. Fixed effects were added for weather (Level 1 fixed) and parcel characteristics (Level 2 fixed)., then temperature response was allowed to vary by household (Random slopes). Finally, neighborhood-scale variables and a cross-level interaction were specified (Full Model) Comparative model fit with Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC) and deviance statistics. Log-likelihood ratio tests indicated significant improvement in model fit in the full model.

Metric	Empty model	Level 1 fixed	Level 2 fixed	Random slopes	Full model
AIC	49082	47854	47749	47581	47571
BIC	49116	47913	47833	47682	47698
deviance	49074	47840	47729	47557	47541
Number of parameters	4	7	10	12	15
Log-likelihood vs. previous (p-value)	-	0	0	0	0.0013

5.4.3. Confidence intervals and hypothesis tests

Confidence intervals constructed from the likelihood profile can be used to assess whether estimated coefficients are significantly different from zero. Results are presented in Figure 5.6. As expected based on bootstrapped p-values, the coefficients for lot area and % neighborhood low vegetation were not significantly different from zero. All other fixed effects were deemed significant.

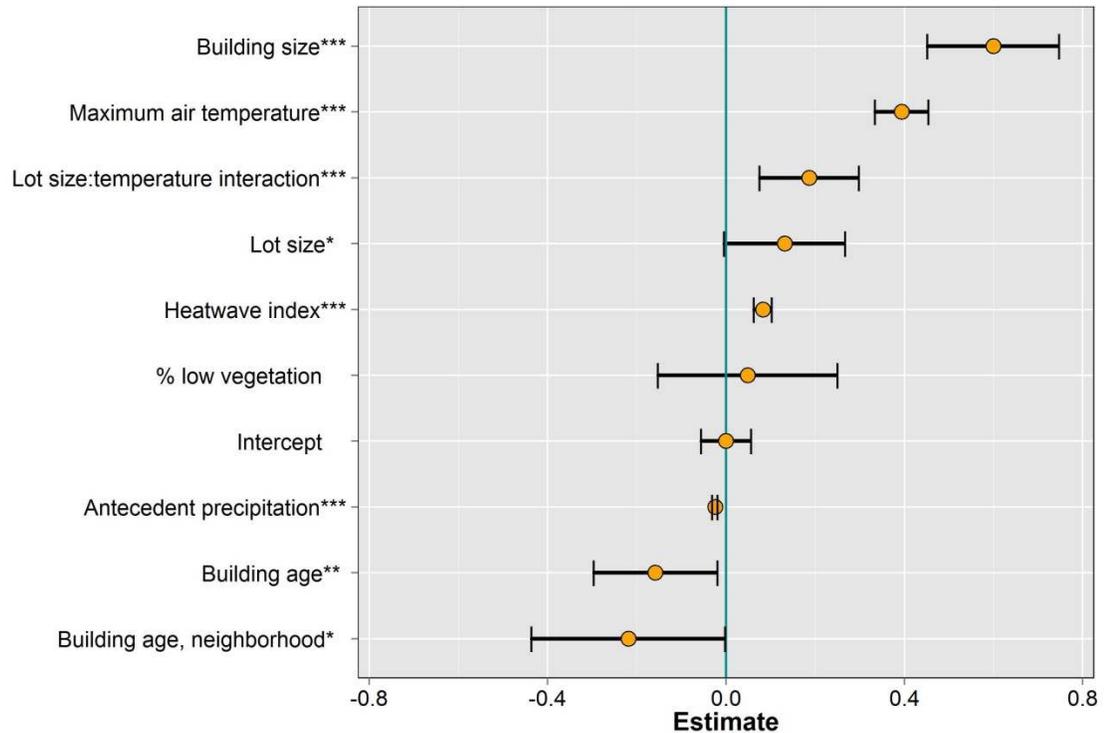


Figure 5.6: Confidence intervals for fixed effects in full model. Estimates are interpreted as the partial elasticity of water use with respect to each predictor variable. Confidence intervals for lot area and neighborhood % low vegetation are not significantly different from zero. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

5.4.4. Predictive power of the multilevel model

To assess the predictive power of a multilevel approach, a pooled ordinary least squares (OLS) model were specified with the same fixed effects. Estimated coefficients and confidence intervals were not significantly different from the full multilevel model. The main difference lay in the nonstationarity of multilevel coefficients and the partitioning of variance across scales. The utility of these model features is revealed when the model is used for prediction. The sample data was subset into a calibration dataset (2001-2004) and a validation dataset (2005). Pooled OLS and multilevel models were estimated using the calibration data. Parameter estimates were very similar to results

presented above. Figure 5.7 compares model predictions to observed 2005 values for water use. The pooled OLS model only explained 22% of variance in water use, commensurate with the marginal R^2 reported above. The multilevel model performed significantly better, explaining 59% of total variance. By decomposing variance across scales and allowing for household heterogeneity, a multilevel approach respected the nested structure of repeat-measure household water use data, which considerably improved the accuracy of model predictions.

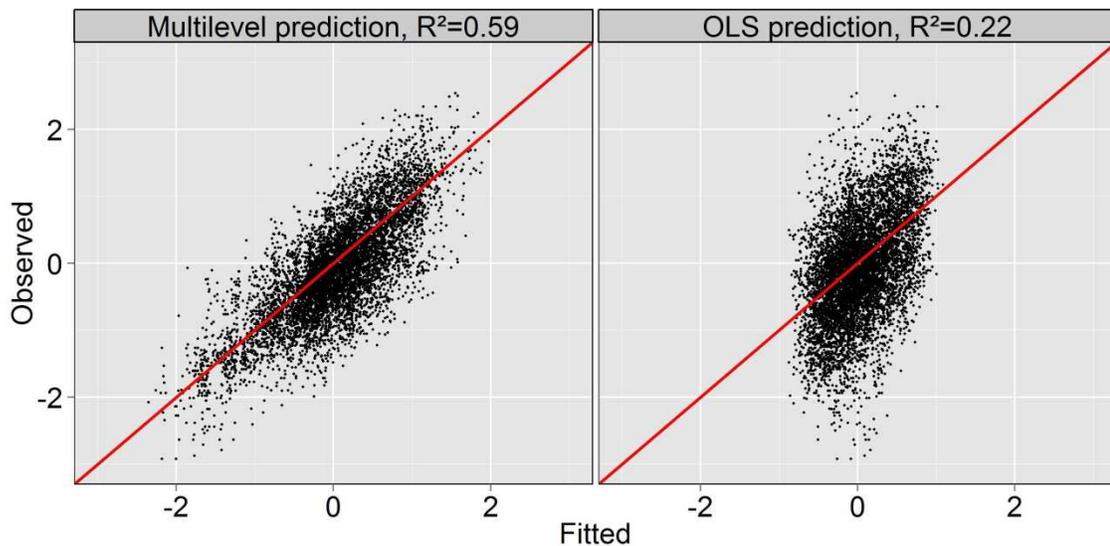


Figure 5.7: Predicted water use from the multilevel model. Multilevel and pooled OLS models were calibrated to 2001-2004 data and validated on 2005 data. The full multilevel model predicted 59% of variance in observed water use 2005 (left), while an OLS model with the same fixed effects structure explained only 22% of variance (right). Decomposing variance across scales and allowing for household heterogeneity were key to improving predictive power of a repeat-measure household water use data model.

Chapter 6: Discussion

This chapter situates multilevel model results in the context of previous research and explores implications for stabilizing peak urban water demand amid combined climate and land use change. Section 6.1 teases out the roles of weather variables in driving temporal variation in water use. Section 6.2 interprets the effects of building structural variables on water use. Sections 6.3 and 6.4 explore possible explanations for why the effects land cover and between-neighborhood variability, respectively, were insignificant in the full model. Section 6.5 concludes the chapter by highlighting the utility of fine-scale analysis for accurate water use projections, while also substantiating calls for greater coordination among water resource managers and land use planners in the interests of long-range water resource sustainability.

6.1. Decomposition of weather sensitivity

Response of SFR water use to temperature fluctuations can be decomposed into four components: a fixed effect, a household-specific random effect, an extreme heat effect, and a land use effect that involved the cross-scale interaction of lot size and temperature (Figure 6.1). The fixed effect of temperature reflected the general tendency for SFR water use to increase on hotter days as households replace water lost to ET. The estimated temperature elasticity of 0.39 was lower than estimates of 1.0 obtained in Seattle, Washington (Polebitski and Palmer 2010) and 0.6 based on data from Phoenix, Arizona (Balling and Gober 2007). Both of these studies relied on data aggregated to coarser spatial and temporal resolutions, which may have lead to conflation of the average effect and household-specific deviations from the average. The fine scale of

analysis used for the present study revealed that households exhibited a wider range temperature responses, with a standard deviation of 0.37 from the average elasticity (Figure 5.6). In the full model, household-specific slopes and intercepts were positively correlated ($\tau_{01} = 0.3$), consistent with previous findings that households with higher water consumption levels also tended to have more seasonal variation in water use (Polebitski and Palmer 2010). Apart from providing insights into between-household heterogeneity, allowing temperature response to vary across households significantly improved model fit (Table 5.5) and predictive power (Figure 5.7). Calibration and validation results indicated that the full multilevel model predicted 59% of variance in water use, a remarkable degree of accuracy given the it is not uncommon for residential water use models to account for less than 40% of total variance (Renwick and Archibald 1998; Ramachandran and Johnson 2011), although the literature includes notable exceptions (Shandas and Parandvash 2010).

Temperature was a more robust predictor of water use than precipitation. The weak linkages between water use and precipitation may have resulted from the fact that rainfall is rare in summer (Balling and Cubuque 2009) and that many automatic irrigation devices are not calibrated to weather fluctuations (personal communication, Steven Carper, Tualatin Valley Water District, July 2013). Antecedent precipitation conditions were more predictive than actual rainfall depth, reflecting the role of soil moisture dynamics in mediating water balances, particularly water lost to ET over time (Berthier et al. 2004). Persistent effects of rainfall and extreme heat are consistent with previous studies (Miaou 1990), although the temperature that elicited an ‘extreme’ heat effect was

relatively low in Portland (32°C). For example, Zhou et al. (2001) defined extreme heat events above 39°C in Melbourne, Australia. Maximum air temperature in Portland exceeded this threshold on only two occasions during the study period.

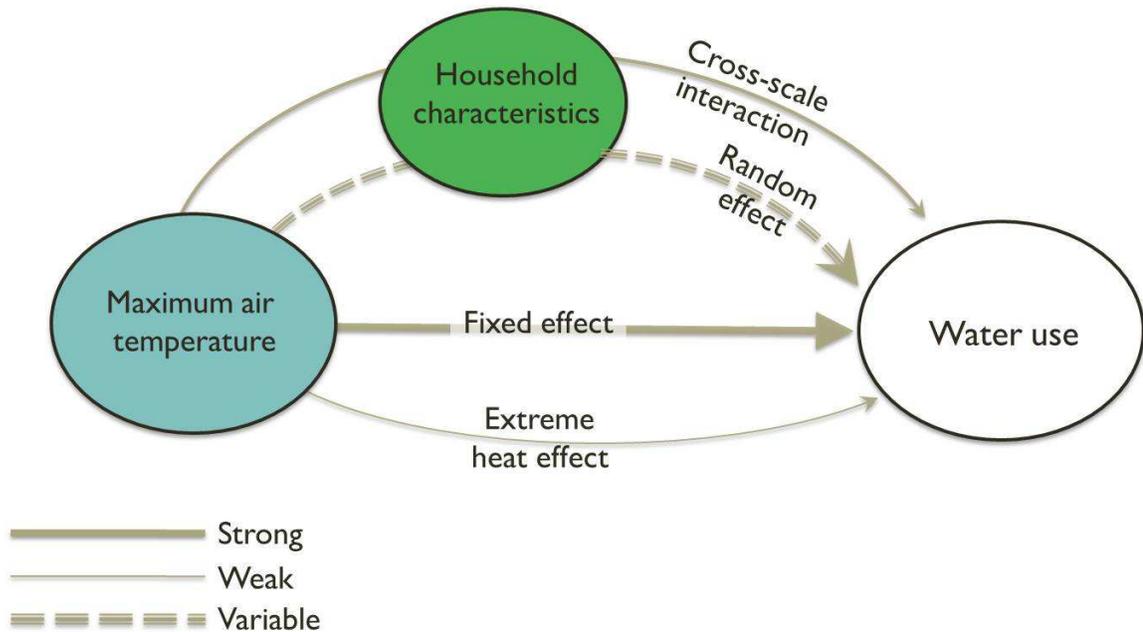


Figure 6.1: Decomposition of temperature response. Four components of single-family residential response were to temperature fluctuations were identified: a fixed effect of maximum air temperature (common to all households), a random effect of water use (household-specific deviation from the fixed effect), a separate extreme heat effect, and a cross-level interaction between temperature and lot size. Width of arrow indicates the strength of the relationship.

6.2. Linkages between land use and water use

Linkages between land use and water use were found at multiple scales (Figure 6.2). Building size and building age were significant predictors of water use. Water use was most elastic with respect to building size, also consistent with previous studies on water use in the Pacific Northwest, in terms of both direction and magnitude (Chang et al. 2010; Shandas and Parandvash 2010). Polebitski and Palmer (2010) estimated a

comparable building size elasticity of 0.51 in Seattle, Washington. It is possible that larger buildings could reflect larger household sizes, although Chang et al. (2010) found no significant relationship between household size and water use at the block-group scale. Alternatively, building size could be collinear with income, where more affluent households tend to both have larger houses and use more water in summer, consistent with survey evidence that outdoor water use arises from higher-income households seeking to display their social status (Askew and McGuirk 2004; Domene et al. 2005).

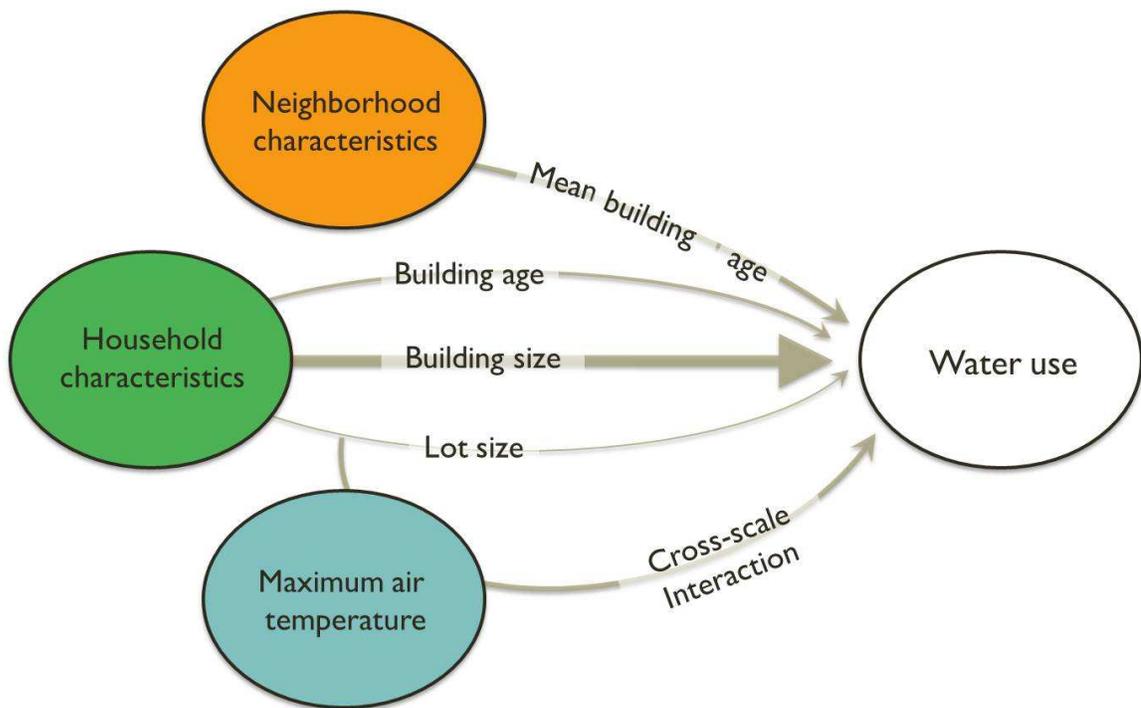


Figure 6.2: Linkages between land use and water use across scales. Building size exhibited the strongest linkage with water use, suggesting possible covariance of high water use patterns and high incomes or household sizes. Building age was negatively correlated with water use at both the household and neighborhood, indicating a suburban effect at work, since the newest construction is located in low-density suburban areas. Although weakly related to water use as a fixed effect, lot size was involved in a robust (but temporally contingent) cross-scale interaction with temperature.

Lot size is commonly thought to drive seasonality of water use, and has been identified as a significant predictor in aggregated studies (House-Peters and Chang 2011). However, at the fine scale used for the current analysis, the linkage between water use and lot area was relatively weak (0.13) and statistically insignificant. Instead of operating as a fixed effect, lot area was a component of a robust cross-scale interaction with temperature, where larger lots tended to amplify the effect of temperature on water use (Figure 6.1). Importantly, the analysis did not detect a similar interaction between weather and building size. This suggested that building size and lot size have different types of effects on water use. Building size had a fixed effect, one that operates with temporal consistency across households at multiple scales, which explains its importance at the parcel (Shandas and Parandvash 2010), block-group (Chang et al. 2010), and tract scales (Polebitski and Palmer 2010). In contrast, the effect of lot size has a temporally inconsistent effect, being contingent on weather fluctuations, with a magnitude jointly determined by its interaction with temperature (Figure 6.3). The model did not detect between-neighborhood heterogeneity in the effect of any land use variable, suggesting that these parameters affect water use similarly across the study area. Once again, fine-scale analysis was required to draw out these distinctions.

The effect of building age appeared particularly strong because it was selected at both the household and neighborhood scale. Chang et al. (2010) also found a significant spatial covariance of average building age and water use at the block-group scale, although Polebitski and Palmer (2010) found that the effect of building age was sensitive to the season and the type of model specified. Because of changes in building code in

1992, older buildings are more likely to be equipped with water-inefficient technologies, such that a positive relationship would be expected between indoor water use and

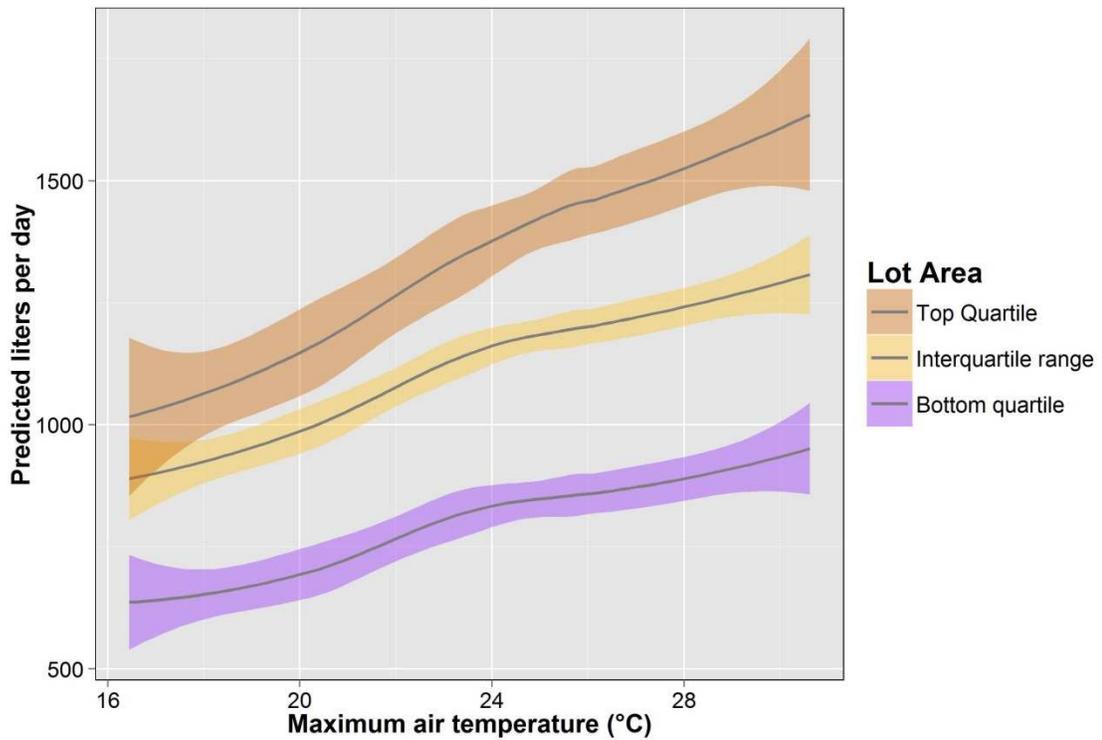


Figure 6.3: The contingent effect of lot size. Predicted water use from a multilevel model increases with temperature, but the magnitude of the increase depends on lot size. At low temperatures, there is little difference in water use patterns between water use from households on large lots (top quartile) and the interquartile range. Interaction between weather and lot size becomes prominent above 22°C. Water use generally tends to be lower and less variable on smallest lots (bottom quartile).

building age. Finding an inverse relationship between summer water use and building age suggested that the land use factors driving high outdoor water use dominated any conservations effects from indoor water use efficiencies. Building age was moderately correlated with building size (Pearsons correlation = -0.31) and lot size (Pearsons correlation = -0.21), but not enough to suggest problems with multicollinearity. Although more significant at the parcel scale, the effect of building age was more pronounced at this scale (-0.22), indicating that a separate age effect is at work at the neighborhood

scale, even after controlling for parcel-scale building age. This result implied that household on two otherwise similar parcels would use water differently depending on the average building age in the neighborhood. This interpretation would be consistent with survey evidence showing that outdoor water use serves to display status and convey social distinction, particularly in newer suburban residential developments (Askew and McGuirk 2004; Domene et al. 2005).

6.3. Why was land cover not more predictive?

Land cover explained almost no variation in SFR water use. On its face, this result was surprising because it would seem that the water demands of vegetation would drive the underlying process that generates seasonal water use. However, the result is not without precedent. Previous research has uncovered a relatively weak relationship between water use and vegetation (Wentz and Gober 2007; Breyer et al. 2012). This finding could have arisen from household misunderstanding of vegetation water requirements. Some evidence suggests that households can be poor estimators of ET and agronomical requirements, over-watering (Balling and Gober 2007) or under-watering (Domene et al. 2005) as a result. Another possible explanation is that building structural variation captured all the relevant information, making land cover variation redundant. Alternatively, the low correlation could have resulted from the data resolution— water use may be highly sensitive to particular types of irrigated vegetation that were not captured by the land use classification. The land cover dataset contained only four land cover classes, which may have been too coarse to distinguish between thirsty lawns and drought-tolerant native shrubbery, since these fine-scale differences may be central to

accounting for why one patch of earth was watered while the other was not (House-Peters and Change 2011b). However, investigation with a more detailed land cover classification (The Intertwine 2012) that included a class for lawns did not improve model results.

A more likely explanation is that the presence of a lawn did not necessarily mean that the lawn is being irrigated. Portland Water Bureau (2010) provided anecdotal evidence to substantiate this claim. According to PWB, declining rates of summertime lawn irrigation stemmed from outdoor water restrictions imposed during a drought in 1992. Many households reported a change in attitude towards lawn irrigation after the drought as they “realized that if they didn’t water their lawn, they didn’t have to spend time mowing it, and the grass grew back green as soon as the fall rains began” (Portland Water Bureau 2010, 3-1). Two decades later, lawn watering appears to have ceased in some neighborhoods, but for a few holdouts, as residents defect en masse from turfgrass irrigation. In other neighborhoods, however, irrigated turfgrass remains the norm. This heterogeneity, in concert with the lack of detail in land cover types, could account for why the low vegetation was selected through log-likelihood ratio testing but was ultimately not significant in the final model. The effect of vegetation was more likely, given the data, but its place-based effect, coupled with lack of differentiation between water-intensive and drought-tolerant vegetation, meant that the estimated coefficient was not significantly different from zero.

6.4. Variance decomposition

One of the main strengths of a multilevel approach lies in the decomposition of variance across scales. However, the portion of total variance attributable to between-neighborhood factors fell sharply once the model controlled for parcel characteristics (Table 5.3). This decline reflected Fotheringham and Brundson's (1999) point regarding the difference between nonstationarity and model misspecification—nonstationarity in neighborhood intercepts in the full model was tiny once the model accounted for relevant sources of between-neighborhood variance. Log-likelihood ratio testing indicated that the full model was more likely given the data (Table 5.5), although the significance of this test ($p = 0.0013$) was smaller than other model pairs ($p < 2.2e-16$). Bootstrapped p -values also indicated that between-neighborhood variability was statistically different from zero, but small. Grouping households by neighborhood had the effect of improving overall model performance, in terms of both model fit and reducing correlation between model residuals and covariates (Figure 5.6). This would suggest that the processes driving high water use were primarily located at the household scale, but that similar households tended to be closer together—this would correspond with correlated neighborhood effects as defined by Manski (1993). By contrast, the effect of neighborhood building age would seem to operate as a contextual effect in Manski's framework, since it captured a separate source of meso-scale variation, after controlling for parcel building age. The analysis concluded that neighborhood context was associated with a very modest but statistically significant effect on water use, controlling for weather and building structural characteristics.

6.5. Policy implications

The decomposition of temperature sensitivity has immediate implications for vulnerability of urban water provisioning systems under the combined effects of climate and land use change. A wide range of climate models project hotter, drier summers, while urbanizations models project ongoing densification. Regarding climate, Mote and Salathé (2010) projected that summer temperatures across the Pacific Northwest would increase 1.9 to 2.7 °C by 2040 and that mean summer rainfall would decrease 5.1% to 11.2%, relative to the 1970-1999 mean. The INSIDE Idaho gridded dataset projected the following changes to summer climate over the period 2025-2055 in the study area, relative to the 1970-1999 mean: 0.24 to 1.43°C increases in summer temperatures, decreases in cumulative five-day precipitation by 4.2% to 34%, and increases of 71.0% to 140% in the extreme heat index (Figure 6.4). At the same time, Metroscope, a regional land use and transportation model, projected ongoing urban densification as population in the Portland area continues to grow within the urban growth boundary. Assuming SFR capacity is exhausted by 2040, median lot size is projected to decrease an average of 160m² (-22%) with decreases of up to 300m² in low-density, high water-consuming areas in proximity to transit (Figure 6.5).

6.5.1. Triple exposure

Taken together, results indicated that SFR water use tended to be higher and more temperature sensitive in residential areas characterized by larger, newer buildings on larger lots. In other words, sprawl drove increases in summertime water use that are likely attributable to increased outdoor use. It has already been noted that the built environment associated with sprawl encourages particularly high levels of resource

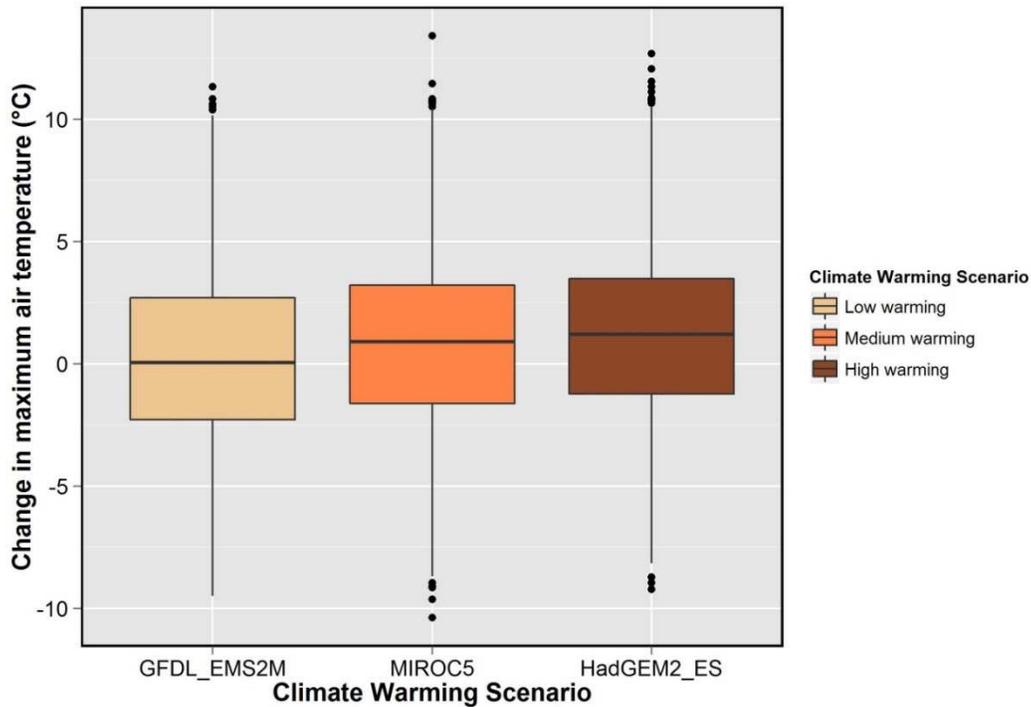


Figure 6.4: Projected deviation in summer maximum air temperatures, 2025-2055. Projections are shown for three emissions scenarios and are expressed relative to the reference period of 1970-1999. The Geophysical Fluid Dynamics Laboratory Earth System Model (GFDL_EMS2M) is the most cool and wet warming scenario, while the Hadley Global Environment Model 2 (HadGEM2_ES) is the most hot and dry scenario. The Model for Interdisciplinary Research on Climate (MIROC5) lies between these extremes. All scenarios project higher temperatures and less precipitation in summer. Data source: INSIDE Idaho.

consumption (Norman et al. 2006). Fine-scale linkages between the built environment and high levels of outdoor water use have important consequences for reliability of urban water supply in the aggregate. For example, rapid suburban growth throughout inland California, United States, magnified seasonality of water use, jeopardizing water security in the region (Hanek and Davis 2006). Similarly, increases in water consumption for newly-built pools and gardens in so-called ecotourist enclaves have stressed local water supplies on the Mediterranean island of Mallorca (Hof and Schmidt 2011). Yet, demand-side shortages are not unique to semi-arid regions—a mismatch between fixed supply and increasing demand can stress water resources in humid climates as well (Lyon et al.

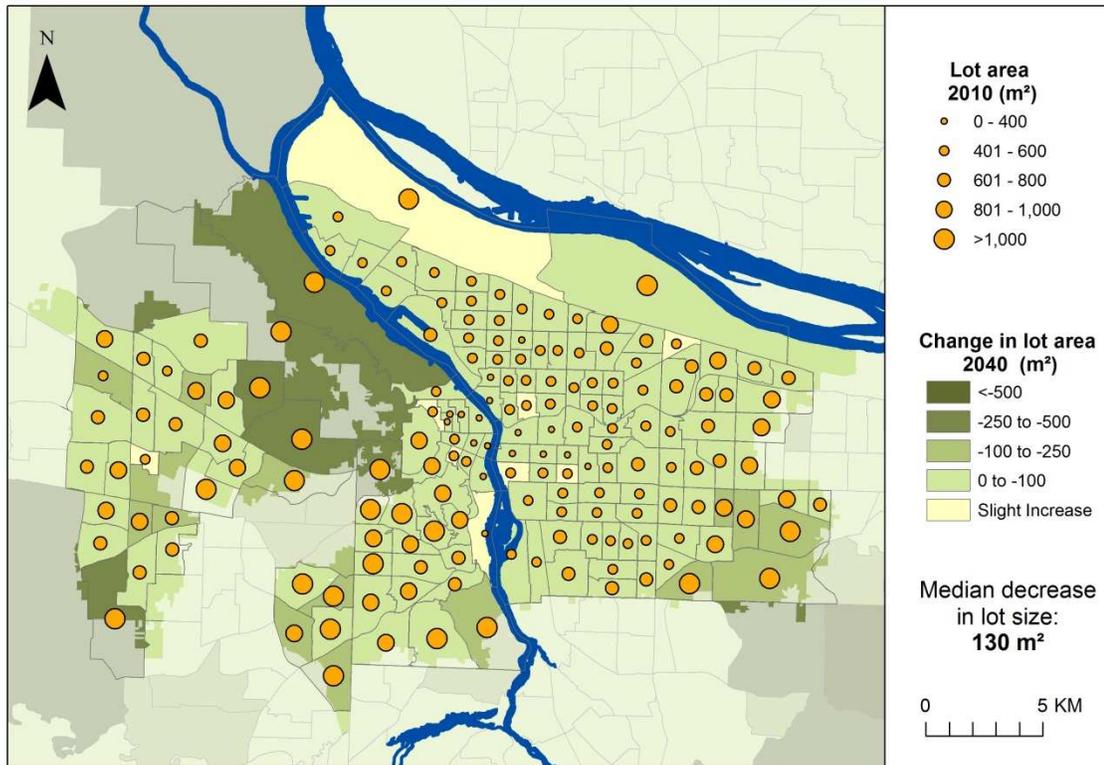


Figure 6.5: Projected change in mean single-family residential (SFR) lot area. Results are aggregated by census tract and refer to projected changes by 2040. This calculation assumed that all capacity for SFR development was exhausted, both in terms of new construction and infill. Densification is greatest in suburban tracts in proximity to transit. Data source: Metroscope.

2006). Hill and Polsky (2007) showed that summertime outdoor watering restrictions in suburban Massachusetts, United States, resulted not from drought-induced shortage of supply but from excessive demand because the low-density urban form promoted intensive outdoor use. In the context of climate change, suburbanization may lead to not double but triple exposure to risk of water shortages by increasing aggregate water demand (population growth) and amplifying the seasonality of that demand (land use), even as temperature drives demand increases while also shrinking the hydrological resource base (climate change).

However, results also point to the capacity for urban form to soften the impact of high temperatures on water use, with urban density serving as a pathway to climate

resilience and, potentially, more sustainable urban water provisioning systems. In Portland, the groundwork for urban density has already been laid out in state-level land use regulations that constrain sprawl within an urban growth boundary. Ongoing densification is projected for the Portland metropolitan region, with the largest changes in median lot size anticipated in low-density, currently suburban areas, particularly those areas that are well-served by light rail (Figure 6.5). Across the region, multifamily developments with limited opportunities for outdoor water use are expected to comprise an increasing share of the housing stock (personal communication, Dennis Yee, Metro Regional Government, June 2013). From a water resources standpoint, this trend appears to double as a form of urban climate change adaptation.

6.5.2. Vegetation feedbacks and landscape regulations

Although densification appears to be a mechanism for resilience from a water demand standpoint, a more comprehensive view would also account for complex feedbacks between land use and water use related to urban heat island (UHI) effects associated with dense areas covered by impervious surfaces. Households may, to some extent, increase water use to mitigate UHI effects (Guhathakurta and Gober 2010), although water use tends to be inversely related to impervious surfaces in Portland (Breyer et al. 2012). Indeed, low-density development in Tucson, Arizona was identified as an urban heat ‘sink,’ where evaporative cooling from vegetated surfaces lowered surface temperatures relative to the surrounding desert (Halper et al. 2012). Land cover change towards more vegetated surfaces was found to mitigate UHI under climate change scenarios in Portland (Middel et al. 2011). A massive tree planting campaign is already

underway in Portland. Although intended primarily for stormwater management, this land cover change may ameliorate the worst effects UHI as well.

Many areas are already developed such that patterns of building structure are relatively fixed. For these areas, outdoor water use restrictions become increasingly common to suppress nonessential uses in times of crisis, although this form of demand management fails to address the structural origins of excessive seasonality. Hanek and Davis (2006) have suggested that regulation of vegetation in residential landscapes may be more effective in addressing the root cause of seasonal water stress. Replacement of water-intensive vegetation with drought-tolerant, water-efficient plants would function as a sort of retrofitting of residential landscapes to discourage excessive outdoor water use, just as previous changes in building code mandated water-efficient appliances in remodels. These sorts of passive conservation techniques have been a significant factor in declining per-capita urban water demand since 1992 (Rockaway et al. 2011).

6.5.3. Land use planning as equitable water demand management

Conservation pricing has been proposed as an alternative strategy to manipulate outdoor water use (Rinaudo et al. 2012). However, water use tends to be relatively inelastic with respect to price (Worthington and Hoffman 2008). Some research indicates that high water-consuming households become more sensitive to water prices in peak season (Kenney et al. 2008), while other studies have found that wealthy households, who tended to have the highest water use rates, were less price-sensitive overall (Mieno and Braden 2011), so it is unclear whether price hikes would have the desired effect. Apart from efficacy concerns, price hikes also have equity implications— raising prices

on water would be disproportionately borne by low-income households (Renwick and Green 2000; Domene et al. 2005). To the extent that landscape irrigation serves to display status, there is reason to believe signaling increased scarcity through price would not only burden low-income households, but would also make irrigated residential landscapes more exclusive, thus enhancing the role of outdoor water use as a marker of social status.

The results of multilevel analysis suggest that land use planning may be an effective non-price mechanism for managing peak long-run demand, as land use decisions have water use implications. Manipulating water use through land use would target the behavioral origins of demand-side vulnerability to water stress. It would also signal a structural change to water demand, since price-based demand management implies movement along the water demand curve, while using urban density to reduce seasonality would constitute a downward shift in the demand curve (Saurí 2003). Gober et al. (2013) argued that the intricate linkages between land use and water use call for increased coordination between land use planners and water resource managers. Indeed, the extent to which integrated land and water planning is possible could be viewed as a form of adaptive capacity, increasing system resilience in the face of complex change and deep uncertainty (Larson et al. 2013). Results from the present analysis could inform that coordination, because examining parcel-scale water use isolates the land use effects in a way that is most relevant for land use planning (Shandas and Parandvash 2010). However, it should be noted that these implications are limited to urban water delivery systems such as those considered in this study. Results may not apply to other water

rights configurations, particularly given the complex, historically-contingent legacy of water rights allocations in the American West.

Chapter 7: Conclusions

7.1. Summary of main findings

This analysis used multilevel regression to investigate scalar dimensions of SFR water use. It provided evidence that residential water use arises from a mix of biophysical and social drivers at multiple spatial and temporal scales. Responses to the research questions presented in Section 1.4 are as follows.

Maximum air temperature drove temporal variation in water use through multiple pathways (Figure 6.1). On average, a 10% increase in maximum air temperature lead to a 3.9% increase in water use. Household sensitivity varied widely from this average effect, ranging from no response to a 10% increase. Antecedent rainfall varied inversely with water use and was more predictive than precipitation depth and, although the magnitude of the effect (0.25%) was less pronounced than temperature. The analysis also revealed a separate, time-lagged effect associated with extreme heat (0.9%). Although heterogeneous responses to maximum air temperature were detected, households exhibited little to no variation in responses to the persistent effects of rainfall and extreme heat.

Between-household variability explained half of the variance in SFR water use, and was more important than weather fluctuations in explaining water use variation. As a result, allowing intercepts to vary by household considerably improved the accuracy of predicted water use. In terms of parcel characteristics, water use was most elastic with respect to building size than any other predictor, including temperature, with a 10% increase in building size associated with a 6% increase in water use, on average. Building

age varied inversely with water use (-1.6%). Water use tended to increase with lot size, although the magnitude of lot size as a fixed effect was surprisingly weak (1.3%). The effects of parcel characteristics were found to be constant across neighborhoods.

Mean neighborhood SFR building age was inversely related to water use, with an effect that was more prominent than household building age (-2.2%), although the significance of this effect was weaker. Controlling for parcel characteristics, variation between neighborhoods was very small, but statistically different from zero. Although most of the variability in water use arose from household-scale processes, a model that accounted for between-neighborhood variation was deemed more likely, given the data, and improved model performance (Figure 5.5). The analysis concluded that neighborhood effects are present in household water use, although the magnitude of these effects is weak at best.

A significant cross-scale interaction was detected between temperature and lot size. Unlike building size, which operated as a fixed effect, the effect of lot size was contingent on weather fluctuations; larger lots tended to magnify temperature-induced increases in water use, but only on hotter days (Figure 6.3). This finding highlighted the utility of multiscale analysis. Lot size has been identified as a significant driver of water use in coarser-scale studies, likely because its weak fixed effect and strong weather-contingent effect became conflated with aggregation. These nuances, which improved the predictive power of the model, may not be evident from a single-scale perspective. Single-family residential land use and water use are bound in a complex and multiscale relationship (Figure 6.2).

7.2. Addressing issues of scale and aggregation

A multilevel approach was selected to jointly model the micro-foundations and meso-scale context of household water consumption so as to avoid both ecological and atomistic fallacies. In this framework, statistical dependence and between-group heterogeneity were not merely sources of standard error bias to be controlled, but important artifacts of the data warranting investigation in their own right. The research design sought to address critiques of multilevel approaches to geographical hierarchies (Fotheringham and Brundson 1999; Chaix et al. 2005). Biases arising from the uncertain geographic context problem (Kwan 2012) were considered by examining alternative neighborhood boundaries. Neighborhood association and community planning organization boundaries were determined to be most contextually relevant. In light of concerns regarding the discrete, fragmented spatial correlation structure implied in multilevel models, clustering of residuals was carefully examined at multiple scales—no clustering was detected in the full model.

7.3. Climate change, urban form, and resilience

One of the main contributions of this research was the decomposition of temperature response into four components: a fixed or average effect, a household-specific deviation from this average, an extreme heat effect, and a land use effect, encapsulated by the cross-scale interaction between lot size and temperature (Figure 6.1). These findings underscored the demand-side vulnerability of urban (and especially suburban) water provisioning systems to water stress stemming from higher summer temperatures and extreme heat events, suggesting the possibility of triple exposure to the

combined effects of population growth, climate change, and suburbanization. However, the same findings also provided evidence for a conservation effect associated with urban density, implying that dense land use patterns may also be areas of resilience (House-Peters and Chang 2011b). Moreover, these findings pointed to land use planning as a potentially robust and equitable non-price mechanism to target unsustainable levels of outdoor water use.

7.4. Future work

This research has focused on the roles of weather and land use variables in shaping SFR water use patterns. Future work could link these patterns to household characteristics, particularly attitudes, income, household size, lawn watering practices, and presence of drought-tolerant vegetation. Evidence is compelling that these factors could meaningfully influence water consumption (Domene et al. 2005; Domene and Saurí 2006). Household water use is generally on the decline, but the average household size is shrinking, too, which has important implications for resource use (Liu et al. 2007). Connecting water use patterns with survey data could reveal important cultural and perceptual dimensions to water use.

Future research could also explore the relationships among land cover, water use, and the MAUP. The land cover data used for this analysis was at a relatively fine scale (3 ft pixels), but the classification itself was fairly coarse, containing only four land cover types. As a consequence, no statistically significant relationship between land cover and water use was detected. Future research could develop a more robust land cover classification to distinguish between irrigated and non-irrigated vegetation. Furthermore,

the MAUP was considered but not rigorously addressed in this analysis. Future research could leverage this detailed classification scheme to develop a more robust treatment of the MAUP, perhaps by using land cover to delineate neighborhood. Results may clarify the relationship between the spatial extent of irrigated vegetation and summertime water use patterns, while avoiding biases associated with scale and unit of aggregation.

Results from the multilevel model can also be incorporated into urban water demand models to project water use under combined climate and land use change scenarios (House-Peters et al. 2010; Polebitski et al. 2011). Conservation effects could also be explored, both in terms of ongoing per-capita reductions in water use as well as the countervailing trend of demand hardening, especially in winter, where marginal changes in water conservation become increasingly difficult because the easiest changes have already been made. Conservation can extend the life of existing infrastructure and is increasingly valued according to the cost of adding new water supply infrastructure in lieu of reducing demand, so accurate projections have a very concrete monetary value.

This analysis considered multiscale dynamics, time lags, and spatial characteristics of residential water use as a coupled human-natural system. As such, the research responds to calls to develop a deeper understanding of how social and ecological processes form systems (Redmon et al. 2004). Residential water use is itself nested within a broader urban water provisioning system with its socio-ecological dynamics. Future research could tie household water use to other water-consuming sectors and situate the aggregate demand for water in relation to the economic and ecological processes that shape water supply.

References

- Abatzoglou, J. T. 2013. Development of gridded surface meteorological data for ecological applications and modelling. *International Journal of Climatology* 33 (1): 121–131.
- Adamowski, J., and C. Karapataki. 2010. Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms. *Journal of Hydrologic Engineering* 15 (10): 729–743.
- Aitken C., H. Duncan, and T. A. McMahon. 1991. A cross-sectional regression-analysis of residential water demand in Melbourne, Australia. *Applied Geography* 11 (2): 157–165.
- Alcamo, J., M. Flörke, and M. Märker. Future long-term changes in global water resources driven by socio-economic and climatic changes. *Hydrological Sciences* 52 (2): 247-275.
- Akuoko-Asibey, A., L. C. Nkemdirim, and D. L. Draper. 1993. The impacts of climatic variables on seasonal water consumption in Calgary, Alberta. *Canadian Water Resources Journal* 18 (2): 107–16.
- Alker, H. A. 1969. A typology of ecological fallacies. In *Quantitative Ecological Analysis*, eds. M. Dogan and S. Rokkan, 69–86. Cambridge: Massachusetts Institute of Technology.
- Alcott, H. 2009. Social norms and energy consumption. *Journal of Public Economics* 95 (9-10): 1082–1095.
- Arbués, F., M. Garcia-Valiñas, and R. Martinez-Espiñeira. 2003. Estimation of residential water demand: A state-of-the-art review. *Journal of Socio-Economics* 32 (1): 81–102.
- Askew, L. E., and P. M. McGuirk. 2004. Watering the suburbs: Distinction, conformity, and the suburban garden. *Australian Geographer* 35 (1): 17:37.

Baayen, R. H., D. J. Davidson, and D. M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59 (4): 390-412.

Balling, R. C., and P. Gober. 2007. Climate variability and residential water use in the City of Phoenix, Arizona. *Journal of Applied Meteorology and Climatology* 46 (7): 1130-37.

Balling, R. C., P. Gober, and N. Jones. 2008. Sensitivity of residential water consumption to variations in climate: An intraurban analysis of Phoenix, Arizona. *Water Resources Research* 44 (10): doi:[10.1029/2007WR006722](https://doi.org/10.1029/2007WR006722).

Balling, R. C., and H. C. Cubaque. 2009. Estimating future residential water consumption in Phoenix, Arizona based on simulated changes in climate. *Physical Geography* 30 (4): 308-323.

Bates, D. 2006. lmer, p-values, and all that. Available at <https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html> (last accessed 11 November 2013).

———. 2010. lme4: Mixed-effects modeling with R. Unpublished manuscript. Available at <http://lme4.r-forge.r-project.org/IMMwR/lrgprt.pdf> (last accessed 11 November 2013).

Bates, D., M. Maechler, B. Bolker and S. Walker. 2013. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.0-4. Available at <http://CRAN.R-project.org/package=lme4> (last accessed 11 November 2013).

Berthier, E., H. Andrieu, and J. Creutin. 2004. The role of soil in the generation of urban runoff: Development and evaluation of a 2D model. *Journal of Hydrology* 299 (3-4): 252-66.

Bolster, A., S. Burgess, R. Johnston, K. Jones, C. Propper, and R. Sarker. Neighbourhoods, households and income dynamics: A semi-parametric investigation of neighbourhood effects. *Journal of Economic Geography* 7 (1): 1-38.

Breyer, B., H. Chang, and G. H. Parandvash. Land-use, temperature, and single-family residential water use patterns in Portland, Oregon and Phoenix, Arizona. *Applied Geography* 35 (1-2): 142-151.

Caughy, M. O., K. L. Hayslett-McCall, and P. J. O'Campo. 2007. No neighborhood is an island: Incorporating distal neighborhood effects into multilevel studies of child developmental competence. *Health & Place* 13 (4): 788-98.

Caughy, M. O., T. Leonard, K. Beron, and J. Murdoch. 2013. Defining neighborhood boundaries in studies of spatial dependence in child behavior problems. *International Journal of Health Geographics* 12 (1): 24-36.

Chaix, B., J. Merlo, and P. Chauvin. 2005. Comparison of a spatial approach with the multilevel approach for investigating place effects on health: The example of healthcare utilisation in France. *Journal of Epidemiology and Community Health* 59 (6): 517-26.

Chang, H., G. H. Parandvash, and V. Shandas. 2010. Spatial variations of single-family residential water consumption in Portland, Oregon. *Urban Geography* 31 (7): 953-972.

Chen, D.-R., and T.-H. Wen. 2010. Socio-spatial patterns of neighborhood effects on adult obesity in Taiwan: A multi-level model. *Social Science & Medicine* 70 (6): 823-33.

Chowdhury, R. R., K. Larson, M. Grove, C. Polsky, E. Cook, J. Onsted, and L. Ogden. 2011. A multi-scalar approach to theorizing socio-ecological dynamics of urban residential landscapes. *Cities and the Environment* 4 (1): 6-19.

Cook, E. M., S. J. Hall, and K. L. Larson. 2011. Residential landscapes as social-ecological systems: A synthesis of multi-scalar interactions between people and their home environment. *Urban Ecosystems* 15 (1): 19-52.

Coomes, P., T. Rockaway, J. Rivard, and B. Kornstein. 2010. Residential water use trends in North America. *JAWRA Journal of American Water Works Association* 103 (2): 76-89.

Cressie, N. 1998. Aggregation and interaction issues in statistical modeling of spatiotemporal processes. *Geoderma* 85 (2-3): 133-40.

Crowder, K., and S. J. South. 2011. Spatial and temporal dimensions of neighborhood effects on high school graduation. *Social Science Research* 40 (1): 87–106.

Domene, E., D. Saurí, and M. Parés. 2005. Urbanization and sustainable resource use: The case of garden watering in the metropolitan region of Barcelona. *Urban Geography* 26 (6): 520–35.

Domene, E., and D. Saurí. 2006. Urbanisation and water consumption: Influencing factors in the metropolitan region of Barcelona.” *Urban Studies* 43(9):1605–23.

Durlauf, S. N. 2004. Neighborhood effects. In *Handbook of Regional and Urban Economics Volume 4*, eds. J. V. Henderson and J.-F. Thisse, 2173-2242. Amsterdam: North Holland.

Enders, C. K., and D. Tofighi. 2007. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological methods* 12 (2): 121–38.

Entwisle, B. 2007. Putting people into place. *Demography* 44 (4): 687–703.

Fotheringham, A. S., and C. Brunson. 1999. Local forms of spatial analysis. *Geographical Analysis* 31 (4): 340–58.

Gato, S., N. Jayasuriya, and P. Roberts. 2007. Temperature and rainfall thresholds for base use urban water demand modeling. *Journal of Hydrology* 337 (3-4): 364–376.

Gelman, A., and J. Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gleick, P. H. 2003. Water use. *Annual Review of Environment Resources* 28:275–314.

Gober, P., K. L. Larson, R. Quay, and C. Polsky. 2013. Why land planners and water managers don't talk to one another and why they should! *Society & Natural Resources* 26 (3): 356–364.

Goldstein, H. 2003. *Multilevel statistical models*. London: Oxford University Press.

Goodman, M. K., D. Goodman, and M. R. Redclift. 2010. Situating consumption, place, and space. In *Consuming space: Placing consumption in perspective*. Working Paper Series. London: Department of Geography, King's College London.

Grafton, R. Q., M. B. Ward, H. To, and T. Kompas. 2011. Determinants of residential water consumption: Evidence and analysis from a 10-country household survey. *Water Resources Research* 47 (W08537) doi:10.1029/2010WR009685.

Guhathakurta, S., and P. Gober. 2010. Residential land use, the urban heat island, and water use in Phoenix: A path analysis. *Journal of Planning Education and Research* 30 (1): 40–51.

Halper, E. B., C. A. Scott, and S. R. Yool. 2012. Correlating vegetation, water use and surface temperature in a semi-arid city: A multi-scale analysis of the impacts of irrigation by single-family residences. *Geographical Analysis* 44 (3): 235–257.

Hamlet, A. F. 2011. Assessing water resources adaptive capacity to climate change impacts in the Pacific Northwest region of North America. *Hydrology and Earth System Sciences* 15 (5): 1427–1443.

Hanak, E., and M. Davis. 2006. *Lawns and water demand in California*. California Economic Policy. San Francisco: Public Policy Institute of California.

Hill, T., and C. Polsky. 2007. Suburbanization and drought: A mixed methods vulnerability assessment in rainy Massachusetts. *Environmental Hazards* 7 (4): 291–301.

Hof, A., and T. Schmitt. 2011. Urban and tourist land use patterns and water consumption: Evidence from Mallorca, Balearic Islands. *Land Use Policy* 28 (4): 792–804.

House-Peters, L., B. Pratt, and H. Chang. 2010. Effects of urban spatial structure, sociodemographics, and climate on residential water consumption in Hillsboro, Oregon. *JAWRA Journal of the American Water Resources Association* 46 (3): 461–472.

House-Peters, L., and H. Chang. 2011a. Urban water demand modeling: Review of concepts, methods, and organizing principles. *Water Resources Research* 47 (5): 1–15.

———. 2011b. Modeling the impact of land use and climate change on neighborhood-scale evaporation and nighttime cooling: A surface energy balance approach. *Landscape and Urban Planning* 103 (2): 139–155.

Howe, C. W., and F. P. Linaweaver. 1967. The impact of price on residential water demand and its relation to system design and price structure. *Water Resources Research* 3 (1): 13–32.

The Intertwine. 2012. *Regional conservation strategy for the Greater Portland-Vancouver region*. Available at <http://theintertwine.org/Conservation> (last accessed 11 November 2013).

Johnston, R., K. Jones, S. M. Burgess, C. Propper, R. Sarker, and A. Bolster. 2004. Scale, factor analyses, and neighborhood effects. *Geographical Analysis* 36 (4): 350–68.

Johnston, R., C. Propper, R. Sarker, K. Jones, A. Bolster, and S. Burgess. 2005. Neighbourhood social capital and neighbourhood effects. *Environment and Planning A* 37 (8): 1443–59.

Jones, K. 1991. Specifying and estimating multi-level models for geographical research. *Transactions of the Institute of British Geographers* 16 (2): 148–59.

Jones, K., and C. Duncan. 1995. Individuals and their ecologies: Analysing the geography of chronic illness within a multilevel modelling framework. *Health & Place* 1 (1): 27–40.

Kenney, D., R. A. Klein, and M. P. Clark. 2004. Use and effectiveness of municipal water restrictions during drought in Colorado. *JAWRA Journal of the American Water Resources Association* 40 (1), 77-87.

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

Kreft, I., and J. de Leeuw. 1998. *Introducing multilevel modeling*. Thousand Oaks: Sage.

- Kwan, M., and T. Schwanen. 2009. Quantitative revolution 2: The critical (re)turn. *Professional Geographer* 61 (3): 283–291.
- Kwan, M. 2012. The uncertain geographic context problem. *Annals of the Association of American Geographers* 102 (5): 958–968.
- Larson, K. L., C. Polsky, P. Gober, H. Chang, and V. Shandas. 2013. Vulnerability of water systems to the effects of climate change and urbanization: A comparison of Phoenix, Arizona and Portland, Oregon (USA). *Environmental Management* 52 (1), 179–95.
- Liu, J., T. Dietz, S. R. Carpenter, C. Folke, E. Moran, A. N. Pell, P. Deadman, T. Kratz, J. Lubchenco, E. Ostrom, Z. Ouyang, W. Provencher, C. L. Redman, S. H. Schneider, and W. W. Taylor. Coupled human and natural systems. *Science* 317 (5844): 1513–1516.
- López-Carr, D., J. Davis, M. M. Jankowska, L. Grant, A. C. López-Carr, and M. Clark. 2012. Space versus place in complex human–natural systems: Spatial and multi-level models of tropical land use and cover change (LUCC) in Guatemala. *Ecological Modelling* 229: 64–75.
- Lyon, B., N. Christie-Blick, and Y. Gluzberg. 2005. Water shortages, development, and drought in Rockland County, New York. *JAWRA Journal of the American Water Resources Association* 41 (6): 1457–1469.
- Maidmont, D., S. -P. Miaou, and M. Crawford. 1985. Transfer function models of daily urban water use. *Water Resource Research* 21 (4): 425–432.
- Maidmont, D. R. and S. -P. Miaou. 1986. Daily water use in nine cities. *Water Resources Research* 22 (6): 845–851.
- Manski, C. F. 1993. Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies* 60 (3): 531–42.
- March, H., and D. Saurí. 2009. What lies behind domestic water use? A review essay on the drivers of domestic water consumption. *Boletín de la Asociación de Geógrafos Españoles* 50: 297–314.

- . 2010. The suburbanization of water scarcity in the Barcelona metropolitan region: Sociodemographic and urban changes influencing domestic water consumption. *Professional Geographer* 62 (1): 32–45.
- Mayer, P. W. and W. B. DeOreo. 1999. *Residential end uses of water*. Denver, Colorado: American Water Works Association Research Foundation.
- Medd, W., and H. Chappells. 2007. Drought, demand and the scale of resilience: Challenges for interdisciplinarity in practice. *Interdisciplinary Science Reviews* 32 (3): 233–249.
- Miaou, S.-P. 1990. A class of time series urban water demand models with nonlinear climate effects. *Water Resources Research* 26 (2): 168–78.
- Middel, A., A. Brazel, P. Gober, S. W. Myint, H. Chang, and J.-D. Duh. 2011. Land cover, climate, and the summer surface energy balance in Phoenix, AZ, and Portland, OR. *International Journal of Climatology* 32 (13): 2020–2032.
- Mieno, T., and J. B. Braden. 2011. Residential demand for water in the Chicago metropolitan area. *JAWRA Journal of the American Water Resources Association* 47 (4): 713–23.
- Miller, K., and D. Yates. 2006. *Climate change and water resources: A primer for municipal water providers*. Denver, Colorado: American Water Works Association Research Foundation.
- Mote, P., and E. Salathé. 2010. Future climate in the Pacific Northwest. *Climatic Change* 102 (1-2): 29–50.
- Muthén, B., and A. Satorra. 1995. Complex sample data in structural equation modeling. In *Sociological methodology*, ed. P. V. Marsden, 267–316. Oxford: Blackwell.
- Nakagawa, S., and H. Schielzeth. 2013. A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 4 (2): 133–142.
- Norman, J., H. MacLean, and C. Kennedy. 2006. Comparing high and low residential density: Life-cycle analysis of energy use and greenhouse gas emissions. *Journal of Urban Planning and Development* 132 (1): 10–21.
- Openshaw, S. 1984. *The modifiable areal unit problem*. Norwich, England: GeoBooks, Regency House.

Palmer, R. N., and M. Hahn. 2002. *The impact of climate change on Portland's water supply: An investigation of potential hydrologic and management impacts on the Bull Run system*. Report for the Portland Water Bureau. Seattle: Department of Civil and Environmental Engineering, University of Washington.

Palmer, R. N., E. Clancy, N. T. Vanrheenen, and M. W. Wiley. 2004. *The impacts of climate change on the Tualatin River Basin water supply: An investigation into projected hydrologic and management impacts*. Report for Clean Water Services. Seattle: Department of Civil and Environmental Engineering, University of Washington.

Perveen, S., and L. A. James. 2010. Scale invariance of water stress and scarcity indicators: Facilitating cross-scale comparisons of water resources vulnerability. *Applied Geography* 31 (1): 321–328.

Polebitski, A. S., and R. N. Palmer. 2010. Seasonal residential water demand forecasting for census tracts. *Journal of Water Resources Planning and Management* 136 (1): 27–37.

Polebitski, A. S., R. N. Palmer, and P. Waddell. 2011. Evaluating water demands under climate change and transitions in the urban environment. *Journal of Water Resources Planning and Management* 137 (3): 249–257.

Portland Water Bureau. 2010. *Water management and conservation plan for the City of Portland, Oregon*. Available at <http://www.portlandoregon.gov/water/46238> (last accessed November 8, 2013).

Praskievicz, S., and H. Chang. 2009. Identifying the relationships between urban water consumption and weather variables in Seoul, Korea. *Physical Geography* 30 (4): 324–337.

Protopapas, A. L., S. Katchamart, and A. Platonova. 2000. Weather effects on daily water use in New York City. *Journal of Hydrologic Engineering* 5 (3): 332–38.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www.R-project.org/> (last accessed 11 November 2013).

Ramachandran, M., and R. Johnson. 2011. *Quantitative restrictions and residential water demand: A spatial analysis of neighborhood effects*. Working Paper 2011-19. Clark University, George Perkins Marsh Institute.

Redman, C. L., J. M. Grove, and L. H. Kuby. 2004. Integrating social science into the long-term ecological research (LTER) network: Social dimensions of ecological change and ecological dimensions of social change. *Ecosystems* 7: 161–171.

- Renwick, M. E., and S. O. Archibald. 1998. Demand side management policies for residential water use: Who bears the conservation burden? *Land Economics* 74 (3): 343–359.
- Renwick, M., and R. Green. 2000. Do residential water demand side management policies measure up? An analysis of eight California water agencies. *Journal of Environmental Economics and Management* 40 (1): 37–55.
- Rinaudo, J., N. Neverre, and M. Montginoul. 2012. Simulating the impact of pricing policies on residential water demand: A Southern France case study. *Water Resources Management* 26 (7): 2057–2068.
- Robbins, P. 2007. *Lawn people: How grasses, weeds, and chemicals make us who we are*. Philadelphia: Temple University Press.
- Roberts, J. K., and J. P. Monaco. 2011. Explained variance in multilevel models. In *Handbook of advanced multilevel analysis*, eds. J. J. Hox and J. K. Roberts, 219–230. New York: Routledge.
- Robinson, W. S. 1950. Ecological correlations and the behavior of individuals. *American Sociological Review* 15 (3): 351–357.
- Rockaway, T., P. Coomes, J. Rivard, and B. Kornstein. 2011. Residential water use trends in North America. *Journal of American Water Works Association* 103 (2): 76–89.
- Saurí, D. 2003. Lights and shadows of urban water demand management: The case of the metropolitan region of Barcelona. *European Planning Studies* 11 (3): 229–243.
- Scherbaum, C. A., and J. M. Ferrerter. 2009. Estimating statistical power and required sample sizes for organizational research using multilevel modeling. *Organizational Research Methods* 12 (2): 347–367.
- Shandas, V., and G. H. Parandvash. 2010. Integrating urban form and demographics in water-demand management: An empirical case study of Portland, Oregon. *Environment and Planning B: Planning and Design* 37 (1): 112–128.
- Singer, J. D., and J. B. Willett. *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford: Oxford University Press.
- Snijders, T. A. B. 2005. Power and sample size in multilevel linear models. In *Encyclopedia of Statistics in Behavioral Science Volume 3*, eds. B.S. Everitt and D.C. Howell, 1570–1573. Chichester: Wiley.

Snijders, T. A. B., and R. J. Bosker, R. J. 2012. *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London: Sage.

Subramanian, S. V., K. Jones, A. Kaddour, and N. Krieger. 2009. Revisiting Robinson: The perils of individualistic and ecologic fallacy. *International journal of epidemiology* 38 (2): 342–60.

Tanverakul, S. A., and Lee, J., 2012. Historical review of US residential water demand. In *World Environmental and Water Resources Congress 2012: Crossing Boundaries*, 3122–3136. doi: 10.1061/9780784412312.313.

Tremblay, A., and J. Ransijn. 2013. Package ‘LMERConvenienceFunctions.’ Available at <http://cran.r-project.org/web/packages/LMERConvenienceFunctions/index.html> (last accessed 11 November 2013).

Tualatin Valley Water District. 2005. *Water Supply Shortage Plan*. Available at http://www.tvwd.org/media/31436/water_supply_shortage_plan.pdf (last accessed 11 November 2013).

———. 2012. *2012 Annual Report*. Available at http://www.tvwd.org/media/2836/annual_report.pdf (last accessed 11 November 2013).

———. 2013. *Technical Memorandum 6 – Evaluation of Supply Options*. Available at <http://www.tvwd.org/media/152813/tm6%20evaluation%20-%20final.pdf> (last accessed 11 November 2013).

Turner, V. K., and D. C. Ibes. 2011. The impact of homeowners associations on residential water demand management in Phoenix, Arizona. *Urban Geography* 32 (8): 1167–1188.

Vano, J. A., N. Voisin, L. Cuo, A. F. Hamlet, M. McGuire Elsner, R. N. Palmer, A. Polebitski, and D. P. Lettenmaier. 2010. Climate change impacts on water management in the Puget Sound region, Washington State, USA. *Climatic Change* 102 (1-2): 261-286.

Wentz, E. A., and P. Gober. 2007. Determinants of water consumption for the City of Phoenix, Arizona. *Water Resources Management* 21 (11): 1849-1863.

Wong, J. S., Q. Zhang, and Y. D. Chen. 2010. Statistical modeling of daily urban water consumption in Hong Kong: Trend, changing patterns, and forecast. *Water Resources Research* 46 (W03506) doi:10.1029/2009WR008147.

Worthington, A. C., and M. Hoffman. 2008. An empirical survey of residential water demand modelling. *Journal of Economic Surveys* 22 (5): 842–71.

Zhou, S., T. McMahon, and Q. Wang. 2001. Frequency analysis of water consumption for metropolitan area of Melbourne. *Journal of Hydrology* 247 (1-2): 72-84.

Zuur, A. F., E. N. Ieno, N. Walker, A. A. Saveliev, and G. M. Smith. 2009. *Mixed effects models and extensions in ecology with R*. New York: Springer.