Title
Planning for water efficient cities: Landscape, microclimate, and heterogeneity in residential water demand

Author
Lassiter, Allison

Publication Date
2015

Peer reviewed|Thesis/dissertation

Planning for water efficient cities:
Landscape, microclimate, and heterogeneity in residential water demand

by

Allison Blythe Lassiter

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Landscape Architecture and Environmental Planning

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor John Radke, Chair
Professor Louise Mozingo
Professor Michael Hanemann

Spring 2015

Planning for water efficient cities: Landscape, microclimate, and heterogeneity in residential water demand

by Allison Blythe Lassiter

Abstract

Planning for water efficient cities: Landscape, microclimate, and residential water demand

by

Allison Blythe Lassiter

Doctor of Philosophy in Landscape Architecture and Environmental Planning

University of California, Berkeley

Professor John Radke, Chair

California is confronting its largest drought in recorded history, which may signal the onset of a megadrought. An executive order mandates reduction in residential water consumption. At the same time, the state's population continues to grow. Reducing residential outdoor water use is a critical management objective to adapt to scarce water resources. This dissertation asks, to what degree can land planning contribute to outdoor water demand management? Can principles of compact development and strategic growth mitigate use? First, the dissertation reviews the literature of water demand models that incorporate landscape variables, thematically characterizing the variables and examining model spatiotemporal resolution. Next, the dissertation examines methods of quantitatively characterizing the urban environment using land cover, weather, and landform variables. It develops variables by parcel, and then defines microclimate zones of similar parcels by clustering on each parcel's microclimate signature. In the fourth chapter, it examines the contributions of landscape and microclimate to household water consumption in the East Bay Municipal Utility District of California. It evaluates 26 million observations of monthly water use data from 2005-2011 recorded at over 300,000 single family residences. It analyzes the data as a whole, and then subsets the full population by quantile of water user and by microclimate zone. Results reveal heterogeneous water demand profiles, but all models indicate that landscape type and the presence of a pool are important predictors of consumption. Less important are lot size and microclimate variables. With both high and low water users spatially distributed throughout the study area and across microclimate zones, there is evidence that many different development styles in many different locations can be water efficient.

Table of Contents

Acknowledgments

This dissertation is an outcome of the combined support from many different people and institutions.

Thank you to my advisors and mentors at Berkeley. Thank you, John Radke, for encouraging me to think critically about patterns, boundaries, and surfaces. I am especially indebted to you for allowing me to camp out on your computer over the past semester. Thank you, Michael Hanemann, for leading me into the world of water demand—from econometrics to policy—and continuing to try (and try, and try) to help become wiser, not just more knowledgeable. Thank you, Louise Mozingo, for helping me to straddle computation and landscape theory, and for allowing me to drop into your office as often as I did.

Without the East Bay Municipal Utility District, this research never would have happened. Thank you, Richard Harris, for going back and forth with Legal, and finding a way to share your data. Thank you, Charles Bohlig, for patiently and quickly answering all my questions with an unending sense of humor. Thank you, Colleen Swain, for skillfully assembling the complicated database that was the basis of this work.

Thank you to the community of PhD students in LAEP. In particular, thank you to Sara Carr for staying one step ahead and giving me someone to follow—I missed you when you left.

Finally, thank you to my family for your unending encouragement. I am so lucky to have you.

Chapter 1. Introduction: Reconciling water scarcity and urban growth

Regions across the globe are contending with municipal water demands that exceed traditional supplies, making it necessary to re-think water consumption habits and management practices. California is one of these places. Faced with expanding urban regions and water scarcity, California's water providers are required to reduce per capita consumption by at least 25 percent before March 1, 2016 (Brown 2015). As water providers grapple with conservation mechanisms, of particular interest is outdoor water, the single largest discretionary residential water use.

Residential outdoor water includes water used for landscaping, pools, and activities like car washing—all water consumed outside the house itself. On average outdoor water use comprises 50 percent of total household consumption (Brown 2015), though there is tremendous variation across the state. Reports of household use range from approximately 35 gallons per capita per day (gpcd) to over 600 gpcd (State Water Resources Control Board 2015a), with the differences largely attributed to outdoor water. In some regions, households are reported to use as much as 80 percent of water outside (State Water Resources Control Board 2015b).

As water agencies engage with questions of outdoor water, they underscore connections between water consumption, landscape, and land planning. Are there patterns of development that are more or less water efficient? To what extent are planners making decisions on water when designing or approving development? This dissertation examines relationships between land and water through quantitative spatial and statistical models, using high resolution data to reveal insights on household water consumption, with implications for future growth.

Unreliable supplies and increasing demand

The connection between land and water is receiving increasing amounts of attention in both practice and the literature. In a survey of land planners and water managers, Gober et al writes (2013, 360) "both groups acknowledged the importance of water issues for land planning and land planning for water issues." Furthermore, both land planners and water managers believe outdoor water use and climate variability are critical concerns in municipal water reliability (Gober et al. 2013). These themes of climate variability and growing outdoor water demands are developed below, with specific focus on the context of California.

Climate change and drought

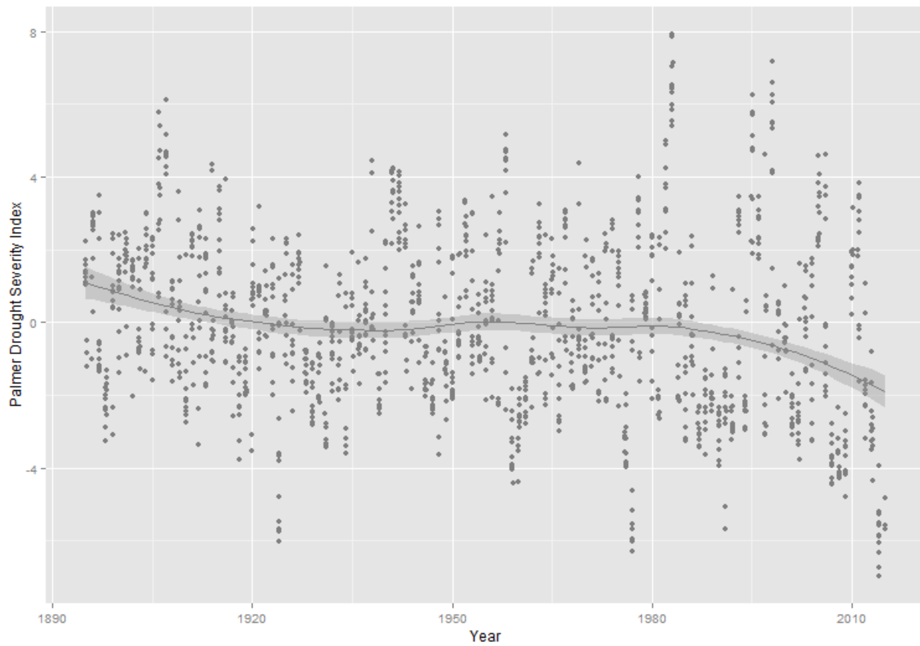
California has lived with water scarcity through its history, but the drought beginning in 2012 is its most severe on record. It may be the most severe in 1200 years (Ault et al. 2014). There are frequent news headlines concerning how today's weather deviates from "normal" and marks extremes compared to past records: January 2015 was the first year in 165 years with no recorded rain in San Francisco (Golden Gate Weather Services 2015); the Sierra Nevada snowpack, as measured on April 2015, was just 5 percent of normal (California Department of Water Resources 2015); calendar year 2014 was the state's warmest on record (National Climatic Data Center 2015). Normal is typically defined against California's formally recorded weather data, beginning in 1849, but some suggest that the climate is shifting—that the next 165 years will be different than the last.

Searching for normal in California weather may be impossible. All evidence indicates California's climate is dynamic and ever-changing. There is annual variation, marked by a rainy winter and dry summer. There is decadal variation, driven by La Niña cooling and El Niño warming of the east-central equatorial Pacific Ocean. There are centennial-scale cycles of change. Paleoclimatogists, studying historic sediments and tree core samples, found evidence of a substantial medieval drought that descended on California from approximately 900-1400, consisting of two, 200-year megadroughts broken by a wet period from 1100-1200 (Ingram and Malamud-Roam 2013). The transition between wet and dry eras can be rapid and unexpected. Modern California most likely developed during a wet era. Now, there is concern whether the state is headed for a dry turn. Recent literature indicates the Western United States may be experiencing the onset of another profound and protracted megadrought (Cook, Ault, and Smerdon 2015).

One method of examining modern climate trends is through the Palmer Drought Severity Index (PDSI). The PDSI combines latent soil moisture, potential evapotranspiration, and rainfall to calculate a monthly soil water balance. The water balance is scored relative to local, average conditions to define a consistent metric that can be used to compare drought across space and time. Monthly PDSI values have limitations,[1] but are useful indicators of the fluctuations in California's weather.

Figure 1 charts PDSI scores from January 1895 to December 2014. In aggregate, PDSI values are trending negative, indicating that California is become drier. Examining the wet and dry years separately, the trends diverge—the wet years are wetter, while the dry years are drier. This indicates that California's climate is drier in aggregate, but is becoming more variable and extreme. PDSI findings adhere to current climate science, which indicates California's future is likely to include deep, dry spells marked by fewer, more intense storms (Andrew 2015).

[1] The PDSI is critiqued for a number of reasons, including: the "arbitrary designation of drought severity classes" (Alley 1984, 1105); whether or not the evapotranspiration calculations truly represent on–the-ground conditions (Alley 1984); and the computational complexity and opacity of the index (Jacobi et al. 2013). Nonetheless, PDSI data is distributed by the National Climactic Data Center, widely referenced, and used as a tool by the federal government in drought-relief programs.



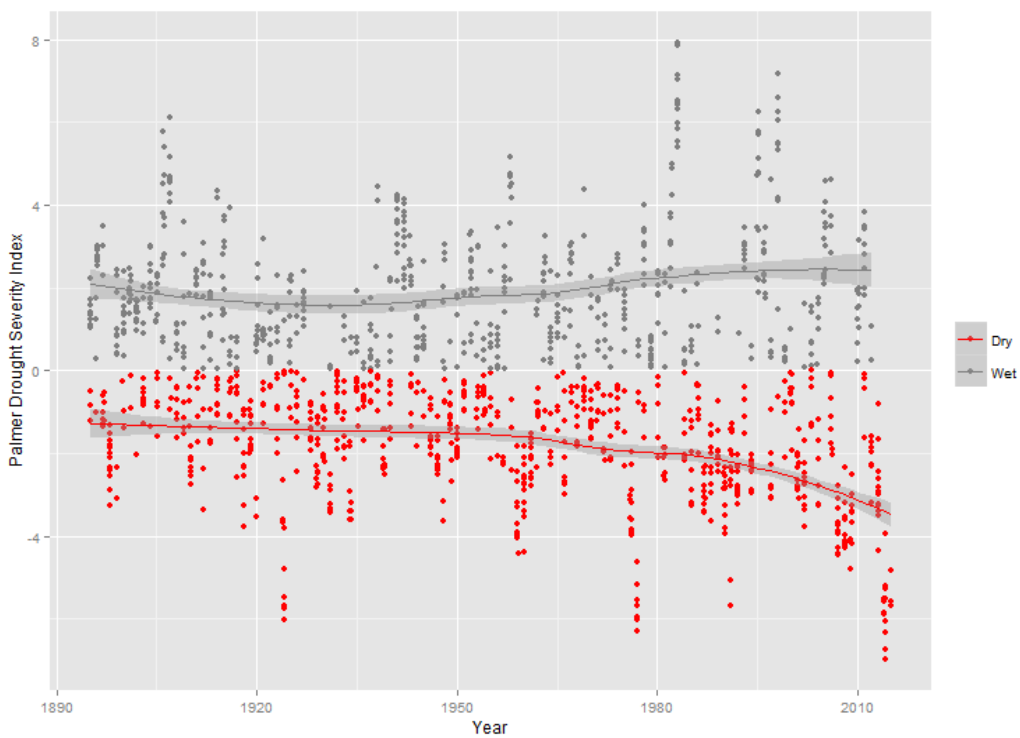


Figure 1. The PDSI values in California range from approximately -6 to 7. Negative values indicate a more severe drought, while positive values indicate very wet conditions. Note that PDSI values can change rapidly. A dry month may end quickly, if there is a large amount of rain. As a result, the data appear noisy. The top chart indicates PDSI values are trending drier. The bottom chart, with negative and positive values fitted separately, show the curves diverge. In 2014, 9 out of 12 months were the driest ever recorded.

Urban population growth

As of 2015, there are more than 38 million people living in California (California Department of Finance 2014a). Though rates of growth are slowing, the population continues to expand nearly 1% per year (California Department of Finance 2014b). This amounts to large numbers of people joining the state: 3.5 million additional people in the next ten years; 6.9 million more in twenty years. Forty years from now, the state population is expected to exceed 50 million people. Many of the state's newest residents are locating inland, in California's hottest and driest regions (see Figure 2).

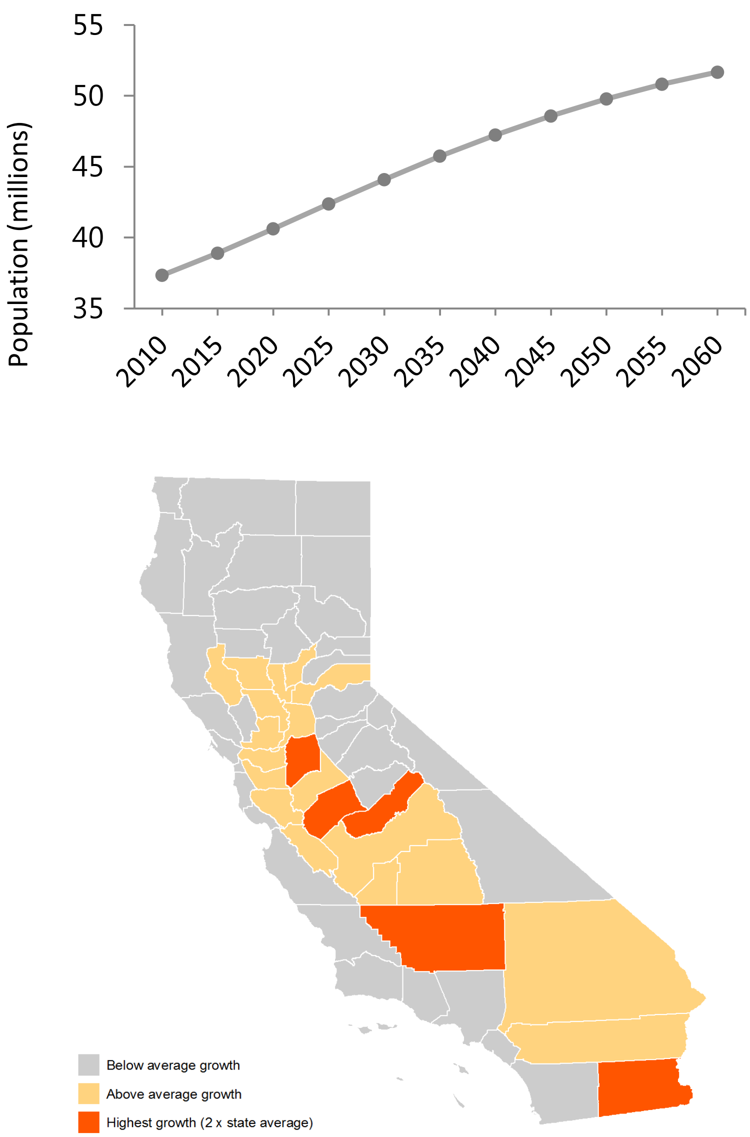


Figure 2. California's population growth. While coastal areas of the state are relatively built out, the inland areas of the state are growing. Through 2060, the expected average growth across the state is approximately 0.8%. Those highlighted in dark are expected to average are over 1.6%. The fastest growing county will like be Kern County, at 2.2%. Source: California Department of Finance 2014b.

The regions with the highest growth rates include many of the water districts reporting the highest per capita use. California water experts Christian-Smith and Heberger (2015, 108) write, "When it comes to water, it doesn't just matter how much we grow, it also matters where and how we grow." But, exactly how much does the environment surrounding a house matter? What types of environmental variation are the largest contributors to residential consumption? Often, it is difficult to tease apart the contributions of local development styles, behaviors, and the environmental setting (see Figure 3).

Figure 3. Two styles of development observed in the East Bay Municipal Utility District. Is one form of growth more water-consumptive than the other?

Reducing demand through efficiency

Reduction in per capita consumption has not always been a policy objective of the state. The following section documents the rise of water efficiency in California and its more recently identified connections to land planning.

The rise of water efficiency in California

The ethos of importing water to support population growth and economic development is as old as the state. Large irrigation projects appropriating water from afar were championed by California's first State Engineer, William Hammond Hall. In an 1889 address to the National Geographic Society he said (1889, 14):

> The interior valley of California will not support, without irrigation, an average of more than 15 to 20 people per square mile. Irrigate it and it will support as many as any other portion of the country – reasonably it will support 200 to the square mile. I have no doubt that the population will run up to ten or twelve millions in that one valley, and there are regions over the is country from the Mississippi to the Pacific, millions of acres, that can be made to support a teeming population by the artificial application of water.

While the early 1900s was characterized by city-scale surface water storage projects, including the Los Angeles aqueduct and San Francisco's Hetch Hetchy reservoir, a plan to implement Hall's vision took shape in the first official *California Water Plan* in 1930. The plan says "[t]here are large inequities in both the geographic and seasonal distribution of the state's water resources, as related to the demands of various purposes" and suggests moving water from "areas of water surplus to areas with insufficient local water to supplies to meet their demands" (Division of Water Resources 1930, 23). Hall's vision was realized by the middle of the 20th century, in what became known as the "Hydraulic Era" (Worster 1985), when extensive, state-scale projects were built (see Figure 5). By the end of the century, the state was woven with aqueducts, canals, levees, reservoirs, and pumps (see Figure 4).

Figure 4. Select waterworks projects in California and naturally occurring water sources. Adapted from (Lassiter 2015).

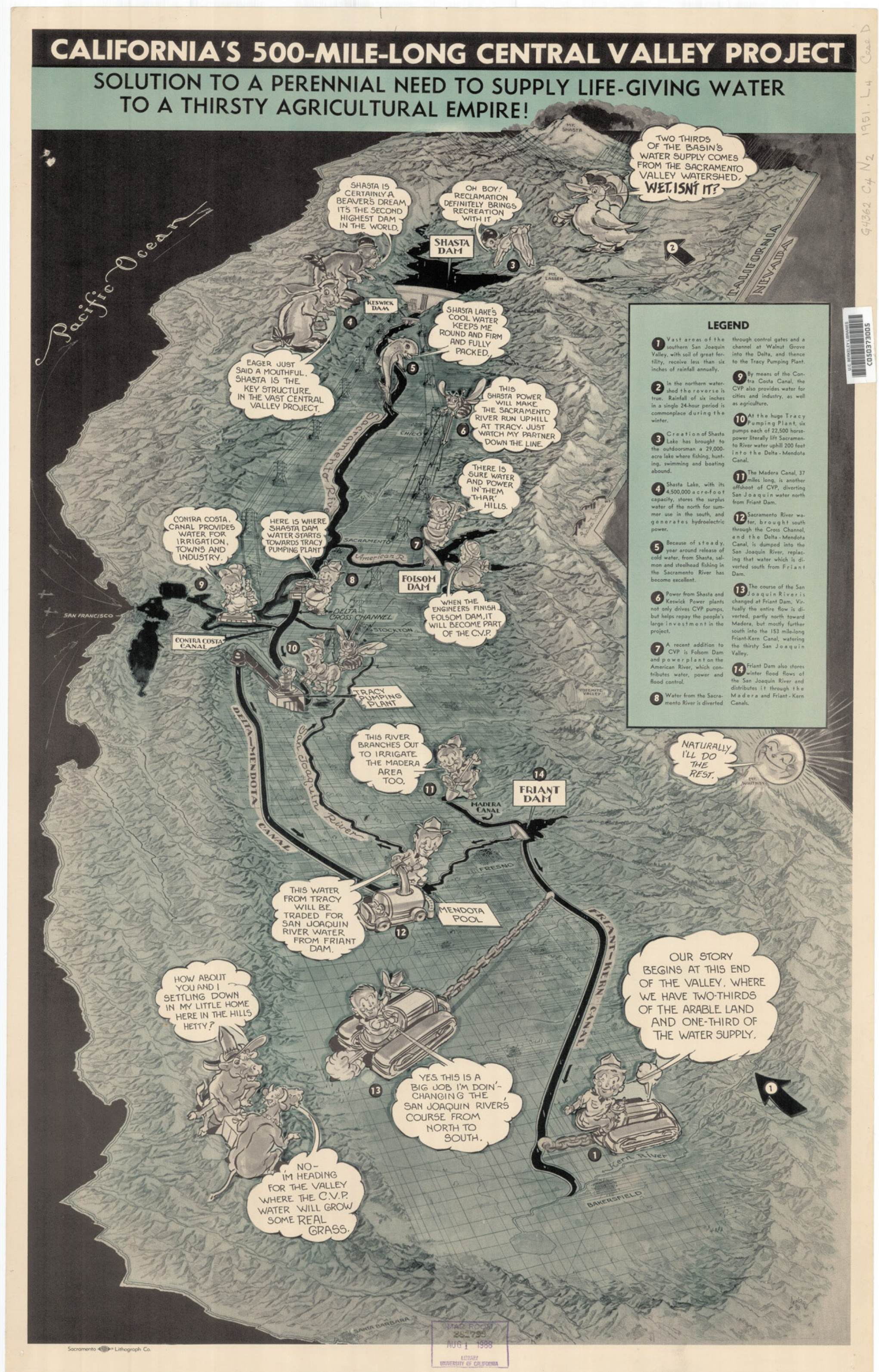


Figure 5. In 1951, a poster supporting the newly proposed Central Valley Project. The poster highlights the engineering feat of the project: "Yes, this is a big job I'm doin' – changing the San Joaquin river's course from North to South." Source: Lenoir 1951. *Courtesy of the Earth Sciences and Map Library, University of California, Berkeley.*

Despite ongoing waterworks construction, popular opinion of water appropriation began to change through the 1960s and 1970s, alongside major environmental lawsuits in Mono Lake and the Owens Valley. When the Peripheral Canal, the newest of the large waterworks projects, was proposed in 1982, it was defeated by voters. Following its defeat, talk of water demand management took hold in the state legislature and the Department of Water Resources. In 1984, the Urban Water Management Planning Act was passed, requiring urban water districts to submit reports to the state that rationalized demands against supplies.

Focus on water demand management and efficiency continued to grow. In 1993, the *California Water Plan* update claimed it "differs from its predecessors by...recognizing and presenting water demand management methods..." (Department of Water Resources 1993, xii). In the 2005 and 2009 updates, DWR shifted more responsibility for securing future water supplies from the state to local water agencies, advocating a portfolio approach to water management that included water efficiency. In 2009, Governor Schwarzenegger pushed forward an urban water conservation initiative mandating a twenty percent reduction in per capita urban water use by 2020, enacted through Senate Bill X7-7. Finally, during record-breaking drought in 2015, Governor Jerry Brown issued a first-ever executive order mandating a 25 percent reduction in urban water use across the state (Brown 2015). Now, reducing urban water use is a central management objective.

Land planning for efficiency

Most residential demand management programs focus on mitigating indoor water use by encouraging fixture and appliance upgrades. Those emphasizing outdoor water use often focus on lawn replacement, as water agencies "buy back" turf grass at a set rate. Both rebates and lawn replacement are components of Governor Brown's executive order. The order specifically mandates the replacement of 50 million square feet (1150 acres) of lawn or ornamental turf with drought-tolerant landscaping (Brown 2015).

Some suggest that there are deeper strategies—that it is possible to reduce future per capita water consumption by growing smarter. Recent literature, including the Land Use Resource Management Strategy in the *California Water Plan Update 2013*, suggests compact development as a method of reducing water use. DWR (California Department of Water Resources 2013, 24-20) writes:

> Traditional suburban-style development landscaped with nonindigenous plants creates high water demand for landscaping. As urban development occurs in hotter regions of the state, this pattern of land use and landscaping is projected to increase water use to a higher amount of residential water demand. More compact mixed-use urban development reduces landscaping-related water demand by minimizing front and back yards and their associated landscape water demands.

DWR refers to a report by the Environmental Protection Agency, *Growing Toward More Efficient Water Use: Linking Development, Infrastructure, and Drinking Water Policies* (2006, 7) that states, "Encouraging compact neighborhood design on smaller lots reduces water demand for landscaping."

Similarly, the State Water Resource Control board implicates density and location as factors influencing consumption on the website for residential water consumption reporting (see Figure 6). They warn that temperature and precipitation influence demand, along with local microclimate factors: "Areas with high temperature and low water need to use more water to maintain outdoor landscaping...even in the same water supply district these factors can vary considerably...". They also indicate that irrigated acreage is an influential factor which is related to development style: "high population densities use less water per person...". This indicates that compact development and strategic growth could reduce use.

IMPORTANT NOTE!

It is not appropriate to use Residential Gallons Per Capita Day (R-GPCD) water use data for comparisons across water suppliers, unless all relevant factors are accounted for. Factors that can affect per capita water include:

- ***Rainfall, temperature and evaporation rates*** – Precipitation and temperature varies widely across the state. Areas with high temperature and low rainfall need to use more water to maintain outdoor landscaping. Even within the same hydrologic region or the same water supply district these factors can vary considerably, having a significant effect on the amount of water needed to maintain landscapes.
- ***Population growth*** – As communities grow, new residential dwellings are constructed with more efficient plumbing fixtures, which causes interior water use to decline per person as compared to water use in older communities. Population growth also increases overall demand.
- ***Population density*** – highly urbanized areas with high population densities use less water per person than do more rural or suburban areas since high density dwellings tend to have shared outdoor spaces and there is less landscaped area per person that needs to be irrigated.
- ***Socio-economic measures such as lot size and income*** – Areas with higher incomes generally use more water than areas with low incomes. Larger landscaped residential lots that require more water are often associated with more affluent communities. Additionally, higher income households may be less sensitive to the cost of water, since it represents a smaller portion of household income.
- ***Water prices*** – Water prices can influence demand by providing a monetary incentive for customers to conserve water. Rate structures have been established in many districts to incentivize water conservation, but the effectiveness of these rate structures to deter excessive use and customers sensitivity to water prices vary.

Figure 6. State Water Resource Control Boart includes the above note on the website distributing per capita residential consumption by water agency. Source: State Water Resources Control Board 2015a

Land planning for reduced resource consumption does not have a long history in water, but is pervasive in the energy efficiency literature. Neighborhood design is sufficiently linked to both energy consumption and subsequent greenhouse gas production that the principles are embedded in California's Sustainable Communities and Climate Protection Act of 2008 (SB 375), the companion bill to the Global Warming Solutions Act of 2006 (AB 32). SB 375 encourages dense development near transit as a mechanism for greenhouse gas mitigation.

With water, state policies focus on individual homes, not neighborhood design. California Green Building Standards Code (California Building Standards Commission 2013), to which all of California's new construction must adhere, includes limits on the flow rates of toilets, showerheads, and faucets. It also mandates weather or soil moisture-based irrigation controllers. The Model Water Efficient Landscape Ordinance (AB 1881) further governs outdoor water use by setting a maximum allowable applied outdoor water for at sites requiring building or landscape permits with greater than 2,500 sq ft of landscaped outdoor space if the owner is a public agency or developer, or 5,000 sq ft if it is a private home owner (California Department of Water Resources 2010). These water budgets are calculated with plant-specific irrigation guidelines provided by the state (Costello and Jones 2014).

If neighborhood design and strategic growth truly influence water consumption, however, there may be opportunities to further leverage existing policy to promote water efficient development. Because water and greenhouse gas production are deeply linked (Wilkinson 2015), it is possible AB 32 could be used as a basis for promoting water efficient development. To date, there has not sufficient evidence to do so, but there is opportunity to establish links between neighborhood planning, design and water. To what degree can planning and design influence consumption?

Research framework

This dissertation responds to the themes of water scarcity, population growth, and efficiency through land planning by examining how the location and form of development relate to water consumption. Using disaggregate, parcel-level data, it focuses on heterogeneity in relationships. Do residents respond differently to their household environments? The analysis considers the endogenous relationships between personal behavior characteristics and environmental characteristics.

Research questions

- Does the location of a parcel relate to its water consumption?
    - How does climate relate to water consumption?
    - How does landform relate to water consumption?
- Does the design of a parcel relate to its water consumption?
    - How does density relate to consumption?
    - How does garden style relate to consumption?
- Can land planning contribute to water efficient development?
    - Does microclimate contribute to high water use zones?
    - Is it possible to manage water demand through compact development or strategic growth?

Study area

This study is situated in California's East Bay Municipal Utility District (EBMUD) (see Figure 7). EBMUD's service area is 331 square miles (803 square kilometer), covering 20 incorporated and 15 unincorporated communities. EBMUD provides water to approximately 1.3 million people (East Bay Municipal Utility District 2015). Approximately 1 million are in single family residences.

EBMUD is notable for its diverse microclimates, housing typologies, and population, allowing for a broad range of conditions to be examined within a single service area.

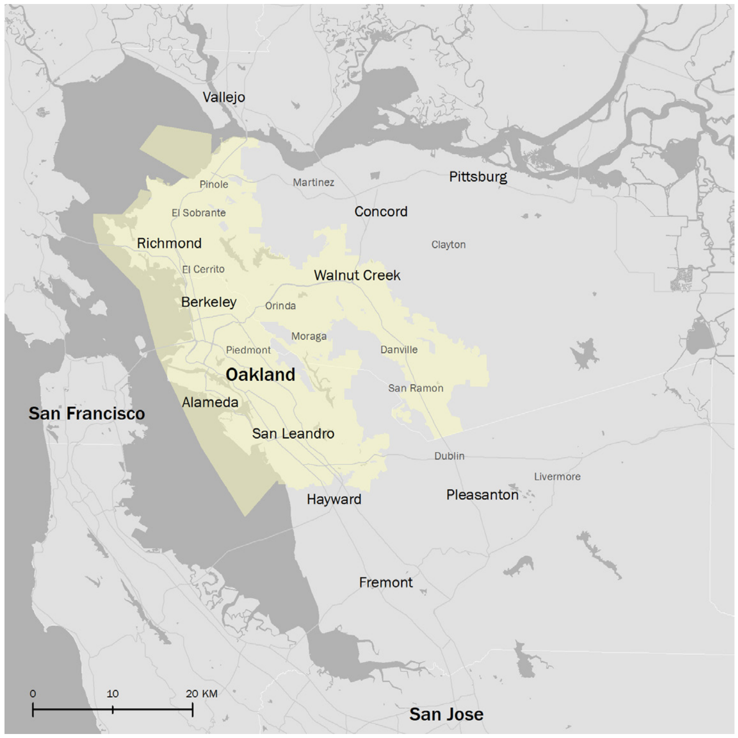


Figure 7. Overview of the EBMUD area. Service boundary denoted in yellow.

Figure 8. The EBMUD service area includes highly urbanized areas , low density development, agriculture, and protected lands. The most urbanized land is visible by its grey hardscape, while the least developed land remain green.

Development patterns
EBMUD includes a dense, urban core on its western border, spanning from Richmond to San Leandro. Development along the eastern side of the study area is less compact as a whole, but includes urban, suburban, and rural development styles (see Figures 7 and 8).

Population growth
The study area intersects two counties. Alameda County is currently growing at approximately 1.5 percent per year. Contra Costa is growing more slowly, at approximately 1 percent per year (California Department of Finance 2014a). Population growth, as it induces new construction, does not directly impact this study because the analysis focuses on households that report water use for the full extent of the study period.

Microclimates
Beyond seasonal variation in climate, there are distinct microclimates in the EBMUD service area. Coarse environmental variation is related to a ridge of hills that extends from northwest of the study area to southeast (see Figure 9). Winds and fogs blow in from the ocean, and then hang on the western side of the hills during the summer months, moderating temperatures. Finer environmental variation is driven by more granular changes in landform, as evidenced in the undeveloped local hillsides (see Figure 10).

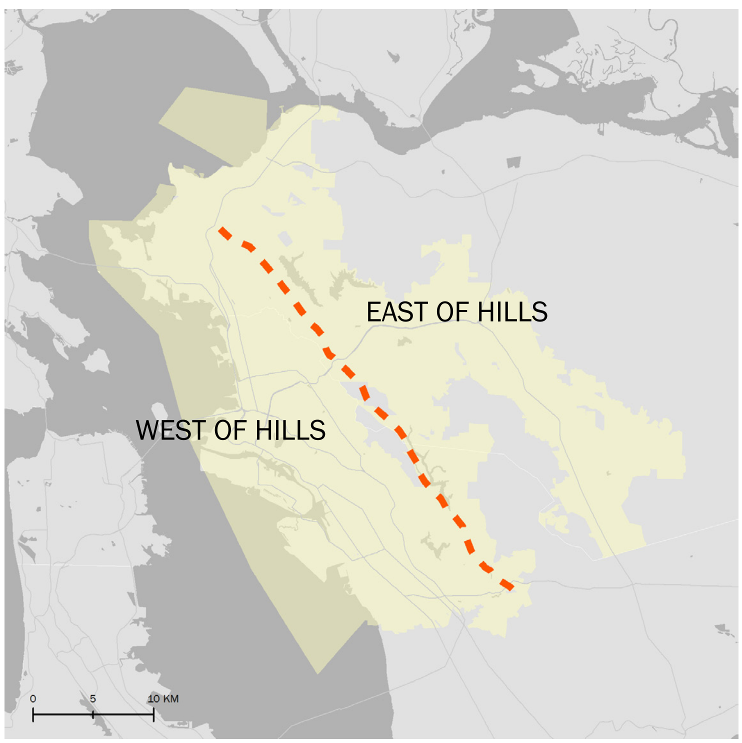


Figure 9. Coarse environmental variation in the study area is related to a ridge of hills extending from the northwest to the southeast of EBMUD. The area to the west of the hills has a more moderate climate than the area to the east of the hills. The east of the hills is cooler in the winter and warmer in the summer.

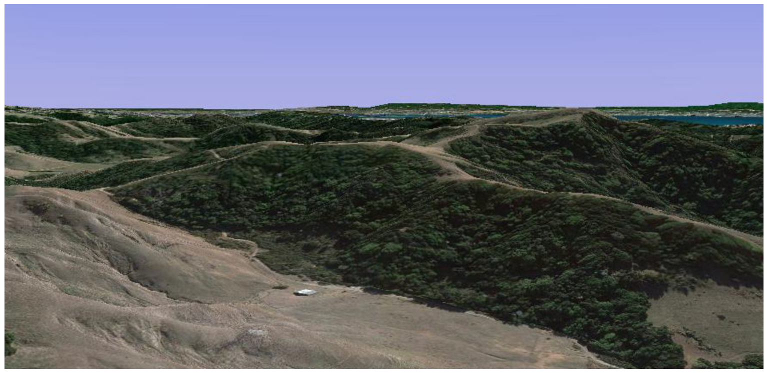


Figure 10. Fine grained environmental variation in the study area is evident in its undeveloped hillsides. In the area pictured above, northerly and easterly slopes are more vegetated than southerly and westerly slopes.

Study period

The study examines monthly water use from 2005 to 2011, 84 months of data. This time period intersects with two major events that likely influenced consumption: the Great Recession of 2007 to 2009; and a significant drought from 2007 to 2009. The drought included conservation campaigns and a fixed drought surcharge to water bills in 2008.

Water use

The study examines all single family residences in EBMUD, approximately 330,000 households, each with an explicit spatial location. All parcels are analyzed individually without aggregation. Using 84 months of data, after data cleaning there are approximately 26 million observations on water use.

During the study period, households consumed an average of 262 gallons per day, or roughly 96 gpcd. Water use exhibits spatiotemporal variability through the study area (see Figure 11). East of the hills exhibits greater change across seasons than the west of the hills.

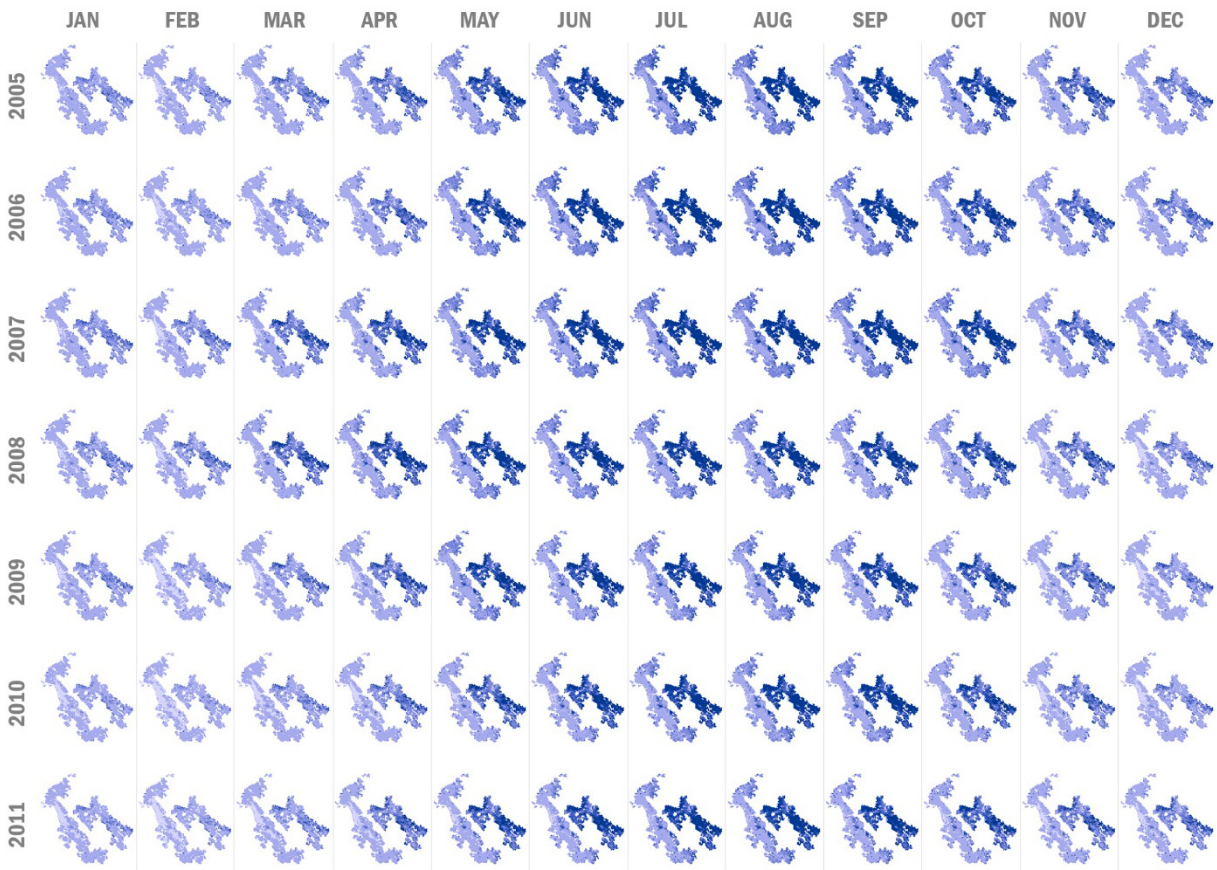


Figure 11. Monthly household water use summarized on a 1 km hexagonal grid to preserve household anonymity. Light blue indicates low water use. Dark blue indicates high water use. There are seasonal effects throughout, but summer use in eastern side of the study area exhibits a larger percentage change than the western side.

Structure of dissertation

The dissertation proceeds in four parts. Chapter 2 reviews the literature of long term residential water demand modeling, giving special focus to water demand models that incorporate land planning characteristics. The chapter characterizes variables commonly implemented in water demand modeling, and then evaluates spatial and temporal resolution. Finally, the chapter assesses gaps in the water demand literature, proposing additional research on policy-relevant variables and disaggregate, high-resolution forecasting.

Chapter 3 examines the challenges of resolving intraurban environmental variation. It begins by discussing data sources and developing microclimate variables at the parcel level. Then, it implements a k-means cluster analysis to characterize microclimates by zone, examining the spatial location of zones when clustering by microclimate characteristics, not location. It identifies several different possible characterizations.

Chapter 4 analyzes 84 months of water data from over 300,000 households in EBMUD. The analysis focuses on the relationship between household water use and local environmental factors – vegetation, temperature, precipitation, and land form. To account for variables acting at different temporal scales, it implements a panel data model with fixed effects, and then regresses the fixed effects on time-invariant variables. First, it examines demand profiles by climate zone. Then, it examines demand profiles for quantiles of water users (low, moderate, and high water users). It asks, to what degree does the physical environment influence demand? Is it possible to identify high water use zones?

Finally, chapter 5 examines urban morphology in the water-efficient city. The chapter suggests opportunities for integrating land planning and water management by broadening the scope of demand management.

Contributions

This dissertation contributes to three bodies of literature:

It advances the water demand literature by analyzing high resolution data on landscape and climate across a large, urban region. It furthers discussion on microclimate, introducing aspect as a model covariate. It examines heterogeneity among households by implementing a quantile regression on disaggregate, parcel-level data.

It contributes to urban climate literature in two respects. Firstly, it develops microclimate characteristics at the parcel-month scale. Though an uncommon scale among climate modelers, it is relevant to socioecological models, since many social data are collected on monthly time step. Secondly, it evaluates urban microclimate zones through a bottom-up clustering processes, allowing regions to emerge from environmental characteristics.

Lastly, it intervenes in ongoing debates in environmental planning concerning efficient urban form with a quantitative, empirical examination of single family residential water demand. The analysis indicates that land cover choices are more important than lot size in determining household water use.

Chapter 2. Integrating water demand management and land planning through models

In California, land planning is legally bound to water supplies through Senate Bill 610, which requires developments with greater than 500 units to show proof of sufficient, assured water rights prior to approval (Waterman 2004). This requires developers to calculate their water needs. How do they do so? The methods vary for each water district. In EBMUD, a single family water budget combines separate calculations for indoor and outdoor use. Indoor use is assumed to be 55 gallons per person per day. Outdoor use varies based on the size of irrigated acreage and the evapotranspiration zone (there are three zones throughout EBMUD, discussed further in chapter 3).

EBMUD's calculations are more fine tuned than many water districts. Even so, it treats all single family growth as roughly equal when determining use, beyond evapotranspiration zone. Yet, there are large differences in observed water consumption, even among households with the same occupancy and the same irrigated acreage (discussed further in chapter 4). Does the form or location of the new development contribute to its water use? If so, is there a more sensitive mechanism for predicting demand?

This chapter reviews the literature of water demand modeling that includes variables related to land planning. First, it examines model resolution and selection, giving special attention to disaggregate models. Next, the chapter synthesizes the literature—crossing the disciplines of economics, geography, and social psychology—by thematically characterizing its variables. Finally, it identifies limitations and possibilities for integrated land-water models.

Frameworks for modeling demand

There are several different disciplines engaged in long term demand modeling. The vast majority of studies come out of economics, seeking to identify demand elasticities under different rate structures (Arbués, García-Valiñas, and Martínez-Espiñeira 2003). Since price is not a universally popular method of mitigating demand (Olmstead and Stavins 2009; Renwick and Green 2000), geographers and social psychologists are beginning to introduce more non-price contributors. Geographers are importing spatial analysis techniques and covariates more common to landscape and urban studies (House-Peters and Chang 2011a). Social psychologists are introducing more nuanced methods of analyzing the attitudes and behaviors behind demand and conservation (Jorgensen, Graymore, and O'Toole 2009; Russell and Fielding 2010). As the disciplinary scope increases, the types of models and factors under consideration are increasing, too.

In all disciplines, models are beginning to better reflect the complexity of urban water systems by including social, infrastructural, and environmental components that interact over space or time (e.g., Arbués, Villanua, and Barberan 2010; Chang, Parandvash, and Shandas 2010; Polebitski and Palmer 2010). Analyses are becoming increasingly data-rich and disaggregate. With heightening model resolution and complexity, researchers are able to more accurately identify many of the contributors to demand.

Model resolution

Disaggregate studies—that is, studies analyzing individual units, instead of pooled data from a whole population or over multiple time steps—are increasingly prevalent. Recent innovations in technology, data availability, and methods for processing larger datasets are making it possible for researchers to reveal intraurban variation in consumption, a better depiction of real-world conditions than metropolitan averages (Troy and Holloway 2004).

Figure 1 maps the temporal and spatial aggregation of recent urban analyses that relate to long term residential water management. Analyses of household behaviors are increasingly common (Hewitt and Hanemann 1995; Arbués, Barberán, and Villanúa 2004; Arbués, Villanúa, and Barberán 2010; Fox, McIntosh, and Jeffrey 2009; Olmstead, Hanemann, and Stavins 2007). There are also a growing number of analyses that include parcel, census block group and tract characteristics (Wentz and Gober 2007; Balling, Gober, and Jones 2008; Lee S.-J. and Wentz E.A. 2008; Chang, Parandvash, and Shandas 2010; Polebitski and Palmer 2010).

Though the number of disaggregate analyses are growing, relative to the whole body of water demand studies they remain rare. Many utilities are reticent to release household information and some states explicitly prohibit utilities from sharing disaggregate data due to privacy concerns (Breyer, Chang, and Parandvash 2012). In Figure 1, studies analyze Portland, Phoenix, Seattle, Texas and Colorado (Wentz and Gober 2007; Balling, Gober, and Jones 2008; Lee S.-J. and Wentz E.A. 2008; Chang, Parandvash, and Shandas 2010; Polebitski and Palmer 2010; Hewitt and Hanemann 1995; Kenney et al. 2008). One study focuses on the southeastern United States and another examines eleven urban areas in the United States and Canada (Sohn 2011; Olmstead, Hanemann, and Stavins 2007). The remaining studies evaluate Australia, Spain and Germany (Jorgensen et al. 2013; Arbués, Villanúa, and Barberán 2010; Schleich and Hillenbrand 2009). Because conclusions from one study area are almost never extended to other locations, the available disaggregate analyses offer insight into few places in the world.

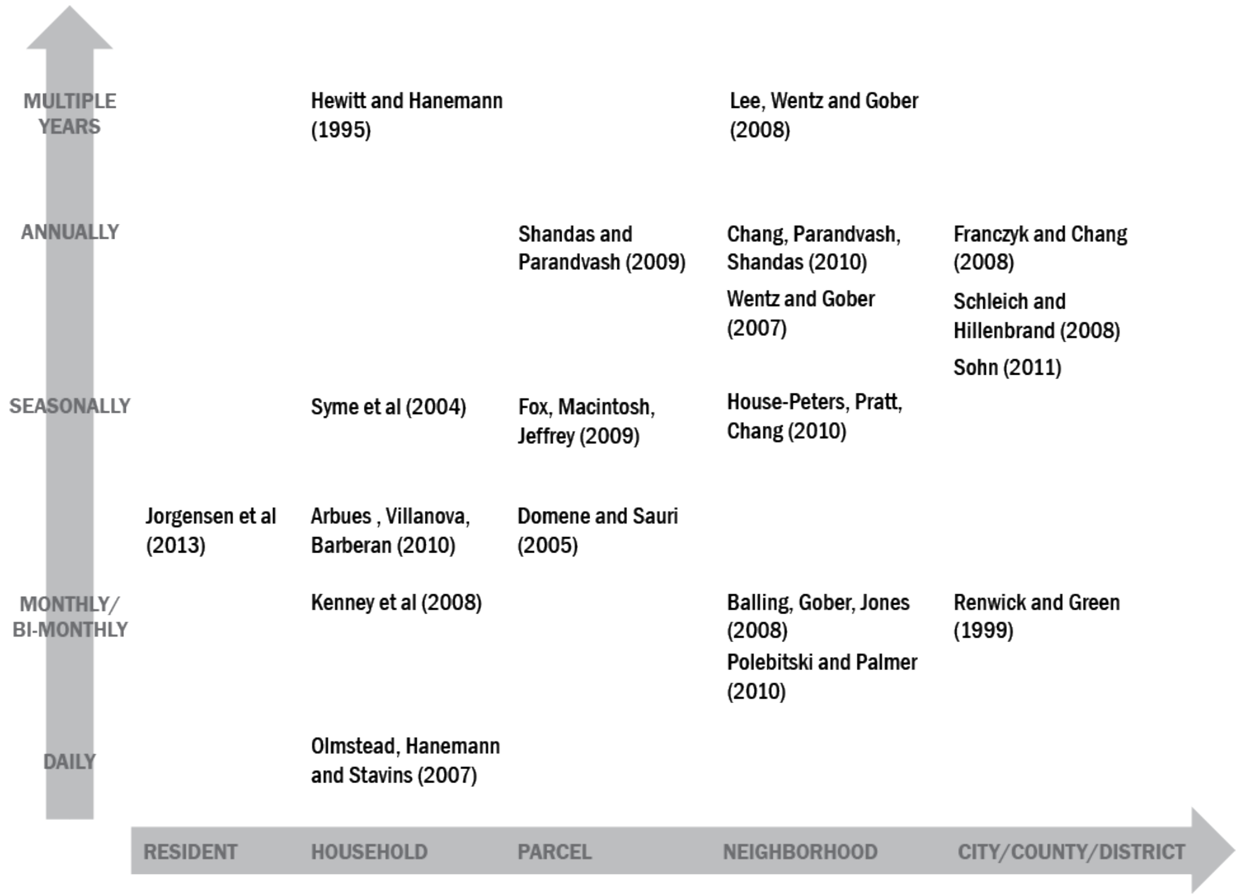


Figure 1. Temporal and spatial aggregations of analyses in the water demand literature. Resident includes analyses that examine individual behavior or attitudes. Household examines aggregate household behavior/attitude and demographic characteristics. Parcel includes analyses are those where the emphasis is on contribution of the built structures, landscape and environment of the parcel, rather than on household behavior or demographic characteristics, like income.

Model selection

Just as studies use different data aggregations, they also employ many different types of models. Researchers most often seek approaches that best suit their data and conceptual framework of the water demand system. In the long term demand literature, approaches are typically stochastic and cross-sectional (Arbués, García-Valiñas, and Martínez-Espiñeira 2003).

Economic studies, seeking to identify demand elasticities, use econometric methods such as instrumental variables and two- and three- stage least squares. To address endogeneity and simultaneity pricing problems of some rate structures, researchers also rely on methods like dynamic two stage models and Continuous Discrete Choice models (Arbués, Villanúa, and Barberán 2010; Hewitt and Hanemann 1995; Olmstead, Hanemann, and Stavins 2007). Panel data approaches are less common due to data limitations, though increasingly available (Arbués, García-Valiñas, and Martínez-Espiñeira 2003; Polebitski and Palmer 2010).

More concerned with spatial patterns in consumption, geographers are introducing regression techniques like spatial error and spatial lag modeling (Chang, Parandvash, and Shandas 2010; House-Peters, Pratt, and Chang 2010; Franczyk and Chang 2009). One study uses Geographically

Weighted Regression to estimate unique regression equations for distinct geographic units, revealing how the contributors to water consumption vary over space (Wentz and Gober 2007). Another uses Principal Components Analysis to identify the effects of a combined set of collinear climate variables (Balling, Gober, and Jones 2008). All techniques seek regression coefficients, indicating the contribution of each covariate to consumption.

Modeling attitudes and perceptions typically follow different frameworks. Social psychologists are introducing Structural Equation Modeling, a technique used to identify causal relationships (Syme et al. 2004; Jorgensen, Graymore, and O'Toole 2009). The outputs from SEM models include path diagrams and path coefficients, which indicates the order of causality and the degree to which each factor contributes to another.

The models from each discipline – economics, geography and social psychology – all have different outputs: demand elasticities, regression coefficients, and causal paths. This makes it difficult to synthesize the results of studies from the different literatures.

Contributors to demand

Across all studies, the variables can be characterized by six major themes: price, demography, behaviors and attitudes, dwelling characteristics, landscaping and outdoor characteristics, and weather or climate. Following is a detailed discussion of findings from these six categories.

Price

Measuring the relationship between the price of water and consumption is the oldest area of study in the economic demand literature. In the past, municipal-level studies gave unclear indications of consumer responses. For many years, people believed that demand was inelastic with respect to price (Howe and Linaweaver 1967). More recent studies agree that "price can play a crucial role in demand management as long as the elasticities are different from zero" (Arbués, García-Valiñas, and Martínez-Espiñeira 2003, 84). A meta-analysis of 296 studies found a mean price elasticity of -0.41, or a 1% increase in price will reduce demand by 0.41%, if all else is held constant (Dalhuisen et al. 2003).

The relationship between price and consumption is not widely agreed upon. The same meta-analysis found a standard deviation of 0.86, a minimum of -7.47 and a maximum of 7.90 (Dalhuisen et al. 2003). This may be due to differences in model selection or biases in the sampled populations, but it is also possible that different locations truly exhibit different elasticities. A 2007 study of North American cities found a price elasticity of -0.33, approximately one sixth the magnitude of a Texas study using a comparable model specification (Olmstead, Hanemann, and Stavins 2007; Hewitt and Hanemann 1995).

Demography

Demographic variables are commonly used as covariates. In particular, income effects are widely studied. In a meta-analysis of income elasticities, researchers found a mean elasticity of 0.43, indicating a 1% increase in income will increase demand by 0.43% (Dalhuisen et al. 2003). The role of income, however, may be overstated. It is possible income is a proxy for several other aspects of consumption that are not explicitly accounted for in many models, such as house size, garden size, or landscape style. As more datasets are available that help decompose income effects into constituent parts, researchers may find that income has less of a contribution to demand than previously thought (Olmstead, Hanemann, and Stavins 2007; Domene and Saurí 2006; Wentz and Gober 2007).

Behavior and Attitude

One of the intrinsic difficulties of water demand modeling is that, even if all other variables are held constant, individual behaviors alter consumption. Behavioral responses are often tested within the economics literature through the specification of price variables. By examining the average price of water, marginal price, and permutations on lagged price, researchers seek to identify which price signals elicit consumer responses (Arbués, García-Valiñas, and Martínez-Espiñeira 2003). The most appropriate method of representing price is not resolved (Olmstead, Hanemann, and Stavins 2007).

Outside of price-related behaviors, the literature on attitudes, behaviors, and the link between the two is nascent. Thus far, however, research teams have not found that a positive attitude toward conservation denotes low water use (Harlan et al. 2009; Syme et al. 2004), contradicting expectations.

Dwelling

Unlike demographic and attitudinal variables, which are more difficult to control, factors relating to the physical structure of a household may help planners develop strategies for water efficient development. The literature examines housing typologies, housing size (e.g., total square footage) and indoor attributes, such as number of bedrooms or number of bathrooms.

Of the studies that explicitly examine housing typology, the findings are relatively consistent. Studies in Spain and England both indicate that indoor water consumption does not change with typology, but houses exhibit different consumption rates based on the presence of a garden (Fox, McIntosh, and Jeffrey 2009; Domene and Saurí 2006). A study in England finds that homes with gardens use as much as 152 liters (40.2 gal) more per day than homes without (Fox, McIntosh, and Jeffrey 2009). Detached, single family homes tend to have gardens and thus consume more water—there are not intrinsic differences in the dwelling structure itself.

Seeking relationships to building architecture, there are an increasing number of researchers relating consumption to the number of bedrooms or bathrooms in a household. Thus far, findings on the relationship between bedrooms and consumption are mixed. While one study found that consumption increases with the number of bedrooms (up until 5 bedrooms or greater), another finds that number of bedrooms is not a significant contributor to demand (Fox, McIntosh, and Jeffrey 2009; Kenney et al. 2008). Bathrooms should theoretically be more significant since fixtures often leak. Yet, few studies examine bathroom counts due to data access. One study that included "number of fixtures" dropped this variable from the demand model because of collinearity with house size (Domene and Saurí 2006). Other studies discuss bathroom behaviors, such as showering frequency, but not the number of bathrooms (Jorgensen et al. 2013; Schleich and Hillenbrand 2009).

Landscaping

To explain outdoor water consumption, studies often include variables that relate to parcel size or neighborhood density (Shandas and Parandvash 2010; Polebitski and Palmer 2010). Some studies additionally quantify land covers. One study calculates a variable called "garden necessities" which indicates the theoretical agronomic water requirements of a household's garden (Domene and Saurí 2006). This study found that "garden necessities" was a significant predictor of consumption, while garden size was not —in other words, landscaping choices are more important than total area.

Quantifying landscaping can be difficult. One study wished to include a continuous variable of garden area, but instead used a presence/absence variable due to lack of data (Fox, McIntosh, and Jeffrey 2009). In the recent literature, remote sensing is proving to be a useful method of evaluating urban land cover in relation to water demand (Wentz and Gober 2007; Guhathakurta and Gober 2007; House-Peters and Chang 2011b). A Phoenix study used Landsat imagery to capture mesic (i.e., irrigated) landscaping with approximately 86% accuracy (Wentz and Gober 2007). In doing so, they found a 1% increase in mesic landscaping increased a household's consumption by 3,797 L/year (14,370 gal/year), or approximately 6% of the average Phoenix household budget.

Irrigation practices are often considered to be a key component of household demand, but explicit irrigation data is typically inaccessible. Some survey-based studies ask questions that relate to irrigation, such as whether the household uses sprinklers or drip (e.g., Syme et al. 2004). One study determined that programmed watering was used in 70% of high-income households and 31% of low-income households, but the impacts of programmed watering on consumption were not isolated (Domene, Sauri, and Pares 2005). The same study asked participants behavioral questions about watering, such as whether or not they selected plants according to climate, irrigated according to species demand, or watered according to weather.

The literature is only beginning to explore combined effects of landscaping choices and climate on water demand (Wentz and Gober 2007; Guhathakurta and Gober 2007; House-Peters and Chang 2011b). Results indicate that there are non-linear system dynamics with feedbacks between evapotranspiration, urban heat island and diurnal temperature differences (Guhathakurta and Gober 2007).

Weather and Climate

The impacts of climate on water demand are widely discussed. Hotter and drier areas are generally assumed to consume more water than cooler and wetter areas, all else being equal. There is supporting evidence in the agronomy literature and in studies exploring climate across broad spatial extents and non-urban environments (Franczyk and Chang 2009).

Impacts of the individual factors that compose climate — such as precipitation and temperature — are not universal. A study in Germany finds that residences are slightly sensitive to precipitation, but are not sensitive to temperature (Schleich and Hillenbrand 2009). An OLS specification reveals a 1% increase in rainy days decreased per capita water consumption by approximately 0.15%, but the relationship is not significant with an Instrumental Variables specification. A study in the Czech Republic found no relationship between consumption, precipitation and temperature (Slavikova et al. 2013). A comparative study between Phoenix and Portland found that temperature sensitivity was positively correlated with shrubs and other low vegetation (Breyer, Chang, and Parandvash 2012).

The impacts of drought on consumption are also explored in the literature. A study in Portland examines the sensitivity of homes to drought conditions, finding that some blocks respond to drought with higher water use, but most do not (House-Peters, Pratt, and Chang 2010). Those that do respond to drought with increased consumption are larger and newer with higher property values. The residents are also more affluent and well-educated. A study in Phoenix uses the Palmer Hydrological Drought Index to quantify degrees of drought (Balling, Gober, and Jones 2008). This study similarly finds that census tracts with large lots, many pools, high proportions of irrigated landscaping and affluent residents increased their water consumption with drought.

The future of integrated land-water models

While there are a growing number of studies incorporating landscape variables, there are several challenges of using existing water demand models to examine land planning strategies. Firstly, the majority of water demand models focus on price effects. The impacts of non-price contributors to demand, such as landscape and climate, are less explored. Secondly, most models do not reveal intraurban variation in demand, which helps link water use with fine grained land cover decisions. Thirdly, results from most studies do not scale well over time or space. There are few extensible rules on water consumption, making it difficult to forecast for future times at new locations or generalize rules beyond the study area.

Policy-relevant variables

Across the literature, many analyses focus on variables that water agencies cannot control. One exception is Kenney at al's 2008 study, which examines several aspects of conservation programs: water restrictions, effectiveness of programs by customer group, drought and non-drought responses to conservation programs, and results of combinations of simultaneous conservation programs. The study specifically separates "items under utility control" and "items outside of utility control" (Kenney et al. 2008). Items under utility control include price, rate structures, water restrictions, rebates and water-smart readers. Items outside of utility control includes seasonality of outdoor water demand.

Though seasonality of outdoor demand may be beyond water utility control, some of the factors which contribute to seasonal demand may be within the realm of land planning. Developing the most appropriate, policy-relevant land cover variables will require ongoing research, however. Since the strength of the contribution of a particular covariate, both in significance and magnitude, is highly dependent on the model selection and the full set of covariates examined within the model, as land planners and water managers mix methods and covariates, understanding of the contributors to demand will likely shift, exposing new covariates of interest.

Disaggregate, extensible forecasting

Demand management may include approving or conditioning proposed developments (Waterman 2004). This requires the ability to forecast for future times and new locations. None of the existing forecasting methods in the long term demand literature are capable of forecasting consumption at locations outside of the initial study area. This means that, with the existing methods, it remains challenging to forecast use of new urban growth at the urban fringe and beyond.

Most long term demand studies examine past years without forecasting future conditions. One exception is Polebitski and Palmer's (2010) panel data study. Using the relationships defined through analysis with pooled, fixed effects and random effects specifications, Polebitski and Palmer forecast single family residential water demands by census tract at a bimonthly time step. Forecasting accuracy is evaluated by omitting the final two years of data from the model during coefficient estimation, running the forecasting model over those years, and then calculating the root mean square error for each census tract's forecasted versus measured water consumption. They find a RMSE of -7% to -12%, depending on season and forecasting method.

Disaggregate, extensible spatial models are a natural extension of existing themes in the literature. Continued development of this course of research will be a valuable addition to applied water demand management in expanding urban areas.

Conclusions

There are opportunities to further develop disaggregate models reveal intraurban variation in water demand. The question of how land planning can contribute to demand management is still unresolved, beyond the coarse filter of assured supply. What does the water efficient city look like? Can principles of compact development or strategic growth contribute to efficiency?

Chapter 3. Quantifying urban and suburban microclimates

As high resolution data on human behaviors with spatial identifiers are increasingly available, there is opportunity to develop greater insight into the heterogeneous relationships between people and their surrounding, physical environment. But, what is the best representation of the nearby environment? In the case of the East Bay Municipal Utility District, data on parcel-level water use allows consumption to be tied to a home and garden—a parcel's unique microclimate signature. This chapter analyzes microclimate by parcel, and then discusses other representations of microclimate in EBMUD, ultimately suggesting that there is not a singular solution to representing the urban environment.

Microclimate classifications are currently used by EBMUD to develop residential water budgets. These water budgets are relied on internally to project demand and shared with consumers to provide a theoretical benchmark for their household water use. Evapotranspiration is an important factor in determining the water budgets. EBMUD assigns each household to a climate zone, which is associated with an evapotranspiration rate. There are three climate zones, and three evapotranspiration values employed by EBMUD (see Figure 1).

Classifying climate—a continuously varying, multivariate phenomenon—raises the widely acknowledged Modifiable Area Unit Problem (MAUP). Named by geographers Openshaw and Taylor (1979), the MAUP occurs when it is possible to represent areal units of analysis at different levels of aggregation (the *aggregation effect*) or as different shapes (the *zoning effect*). Ultimately, either aggregation effects or zoning effects may lead to cases of ecological fallacy—that is, results of aggregate data analysis differ from analysis of individual data (Gotway and Young 2002).

As statisticians Gotway and Young (2002, 632) state "[t]he choice of an appropriate scale for the study of spatial processes is an extremely important one because mechanisms vital to the spatial dynamics of a process at one scale may be unimportant or inoperative at another. Moreover, relationships between variables at one scale may be obscured or distorted when viewed from another scale." Similarly, Moellering and Tobler (1972) make the distinction that while geographic processes are often characterized as operating at local or global scales, in fact the same process can operate simultaneously at multiple spatial scales. A critical question is identifying the scale "where the action is" (Moellering and Tobler 1972, 36).

Figure 12. Climate zones delineated by East Bay Municipal Utility District. These zones are used for calculating expected residential water budgets. The basic pattern of the zones follows evapotranspiration data supplied by the California Irrigation Management Information System (CIMIS), but was updated by EBMUD from observed evapotranspiration at EBMUD-owned weather stations. The precise boundaries of the zones follow zip codes. Source: EBMUD.

This chapter begins a greater inquiry on spatiotemporal scale, ecological fallacy, and individual analysis in residential water demand. The analysis proceeds in two parts. First, it develops land cover, weather, and landform characteristics by parcel, examining opportunities and limitations of available data sources. Next, it clusters households with similar microclimate signatures using the k-means clustering algorithm. By exploring multiple clustering schemes, it reveals microclimate zones within EBMUD.

Microclimate covariates by parcel

The variables in this chapter are commonly used in agronomic and hydrologic examinations of water. Each theoretically influences local consumption by altering evapotranspiration and runoff (see Figure 2). All data are publicly available. Because of the broad spatial extent of the study area, all processing techniques were intentionally selected to be largely automated.

For each variable, the chapter discusses the primary issues surrounding parcel quantification. Parcel-level environmental analysis exposes the common challenge of associated with spatial data of trading off spatial resolution and temporal resolution. The most granular spatial data is temporally coarse. The most granular temporal data is spatially coarse. Environmental data that varies at the same spatiotemporal scale as the available water use data (parcel-month) is not currently publicly available.

Though the challenges differ for each variable, together they reveal that if social processes are to be examined alongside environmental conditions in socioecological models, there is considerable opportunity to develop variables that match the spatiotemporal scale of behavioral data streams, including urban water use.

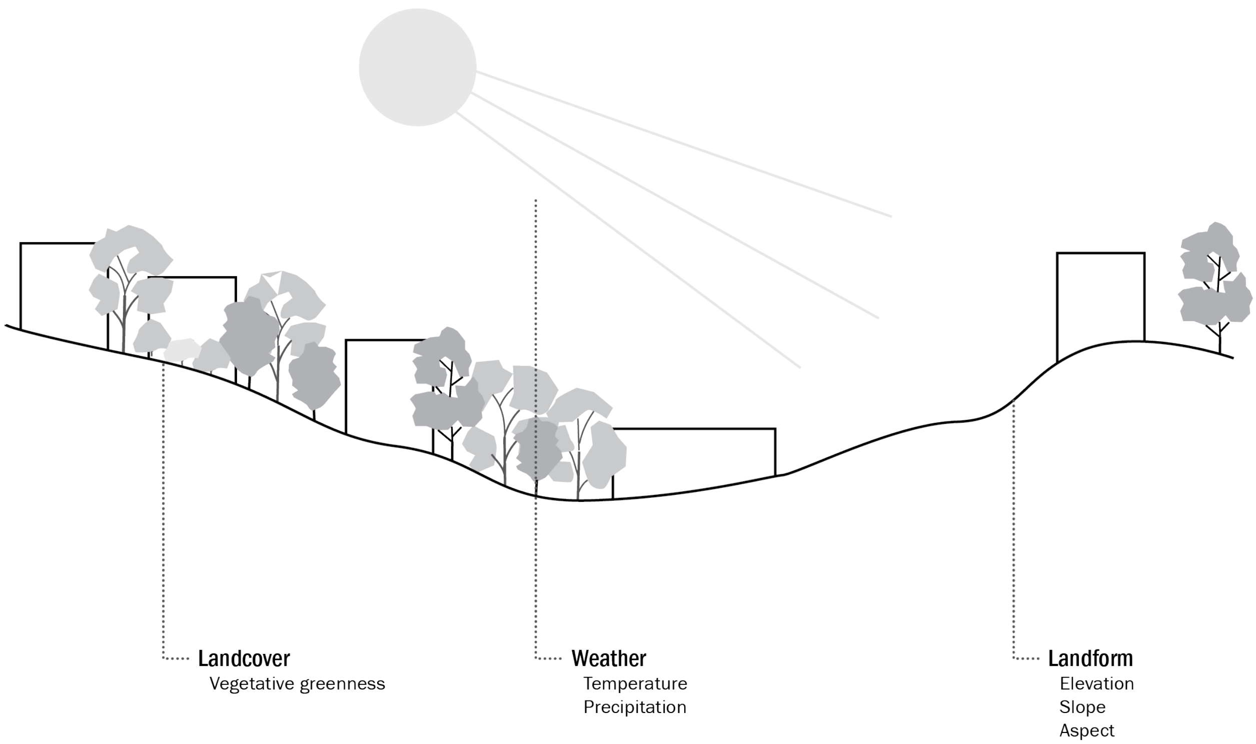


Figure 13. Variations in land cover, climate, and landform contribute to heterogeneity in evapotranspiration.

Land cover

Vegetative cover should influence water demand based on its agronomic water requirements (Domene, Sauri, and Pares 2005). To capture garden types (mesic versus xeric), the Normalized Difference Vegetation Index (NDVI) is often used in the water demand literature (see chapter 2). NDVI is calculated from the red and near infrared bands of any multispectral imagery source.

Landsat, the most common satellite imagery provider, makes NDVI data directly available for download. Landsat's key advantage is temporal granularity. Landsat satellite imagery has been collected globally since 1999 and revisits the same site approximately every 16 days. While satellite images may occasionally be obscured by cloud cover, Landsat's consistent and calibrated imaging is valuable for capturing change. For many studies, this temporal frequency is critical. Landsat has a key disadvantage in the urban and surburban realm, however: its 30 meter resolution is too coarse to resolve differences among parcels.

There are spatially higher resolution alternatives to Landsat, but these sources have significant drawbacks. The most granular images available are from DigitalGlobe (including GeoEye, IKONOS, and Worldview satellite products). They can be very expensive. One commonly used distributer, Apollo Mapping, sells DigitalGlobe for $9.10 to $17.50 per square kilometer (personal communication, July 31, 2014). Because EBMUD's current service boundary is approximately 1055 square kilometers, one month of data costs a minimum of $9,000.

The highest resolution source of publicly available multispectral imagery is from the National Agricultural Imagery Program (NAIP). Nationally, NAIP is flown by plane approximately every 2-3 years. In California, multispectral images are available during the study period for June 2005, June 2009, and May 2010. All years have a 1 meter resolution.

LANDSAT | LANDSAT BY PARCEL | NAIP BY PARCEL

Figure 14. Landsat imagery is coarse relative to parcels. Often, a parcel is represented by one or two 30 meter pixels. In contrast, NAIP imagery is at a 1 meter resolution, making it possible to resolve garden types.

The disadvantage of NAIP is that, unlike Landsat, it is not designed to be compared over time. It does not include surface reflectance corrections. Appendix 1 evaluates whether it is possible to calibrate NDVI imagery in order for it to be compared over time. Results indicate that it is not possible to use NAIP in time series at the parcel level. This means that, at present, there are no sources of public imagery that capture vegetation change by parcel over time. For this reason, this study considers landscape type to be time invariant—an assumption that is true for most households, but not all.

Weather

Local weather conditions are key contributors to theoretical evapotranspiration (Monteith 1965). In the water demand literature, weather data is most commonly represented through a data product known as PRISM, provided by the University of Oregon. PRISM models temperature and precipitation at a monthly time step. The data are convenient and accessible, but have some disadvantages in the context of this study.

Like many weather data products that draw from 30-year normals, PRISM reaches back in time for 30 years to identify the spatial relationships that underlie the final models (Daly et al. 2008). PRISM assumes that the relationships among the weather data for a given month will remain consistent over time: all Januarys will show similar patterns, all Februarys will show similar patterns, et cetera. This temporal stationarity embedded in PRISM may not be an appropriate assumption for the Bay Area, with its ever-changing climate (see Chapter 1).

PRISM data are spatially coarse relative to parcels. PRISM is available at a resolution of 4 kilometers for free, or 800 meters for a small fee. Both resolutions are sufficiently dense for many regional or rural studies, but coarse in an urban context. Though it will never be possible to interpolate weather with certainty at every parcel, it is possible to explicitly model by parcel and directly capture associated uncertainty.

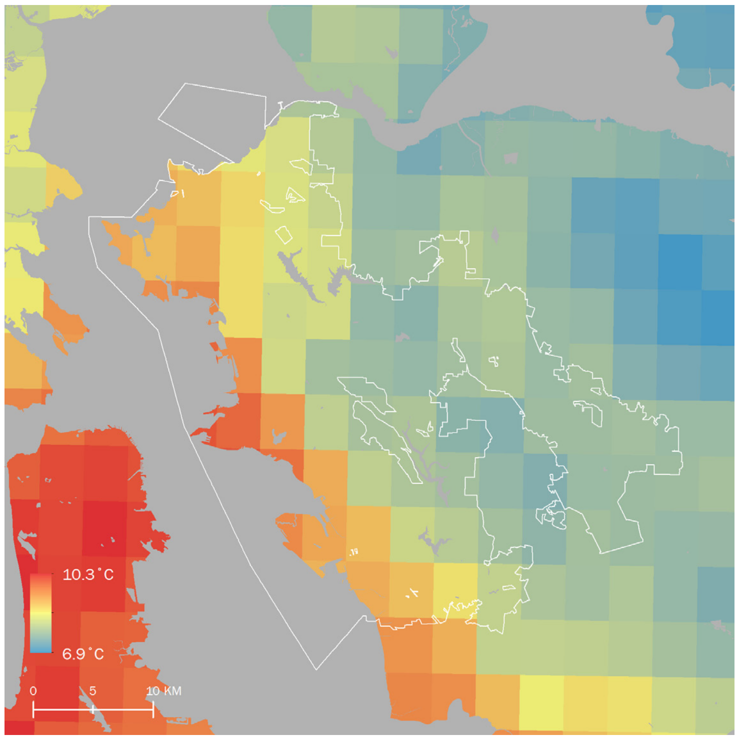


Figure 15. PRISM temperature data in the bay area for January 2005 on a 4km grid. Relative to the parcelization of EBMUD, the dataset is coarse.

To create a higher resolution model, this dissertation examines data from the National Weather Service. NWS weather data are used as an input to the PRISM model. They are also have precedent in the water demand literature: Brent (2013) interpolates NWS data by nearest neighbor allocation; Mini interpolates NWS data through Inverse Distance Weighting(Mini, Hogue, and Pincetl 2014). Appendix 2 and Appendix 3 develop parcel-level Temperature (see Figure 5) and Rainy Day (see Figure 6) models by Kriging. This method exposes large amounts of uncertainty in intraurban climate characterizations using available data products.

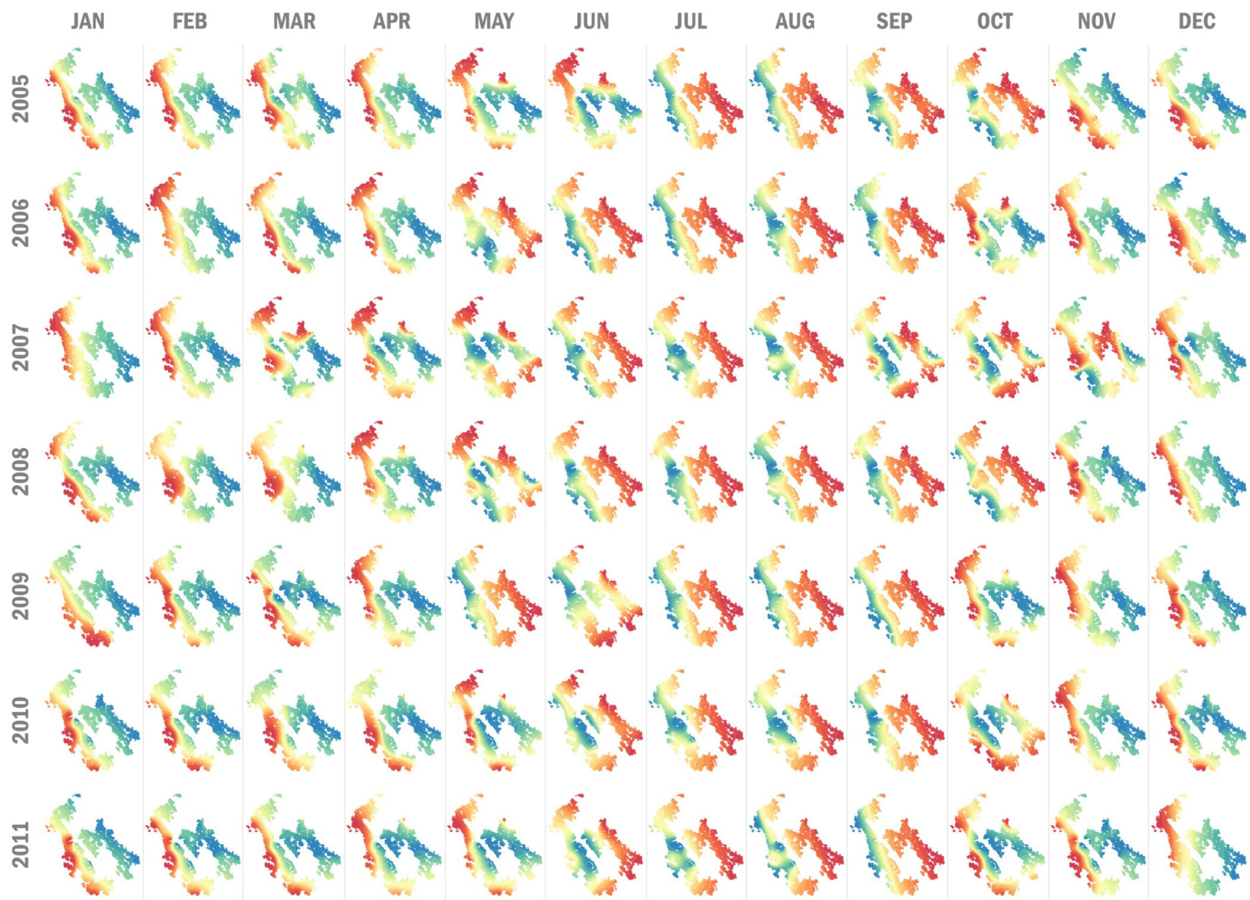


Figure 16. Changing patterns in temperature by month. In the winter, the warmest region is along the San Francisco Bay, on the western side of the study area. In the summer the warmest region is inland, along the eastern side of the study area. The temperature gradient flips each spring and fall, but the timing of the inversion varies by year.

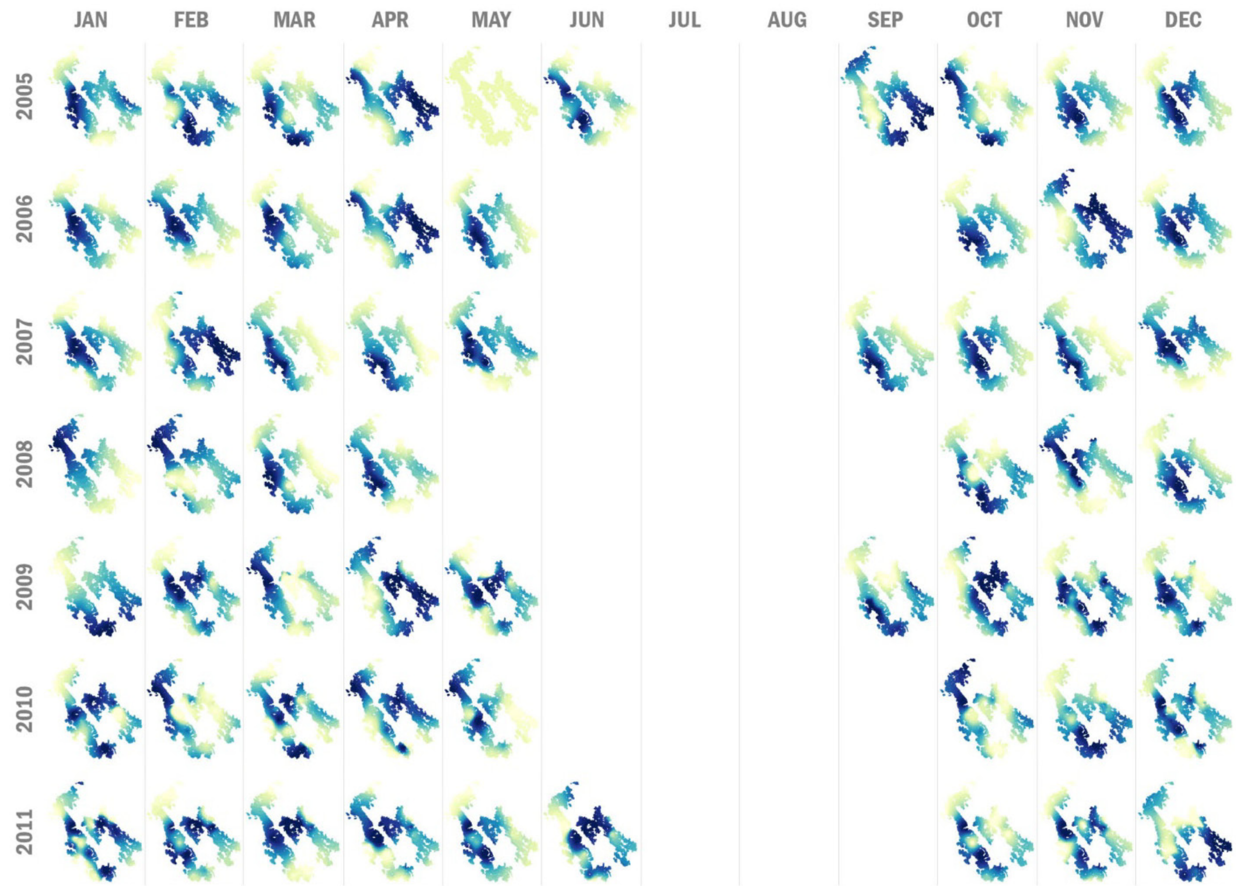


Figure 17. Changing patterns in the number of rainy days (greater than 0.1 inches of rain) by month. Summer months may not have any rainy days—there were no rainy days in July or August during the study period. During the seven year study period, there is little evidence of consistent spatial trends in the number of rainy days.

Landform

While vegetation and weather variables are increasingly prevalent in the literature, landform variables are less explored. This dissertation examines elevation, slope and aspect (see Figure 7). Elevation data is distributed by the United States Geological Service at a 3 meter resolution. The remaining variables are easily derived from elevation using ArcGIS with Spatial Analyst. Aspect is introduced to urban water demand modeling in this dissertation, though it is used frequently in the agronomic literature since solar exposure relates to evapotranspiration rates.

In contrast to the other variables examined in this chapter, landform data is finer than the parcel at it is necessary to aggregate the data. For elevation and slope, I calculated the mean value for each parcel. For aspect, I categorized the data into north, east, south, and west-facing slopes, and then calculated the most common value per parcel.[2] These values are assumed

[2] Aspect values range from 0-360˚, with both the smallest and largest numbers in the range indicating a north-facing slope. If averaging values over a slope that is primarily north-facing, included values on both ends of the spectrum, the results would be close to 180˚, speciously indicating a southerly slope. As a result, categorizing and take the mode provides more accurate results on the general trend of the parcel.

invariant during the study period due to the temporal resolution of the data. This is most typically reasonable, since earthworks projects impacting existing homes (i.e., those reporting water use for the full extent of the study period) are rare.

Defining microclimate zones

While each individual parcel has a unique combination of temperature, rainy days, vegetative greenness, elevation, slope, and aspect, there are similarities among groups of houses. Aggregating houses by microclimate signature may help better capture unobserved variables or interaction effects among the variables.

Defining data by parcel, and then aggregating, presents unique benefits. In particular, while microclimate zones are often defined by spatial location, this analysis excludes location and zones emerge through the combination of variables that indicate a microclimate signature. This bottom-up approach may help mitigate some aggregation and zoning effects that would otherwise be introduced when drawing spatial boundaries.

K-means clustering

Data points can be assigned to clusters through several possible unsupervised learning techniques. On larger datasets, k-means is often used because, unlike hierarchical clustering methods, it does not require a computationally expensive dissimilarity matrix (also known as distance or difference matrix). Instead, with k-means, the number of clusters are assigned *a priori*. The algorithm creates k centroids among data and calculates the distance between each observation and the centroids. Observations are assigned to the nearest centroid, creating the cluster. The position of the centroid is then recalculated to minimize the sum of squared errors within the cluster. After, each observation is reassigned to the closest centroid, given its new location. The algorithm repeats until there is no reassignment. Rather than solving a dimensionally reduction problem, k-means converges on a solution.

There are two notable disadvantages to k-means. Firstly, k-means may cluster the data points slightly differently on each model run based on the initial locations of the centroids or the order of the data. Secondly, choosing the number of clusters can requires interpretation. While some data have very clear clusters, most data can be interpreted through multiple clustering schemes. There is rarely a single solution and it is difficult to compare goodness of fit, since k-means does not generate dendograms or fit statistics like hierarchical clustering methods. Some researchers use methods like silhouette plots or the Dunn Index to evaluate fit with k-means, but these require the very dissimilarity matrix k-means usefully avoids.

Data plotting can contribute to identifying appropriate cluster selection, in many multivariate clustering problems, the clustering occurs in higher dimensional space. Visualizing the clusters

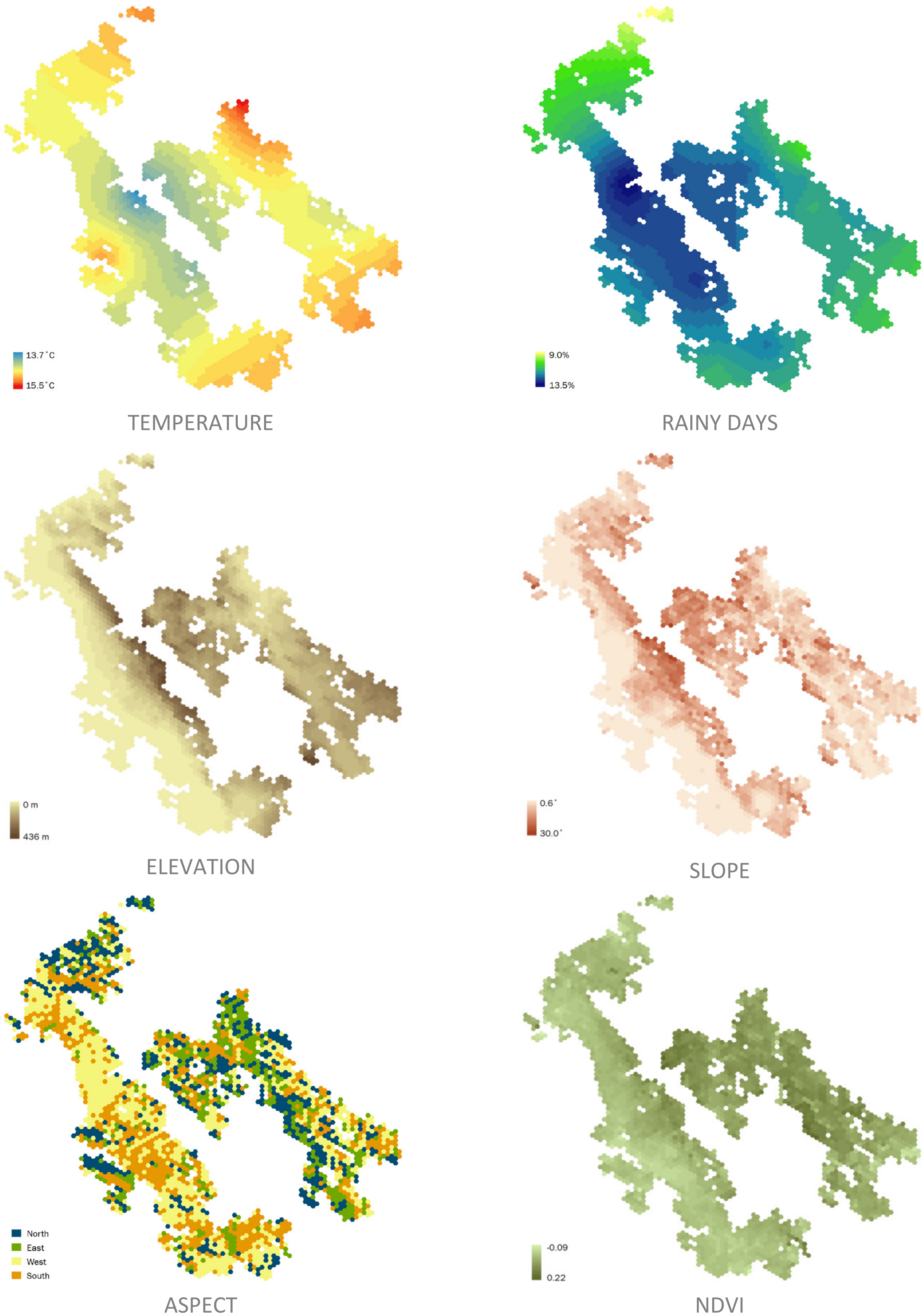


Figure 18. Environmental patterns over space. Data developed at the parcel and visualized by mean values on a 1 km hexagonal grid for clarity. Temperature and Rainy Days are averaged over the course of the study period.

directly is not possible when there are more than three variables. While there are several possible visualization techniques, in the case of the location-specific parcel analysis it is likely most useful to map the clusters back to the parcels and visualize physical variation of the clusters. Microclimates should exhibit physical clustering, since several of the environmental variables contributing to the clusters vary smoothly in space.

Identifying clusters

Because of the large number of observations in this analysis, microclimate zones are derived through k-means clustering. It was not possible to calculate a dissimilarity matrix with current computing resources to leverage many fit statistics. Instead, I used a scree plot to identify the most appropriate number of clusters, looking for the point that minimizes both the sum of squared errors within the cluster and the total number of clusters. In this analysis, the scree plot indicates three to six clusters.

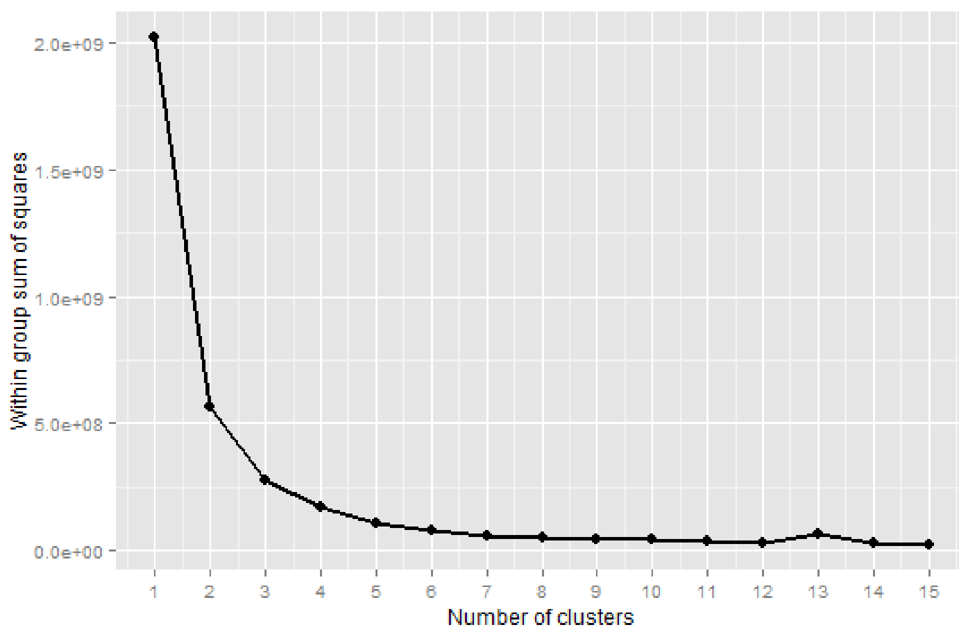


Figure 19. The sum of squared errors within each cluster versus the number of clusters. In some cases, there is a clear bend or "elbow" in the plot, indicating a best choice for the number of clusters. In this case, the plot indicates the best number of clusters is from three to six.

Correlations between vegetation, weather, and landform variables are available in Table 1. Mapped clusters are provided in Figure 9.

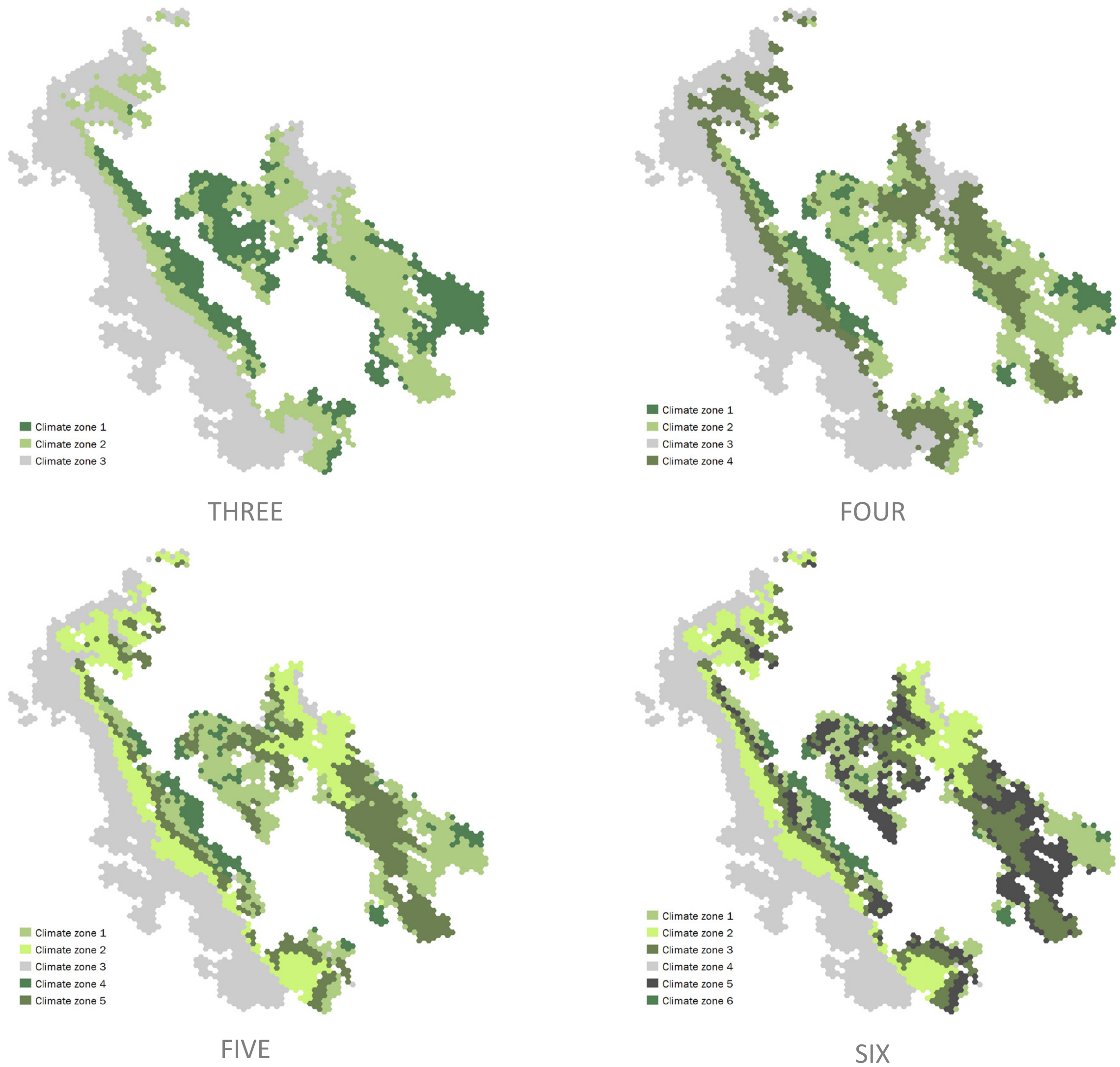


Figure 20. Alternative clustering schemes. Data clustered at the parcel level and visualized by median values on a 1 km hexagonal grid for clarity.

Table 1. Correlations between variables. Temperature shows a strong negative correlation with rainy days. Elevation is positively related to slope. Vegetation is related to both elevation and slope.

	Temperature	Rainy Days	Elevation	Slope	Aspect	Vegetation
Temperature	1	-0.70	-0.15	-0.24	-0.15	-0.07
Rainy Days	-0.70	1	0.11	0.14	0.18	0.08
Elevation	-0.15	0.11	1	0.56	-0.03	0.46
Slope	-0.24	0.14	0.56	1	-0.03	0.42
Aspect	-0.15	0.18	-0.03	-0.03	1	-0.12
Vegetation	-0.07	0.08	0.46	0.42	-0.12	1

Interpreting clusters

The k-means algorithm clustered on microclimate characteristics, without explicitly evaluating spatial location. Yet, there is a strong spatial pattern in the data. This physical grouping of points provides some validation of the results of the clustering algorithm, since microclimates should exhibit physical continuity. A noisy pattern would indicate poor clustering performance.

One of the challenges with the cluster analysis, however, is characterizing the nature of the pattern. What defines a zone? K-means does not directly provide statistics that indicate the variables associated with each cluster. To evaluate cluster membership, the three zone solution is examined more closely in Figure 10. Insets A, B, and C show areas where all three zones are spatially proximate. In these insets, microclimate zone 3 is the coolest, flattest, and most urbanized. Microclimate zone 2 represents the moderate and warm slopes of the lower hills. Microclimate zone 1 is at the highest elevations, in the coolest, most vegetated locations.

Analyzing the clusters against EBMUD climate zones does not help confirm the nature of the clusters. Figure 11 maps the three zone solution on top of EBMUD's climate zones. None of the clusters are fully contained within any of EBMUD's climate zones—all clusters exist in all EBMUD zones.

A higher number of clusters further obscures the identity of each group. Some households, which were likely on the edge of the k-means cluster (marginally closer to one centroid than the other), change group membership between clustering schemes (see Figure 9). This exposes the uncertainty in classifying a multivariate phenomenon—there are many borders that are difficult to visualize. More clusters provide a window through which to see nuance in the data, but also introduce more boundary effects.

There are additional methods available to define clusters. With highly correlated variables, for example, it can be useful to cluster on principle components instead of the data (Ding and He 2004). Because the weather variables in this analysis reveal a strong negative correlation, future analyses may compare the results of clustering on the data versus clustering on principle components. The expense of principle component analysis, however, is that it further obscures the relationships among the variables, making it more difficult to evaluate which factors are contributing to the cluster patterns.

Using k-means clustering to define microclimates has previously been employed at the national scale (Hargrove and Hoffman 2004) and on very fine, simulated data (Häb, Middel, and Hagen 2013). Applications to an urban gradient are new and require continued vetting. One of the challenges with generating all microclimate zone is testing validity. How is it possible to determine if a boundary is correct? Over what time period?

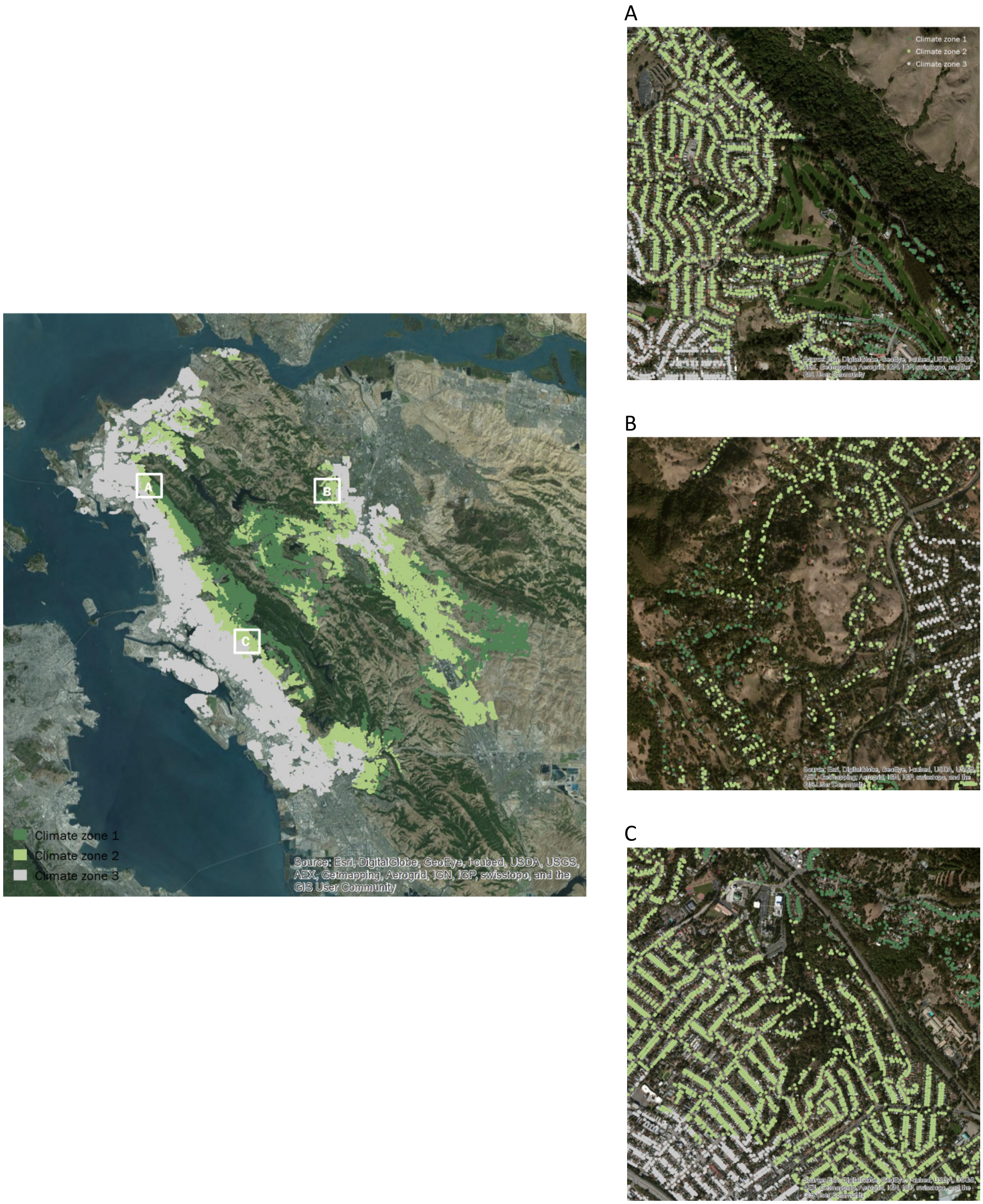


Figure 21. Evaluation of the three cluster solution. Microclimate zone 1 is at the highest elevations, in the coolest, most vegetated locations. Zone 2 is the foothills, on moderate and warm slopes. Zone 3 is the coolest, flattest, and most urbanized.

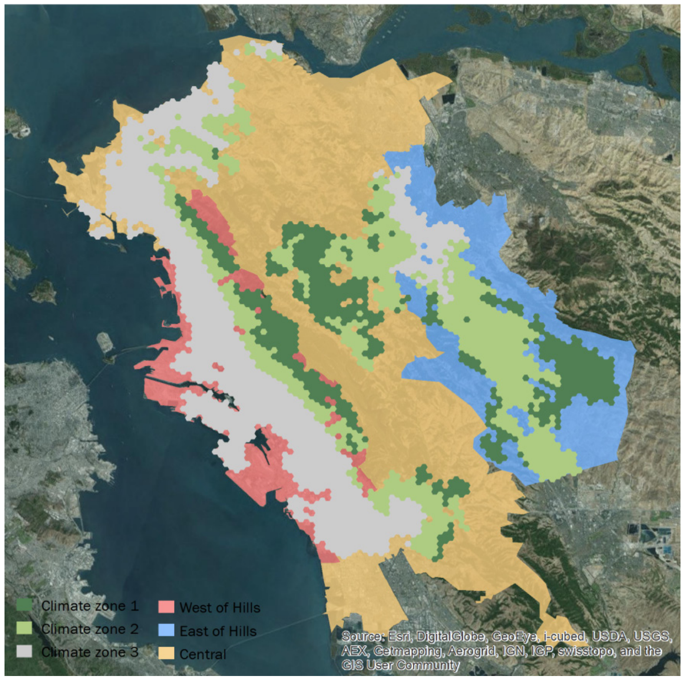


Figure 22. EBMUD evapotranspiration zones (also visualized in Figure 1) overlaid with microclimate zones (also visualized in Figure 9). All three clusters exist in West of the Hills, East of the Hills, and Central.

Conclusions

This chapter examines methods of characterizing the urban and suburban environment. Beginning with parcel-by-parcel classification, it pushes the resolution of many of the available data sources. While this chapter is able to successfully develop parcel characterizations of weather and landform variables, it is not able to resolve land cover change by parcel. Parcel-level, monthly data remain rare. There is a spatiotemporal gap that datasets are moving toward, but variables are progressing at different rates.

Clustering may prove to be a valuable method of interpreting the urban environment with data of different spatiotemporal origins. This bottom-up approach to defining microclimate zones has the added potential of creating fuzzy physical boundaries, where some points along the boundary line have membership in one cluster, and others have membership in the neighboring cluster. Surprisingly, the clustering schemes in this chapter reveal a physical boundary that is finely undulating, but firm. Which data contributes most to these hard transitions? Are there other combinations of data that would reveal urban ecotones?

The question of representation and spatiotemporal scale in the urban environment is continuing to evolve. Even if high resolution data were consistently available, all continuous phenomena must be classified at some level. Any classification, with boundary effects and different possible aggregation schemes, may offer inaccurate characterizations. For this reason, there are no best aggregation—instead, the most appropriate data representation is dependent on the question being asked. Chapter 4 evaluates both parcel microclimate characteristics and microclimate zones in relation to water use.

Chapter 4. Heterogeneity in relationships among water use, microclimate, and landscape

Within the EBMUD service area, there is tremendous variation in household water use. Over time, average monthly use spans from 144 to 429 gallons of water per day (gpd). Across households, there are even larger differences in water use. Average monthly use by household ranges from 25 to more than 3,000 gpd. Figure 1 plots average water use by month and average water use by household. The data reveal that temporal variability is far smaller than variability observed across EBMUD's residents.

Temporal variation in consumption can be explained in part by weather: water use is highest when it is warm out, and lowest when it is cold. It can also be explained in part by prevailing climate conditions, including demand responses related to ongoing drought. These factors do not explain everyone in the service area equally, however, since environmental conditions differ at each parcel.

Beyond exposure to weather variables, what explains the cross-sectional variation observed within a single service area? Within the same time period, why do some houses consume so much more water than others, while some consume so much less? It is likely the observed heterogeneity in water use can be attributed to three factors:

1. Features of the dwelling (influencing indoor use)
2. Features of the parcel (influencing outdoor use)
3. Features of the people living in the house (influencing indoor and outdoor use)

Data on each of the variables related to temporal and cross-sectional variation—weather, dwelling, parcel, and people—move on different time scales. With regard to outdoor water use, the weather varies continuously, while variables like aspect can be considered time invariant. With regard to indoor water use, features of the dwelling are most typically time invariant. Household residents, and what they do, affect both outdoor and indoor water use in ways that are likely both time invariant and highly time varying.

The goal of the analysis is to evaluate variation in water use and decompose it into variables associated with each factor, giving special focus to outdoor water use. To what degree does microclimate and landscape influence demand? Is it possible to identify high water use zones?

Or, are behavior patterns intrinsic to different user groups, rather than location? With disaggregate, monthly water consumption records for all households in EBMUD, this analysis precisely matches parcel covariates to water use, revealing sources of heterogeneity among water users.

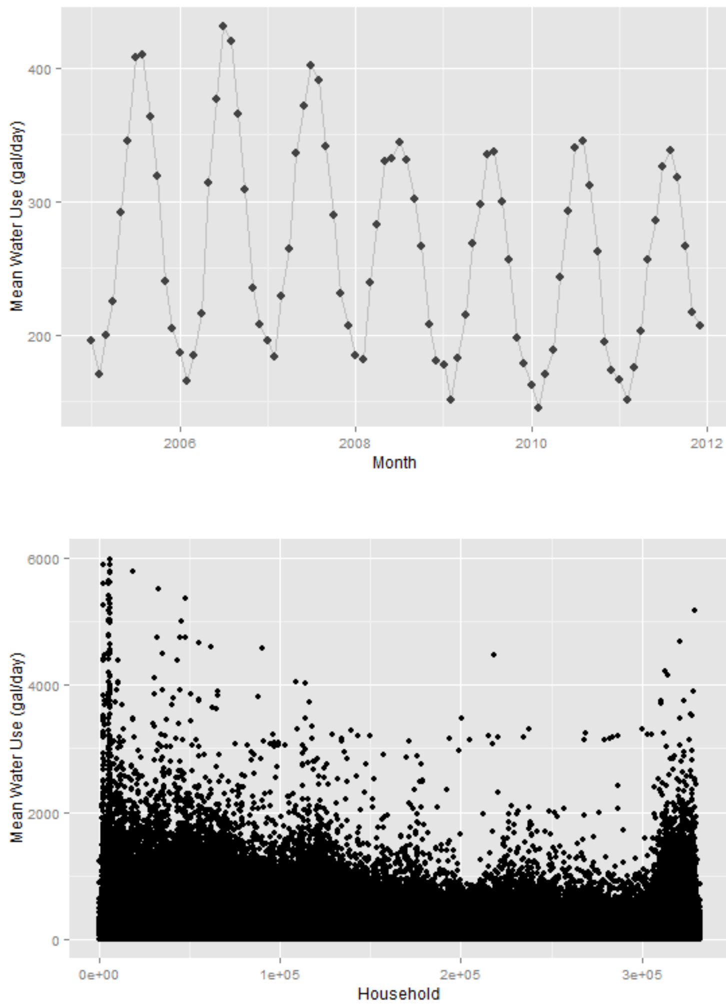


Figure 23. Water use data varies more by house than through time. Average water use over time ranges from approximately 144 to 429 gpd. Water use by household ranges from 25 to more than 3,000 gpd.

Data

The data in this analysis originate from EBMUD, the Alameda and Contra Costa County tax assessors, the National Weather Service, and the United States Geological Service. These variables, their initial spatial resolution, and processing for parcel evaluation are summarized in Table 1.

Table 2. Model covariates and original spatial resolution. Data were processed to evaluate at the parcel-level, alongside household water consumption data. See Chapter 3 for more information on temperature, rainy days, NDVI, elevation, slope, and aspect.

Time step	Covariate	Spatial resolution	Processing for parcel
Monthly	Temperature	~7 km	Interpolated
	Rainy Days	~7 km	Interpolated
Time invariant	Landscape style (NDVI)	1 m	Aggregated (Mean)
	Elevation	3 m	Aggregated (Mean)
	Slope	3 m	Aggregated (Mean)
	Aspect	3 m	Aggregated (Mode)
	Lot size	Parcel	n/a
	Built square feet	Parcel	n/a
	Age of house	Parcel	n/a
	Number of bedrooms	Parcel	n/a
	Number of bathrooms	Parcel	n/a
	Presence of pool	Parcel	n/a
	Household size	Block group	Imputed
	Income	Block group	Imputed

Water use

EBMUD includes 331,957 single family residential households. Observations were recorded monthly from 2005 to 2011. Of the 331,957 accounts, approximately 15,000 were eliminated due to missing observations, errors, or zero values. Missing observations occurred when accounts were not serviced by EBMUD—they were shut off or came online during the middle of the study period. Observations more than four standard deviations above mean were dropped, assumed to indicate a broken meter or a large leak. The upper cutoff was 2,917 gpd. After cleaning, the data had a mean of 269 gpd and median 199 gpd.

Though water use is referenced throughout this document in gallons per day, water use is recorded by the water agency in centum cubic feet (100 cubic feet or 748 gallons) per month. This is the unit of water use in the statistical analysis because it reduces the range of the variable and increases the stability of matrix transformations. In this analysis, water use is not otherwise transformed, though this decision merits further evaluation. Figure 2 plots a histogram of water use in the study area. There is a long right tail to the distribution—it is more skewed than a log-normal distribution. The 95th percentile of water use is 773 gpd, 99th percentile is 1,146 gpd, and 99.5 percentile is 1,795 gpd. A semi-log evaluation of the data may be appropriate when evaluating the whole distribution. One of the challenges in the context of this study, however, is that the data is not always evaluated as a whole—it is also examined through sub-populations. The distributions of many of the sub-populations are not appropriately captured by a semi-log specification.

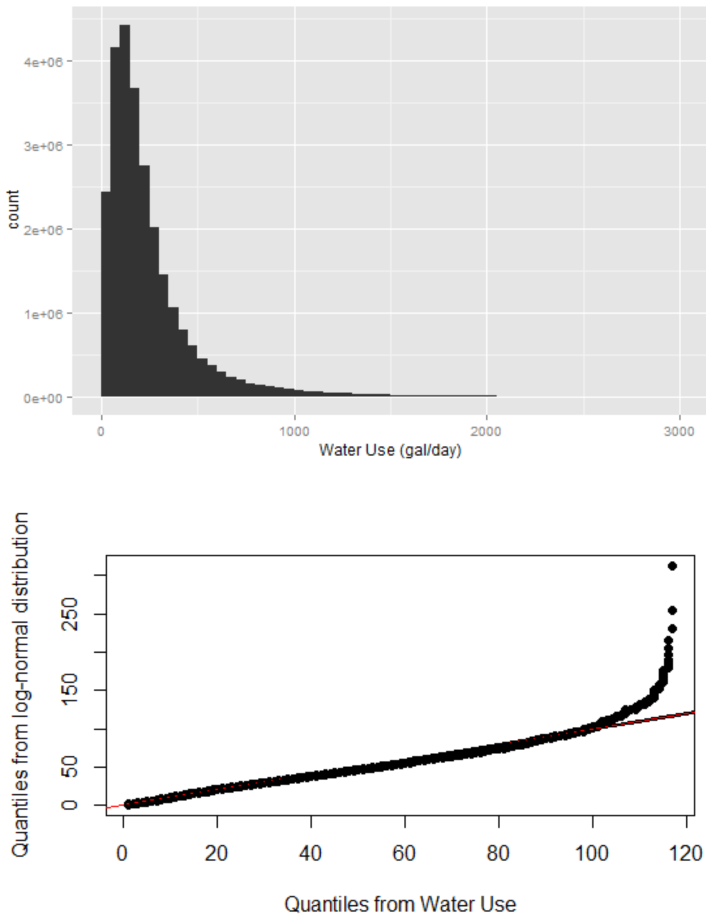


Figure 24. Histogram of household water use in EBMUD. During the study period, mean use was 269 gpd. Median use was 199 gpd. The distribution has a strong right tail, more skewed than a log-normal distribution.

This study relies on total household water consumption as the dependent variable. Other studies that focus on outdoor demand isolate outdoor water use as the dependent variable. In areas lacking separate outdoor water meters, like EBMUD, this variable is typically developed by subtracting winter use from summer use. Winter use is assumed to represent an indoor, base water use that is time invariant. This study does not attempt to isolate outdoor use because some households do not exhibit the sinusoidal shape of demand that is commonly seen in aggregate (see Figure 1). Some households have a relatively flat consumption throughout the year, while others have highs and lows that do not directly related to seasons. Figure 3 plots monthly use from randomly sampled houses, indicating the variety of consumption patterns.

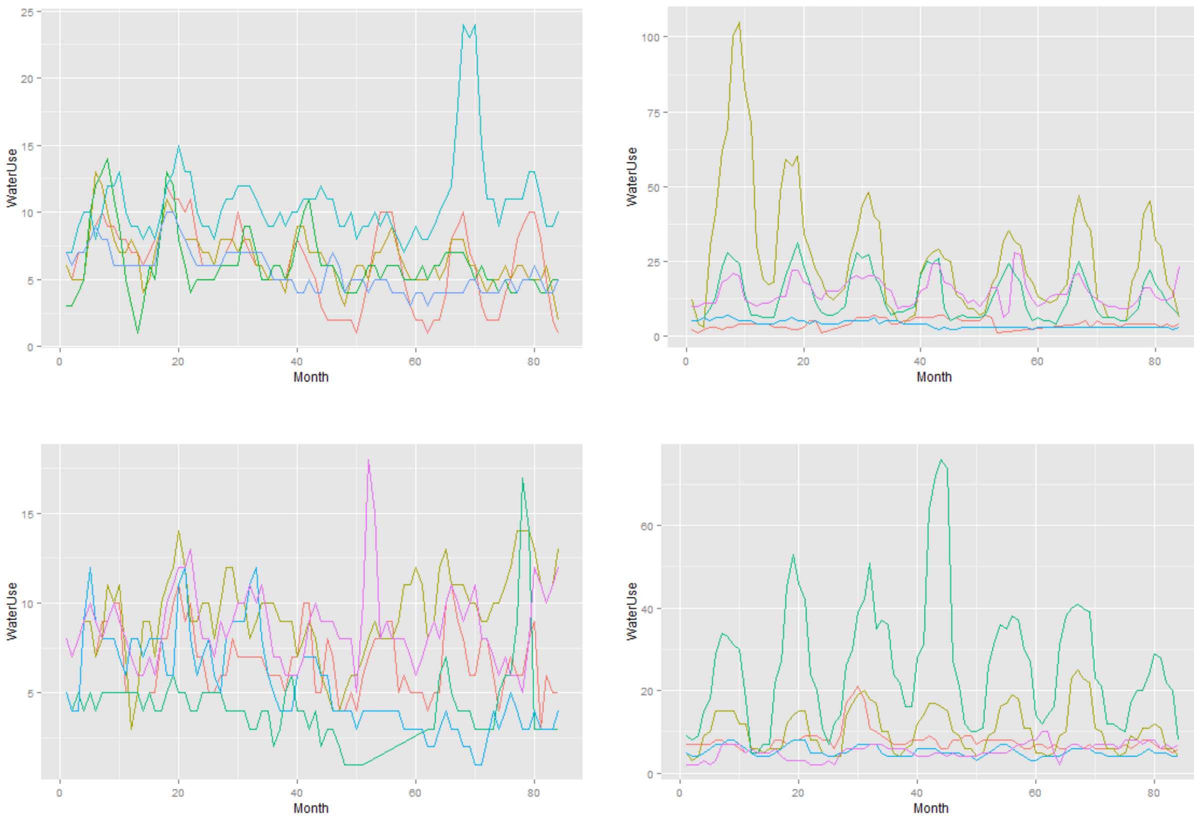


Figure 25. Four examples of water use over the 84 months of the study period for five, randomly sampled houses. In each chart, two to three households exhibit the expected periodicity in seasonal water use, while the others do not. The bottom left chart appears particularly noisy because of the small range of the y-axis, relative to the other charts. Monthly water consumption by user group is available in appendix 4.

Water price

The relationship between water price and consumption is a chief concern of many water demand studies. Yet, including price as a variable in a traditional panel model alongside other covariates can prove problematic. In a block rate structure (see Table 2), price is endogenous to demand—the price a household is exposed to is directly related to the volume of water consumed. In addition, price is correlated to the year because the rate structure changes annually.

Despite the primacy of price, recent literature indicates that price does not strongly influence consumption. Recent research finds consumers respond to average price, not marginal price (Ito 2014; Wichman 2014). This average price response is small. Koichiro Ito identifies a -0.06 to -0.09 price elasticity in household energy consumption, which may have a stronger price signal than water. To isolate price effects in water, a targeted experimental design is necessary (e.g., Klaiber et al. 2014). This study is not intended to evaluate price effects and, for this reason, does not examine price further.

Table 3. Variable water rate schedule for single family houses within the East Bay Municipal Utility District, 2005 - 2011. Water rates are comprised of flat fees and variables fees. The variable fee is increasing block rate structure with three blocks, as described in this table. In 2008, there was an additional variable surcharge in response to drought conditions.

YEAR	VOLUME (gpd)	RATE	OTHER
2005	<172	1.59	
	172-292	1.98	
	>293	2.42	
2006	<172	1.65	
	172-292	2.05	
	>293	2.51	
2007	<172	1.73	
	172-292	2.15	
	>293	2.64	
2008	<172	2.00	All houses assigned a water allocation based on average water use for the equivalent billing cycle in FY 2005-07. Fined $2.00 per ccf consumed in excess of 90% of past use. Does not apply to accounts using less than 100 gpd. (East Bay Municipal Utility District 2008)
	172-292	2.49	
	>293	3.05	
2009	<172	2.00	
	172-292	2.48	
	>293	3.04	
2010	<172	2.15	
	172-292	2.67	
	>293	3.27	
2011	<172	2.28	
	172-292	2.83	
	>293	3.47	

Indoor: Dwelling

Several dwelling structure and parcel characteristics were accessed from the county tax assessors. Relative to the other data sources, this is the most error prone. There are many missing observations: 24,442 parcels do not have information on the year the structure was built; 16,273 lack data on building square footage; 5,588 are missing lot size. After accounting for missingness, the final dataset contains 302,807 parcels.

Distributions of the data are available in Figure 4.

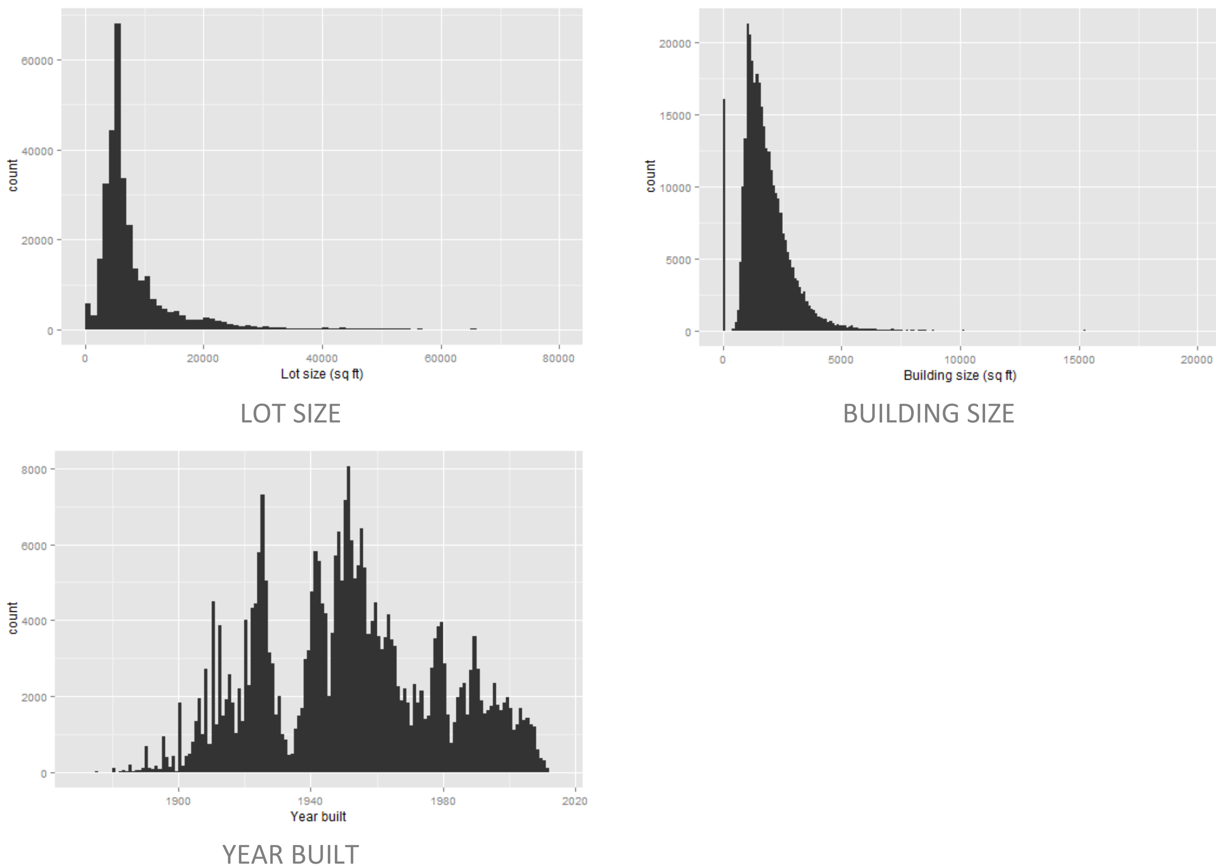


Figure 26. Distributions of covariates relating to indoor water use.

Outdoor: Microclimate and landscape

The development of vegetation, weather, and landform variables is discussed in Chapter 3. In addition, this study includes information on pools. Pools are accessed through the county assessors, which marks the presence of pools (no data indicates no pool). There are approximately 22,000 pools in the study area. Visual inspection against aerial imagery confirms that the variable is reasonably accurate, though some pools are likely not recorded, while other recorded pools have been filled in.

Distributions of the microclimate and landscape data are available in Figure 5.

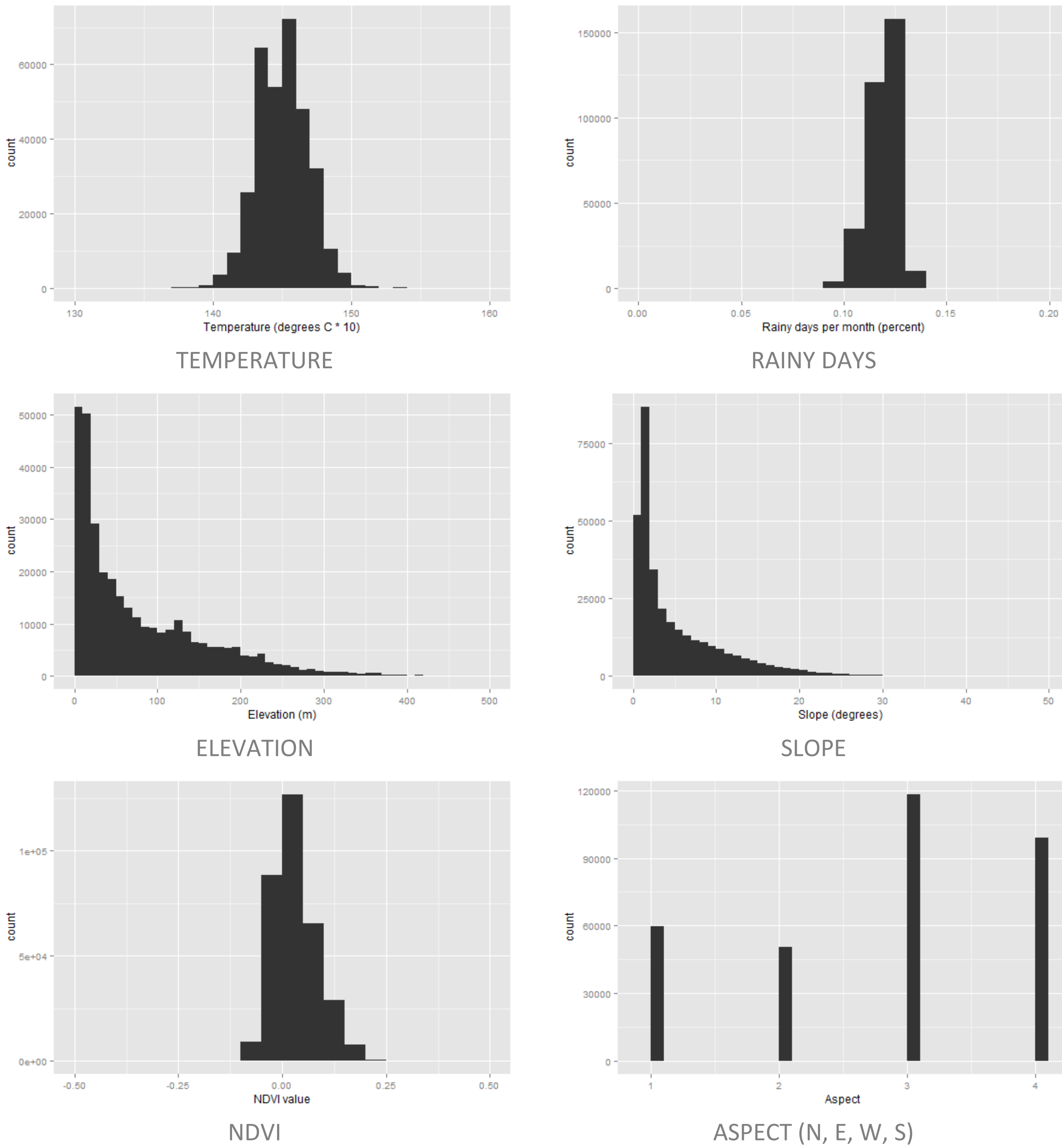
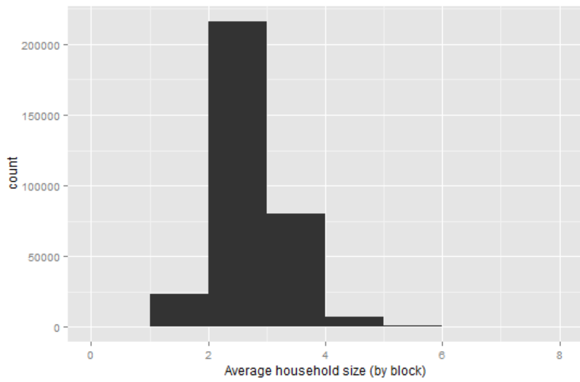


Figure 27. Distributions of covariates relating to outdoor water use.

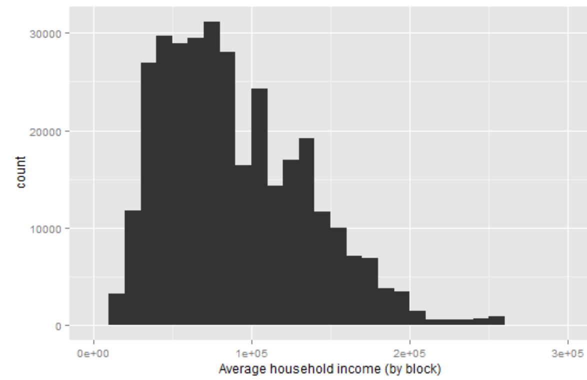
Indoor + Outdoor: Demographic

Personal characteristics will influence water consumed inside and outside the house. The highest resolution information on residents in the study area is block group-level data from the U.S. Census. These are spatially and temporally coarse relative to the demand data.

Distributions are available in Figure 6.



HOUSEHOLD SIZE



INCOME

Figure 28. Distributions of covariates relating to indoor and outdoor water use.

Full covariate set

In addition to the covariates relating to indoor and outdoor water use, an annual factor variable is included in the study. The base year is 2005. The factor variable captures unobserved annual effects. This includes a statewide drought in 2007-2009, and the associated impacts of an aggressive water conservation campaign. It also includes the economic recession in 2008-2009, which anecdotally led to decreased use across active accounts.

The final set of covariates is described in Table 3. Correlations between covariates are available in Table 4. Most variables show little correlation. Exceptions are temperature and rainy days, which are negatively correlated, indicating that hotter areas are also drier. Elevation, slope, and income are positively correlated, echoing the observed distribution of wealthier households on the steeply sloping hillsides within the study area. Presence of a pool, landscape greenness, and income are also positively correlated, indicating that wealthier households may maintain more lush landscape styles.

Table 4. Descriptive statistics of numeric variables.

Variable	Description	Units	N	Mean	Std dev	Min	Max
water	water use	ccf/month	26,315,967	10.80	11.14	1	117
lotsize	lot size	sq ft	322,201	17,343	285,349	49	41,973,458
bldgsize	building size	sq ft	311,635	2,106	3,974	1	147,048
yearblt	year built		303,429	1951	27	1850	2011
elev	mean elevation	meters	327,734	77.07	78.33	-0.57	454.83
slope	mean slope	degrees	327,734	5.18	5.51	0.14	40.38
ndvi	mean NDVI		327,734	0.03	0.05	-0.19	0.25
temp	mean temperature	Celsius	327,734	14.50	0.19	13.67	15.45
rainyday	rainy days	%/month	327,734	11.93	0.75	9.04	13.54
hhsize	mean occupancy	people/ household	327,734	2.73	0.57	0.00	5.55
hhinc	median household income	$/year	327,734	89,300	45,104	0	250,001

Table 5. Correlations for numeric variables. Presence of a pool is treated as numeric, aspect is not included.

	lotsize	bldgsize	yearblt	pool	hhsize	hhinc	Elev	slope	ndvi	temp	rainyday
lotsize	1	0.04	-0.09	0.01	0.02	0.05	0.05	0.04	0.02	0.03	-0.02
bldgsize	0.04	1	0.02	0.05	-0.01	0.07	0.06	0.05	0.02	0.02	0.01
yearblt	-0.09	0.02	1	0.19	0.05	0.04	0.02	0.05	0.10	-0.03	0.02
pool	0.01	0.05	0.19	1	0.06	0.27	0.21	0.06	0.22	0.08	-0.07
hhsize	0.02	-0.01	0.05	0.06	1	-0.03	-0.15	-0.20	-0.15	0.12	-0.30
hhinc	0.05	0.07	0.04	0.27	-0.03	1	0.66	0.40	0.49	-0.03	0.01
elev	0.05	0.06	0.02	0.21	-0.15	0.66	1	0.56	0.46	-0.15	0.11
slope	0.04	0.05	0.05	0.06	-0.20	0.40	0.56	1	0.42	-0.24	0.14
ndvi	0.02	0.02	0.10	0.22	-0.15	0.49	0.46	0.42	1	-0.07	0.08
temp	0.03	0.02	-0.03	0.08	0.12	-0.03	-0.15	-0.24	-0.07	1	-0.70
rainyday	-0.02	0.01	0.02	-0.07	-0.30	0.01	0.11	0.14	0.08	-0.70	1

Estimation strategy

To account for both time varying and time invariant data, this analysis is executed in two steps. The first step evaluates water use through a panel data model with parcel fixed effects. This model explicitly captures the influences of time varying exogenous effects on water consumption—weather and annual effects (e.g., drought and recession). These covariates primarily address effects that are beyond the control of the parcel.

$$water_{it} = \beta_{temp} temp_{it} + \beta_{rainyday} rainyday_{it} + \beta_{year} year_t + FE_i + \varepsilon_{it}$$

The second step regresses parcel fixed effects on the time invariant variables. This model largely explains the role of parcel and demographic characteristics on consumption.

$$\begin{aligned} FE_i = \beta_0 + \beta_{lotsize} lotsize_i + \beta_{bldgsize} bldgsize_i + \beta_{elev} elev_i + \beta_{slope} slope_i \\ + \beta_{aspect} aspect_i + \beta_{ndvi} ndvi_i + \beta_{hhsize} hhsize_i + \beta_{hhinc} hhinc_i \\ + \beta_{yearblt} yearblt_i + \varepsilon_i \end{aligned}$$

This two-stage method has the advantages of integrating variables moving on different time scales, while capturing the serial correlations expected in a single household over time.

It is not a perfect estimation strategy, however. The primary drawback is that the it is not possible to evaluate the combined explanatory power of the model, since the R^2 values from step one and two are not additive. As a result, the models show two R^2 values for each stage of the equation.

This problem of combining data on multiple time scales with unit fixed effects is both recognized and contested in the literature. Plumper and Troeger (2007) propose a method, known as Fixed Effects Variable Decomposition, that executes these first two steps and then combines the two in a third stage by re-estimating with pooled OLS. This method was disputed by Greene (2011), who claimed that the re-estimation was specious and overstates confidence.

Breusch et al (2011) echo this sentiment, asserting that FEVD is a special case of Instrumental Variables and the standard errors should reflect as much.

To sidestep the problem of mixing time varying and time invariant variables, it is possible to avoid fixed effects and evaluate the panel though OLS. In doing so, however, the analysis loses the ability to account for the unobserved factors contributing to serial correlation by parcel. For this reason, we adopt the two-stage approach, but acknowledge its limitations.

Using the two-stage method, we first evaluate all of the data. Then, we compare results from the full data set to sub-populations to examine heterogeneous responses among water users. Firstly, we examine groups by tercile—high, moderate, and low water users. This method is similar to those employed previously by Mini, Hogue and Pincetl (2014) and Kenney et al (2008). The tercile classification is evaluated, mapped, and compared through a mulitnomial logit model in appendix 4. Of note is the mean water use of each user group: low, moderate, and high households consume an average of 110, 218, and 502 gpd, respectively. Appendix 4 reveals how, relative to average users, low water users are in cooler and more rainy areas, they have smaller household sizes, fewer pools, and drier landscape styles. Relative to moderate water users, high water users are in hotter areas, have slightly larger household sizes, have more pools, and more lush landscape styles.

Secondly, we examine groups by climate zone. Miroclimate zone 1 is hilly, cooler, and more vegetated. Zone 2 is the foothills, on moderate and warm slopes. Zone 3 is the coolest, flattest, and most urbanized. (See chapter 3 for more information.) These zones are another method of identifying heterogeneity among households. Do different zones relate to different household water consumption profiles? EBMUD's use of evapotranspiration zones indicates that there should be a larger fraction of water applied outside in some regions of the service area.

Lastly, we examine the covariates through simple linear models to evaluate the regression coefficients of individual variables while controlling for annual effects. This reveals variable loading without covariate interactions. We compare these results to the results from the two-stage panel models.

Results

Regression values are summarized in three tables. Table 5 presents results from analysis by tercile. Table 6 presents results from analysis by climate zone. In both tables, the results from the full data set is available in the first column. Table 7 presents results from the bivariate regressions.

By user group

When analyzing all households, the models exhibit fairly strong explanatory power—17% of variance is explained in the first stage, while 36% of variance is explained in the second stage. There are differences among the sub-populations, howevers. The behaviors of low water users and average water users are poorly explained by the models, while high water users are well explained. The difference in the explanatory power of the model is particularly evident in the second stage. Only 4% and 5% of the variance is explained for low and moderate water users.

Table 6. Water use by quantile. North-facing aspects (factor, Aspect: North) were not present in the Moderate or High terciles. Even so, northerliness did not have a significant contribution to household consumption. Those variables marked as 0.000 had a very small, though non-zero coefficient that is significant.

Variable	All		Low		Moderate		High	
	β	(SE)	β	(SE)	β	(SE)	β	(SE)
temp	0.79^{***}	(0.001)	0.10^{***}	(0.0005)	0.33^{***}	(0.0006)	1.53^{***}	(0.001)
rainyday	-0.01^{***}	(0.0002)	-0.01^{***}	(0.0001)	-0.03^{***}	(0.0002)	-0.07^{***}	(0.0004)
year: 2006	0.13^{***}	(0.005)	-0.23^{***}	(0.004)	0.01^{***}	(0.005)	0.60^{***}	(0.01)
year: 2007	0.18^{***}	(0.01)	-0.58^{***}	(0.004)	-0.30^{***}	(0.005)	1.03^{***}	(0.01)
year: 2008	-0.66^{***}	(0.01)	-0.98^{***}	(0.004)	-1.03^{***}	(0.005)	-0.50^{***}	(0.01)
year: 2009	-1.54^{***}	(0.01)	-1.22^{***}	(0.004)	-1.70^{***}	(0.005)	-2.25^{***}	(0.01)
year: 2010	-1.46^{***}	(0.01)	-1.05^{***}	(0.004)	-1.52^{***}	(0.005)	-2.08^{***}	(0.01)
year: 2011	-1.25^{***}	(0.01)	-0.90^{***}	(0.004)	-1.54^{***}	(0.005)	-2.00^{***}	(0.01)
N	26,315,967		8,226,017		9,910,137		7,838,682	
Adjusted R^2	0.17		0.04		0.13		0.34	
lotsize	0.0000^{***}	(0.0000)	0.0000^{***}	(0.0000)	0.0000^{***}	(0.0000)	0.0000^{***}	(0.0000)
elev	0.01^{***}	(0.0002)	0.0004^{***}	(0.0002)	0.001^{***}	(0.0001)	0.003^{***}	(0.0005)
slope	-0.19^{***}	(0.003)	-0.02^{***}	(0.002)	-0.02^{***}	(0.001)	-0.03^{***}	(0.01)
aspect: north	0.35	(4.76)	-2.45	(1.56)	--	--	--	--
aspect: east	0.63	(4.76)	-2.38	(1.56)	0.03^{*}	(0.02)	0.15^{*}	(0.08)
aspect: south	0.69	(4.76)	-2.42	(1.56)	-0.01	(0.02)	0.73^{***}	(0.08)
aspect: west	0.28	(4.76)	-2.47	(1.56)	-0.05^{***}	(0.02)	0.49^{***}	(0.08)
hhinc	0.0000^{***}	(0.0000)	0.0000^{***}	(0.0000)	0.0000^{***}	(0.0000)	0.0000^{***}	(0.0000)
hhsize	1.07^{***}	(0.02)	0.33^{***}	(0.01)	0.20^{***}	(0.01)	0.52^{***}	(0.06)
bldgsize	0.0000^{***}	(0.0000)	-0.000^{***}	(0.0000)	-0.000^{***}	(0.0000)	0.001^{***}	(0.0000)
yearblt	0.04^{***}	(0.001)	0.01^{***}	(0.0003)	0.01^{***}	(0.0002)	0.02^{***}	(0.001)
pool	10.14^{***}	(0.05)	2.31^{***}	(0.07)	0.90^{***}	(0.03)	6.71^{***}	(0.07)
ndvi	37.49^{***}	(0.29)	2.36^{***}	(0.19)	4.83^{***}	(0.19)	40.03^{***}	(0.58)
Constant	95.27^{***}	(4.87)	-12.94^{***}	(1.67)	-10.42^{***}	(0.47)	-43.97^{***}	(2.34)
N	302,807		94,818		113,157		90,865	
Adjusted R^2	0.36		0.04		0.05		0.32	

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

There are similarities among many of the coefficient values. Income is always statistically significant and near zero, though this is possibly this is an artifact of using the block group-level data. Household size contributes to use, despite its block group resolution. Lot size and building size are both consistently significant and near zero. In all models, the presence of a pool and green landscape styles make the most substantial contribution to use.

Some variables show marked differences among the models. Aspect is not significant in analysis of the full data. Among high water users, however, southerly and westerly aspects—the warmest aspects—become significant. All north-facing parcels are located in the lowest tercile of water users, even though northerliness does not significantly contribute to water consumption.

By climate zone

As with results by tercile, there is evidence of differences in the explanatory power of the model by microclimate zone. Notably different is the explanatory power of the second stage. 49% of the variance is captured in zone 1, 36% in zone 2, and 13% in zone 3. Climate zones are better captured by the model than the terciles.

The coefficient values from the models are similar in zones 1 and 2—the hillier, more vegetated zones—even though the explanatory power differs. Both are sensitive to temperature, aspect, and vegetation. Zone 3—the more urbanized zone—reveals a smaller relationship to NDVI. Water use does not strongly relate to lot size or income in any of the zones.

Table 7. Water use by climate zone

Variable	All		Climate 1		Climate 2		Climate 3	
	β	(SE)	β	(SE)	β	(SE)	β	(SE)
temp	0.79***	(0.001)	1.20***	(0.002)	1.13***	(0.001)	0.35***	(0.001)
rainyday	-0.01***	(0.0002)	-0.04***	(0.001)	-0.04***	(0.0004)	-0.02***	(0.0001)
year: 2006	0.13***	(0.005)	0.23***	(0.02)	0.27***	(0.01)	0.02***	(0.004)
year: 2007	0.18***	(0.01)	0.60***	(0.02)	0.36***	(0.01)	-0.15***	(0.005)
year: 2008	-0.66***	(0.01)	-0.67***	(0.02)	-0.93***	(0.01)	-0.76***	(0.005)
year: 2009	-1.54***	(0.01)	-2.00***	(0.02)	-2.12***	(0.01)	-1.40***	(0.005)
year: 2010	-1.46***	(0.01)	-1.72***	(0.02)	-1.86***	(0.01)	-1.30***	(0.005)
year: 2011	-1.25***	(0.01)	-1.38***	(0.02)	-1.65***	(0.01)	-1.29***	(0.005)
Observations	26,315,967		3,197,364		6,958,328		15,800,619	
Adjusted R^2	0.17		0.26		0.27		0.09	
lotsize	0.0000***	(0.0000)	0.0000***	(0.0000)	0.0000***	(0.0000)	0.0000***	(0.0000)
elev	0.01***	(0.0002)	-0.01***	(0.001)	0.003***	(0.001)	0.01***	(0.001)
slope	-0.19***	(0.003)	-0.32***	(0.01)	-0.16***	(0.01)	-0.10***	(0.004)
aspect: north	0.35	(4.76)	--	--	--	--	0.74	(3.57)
aspect: east	0.63	(4.76)	0.57***	(0.15)	0.27***	(0.09)	0.93	(3.57)
aspect: south	0.69	(4.76)	1.27***	(0.13)	0.75***	(0.09)	0.52	(3.57)
aspect: west	0.28	(4.76)	0.58***	(0.13)	0.52***	(0.08)	0.25	(3.57)
hhinc	0.0000***	(0.0000)	0.0000***	(0.0000)	0.0000***	(0.0000)	0.0000**	(0.0000)
hhsize	1.07***	(0.02)	1.21***	(0.15)	0.37***	(0.08)	0.86***	(0.02)
bldgsize	0.0000***	(0.0000)	0.004***	(0.0001)	0.0000	(0.0000)	-0.0000	(0.0000)
yearblt	0.04***	(0.001)	0.04***	(0.003)	0.08***	(0.001)	0.03***	(0.0005)
pool	10.14***	(0.05)	9.06***	(0.13)	9.10***	(0.08)	7.01***	(0.08)
ndvi	37.49***	(0.29)	60.05***	(1.02)	55.30***	(0.61)	20.43***	(0.33)
Constant	95.27***	(4.87)	-97.05***	(4.95)	-175.5***	(2.76)	-51.53***	(3.69)
Observations	302,807		35,373		80,754		182,712	
Adjusted R^2	0.36		0.49		0.36		0.13	

*Note: *p<0.1; **p<0.05; ***p<0.01*

Bivariate regressions

Results from the bivariate regressions are relatively consistent with covariate estimations found in the two-stage model. Effects of rainy days, year built, presence of pool, and NDVI are all slightly stronger than in the full panel model. The explanatory power is notably lower, however. Regressions on lot size and year built only capture 1% of the variance in water use.

Table 8. Bivariate linear regressions of select variables, controlling for year.

Variable	1	2	3	4	5	6	7	8
temp	0.80^{***} (0.001)							
rainyday		-0.18^{***} (0.0002)						
lotsize			0.0000^{***} (0.0000)					
hhinc				0.0001^{***} (0.0000)				
hhsize					1.18^{***} (0.004)			
yearblt						0.10^{***} (0.0001)		
pool							14.19^{***} (0.01)	
ndvi								59.19^{***} (0.04)
year: 2006	0.14^{***} (0.01)	-0.09^{***} (0.01)	0.05^{***} (0.01)	0.04^{***} (0.01)	0.04^{***} (0.01)	0.05^{***} (0.01)	0.05^{***} (0.01)	0.06^{***} (0.01)
year: 2007	0.26^{***} (0.01)	-0.86^{***} (0.01)	0.14^{***} (0.01)	0.12^{***} (0.01)	0.11^{***} (0.01)	0.14^{***} (0.01)	0.14^{***} (0.01)	0.16^{***} (0.01)
year: 2008	-0.54^{***} (0.01)	-1.79^{***} (0.01)	-0.63^{***} (0.01)	-0.67^{***} (0.01)	-0.71^{***} (0.01)	-0.63^{***} (0.01)	-0.64^{***} (0.01)	-0.62^{***} (0.01)
year: 2009	-1.43^{***} (0.01)	-2.53^{***} (0.01)	-1.64^{***} (0.01)	-1.67^{***} (0.01)	-1.74^{***} (0.01)	-1.64^{***} (0.01)	-1.64^{***} (0.01)	-1.62^{***} (0.01)
year: 2010	-1.39^{***} (0.01)	-1.98^{***} (0.01)	-1.90^{***} (0.01)	-1.93^{***} (0.01)	-2.02^{***} (0.01)	-1.90^{***} (0.01)	-1.90^{***} (0.01)	-1.88^{***} (0.01)
year: 2011	-1.16^{***} (0.01)	-2.52^{***} (0.01)	-1.79^{***} (0.01)	-1.80^{***} (0.01)	-1.90^{***} (0.01)	-1.79^{***} (0.01)	-1.79^{***} (0.01)	-1.95^{***} (0.01)
Constant	-0.20^{***} (0.01)	14.39^{***} (0.01)	11.60^{***} (0.01)	4.57^{***} (0.01)	-184.8^{***} (0.16)	8.42^{***} (0.01)	10.65^{***} (0.01)	9.73^{***} (0.01)
N	26,315,967	26,315,967	25,521,997	25,956,311	24,180,951	25,956,311	25,956,311	25,956,311
Adjusted R^2	0.08	0.05	0.01	0.11	0.06	0.01	0.11	0.08

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Discussion

All models indicate that landscape style (NDVI) and the presence of a pool are the most important contributors to demand. All models indicate that households in hotter and drier regions use slightly more water than cooler and damper regions. All models indicate a small relationship between water use and household size. None of the two-stage models indicate that lot size or income is a large factor in consumption, but the bivariate regression points to a slight relationship to income. These income effects may be absorbed by other covariates in the multivariate models.

The tercile analysis reveals different water consumption behavior patterns among user groups. Low water users, who use little water outside, are poorly explained by the model. High water users are explained very well by the model. They are sensitive to their environmental conditions and use even more water when located in an especially evaporative microclimate. Microclimate factors are not nearly as strong and landscape style and pools, however, indicating that microclimate may cause high water users to consume slightly more, but are not sufficiently influential to cause households to change class of users.

Moderate water users are most challenging to characterize. They are partly explained by the weather and annual effects captured in stage one, but poorly explained by their parcel and demographic characteristics. The descriptive statistics for the moderate users (available in appendix 4) do not indicate a wider range in the covariates.

Climate zones are similarly difficult to interpret. Why is microclimate zone 1 described so well by the model, particularly in the second stage? The ambiguity associated with the climate zones (as described in chapter 3) is only exacerbated when used to spatially subset the EBMUD population. Yet, despite weaknesses in the climate zones, they illuminate that there are many different types of sub-populations within EBMUD, each exhibiting slightly different water demand profiles.

In further evaluating EBMUD, there is potential for continued refinement of both the models and several of the covariates. Initial exploration of standardized coefficients confirms that NDVI and pools are the most important model variables, yet further evaluation of lot size is indicated. In addition to improving explanatory power by transforming the data, a direct evaluation of the relationship between lot size and land cover choice may prove valuable.

Further, NDVI must be critically examined as a model variable in EBMUD. NDVI may be capturing two unintended landscape effects. Its intent is to model mesic versus xeric landscape styles (i.e., turf grass versus drought-tolerant landscaping), which it did with 86% accuracy in a study by Wentz and Gober (2007) in Phoenix. It is possible that NDVI was more effective in Phoenix, however, because irrigated vegetation is more distinct from the surroundings. In EBMUD, NDVI's positive correlation with slope signals that NDVI is capturing the greenness of naturalized landscapes in the hills, like eucalyptus trees, which are not likely irrigated. In addition, in EBMUD, NDVI may not adequately differentiate between poorly maintained turf grass and xeric landscapes. If NDVI is a stronger indicator of landscape maintenance than style, the variable could be endogenous to demand.

Conclusions

This chapter reveals that there are many different methods of identifying heterogeneity in water users. Analysis of the full EBMUD population alongside sub-populations provides several different viewpoints from which to interpret the relationships between household water consumption and variables related to outdoor water use. Though there are some similarities across all users, the analysis reveals that there is not likely a median user that is representative of the full spectrum of EBMUD households.

Disaggregate data makes it possible to tease apart these different patterns. One question the household data reveal is that high water users, spatially distributed in almost every block group and census tract, may be dominating the patterns observed in more aggregate analyses. High water users consume so much more water than low and average users, both in the winter and in the summer, that their patterns may dwarf other households. The variable coefficients for the full data set align more closely with the high water users than the other terciles.

Yet, while there is some indication that coarser aggregations have the potential to introduce ecological fallacy, this requires a thorough investigation. Against which aggregations do inconsistencies arise? Are there ever times at which the aggregate analysis is more accurate than the individual analysis, given the available resolution of model covariates?

Conclusion: Planning for water efficient cities

The imperative to design for resource efficient cities is widely discussed in planning practice and scholarship. How to do so is debated, with solutions crossing policy and design. Urban form is a popular approach, seeking to realize resource efficiency through compact development and strategic growth. Resource efficient design principles underlie many green neighborhood guidelines, such as LEED for Neighborhood Development (Congress for the New Urbanism, Natural Resources Defense Council, and U.S. Green Building Council 2011). They are also suggested by the California Department of Water Resources and the U.S. EPA as a means of reducing water consumption (see chapter 1). This chapter examines the possibilities and limitations of urban form as a tool to manage water use.

Lessons for planning: results from analysis of single family homes in EBMUD

There are an increasing number of empirical analyses that evaluate form and consumption. While most relate to transit, building energy, or land conservation, this chapter gives a rare focus on water use. There are limits to this analysis: it does not offer direct comparisons between single family and multi-family or mixed use development; its identified relationships are specific to the region. Even so, lessons from the case of EBMUD can be extracted and considered in this broader planning context.

1. Density is less important than land cover

In EBMUD, there is evidence that small lots can use a lot of water, while large lots can consume little. In the models presented in chapter 4, the relationship between lot size and consumption is statistically significant, but small. This is the case when lot size is analyzed as the only controlling variable; when the data is analyzed as a whole; when separated into low, moderate, and high users; when analyzed by climate zone; and when analyzed in a multinomial logit model (see appendix 4 for logit results). While it may be possible to identify a stronger relationship between consumption and lot size under another set of conditions, the data reveals that lot size is not as important as landscape type and the presence of pools.

These findings are consistent with other reports in the literature that landscape style matters more than its size (Domene and Saurí 2006). It also echoes findings that stated environmental ethics do not always reflect water consumption habits (Harlan et al. 2009)—some people who identify as being environmentalists enjoy gardening and maintain lush landscapes (Syme et al.

2004). Within EBMUD, there are many smaller lots that are using water intensively, possibly for vegetable gardens, if not turf. There are also some very large lots that use little water because only small portions of the whole lot are irrigated. Most certainly, there are homes with expansive lawns and pools that consume large volumes of water, but neither land cover choices nor subsequent water consumption are universally determined by lot size.

Based on these findings, there is little indication that, in EBMUD, more compact single family residential development would necessarily reduce water consumption. These findings relate to ongoing discussions within planning on the value of compact development, where critics are concerned that the benefits of compact growth may be overstated (e.g., Echenique et al. 2012). Sustainable development advocate Michael Neuman suggests that the compact city is the "default model" and writes (2005, 14):

> Have not planners relied on the compact city, a model inspired mainly by old European and North American cities and towns, because there is no clear alternative? One reason why there is no alternative is that many do not know what to do to be sustainable. Yet we assume it is different, perhaps radically different, from the way we live now. If this is true, then why do urban planners favor the centuries-old model of the compact city? Why do they use old strategies to accomplish a fundamentally new task?

Similarly, architectural historian Richard Bruegmann writes that there is little evidence new suburban forms (i.e., New Urbanist developments) are more resource efficient than existing suburbs (Bruegmann 2005). He suggests that, instead, arguments for compact development are critiques of taste (Bruegmann 2008). Judgments on the form of "sprawl," taken to be the opposite of compact development, abound in planning (Hayden and Wark 2004). The New Urbanist definition of sprawl, "cookie-cutter houses, wide, treeless, sidewalk-free roadways, mindlessly curving cul-de-sacs, a streetscape of garage doors" (Duany, Plater-Zyberk, and Speck 2000), is a judgment on form, not function.

Perhaps analysis of multifamily and mixed use development alternatives will indicate a larger relationship between water use and lot size. New Urbanists suggest that compact developments will reduce resource use by sharing open space (Duany, Plater-Zyberk, and Speck 2000). Empirical research on the subject is nascent. A recent dissertation by Eve Halper (2011) finds some nuance in the trade-offs between household water use the presence of a public park. Halper writes that small, local parks may positively substitute for household use, while large parks may be more consumptive. In her words, "small parks tended to be net water savers, most large parks consumed more water than they suppressed" (Halper 2011, 64). She notes, however, that large parks may irrigate with recycled water, so may be beneficial despite higher water requirements.

There are a range of advantages associated with compact development that do not relate to household water demand management. Supporters of compact development indicate that there will be lower costs in providing and maintaining infrastructure connections, including water and wastewater infrastructure (Burchell et al. 2002). There may be fewer vehicle miles traveled by residents in compact developments (Cervero and Murakami 2010). There may be less building energy used (Ewing and Rong 2008; Randolph 2008). Compact development may even lead to some social gains, with possibilities of increasing inclusion and building community (Bramley and Power 2009).

Even if all of these benefits hold true, however, density is not a sufficient condition to ensure low water use. Land cover choices matter, regardless of lot size. A focus on compact development as a means of reducing use may undermine the most important lesson: selecting drought-tolerant landscape styles and avoiding pools can make most households water efficient.

2. Microclimate influences high water users the most

While many studies observe that climate variables relate to water consumption across broad spatial extents, this study finds that there is a small, but notable relationship between water use and weather variables within a single service area. The landform factors that contribute to microclimates do not significantly relate to water consumption across the whole distribution of water users, however. Instead, each class of water user exhibits a different set of relationships to landform. High water users are most sensitive to all microclimate conditions.

Results from chapter 3 and appendix 4 indicate that low water users consume little water outside. This may signal that they do not maintain their landscaping. Or, it may mean they adapted their landscape style with plants that do not require additional water in their microclimate zone. Though low water users are located at a greater density on the western side of EBMUD, they are distributed throughout the service area.

High water users consume a far larger amount of water outdoors to maintain more lush landscape styles or pools. Warmer, southerly and westerly aspects contribute to high water use, but these factors are generally not strong enough to change a household's usage class. The average winter water use among high water users—the smallest recorded water use for the year, associated with indoor consumption—is higher than the peak summer use for low or moderate water users. If a high water using household were rotated to another aspect, they would likely still be among the high water users. High water users are located at a greater density on the eastern side of the study area, but they are present in all microclimates, dwellings, and demographic categories.

Teasing apart the contributions of microclimate and social norms remains challenging. Within EBMUD, do correlations between water use and weather follow a social topography? In some places, Home Owners Association rules may confine landscape choice. In most places, however, households have the ability to adapt water efficient landscaping styles. There may be unobserved social trends across the study area that contribute to landscape decisions and subsequent household use.

3. There are multiple, simultaneous land capabilities

This dissertation was initially motivated by the regionalist framework for strategic growth embodied by environmental planner Ian McHarg. In McHarg's seminal book, *Design With Nature*, he asserts that nature is not a "uniform commodity" (1994, 56) and promotes growing in the most environmentally suitable locations through a method he calls "physiographic determinism" (1994, 81). He writes, "physiographic determinism suggests that development should respond to the operation of natural processes" (1994, 81) and "If growth responds to natural processes, it will be clearly visible in the pattern and distribution of development —and, indeed, its density" (1994, 161). He suggests that urban morphology should emerge from landscape process.

One of the questions this dissertation sought to resolve is whether there are natural processes that development could respond to, in order to reduce water demand. Results indicate, however, that environmental processes do not singularly define high water use zones. Low, moderate, and high water users exist everywhere. This revealed spatial heterogeneity in consumption indicates that, with respect to water use, land capability has less to do with environmental process, and much more to do with how households relate to those processes. Land capability is relative to the land's user. There is no single land capability, but multiple, simultaneous capabilities.

While strategic growth is an appropriate method of approaching many resource management problems, it may not be as useful in managing outdoor water demand on the scale of the water district. Within EBMUD, microclimate alone is not a strong predictor of water use—placed in the same microclimate and in the same dwelling, two different households will consume differently.

Resource flexibility as a planning target

Results from this analysis indicate that individual household water use is likely too narrow of a scope by which to define the problem of demand management in the context of land planning. There are many different methods of drawing boundaries on water demand, both spatially and temporally, that may allow planners to more effectively engage with residential outdoor water use.

Decreased reliance on imported water

Perhaps less important than reducing residential water use, per se, is reducing total consumed volumes of imported, municipal drinking water. While it may be difficult to invoke land planning tools to manage household consumption volumes, land planning will be a component of larger strategies to reduce the total flux of water through cities and suburbs.

Water supply portfolios that draw from multiple sources are essential in California water planning today. These water supplies can come from the large water projects described in chapter 1, local surface water projects, local groundwater, or local water recycling. Though not developed in this dissertation, planners have long engaged with improving both local surface water and groundwater supplies by managing hydrologic impacts of the city. It is possible to join these efforts with newer advances in recycled water. Methods of locally collecting wastewater, treating this water, and then distributing recycled water will be an important evolution in planning.

If recycled water is to become a meaningful supply in California, it will likely integrate with surface water and groundwater replenishment efforts by developing planned alternatives that combine engineered and natural infrastructures. It is possible the future water efficient city is less about reducing household use than integrating systems to build the local capacity to stretch one gallon further.

Responsive consumption

In EBMUD, there is evidence that patterns in water use respond slowly to water supply conditions. The annual effects assessed through the models in chapter 3 indicate that, during the drought of 2007-2009, water use declined slowly. Even though water restrictions were put in place only in 2008, water use continued to decrease through 2009. It remained low in 2010, and then started to increase again. The observed patterns may also reflect effects of the recession, but nonetheless raise the question of responsive water use.

One of the biggest challenges facing California is not the drought, but variability. California has very wet years and very dry years. Water efficiency is only a challenge half the time—often, the state is contending with problems of too much water, like flooding. During California's wet years, consuming water can even be a good thing. Living on a fluctuating landscape is incredibly difficult from a policy and design perspective, however. How does EBMUD establish a flexible water use pattern that is able to scale back quickly during dry years?

Likely the variable that has the most potential to rapidly influence water consumption is price. If water pricing were adapted to reflect real marginal costs at the moment of consumption, this could very efficiently lead to changes in water use, while allowing households express preferences. The most appropriate pricing structure is not the focus of this analysis (e.g. Zetland 2011), but this dissertation still concludes that price is an essential mechanism for a responsive water system.

Implementing transparent and dynamic pricing—reflecting costs of water in real time, while providing instantaneous feedback—is a real possibility. There are legal hurdles to clear, including a recent battle in California to with respect to pricing and Proposition 218, but such technology is an important goal.

Performance-based metrics

In the future city, resource performance is critical. While unit counts, compact development styles, and regionalist approaches to strategic growth have served as proxies for resource use in the past, new technologies will allow planners to decouple performance and form. In doing so, it will be possible to more effectively evaluate resource management issues, while allowing urban and suburban residents to exert their housing preferences.

Focus on urban form is an understandable tendency. Performance is difficult to measure. Performance requires ongoing data collection, analysis, and adaptation as new information emerges. Form, on the other hand, can be implemented through land planning and design guidelines without further evaluation. Yet, form is also reductive. To create an urban environment that is simultaneously efficient and human-centered, it will be necessary to accommodate social heterogeneity, whenever possible. As Michael Neuman writes (2005, 16), "Livability is not only a matter of urban form, it is also a matter of personal preference." The water efficient city does not need to constructed on a narrow set of design principles.

Future research

This dissertation is fundamentally about identifying patterns in socioecological systems by evaluating heterogeneous, spatiotemporal data and by critically considering scale and classification. Through this examination, it exposes many new research avenues.

Disaggregate, extensible forecasting

Chapter 2 identifies a need for disaggregate, extensible forecasting. This dissertation makes progress toward disaggregate analysis, but does not answer problems of extensible forecasting. In fact, it raises the issue that there is not a median water user whose demand profile is representative of EBMUD, which indicates it is not probable that the median characteristics of EBMUD would be representative of another location at another time.

For this reason, it is not possible to generate precise, parcel-based forecasts. New development that is not yet occupied will be even more difficult to model. Household behavior has a such a significant impact on water consumption that the physical attributes of the property, alone, explain little variation in water use. In the models in chapter 4, less than half of the variance observed among water users is explained, even while including income and household data. This indicates that there are likely unobserved behavioral components that strongly influence consumption patterns.

It may possible to generate forecasts that predict consumption within some degree of probability. Running comparative scenarios with different land cover types or weather parameters could generate trajectories of future water consumption within a band of uncertainty.

With access to additional water consumption data for a new water agency, the models and variables developed in this dissertation could be applied to almost any location in the United States. Do the relationships observed across EBMUD hold true in other regions? Does the median household in EBMUD behave like median households in other regions? If not, why?

Landscape sensing

Moving from form-based to performance-based urban design will require methods of high resolution environmental monitoring and change detection. It also raises the question of scale. At which aggregation is it best to evaluate continuously varying environmental phenomena? Over which time periods is it necessary to detect change? High resolution landscape sensing will lead to more accurate characterizations of socioecological relationships.

Form and resource consumption

This dissertation provides insight into the form of single family residential development in EBMUD. There is additional research necessary on the relationship between form and consumption in more geographic locations and with a broader spectrum of development typologies. While there are many data barriers to evaluating multi-family and mixed use development, analyses of these housing types will best clarify the value of compact development as a method of managing demand.

It is also possible, however, that the resource efficient development principles of compact development and strategic growth are utopian. Planning has a long history of suggesting that new urban forms are necessary to counter the resource management problem of the era. Are new forms the answer now? Continuing to reconcile urban and landscape theory against empirical evidence will be a useful area of research.

Appendix 1. Detecting landscape change by parcel with NAIP imagery

Detecting parcel-level vegetation change within an urban gradient is a developing area of research. Until recently, four-band, multispectral imagery was not widely available at a sufficiently high resolution to examine parcels. As of 2009, the National Agricultural Imagery Program (NAIP) began to regularly collect aerial photography that included the near infrared (NIR) band needed for many surface calculations. This is the first publicly available imagery in the United States with high spatial resolution, broad coverage, and NIR.

Using the red and NIR bands of the multispectral imagery, it is possible to examine vegetative greenness and density. The most common metric is known as the Normalized Difference Vegetative Index (NDVI) and is valuable method of assessing landscape type. NDVI is measured on a scale of -1 to 1. Living plant tissue with active chlorophyll scores positive values. Verdant areas score higher than desert landscaping. Bare rock, soil, and hardscaping score negative values.

To use NDVI as a variable in a time series analysis, it is necessary for NDVI values to align over time: vegetation that stays the same is consistently represented with the same value, while changes in NDVI value indicate that the landscape has, indeed, changed. This presents the most significant challenge in evaluating NDVI with NAIP. NAIP imagery is not standardized against a common benchmark and there are differences in the images from year to year due to atmospheric and seasonal effects. There are also slight image alignment changes from year to year.

This appendix evaluates the utility of NAIP imagery in a time series, parcel analysis. First, it calibrates the imagery for the study area to a common benchmark. Then, it calculates NDVI, comparing mean parcel NDVI values across years. Ultimately, it concludes that NAIP imagery does not accurately reveal parcel-level landscape change.

Calibrating imagery

In California, NAIP with NIR is available for three years of the study period: 2005, 2009, and 2010. A portion of the seasonal effects in NAIP are due to different fly dates from year to year. NAIP is intended to capture vegetation that is "leaf-on." The images were flown in mid-June in 2005, mid-June to mid-July in 2009 (depending on the sub-region), and late May in 2010.

The 2005, 2009 and 2010 images can be calibrated to a common scale using the pixel-based processing technique known as radiometric normalization with pseudo-invariant targets. In this appendix, the calibration proceeds in four steps:

1. Identifying locations (pseudo-invariant targets) across the study area that remain the same over time.
2. Isolating the red and NIR values at the target locations for all years.

3. Using 2005 imagery as a baseline, regressing 2005 values against 2009 values and 2010 values.
4. With the intercept and coefficient from the regression equations, adjusting 2009 and 2010 imagery.

I placed sixty points in locations across study area, choosing targets that appear constant over the duration of the study period based on visual inspection of the aerial imagery (see Figure 1). These features included water (lakes, pools), sand (e.g., baseball diamonds), asphalt, and roof tops, with special emphasis on light and dark colored objects (see Figure 2).

Figure 29. Locations of the sixty pseudo-invariant targets used in imagery calibration.

2005 2009 2010

Figure 30. A sample target site in a stand trap across the available NAIP imagery. The imagery from 2005 did not include a blue band. The images from 2009 and 2010 have red, green, blue and NIR bands. In the case of this sand trap, the point location lined up well with the target across all three years. In some cases, the images were shifted enough that the point did not intersect the same location on all layers.

For each point, I extracted the value of the Red and NIR bands. If there were no surface reflectance changes, the values of the pseudo-invariant would remain constant over time. Instead, the target points exhibit differences between years. The calibration process identifies the relationship between the baseline image (2005) and subsequent images. A linear relationship between images can be assumed.

I regressed the values of the red and NIR bands at the target sites in the 2005 images on the values from 2009 and 2010 (see Figure 3):

$$2005_{red} = \alpha_{2009\text{-}red} + \beta_{2009\text{-}red} * 2009_{red} + \varepsilon_{2009\text{-}red}$$

$$2005_{nir} = \alpha_{2009\text{-}nir} + \beta_{2009\text{-}nir} * 2009_{nir} + \varepsilon_{2009\text{-}nir}$$

$$2005_{red} = \alpha_{2010\text{-}red} + \beta_{2010\text{-}red} * 2010_{red} + \varepsilon_{2010\text{-}red}$$

$$2005_{nir} = \alpha_{2010\text{-}nir} + \beta_{2010\text{-}nir} * 2010_{nir} + \varepsilon_{2010\text{-}nir}$$

Regression results are summarized in Table 1.

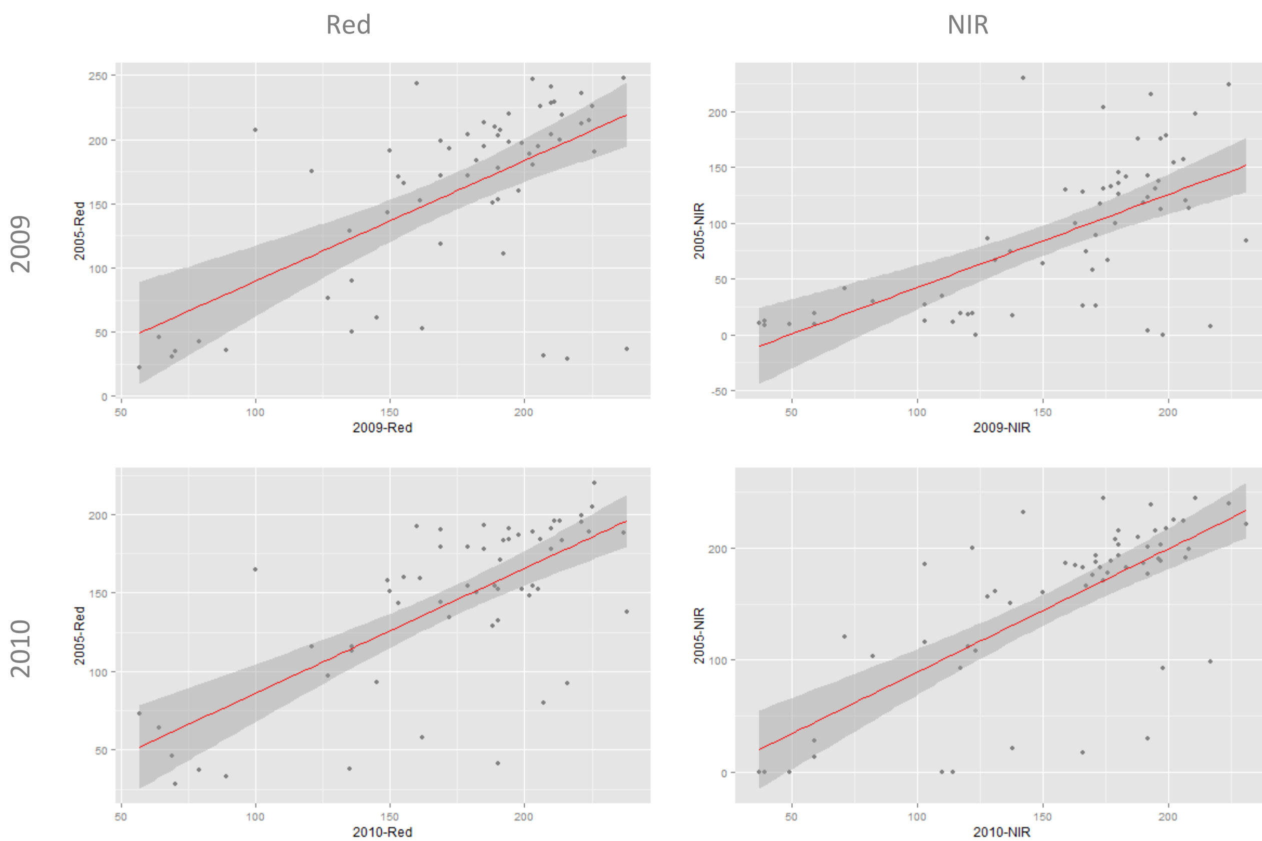


Figure 31. NAIP image calibration regressions. The error in the regressions reveals local atmospheric effects, local seasonal effects, and that the target site is not perfectly capturing the same feature in each year due to alignment issues. In the north-east region of the study area, the images were particularly poorly aligned.

Table 9. NAIP regression results used in image calibration. $p < 0.01$ for all intercepts and coefficients.

	Band	α	β	R^2
2009	Red	110.34	0.40	0.37
	NIR	112.39	0.48	0.40
2010	Red	84.21	0.62	0.49
	NIR	83.00	0.48	0.53

Finally, I adjusted the 2009 and 2010 bands by multiplying each by their respective coefficient (β), and then adding the value of the intercept (α).

Calculating NDVI

I executed NDVI calculations in ArcGIS with ArcPy using the following equation:

$$\text{NDVI} = \frac{(\text{NIR} - \text{Red})}{(\text{NIR} + \text{Red})}$$

The original image, the derived NDVI, and the calibrated NDVI are shown in Figure 4.

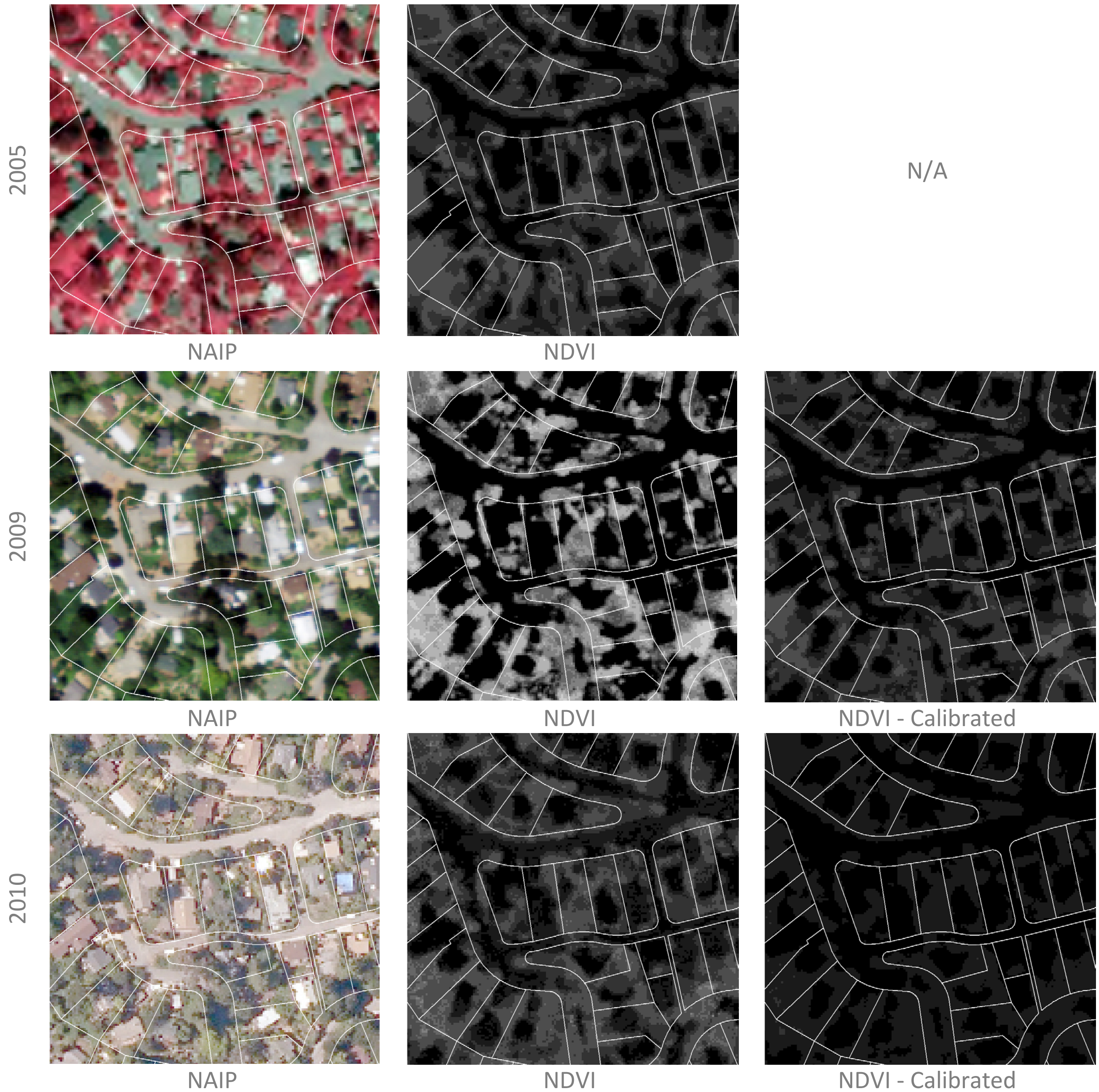


Figure 32. Original NAIP imagery, NDVI, and calibrated NDVI for a set of parcels within the study area.

I calculated the mean NDVI value for each parcel. A large portion of most parcels is occupied by hardscape, resulting in a fairly narrow range in the summarized NDVI values. Figure 5 shows the distribution of mean parcel NDVI before and after image calibration. Table 2 includes summary statistics.

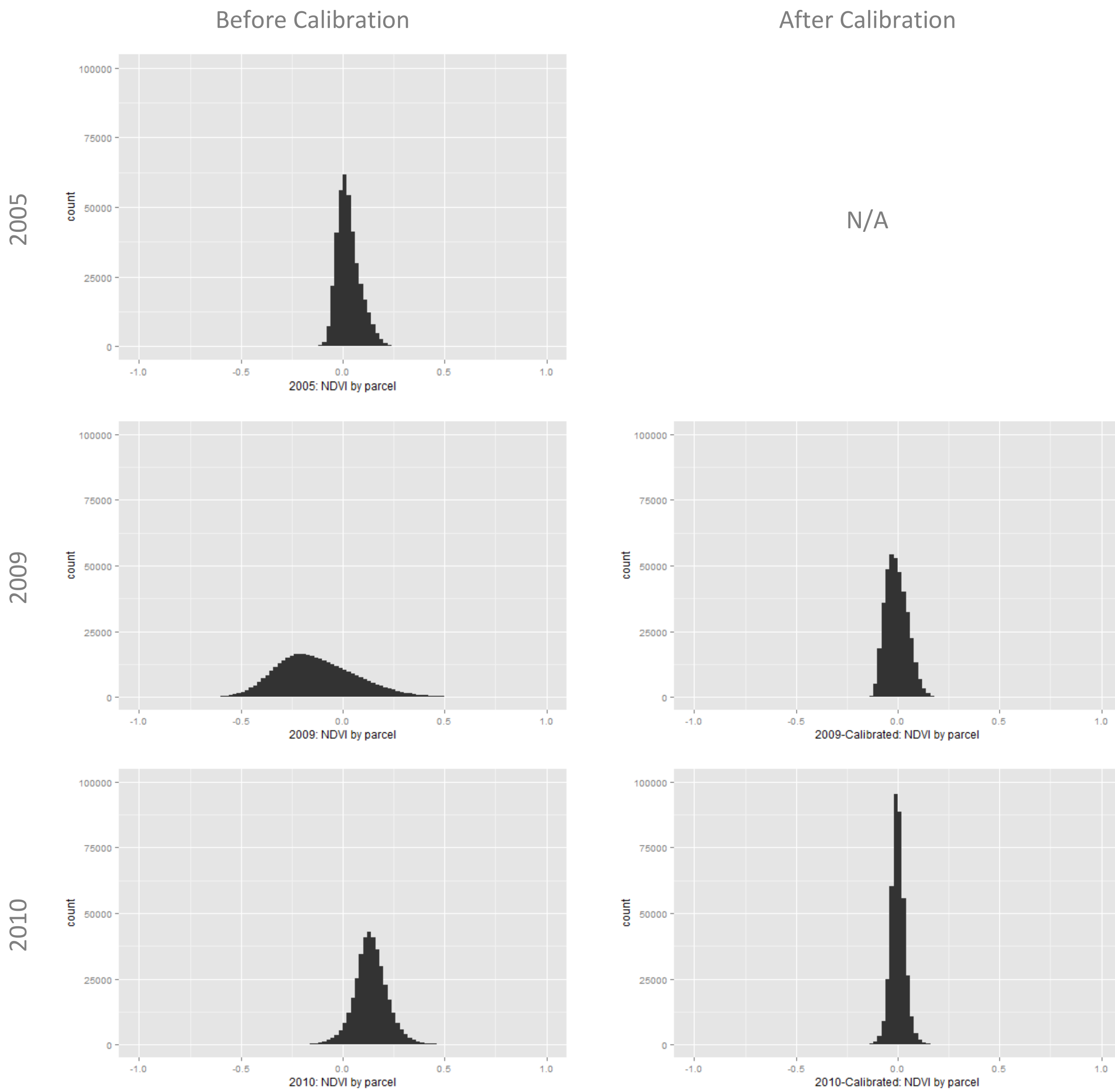


Figure 33. Comparing NDVI values before and after calibration. After, the distribution of NDVI values in 2009 and 2010 becomes narrower, reflecting the distribution of the 2005 imagery.

Table 10. Minimum, maximum and mean values of NDVI data.

	Before Calibration			After Calibration		
	Minimum	Maximum	Mean	Minimum	Maximum	Mean
2005	-0.44	0.30	0.03	--	--	--
2009	-1.0	0.71	-0.13	-0.16	0.22	0.00
2010	-0.73	1.00	0.14	-0.17	0.23	0.00

Detecting change

Calibrating the images to a common benchmark should allow for NDVI to be compared over time. To evaluate the results of the calibration and suitability for time series analysis, I calculated change in NDVI from 2005 to 2009 and 2009 to 2010 for each parcel. Figure 6 shows the distribution of change. Summary statistics are available in Table 3.

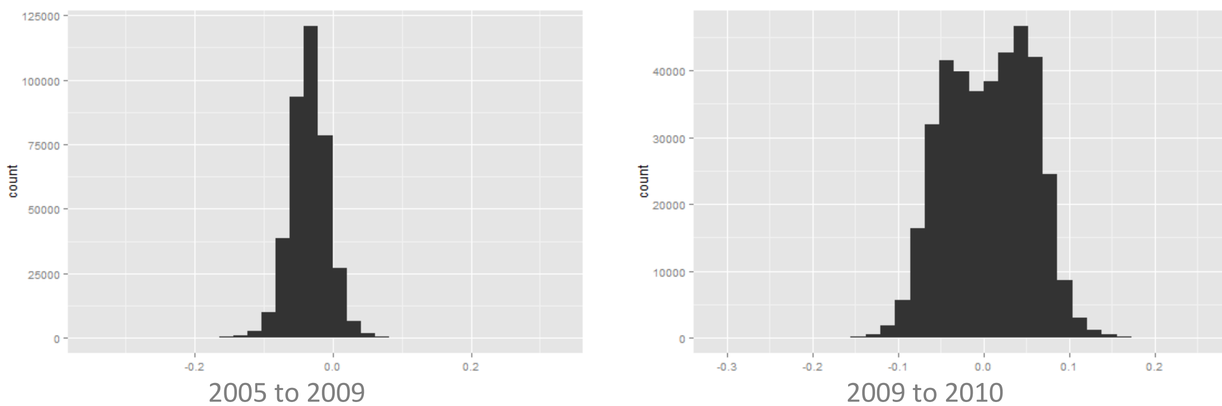


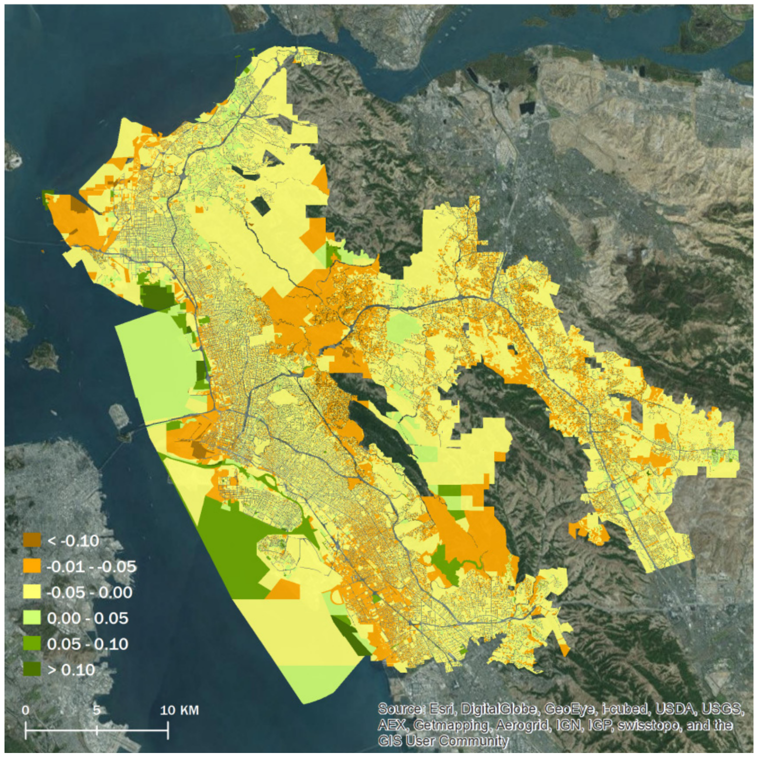
Figure 34. Change in NDVI value from 2005 to 2009 and 2009 to 2010. The change observed from 2005 to 2009 follows a normal distribution with a slight negative shift. This indicates that landscapes as a whole became drier. The change observed from 2009 to 2010 indicates that there may be a bimodal response: while the whole distribution shifted slightly positive (greener) on average, some houses moved negative while others moved positive.

Table 11. Mean with 95% confidence interval, minimum, and maximum parcel change in NDVI.

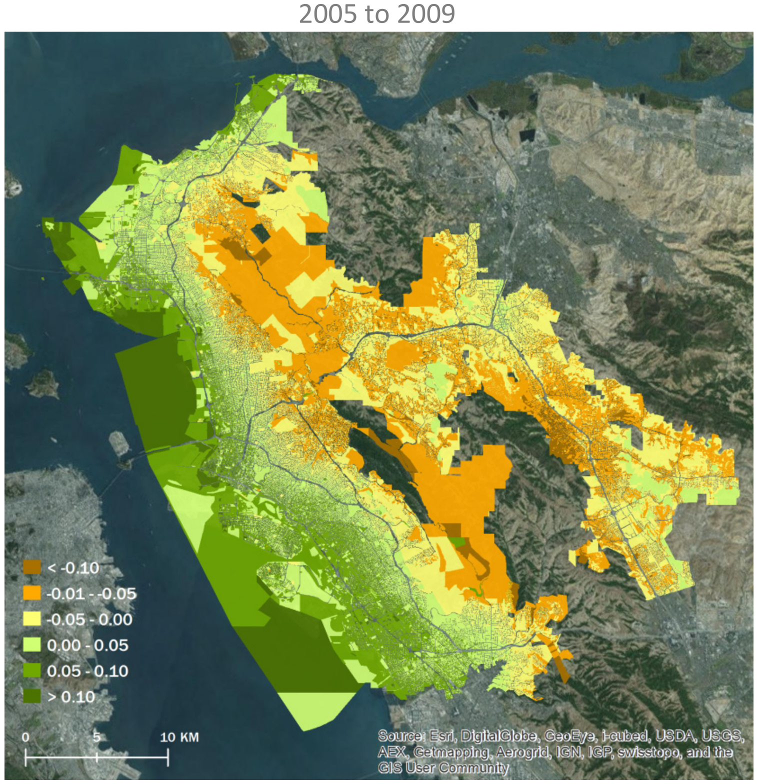
	Mean	Minimum	Maximum
2005 to 2009	-0.034 ($\pm$ 0.034)	-0.308	0.308
2009 to 2010	0.006 ($\pm$ 0.005)	-0.281	0.237

Though the distribution shift is small, t-tests indicate the difference are significant (in both cases, $p < 0.01$). Between 2005 and 2009 there is a very slight shift toward drier landscapes. From 2009 to 2010 there is a very slight shift toward more lush landscapes, though the distribution may indicate bimodality, with some landscapes shifting drier, while more are shifting greener. The parcels exhibiting the largest shifts toward wetter and drier landscapes are scattered through the study area (see Figure 7).

Occasionally, the change in NDVI appears to reflect true shifts in landscape conditions. In some cases, landscapes are becoming greener in new developments because young vegetation is maturing. In a few instances, shifts toward drier landscapes indicated the removal of a tree. For many parcels, the shift toward a drier landscape may have been due to a reduced water scheme—several local baseball fields were visibly more brown in 2009, which was a drought year that included aggressive water conservation campaigns.



2005 to 2009



2009 to 2010

Figure 35. Change in NDVI values mapped by parcel. Negative values indicate a shift toward drier landscape styles. Positive values indicate a shift toward more verdant landscape styles. Parcels on far west of study area intersect the waters of the San

Francisco Bay. The large changes in the value of the water could signal eutrophication, but likely indicates latent error in the surface reflectance corrections.

In most cases, however, the shift in NDVI value does not appear to be a result of a true change in land cover. Instead, the change in NDVI value reveals pathologies in the imagery. Parcels intersecting the water of the Bay exhibited large changes (see Figure 7). While it is possible there is increased vegetation in the Bay, it is more likely this change is a result of uncorrected changes in surface reflectance. In other cases, parcels exhibited angular effects from the low-flying planes that shoot NAIP imagery: depending on the precise location of the plane as an image was captured, the location of a tree may seem to jump from one parcel to another. When this happens cases, the NDVI values indicate that a parcel had a large vegetation change.

Future research: landscape change by parcel

Analysis of the NAIP imagery within EBMUD for 2005, 2009, and 2010 indicates that it cannot accurately capture NDVI change at the parcel level. While the data may reveal real landscape trends in aggregate, by parcel the greatest changes appear to result from image pathologies.

Some improvements to change detection could be made by further refining the imagery calibration. The current regressions have low R^2 values, indicating that it may be possible to improve target selection or introduce additional regression covariates. Yet, seasonal effects will remain challenging to capture because, at times, seasonal change means that there is a larger canopy on a tree, not just a greener canopy. Issues with alignment and angular effects will likely be impossible to correct through calibration, indicating that the parcel resolution may simply be too fine of a spatial unit for NAIP imagery.

The question remains if there is a spatial unit at which NAIP can reliably detect change. To most accurately evaluate parcel level land covers, however, there is a need for high resolution aerial imagery that is designed to be compared over time. Ideally, this imagery would include multiple fly dates per year to capture seasonal change.

Appendix 2. Modeling intraurban temperature

Intraurban temperature variation results from factors acting at multiple spatial scales. One of these factors is climatic temperature, whose role in localized heating and cooling is distinct from contributions of materials, vegetation, and physical orientation (e.g., aspect). To interpolate climactic temperature patterns through an urban and suburban gradient, Kriging is commonly used. This analysis implements Universal Kriging, evaluates the assumptions embedded in the Kriging model, and then suggests future directions for modeling dynamic environmental processes.

Data source

There are twenty-three to twenty-six land-based weather stations proximate to the East Bay Municipal Utility District study site that recorded temperature from January 2005 to December 2011 (see Figure 1). The stations range from approximately 1.4 to 17.6 km apart, with a median distance of 7 km (distances calculated with ArcGIS). Each weather station records several temperature variables. This study analyzes Mean Monthly Temperature.

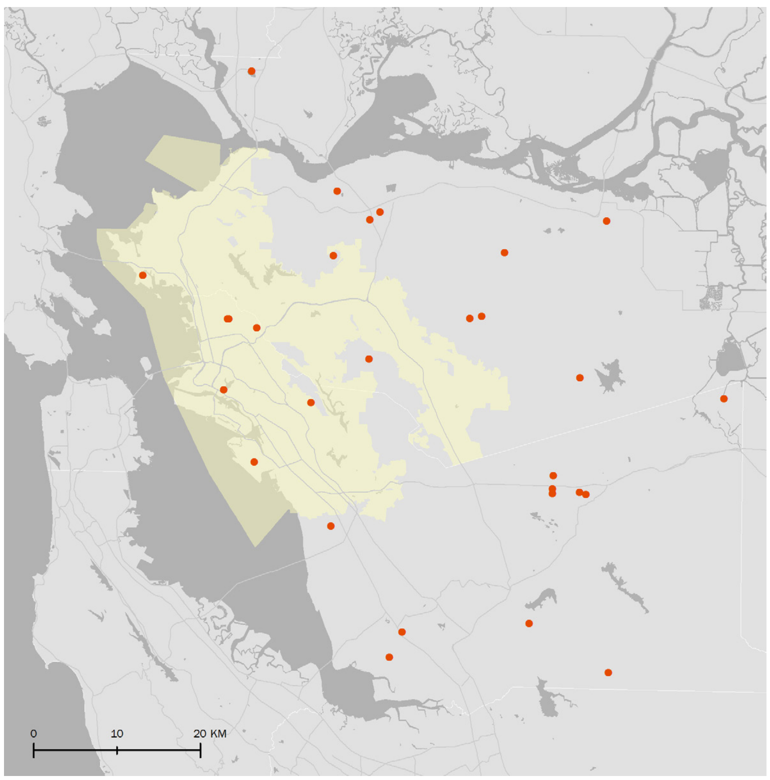


Figure 36. National Weather Service stations recording temperature (orange) in the EBMUD area (yellow) from January 2005 to December 2011. Distributed by the National Climactic Data Center.

Data cleaning

There is error in the reported temperatures. Some data points were much smaller or larger than those recorded at other stations, likely due to equipment malfunctions (see Figure 2). Some errors were more subtle, however. In these cases, clusters of neighboring stations collected temperature readings dissimilar from one another. It was often not possible to verify which readings should be cleaned out of the dataset and which should be kept when the differences were small. All of the readings may, in fact, be objectively correct and reflect the local microclimate (i.e., materials, vegetation, orientation). Both types of cleaning problems – the presence of large and small differences in the data – are illustrated in Figure 3.

All data points greater than four standard deviations from the mean were automatically dropped. Some data points that appeared to be local outliers were dropped after visual inspection. Further iterations of a temperature model may implement local statistics to develop a more standardized method for local cleaning.

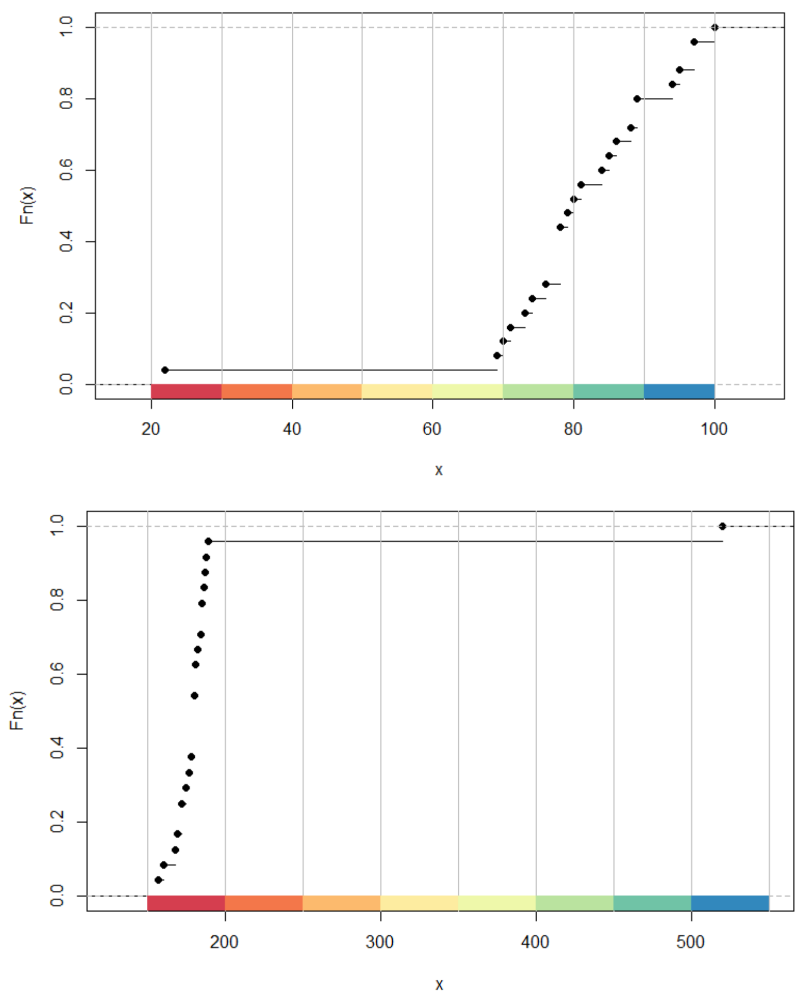


Figure 37. Visualization of the density of temperature readings (in degrees C * 10) from the month of January 2005 (above) and October 2007 (below). Each has a clear outlier.

Figure 38. Temperature recorded (degrees C * 10) at the NCDC monitoring stations in January 2007. Inset 1 shows three nearby stations with different temperature readings. The two closest stations are 1 mile apart, the third station is 3 miles from the other two. It is unlikely that there is such a large difference in climactic temperatures in these areas, and instead may reflect the microclimate surrounding each monitoring station. All three data points remained in the analysis. Inset 2 shows the lowest reading. This station is located on Mt Diablo, at a significantly higher elevation than the surroundings. This reading may not be representative of the climate of the whole area, so this data point was dropped. Inset 3 shows the highest data reading. Compared to the other months in the study area, this inland Oakland location was an unusual place for the highest reading. Though a seemingly unusual location for the highest temperature in the study area, because there was little evidence indicating an error, the data point remained in the analysis.

Model selection

Temperature is widely known to relate to distance from the ocean and latitude. A portion of the variation in Mean Monthly Temperature, as recorded by the NWS monitoring stations, can be explained by the linear model Temperature ~ Easting + Northing. Universal Kriging (also known as External Drift Kriging) examines the variance of the residuals from the linear model within spatial lags. Because of its ability to reveal spatial correlation as the underlying field changes, Universal Kriging is an appropriate method for modeling temperature in the study site.

Using the statistical package R, the interpolation was executed in four steps:

1. Cleaning the data by removing outliers
2. Fitting a variogram to residuals of Temperature ~ Easting + Northing using R's gstat package
3. Interpolating temperature for each parcel within the EBMUD boundary using R's gstat package
4. Taking the square root of the variance to find one standard deviation from the predicted value

The modeled variogram, the interpolated temperatures, and map of the standard deviations are available in Table 1.

Model assumptions

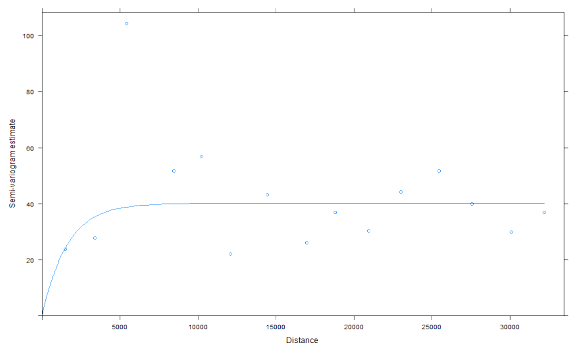
Kriging proceeds in two steps. The variogram is fit to estimate a covariance matrix. Then, the covariance matrix is used as a basis for interpolation between known data points. Embedded in these steps are several assumptions that are discussed below.

Temporal non-stationarity

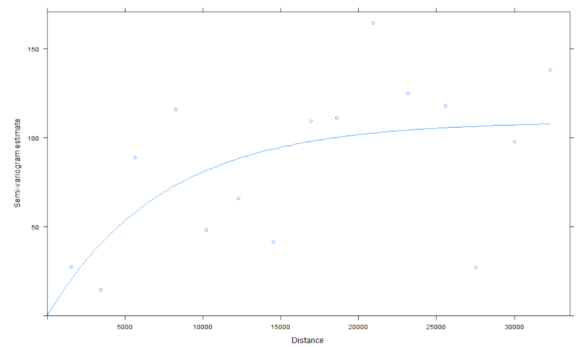
During the study period, the relationships among the temperature data through the study area shift. Some months exhibit reasonably strong spatial relationships, while many months show weak spatial relationships. Variograms from January exemplify the differences in spatial relationships (see Figure 4). In some years, such as 2005, the January data exhibit reasonably well conforming spatial trends. In other years, there is no evidence of spatial correlations. For example, January 2007 is modeled with a range parameter equal to zero, which indicates that there is no spatial autocorrelation among the data.

Within the bounds of this study, it is difficult to determine if the differences are an artifact of sparse data, measurement errors, and an under-specified linear model, or if the spatial relationships truly differ year-to-year. Many climate modelers assume that a given month should exhibit a similar spatial relationships – in other words, all Januarys should have the same spatial relationship among the data. To take advantage of this assumed temporal stationarity, multiple years of data are used to develop the variogram, and a given year's temperatures are interpolated with parameters from the time-averaged variogram. For example, PRISM, one of the most pervasive climate modeling efforts in the West designed for use in areas with few reporting weather stations, uses a 30 year history as the basis of the climate model (Daly et al. 2008).

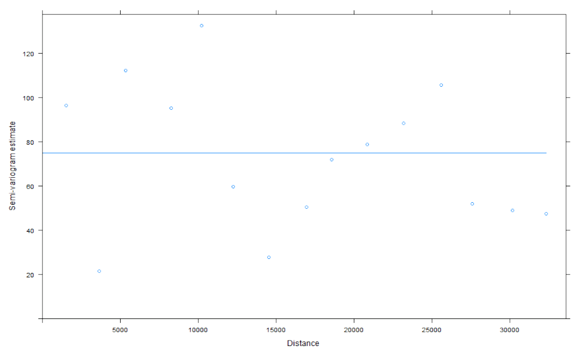
It is possible, however, that spatial relationships may truly be dynamic or changing over time, given the annual variability of the Bay Area's Mediterranean climate and climate change impacts on temperature. For these reasons, this study recalculates the variogram for each month of each year, rather than relying on a time-averaged variogram.



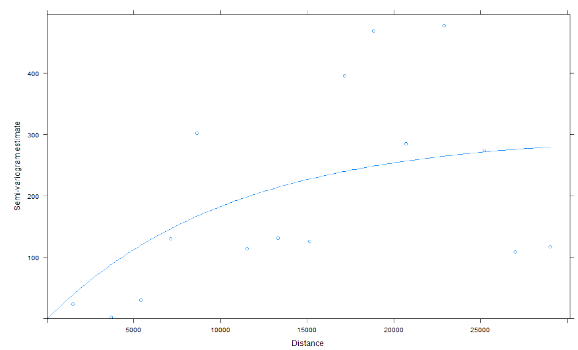
2005



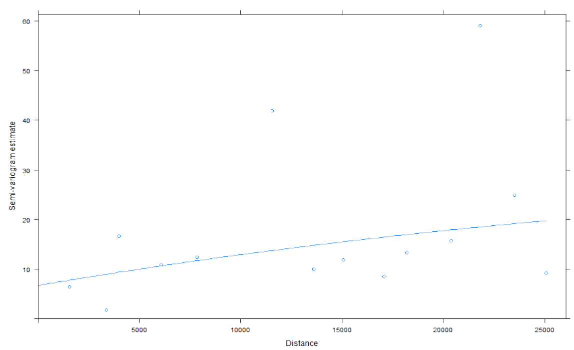
2006



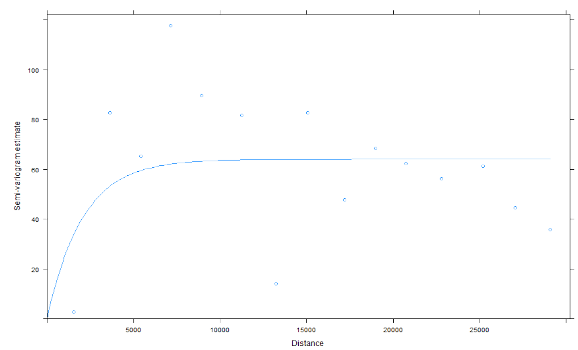
2007



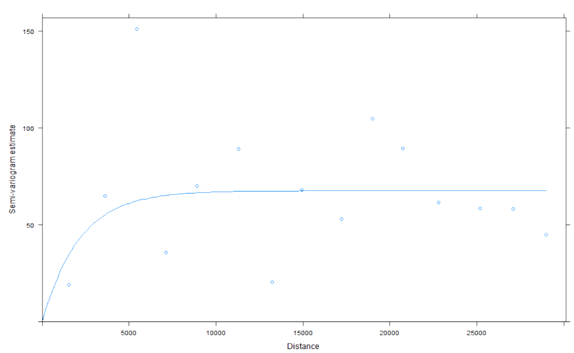
2008



2009



2010



2011

Figure 39. Modeled variograms from the month of January for the duration of the study period. Each exhibits different nugget, range and sill parameters. Subset from Table 1.

Spatial correlation structure

Even though the parameters (nugget, sill, range) of the spatial relationships change from month to month, this study assumes that the correlations among the data points maintain a consistent structure. It is possible to specify many different types of models to capture the shape of the covariance function, but this analysis uses an exponential model for all months within the study period (see Figure 5). This decision is based on initial fitting exercises with several sample months. Determining if an exponential model is always the best representation for all months throughout all years requires further investigation, but is not the primary goal of this research.

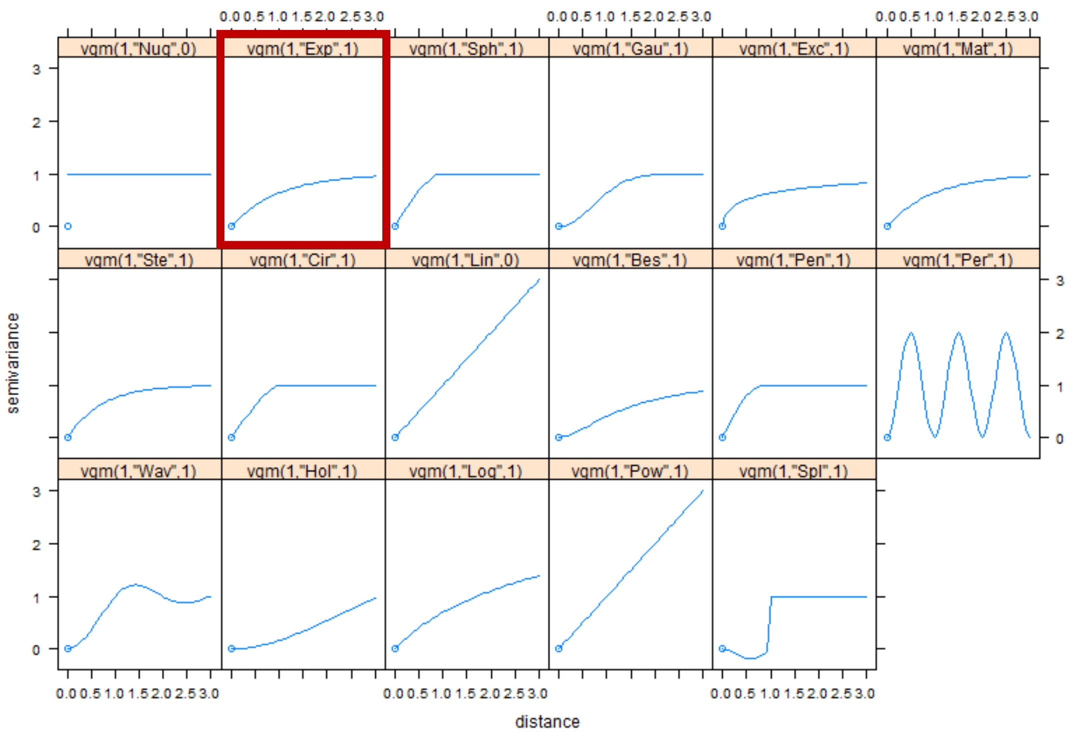


Figure 40. Covariance functions included within R's gstat package. Image provided in gstat. Exponential function highlighted in red.

Spatial stationarity

Implicit to the Universal Kriging approach employed in this analysis is that the spatial relationships are consistent across the study area within any given time period, after accounting for the underlying linear model. Points that are outliers from the modeled variogram are assumed to be errors due to sparse data, failings in the data cleaning process, or unobserved effects not captured in the linear model (see Figure 6).

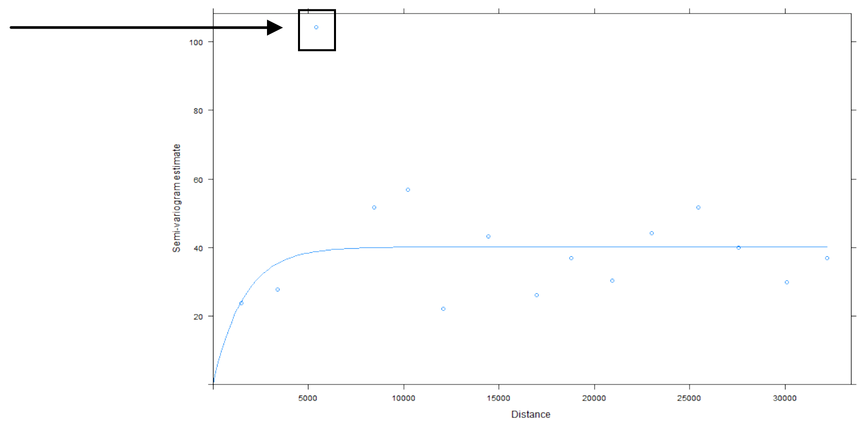


Figure 41. The upper variogram, from January 2005, shows a reasonable pattern, despite the outlying point. The outlying point may indicate an error in the data within the spatial lag, despite data cleaning. Closer analysis is necessary to verify whether or not the variance within the spatial lag is real.

There are two challenges of note. Firstly, the moderating effect of ocean temperature may not taper linearly over distance. This may mean that, within at a given spatial lag, some sets of data points are more closely related than others. Secondly, there is a band of hills within the study area known to cause a discontinuity in temperature, since the hills act as a barrier for incoming coastal fogs and wind. The effect of the hills is so significant that EBMUD internally calculates water use with different evapotranspiration parameters for East of the Hills and West of the Hills. This may mean that weather stations near one another, but located on opposite sides of the ridgeline, are more dissimilar than otherwise expected.

It is important to note, however, that the impacts of the ocean and the hills are contributors to microclimate variation, not climactic temperature patterns. If the underlying linear model were specified to better account for microclimate variation, the spatial stationarity assumption of Universal Kriging may, indeed, be appropriate.

Temporal resolution

Across the study area, there is an east-west temperature gradient that flips each spring and fall. The eastern side of the study area is warmest in the summer, while the western side is warmest during the winter. This inversion is captured in the modeled data every year of the study period (e.g., see Figure 7). The precise month of the inversion varies from year to year (see Table 1). In the spring, the inversion occurs in approximately May or June. In the fall, it is approximately October or November.

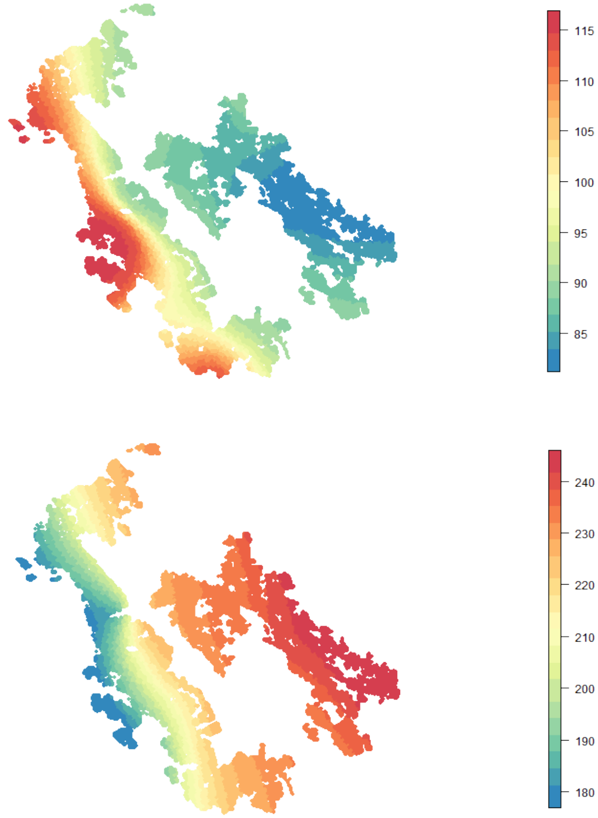


Figure 42. Temperature in January and July 2006 (˚C*10). The temperature pattern inverts in every year in the study period. The month at which the inversion is first documented varies from year to year.

Temperature inversion reveals a limitation in modeling at monthly time intervals. During the months the temperature flips, there is often lower spatial correlation among the data and subsequently lower predictive power, as indicated by a wider standard deviation in the interpolated data (see Table 1). This implies that Mean Monthly Temperature may be too coarse of an aggregation as the temperature gradient is changing. During these months of inversion, spatial relationships may be changing, too.

Future research: modeling temperature in dynamic, heterogeneous environments

This study is a preliminary exploration of temperature variation throughout the study site. In order to best explain the variation among the recorded data, it is necessary to further develop the linear model that underlies the Universal Kriging model. In particular, it would be appropriate to examine variables that account for the impacts of microclimate (e.g., a dummy variable representing West of the Hills versus East of the Hills). It is also likely prudent to further examine the appropriate stationarity and aggregation assumptions, determining if the same stationarity and aggregation assumptions hold true throughout the year.

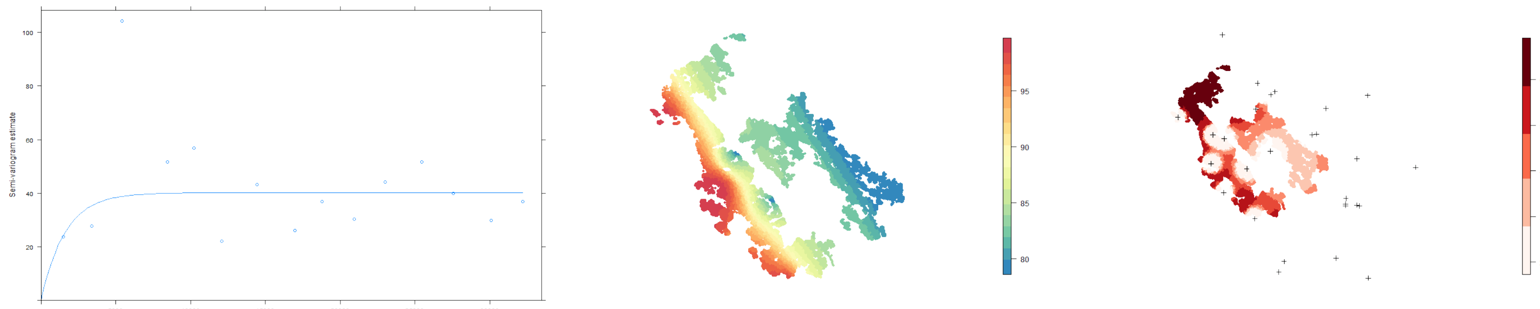
Adjudication of the interpolated data will remain difficult. At a minimum, more temperature sensors are necessary, so it is possible to withhold a subset of data from an interpolation, and then test predicted data again measured data. Furthermore, a greater density of sensors would increase robustness to measurement errors through redundancy. Yet, even with a higher resolution temperature sensor network, it will remain challenging to separate climactic variation from the impacts of microclimate.

Table 12. Modeled variograms, interpolated temperatures, and standard deviation in the East Bay Municipal Utility District from January 2005 to December 2011. Temperature given in degrees Celsius * 10. The map of standard deviations includes the location of reporting weather stations, denoted with a cross.

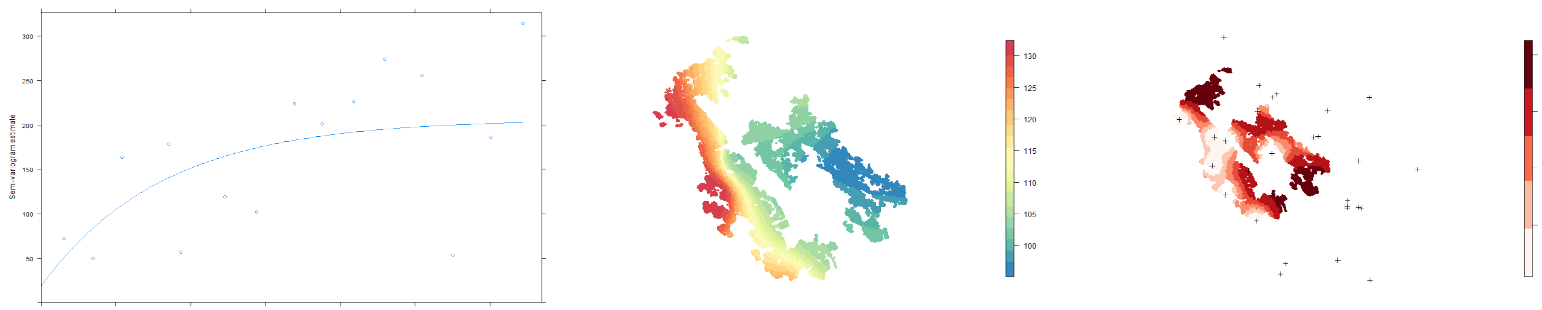
MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS

2005

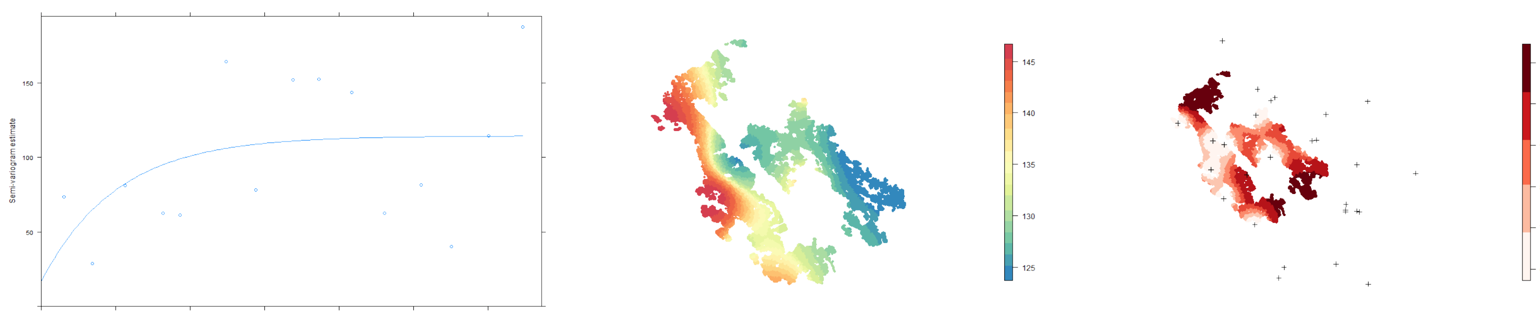
JANUARY

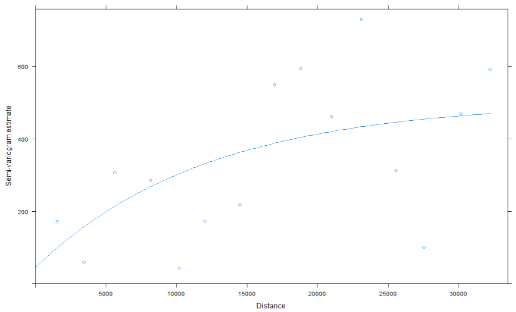
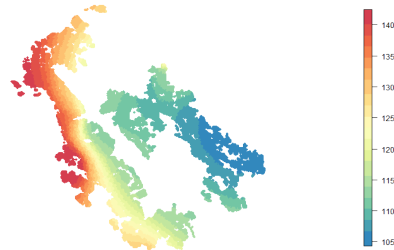
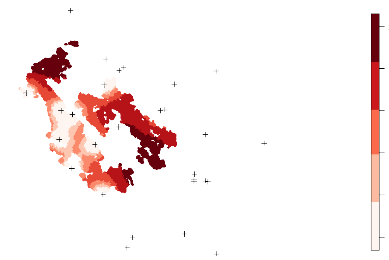
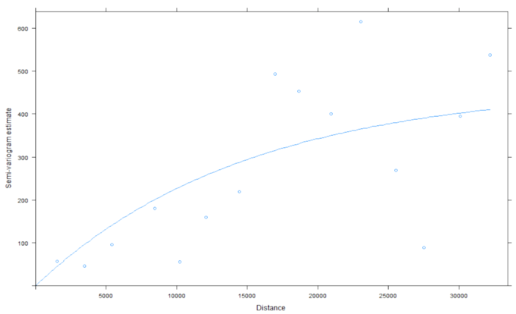
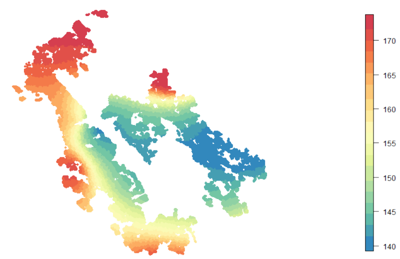
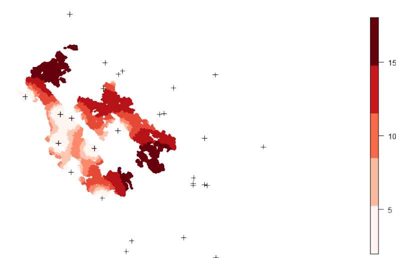
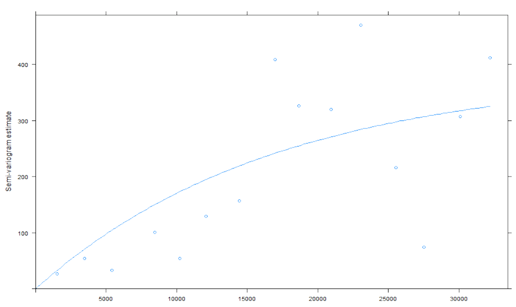
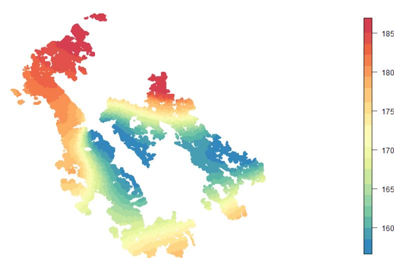
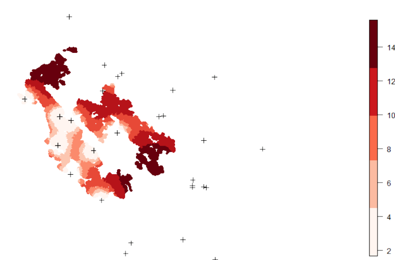


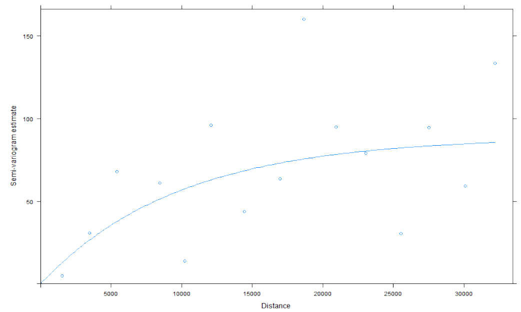
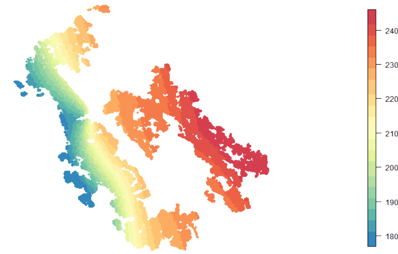
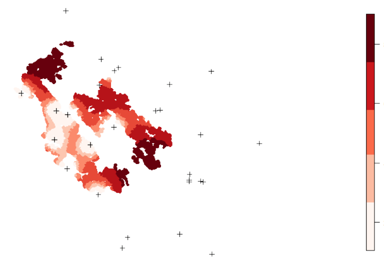
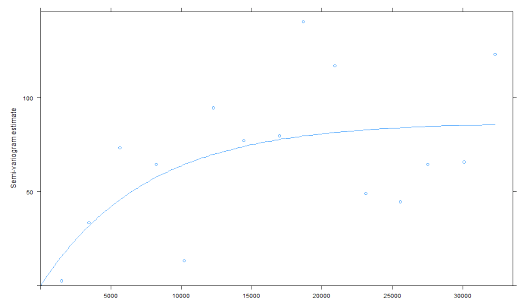
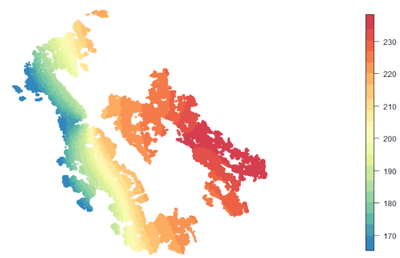
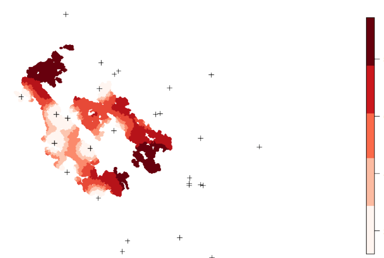
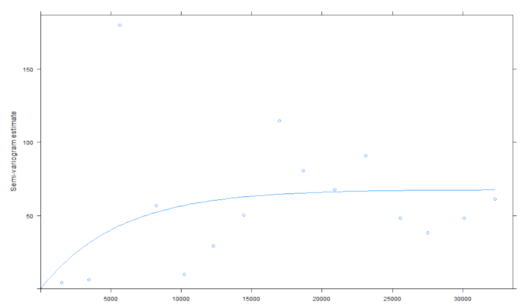
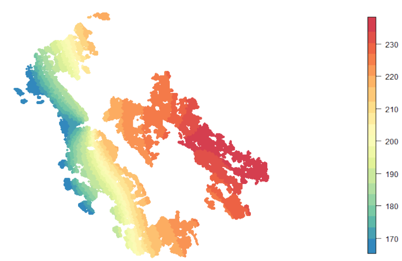
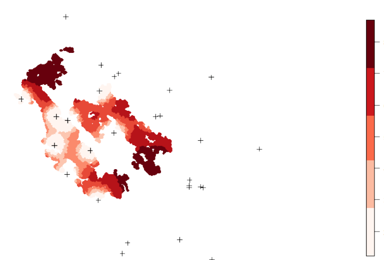
FEBRUARY

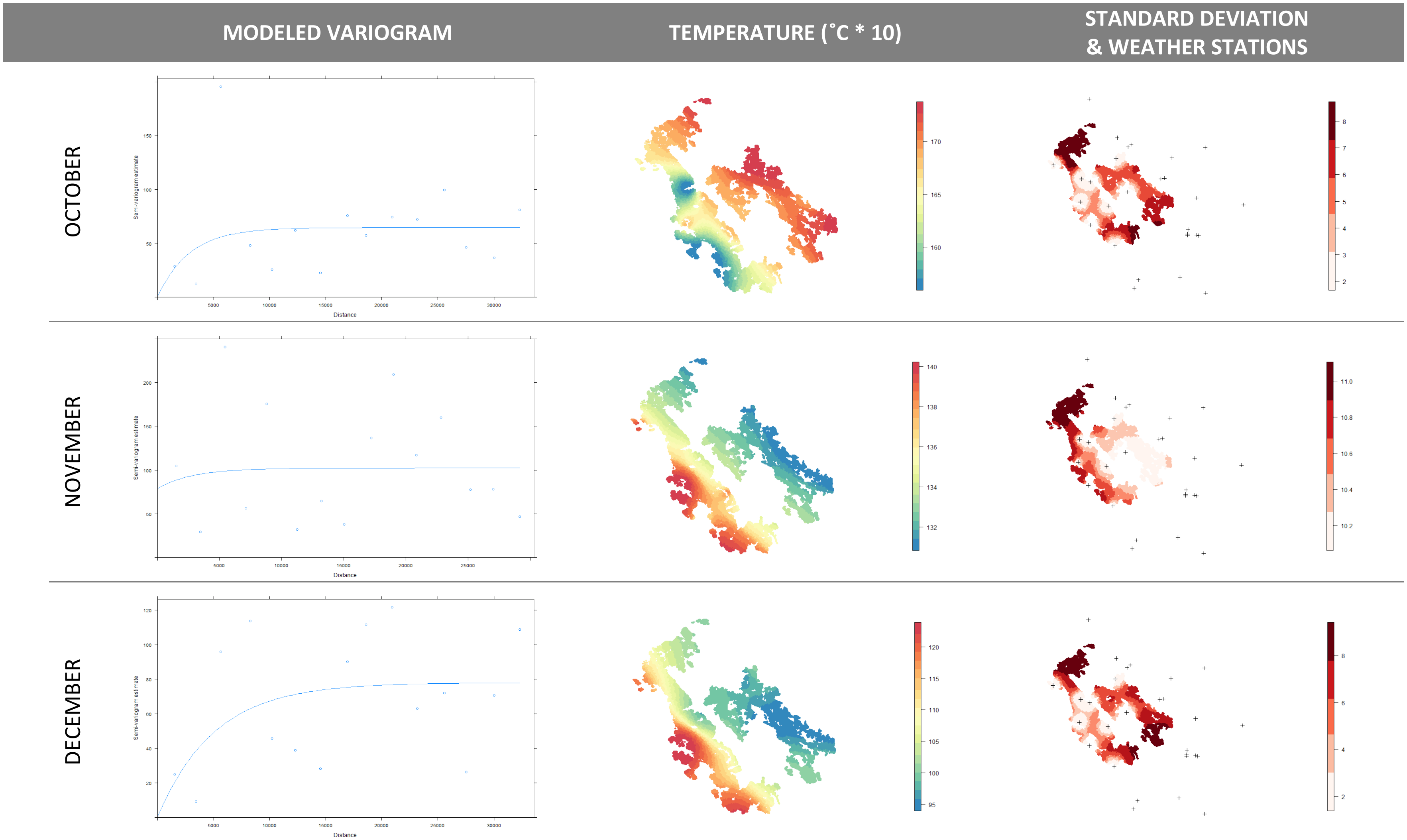


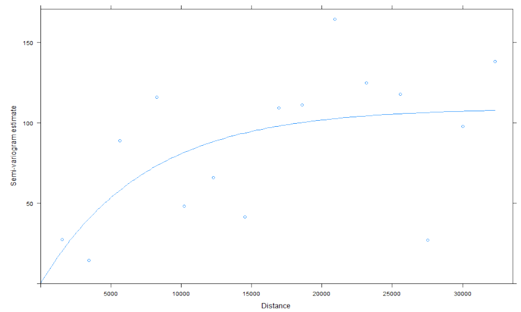
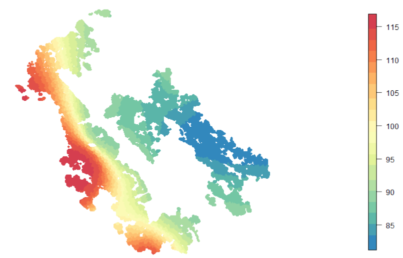
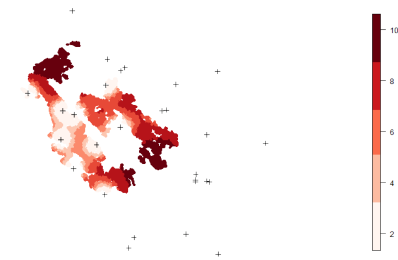
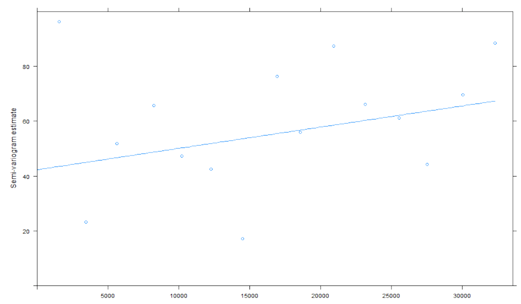
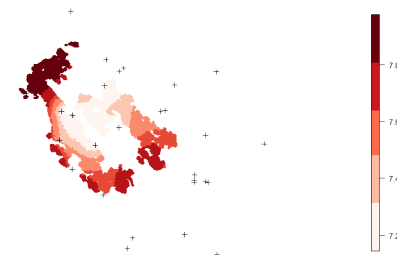
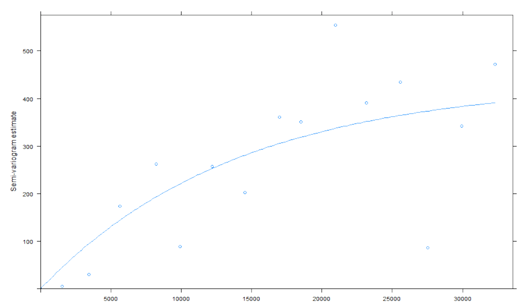
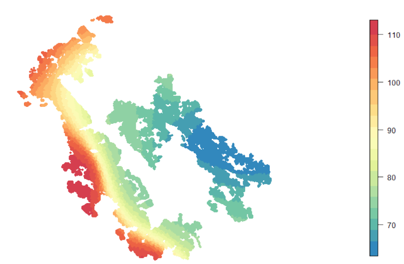
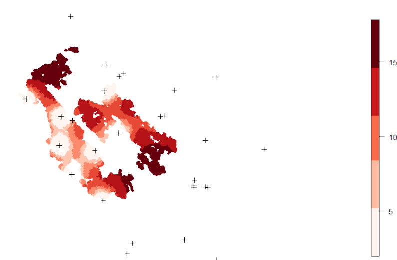
MARCH

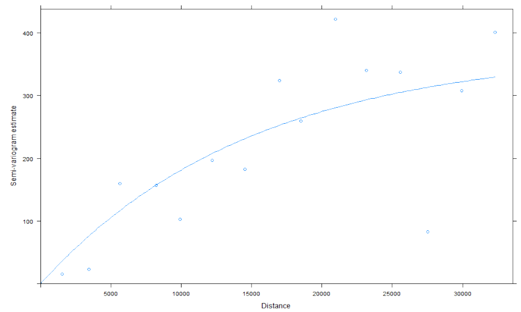
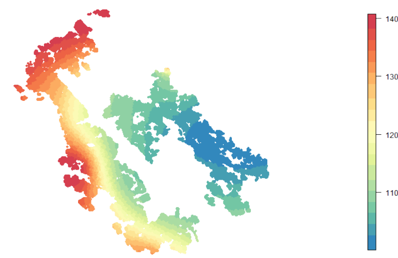
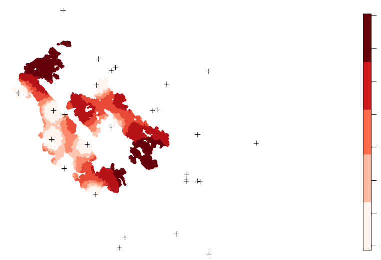
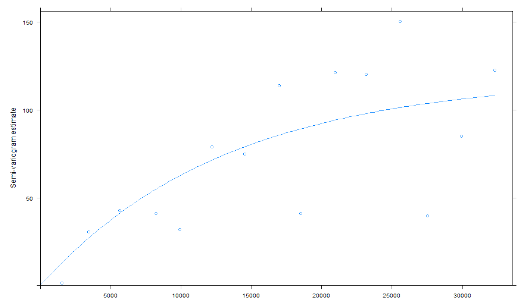
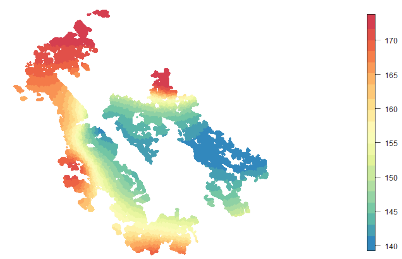
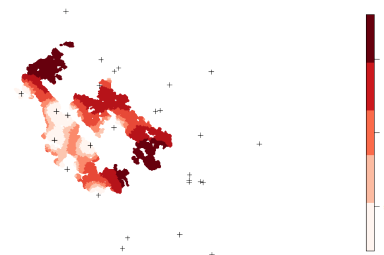
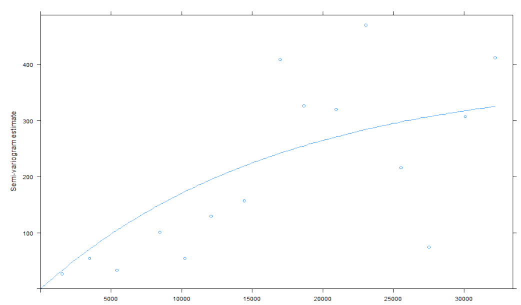
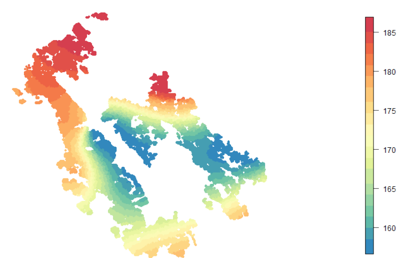
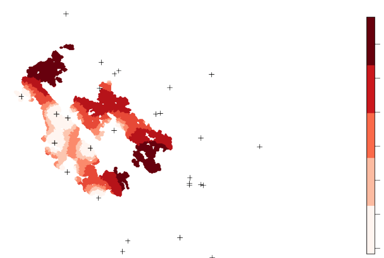


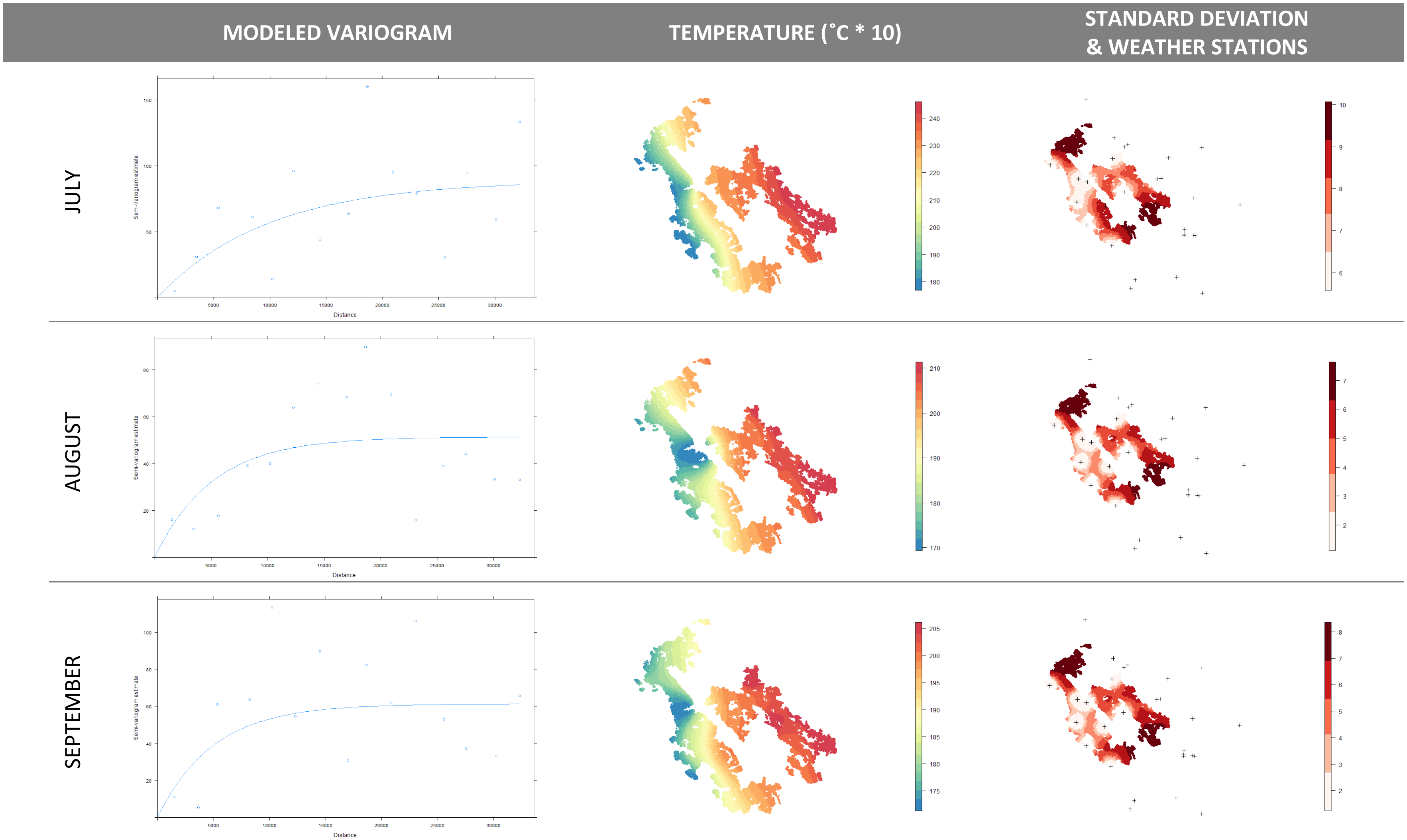
	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
APRIL			
MAY			
JUNE			

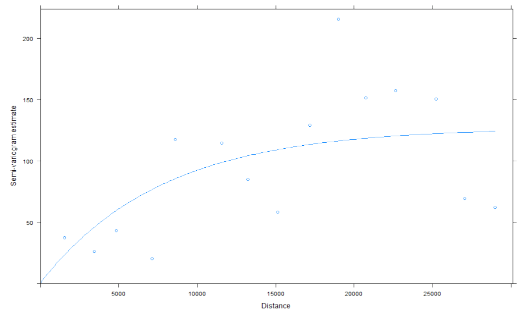
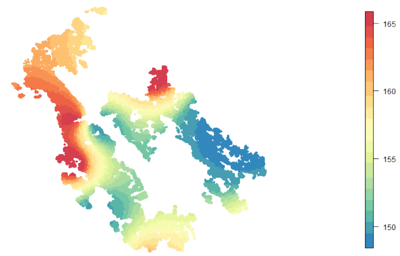
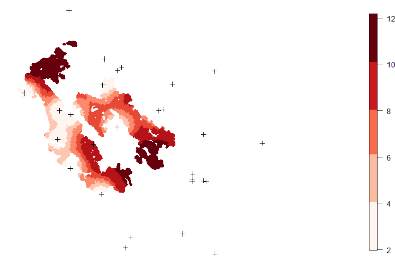
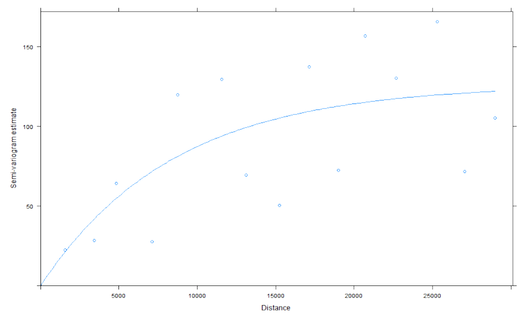
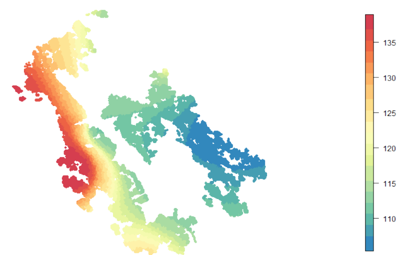
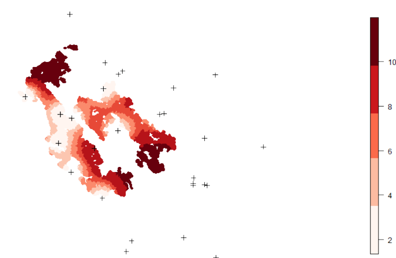
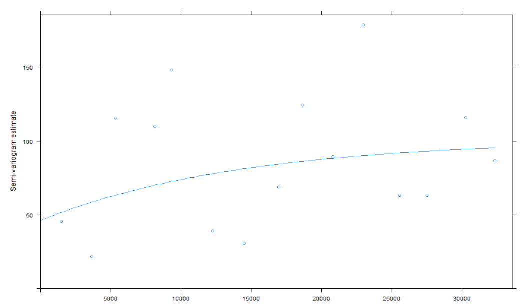
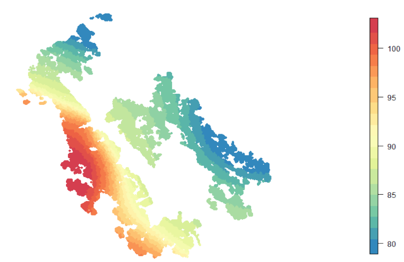
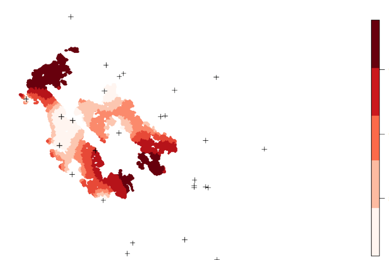
	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
JULY			
AUGUST			
SEPTEMBER			


MODELED VARIOGRAM
TEMPERATURE (˚C * 10)
STANDARD DEVIATION & WEATHER STATIONS
OCTOBER
NOVEMBER
DECEMBER
Semi-variogram estimate
Distance

		MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
	JANUARY			
2006	FEBRUARY			
	MARCH			

	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
APRIL			
MAY			
JUNE			


MODELED VARIOGRAM
TEMPERATURE (˚C * 10)
STANDARD DEVIATION & WEATHER STATIONS
JULY
AUGUST
SEPTEMBER
Semi-variogram estimate
Distance

	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
OCTOBER			
NOVEMBER			
DECEMBER			

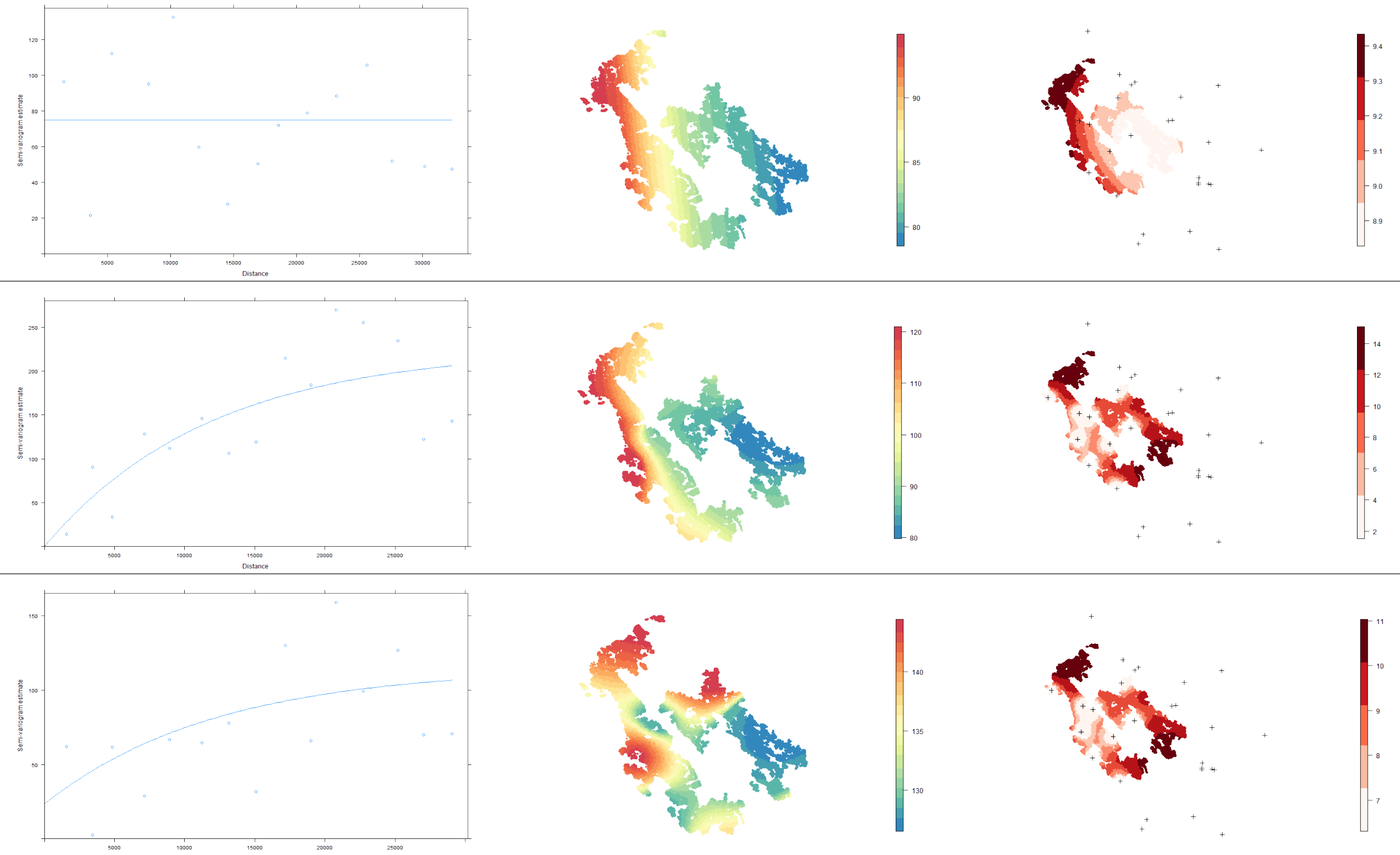
MODELED VARIOGRAM | TEMPERATURE (˚C * 10) | STANDARD DEVIATION & WEATHER STATIONS

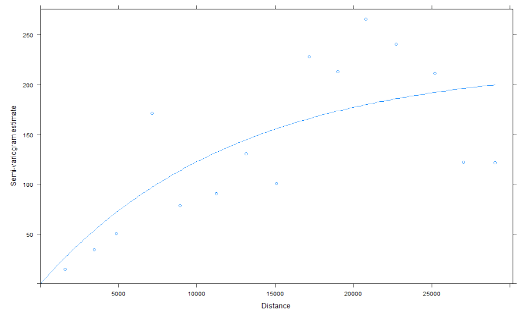
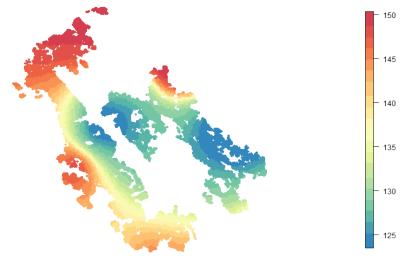
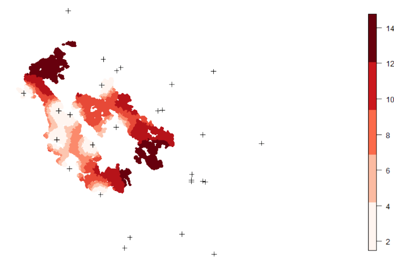
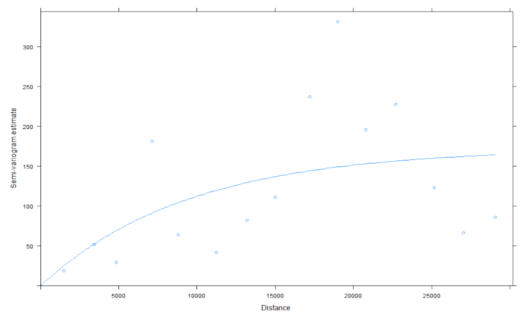
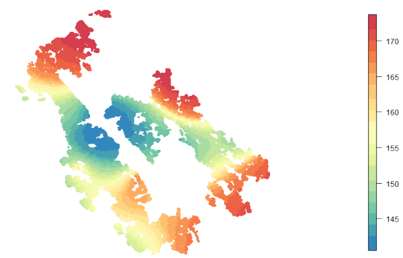
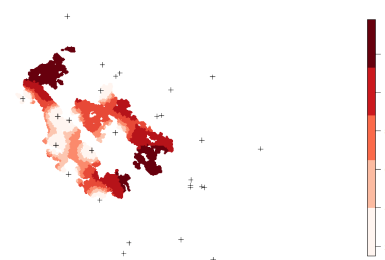
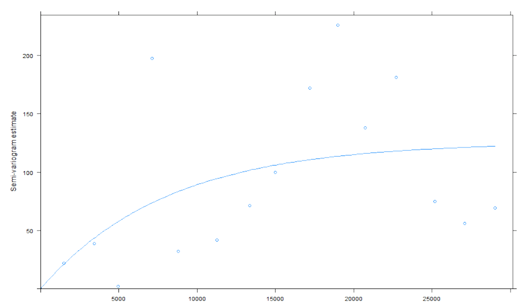
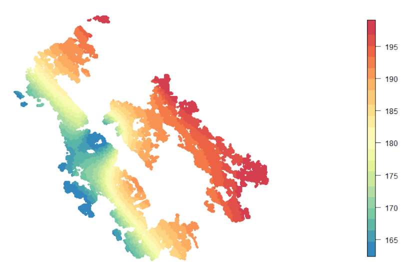
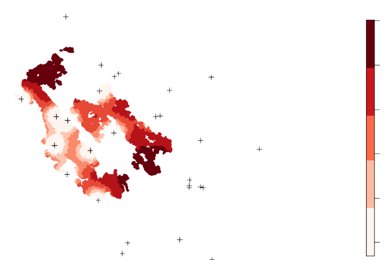
2007

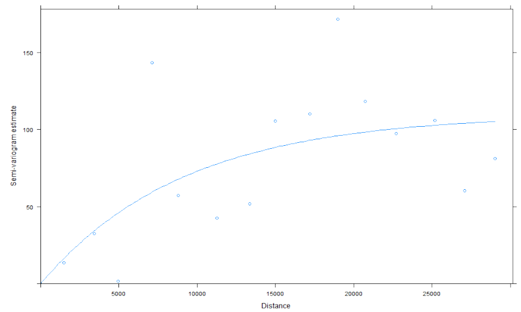
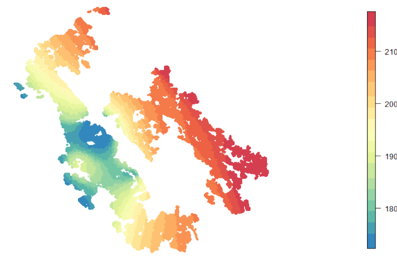
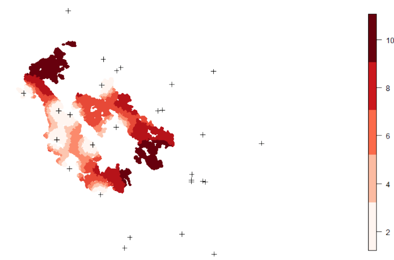
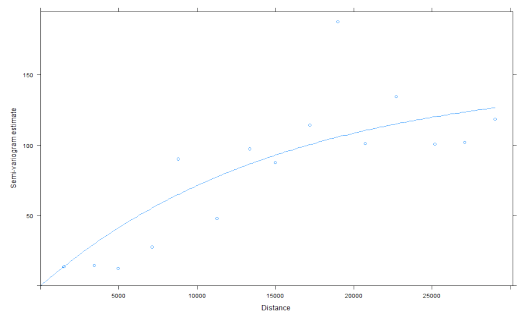
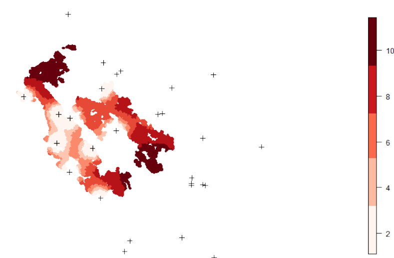
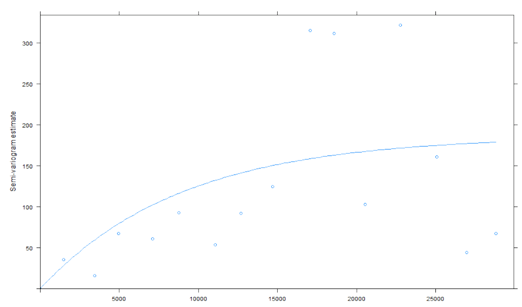
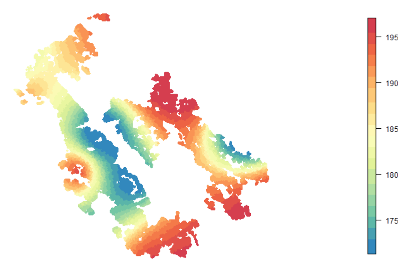
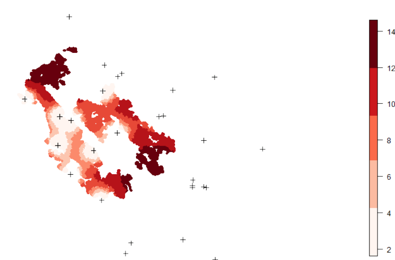
JANUARY

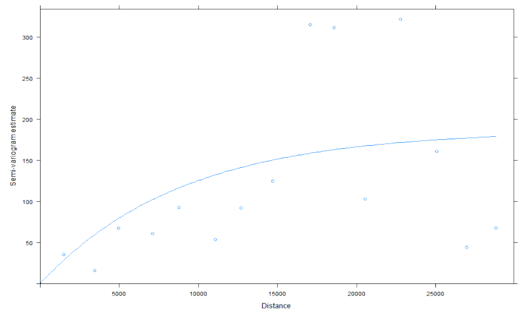
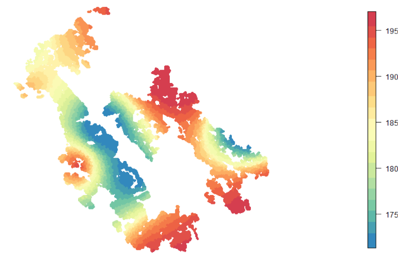
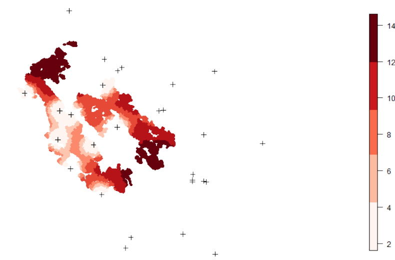
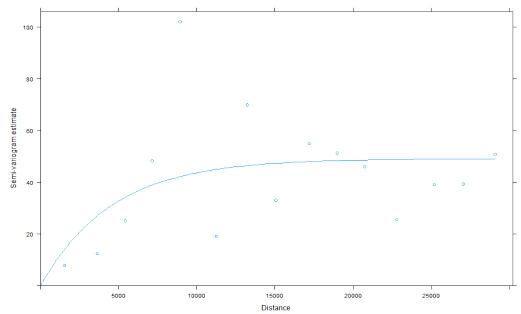
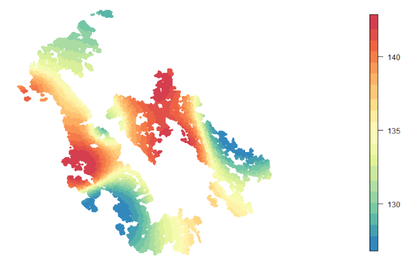
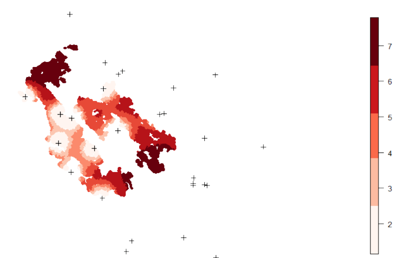
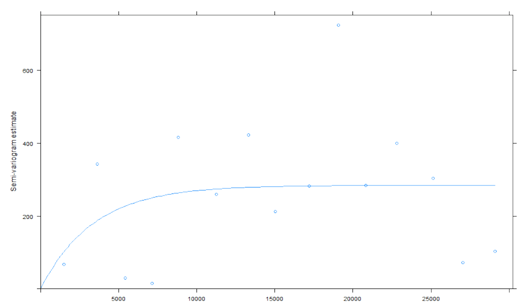
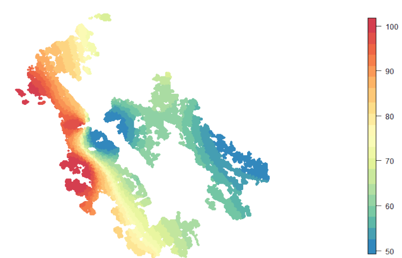
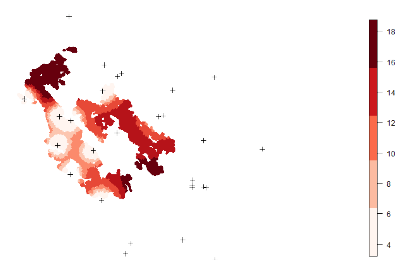
FEBRUARY

MARCH



	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
APRIL			
MAY			
JUNE			

	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
JULY			
AUGUST			
SEPTEMBER			

	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
OCTOBER			
NOVEMBER			
DECEMBER			

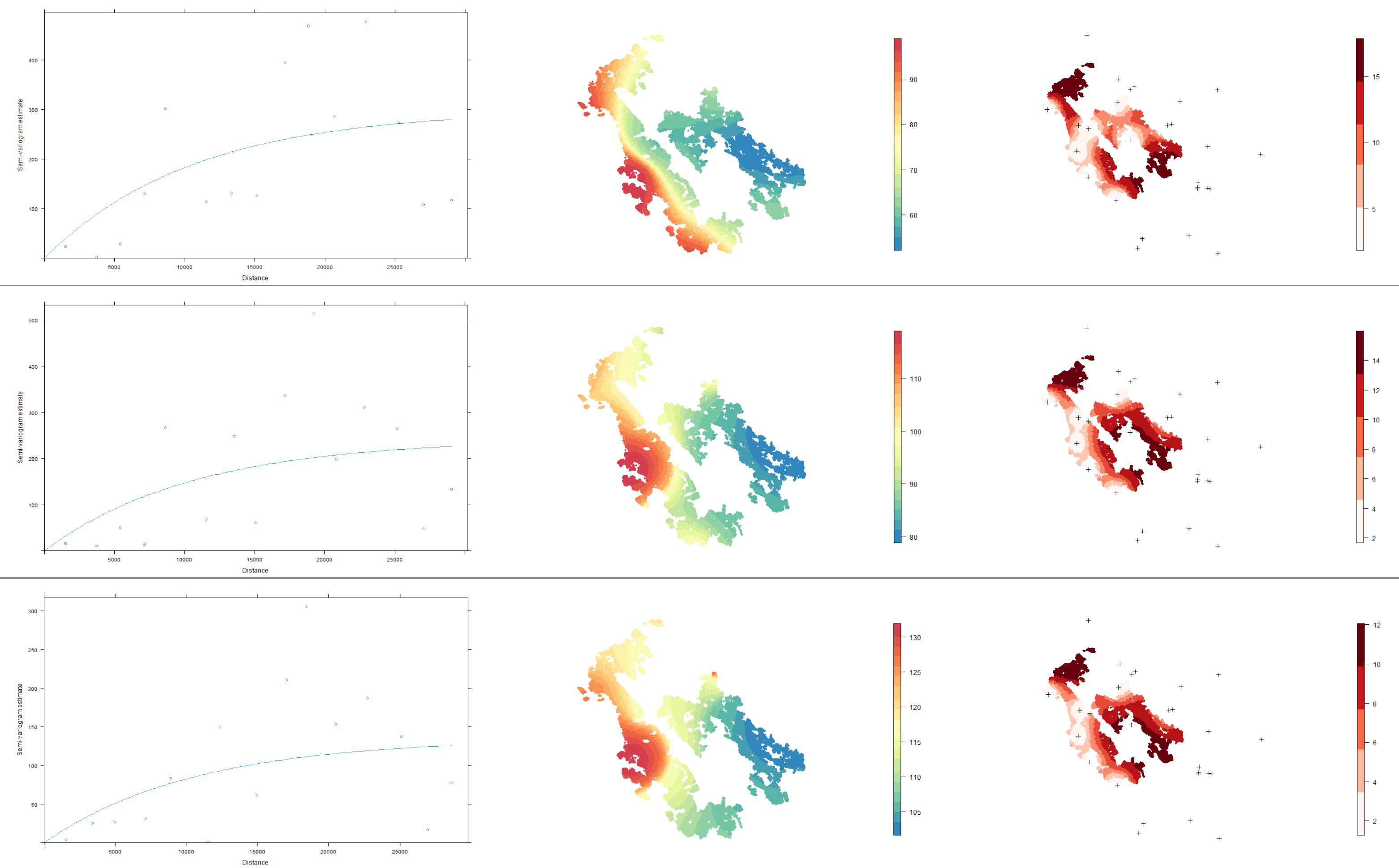
	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS

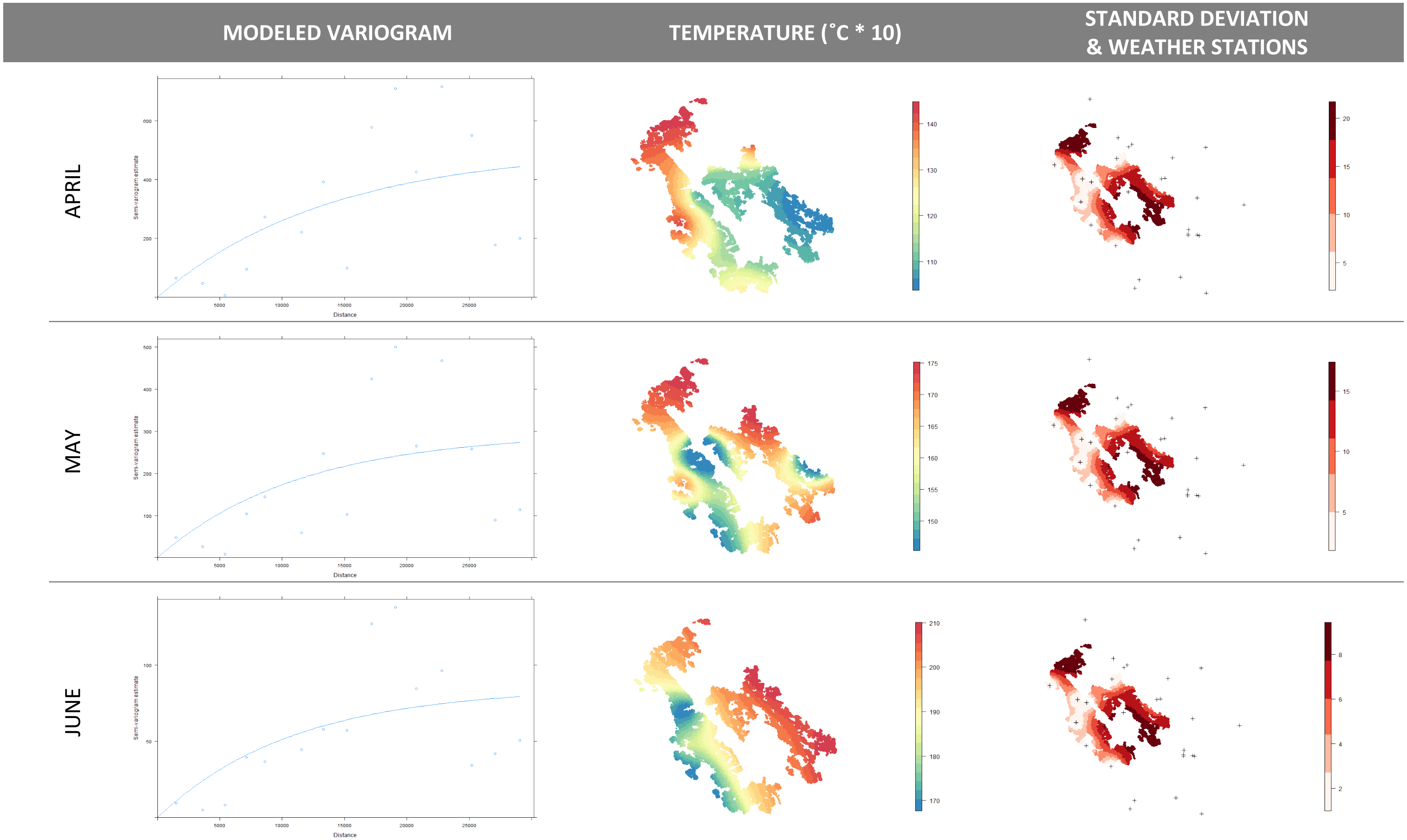
2008

JANUARY

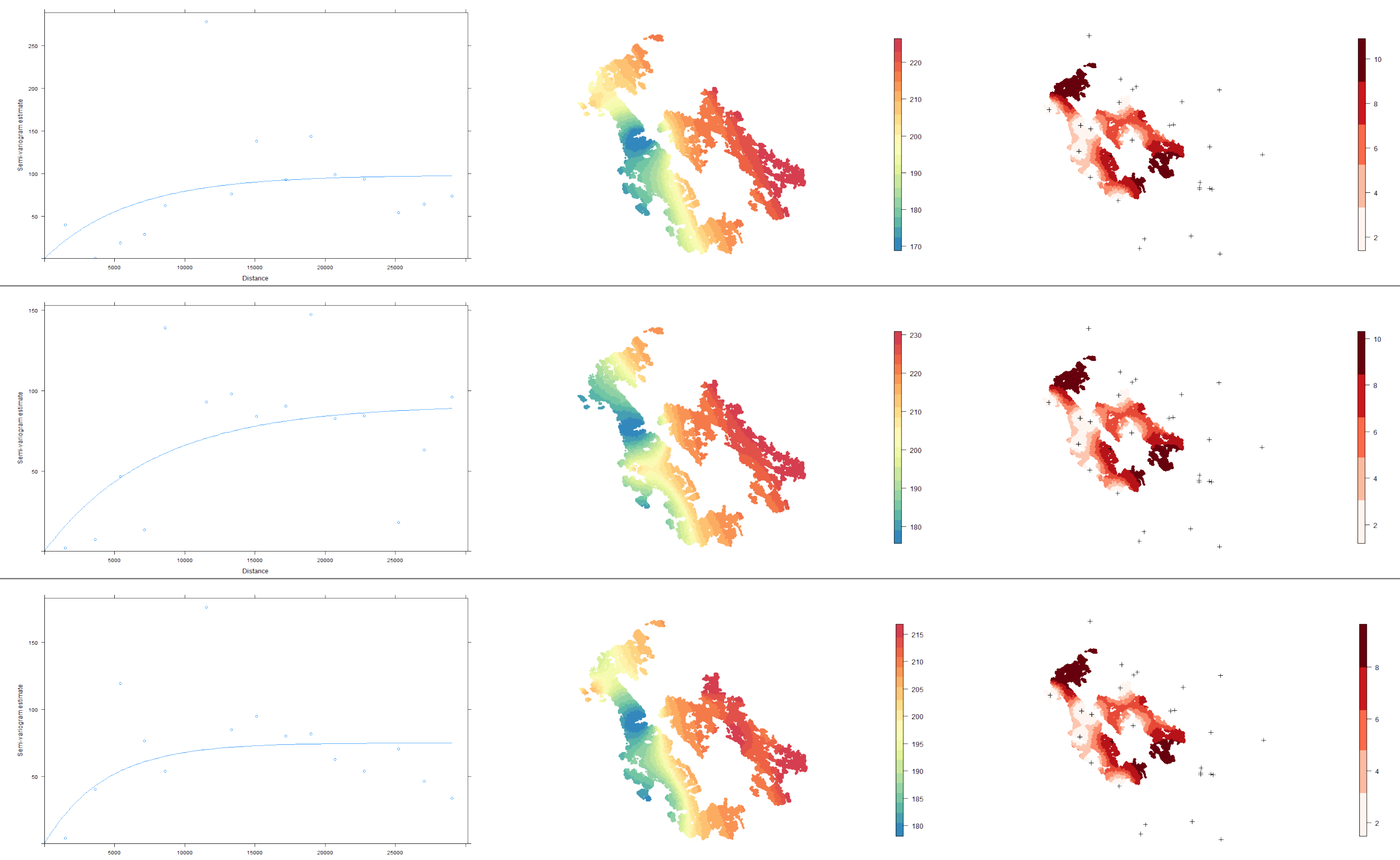
FEBRUARY

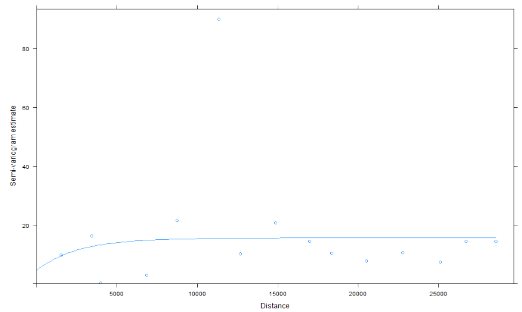
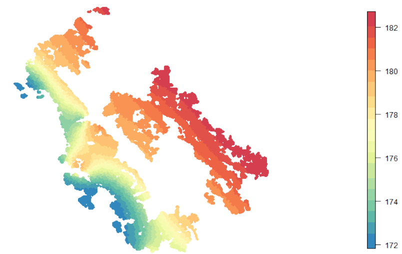
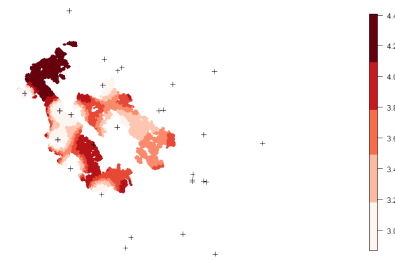
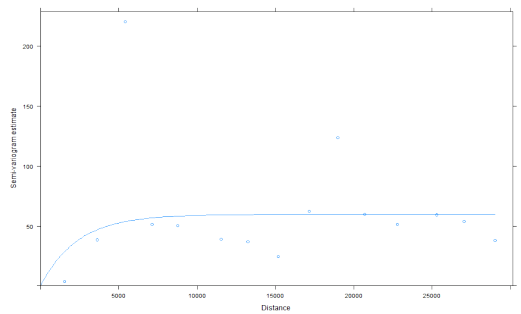
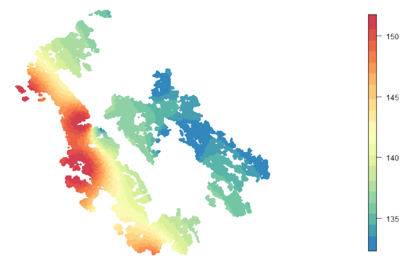
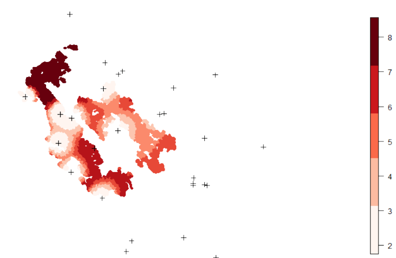
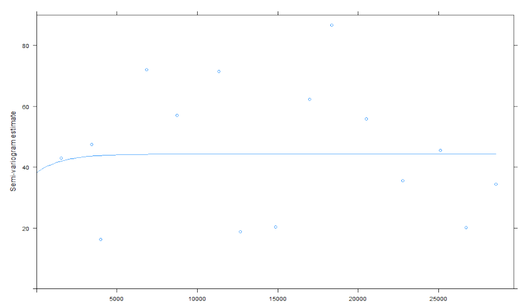
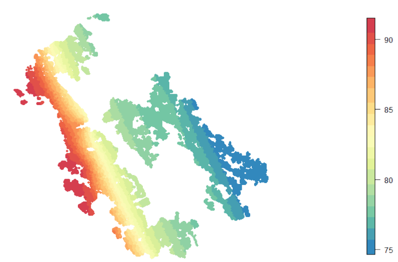
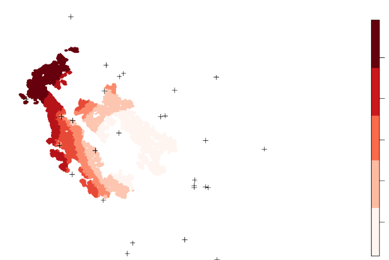
MARCH




MODELED VARIOGRAM
TEMPERATURE (˚C * 10)
STANDARD DEVIATION
& WEATHER STATIONS
APRIL
MAY
JUNE
Semi-variogram estimate
Distance
5000
10000
15000
20000
25000
140
130
120
110
20
15
10
5
175
170
165
160
155
150
210
200
190
180
170
8
6
4
2

	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
JULY			
AUGUST			
SEPTEMBER			

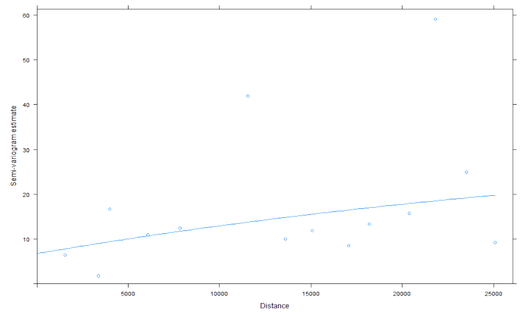


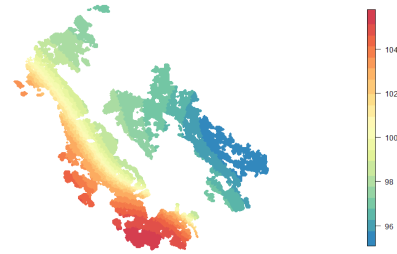
	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
OCTOBER			
NOVEMBER			
DECEMBER			

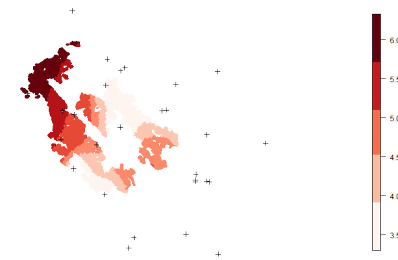
MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS

2009

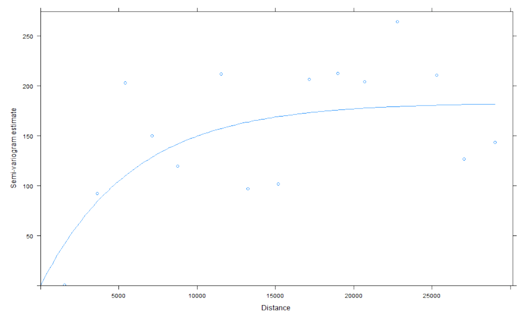
JANUARY

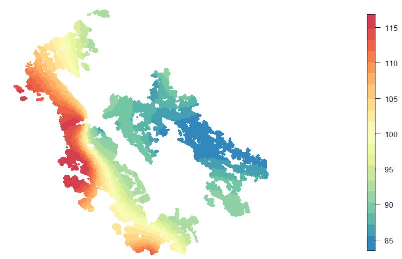


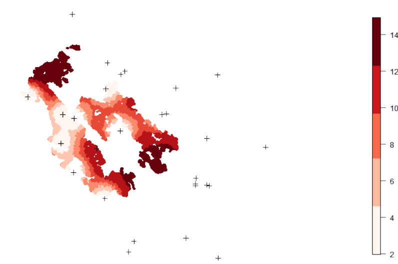




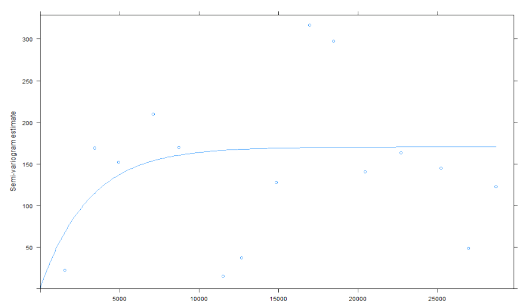
FEBRUARY

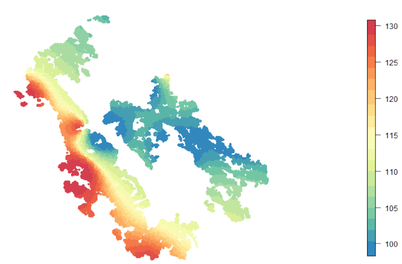


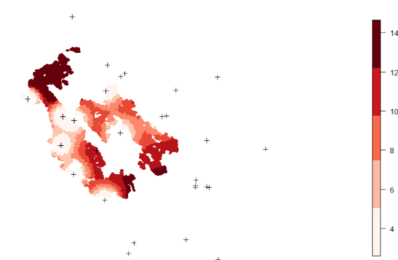


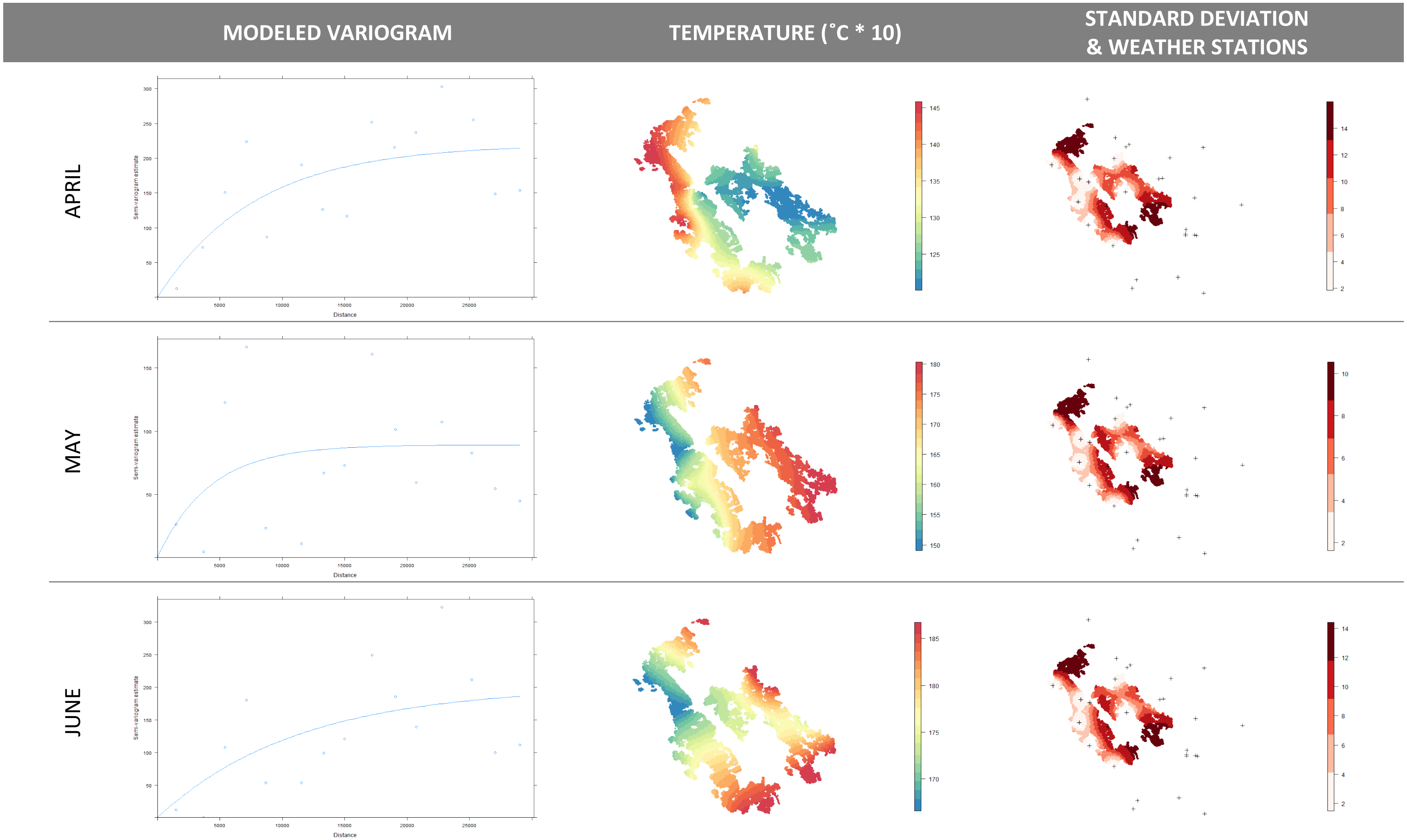


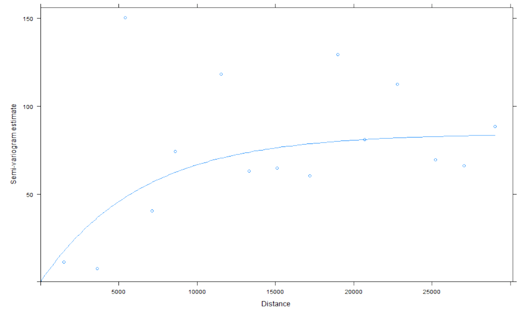
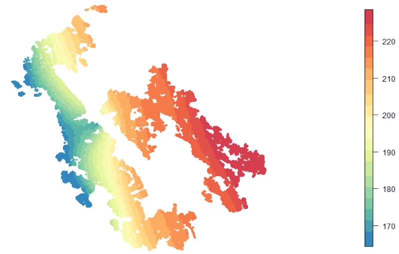
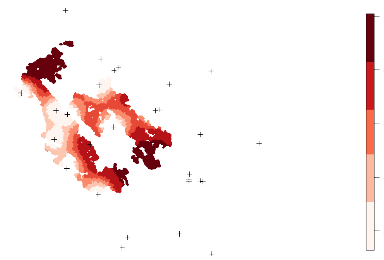
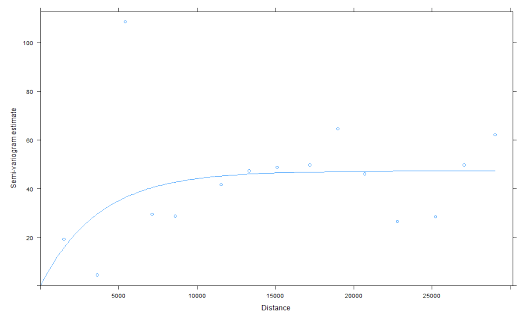
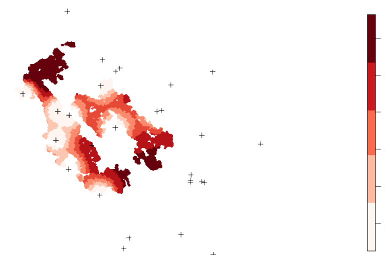
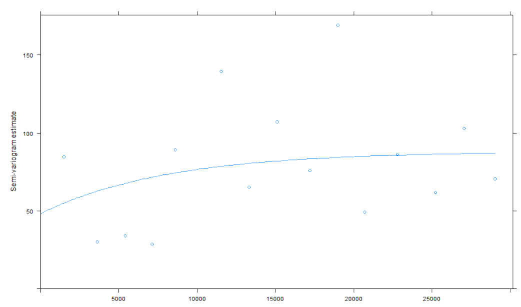
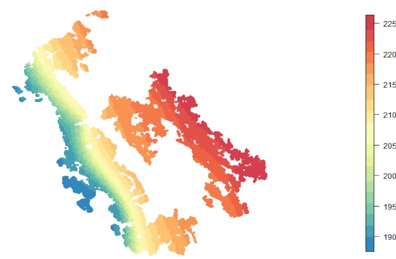
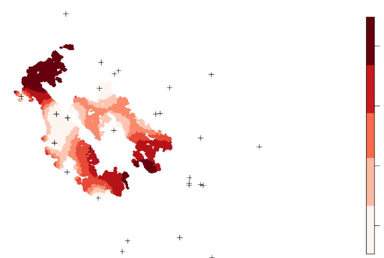
MARCH

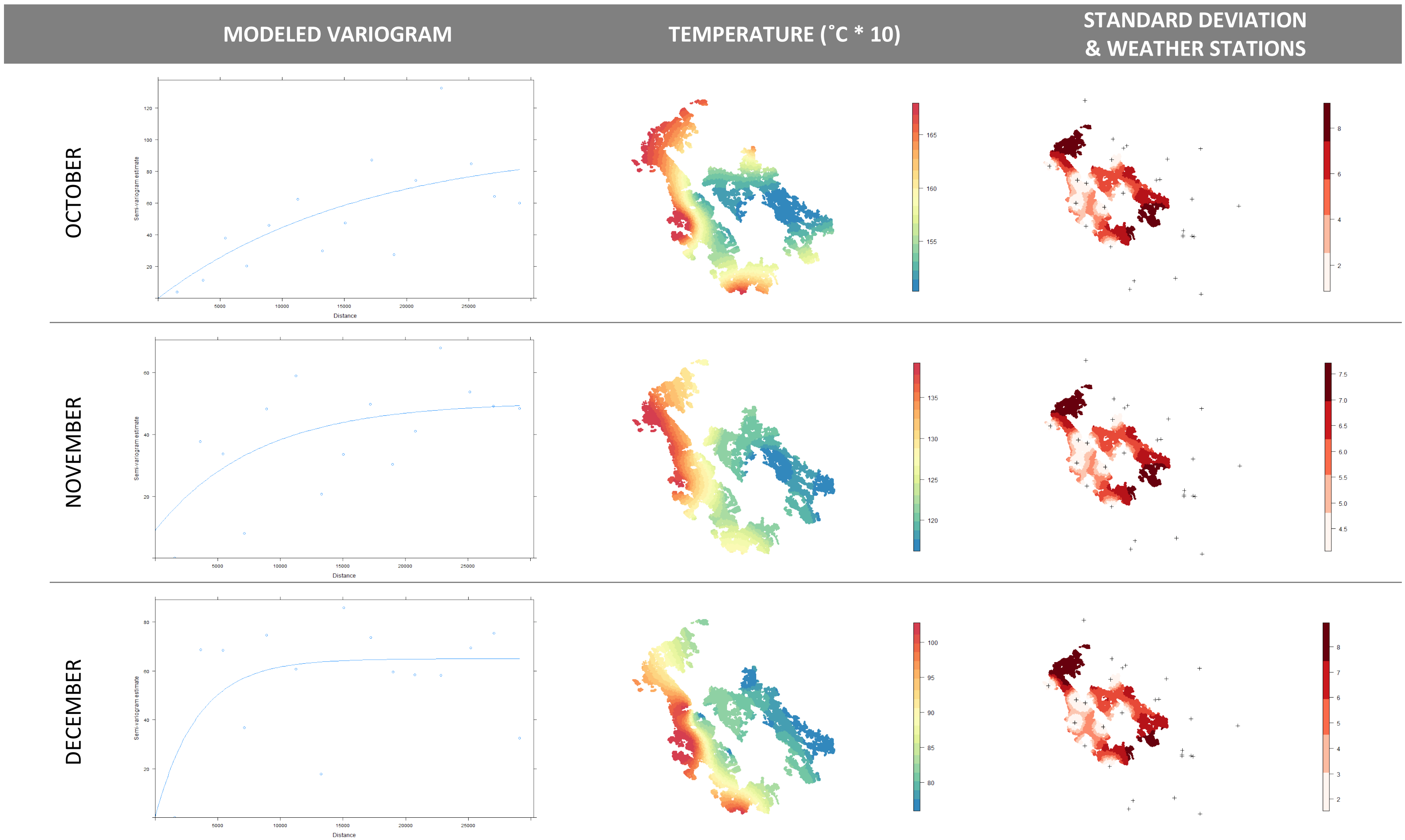


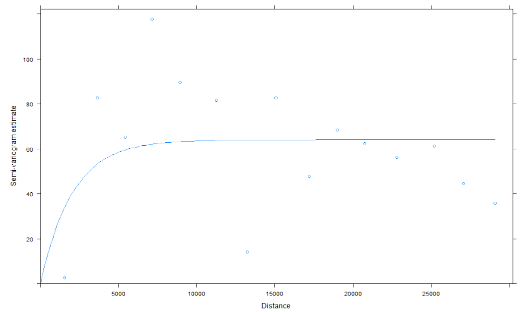
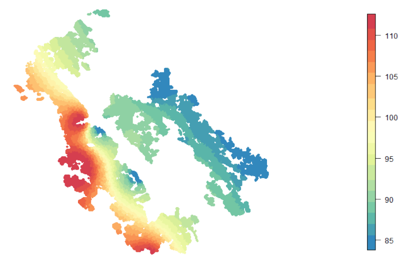
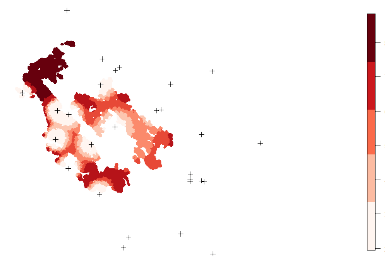
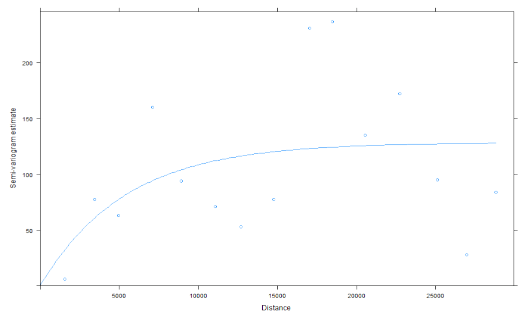
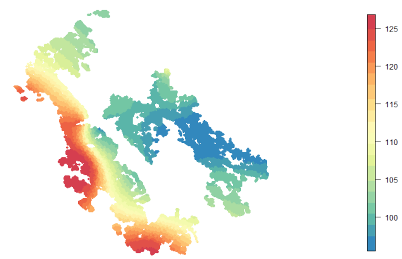
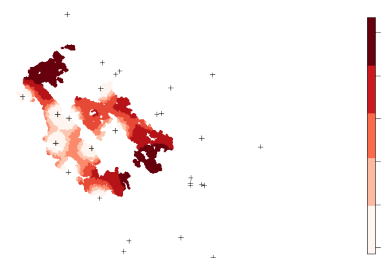
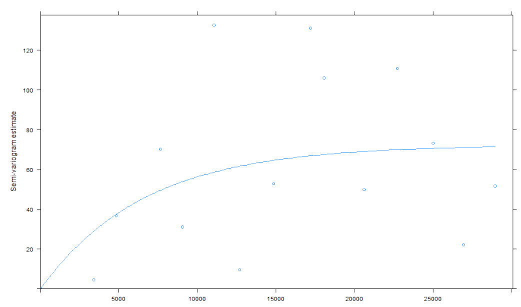
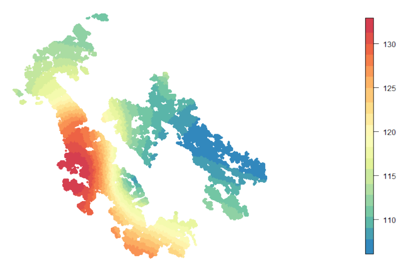
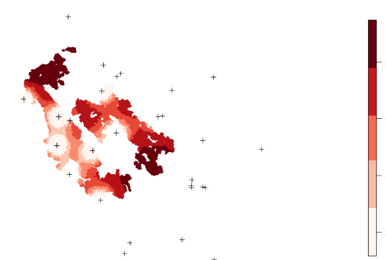


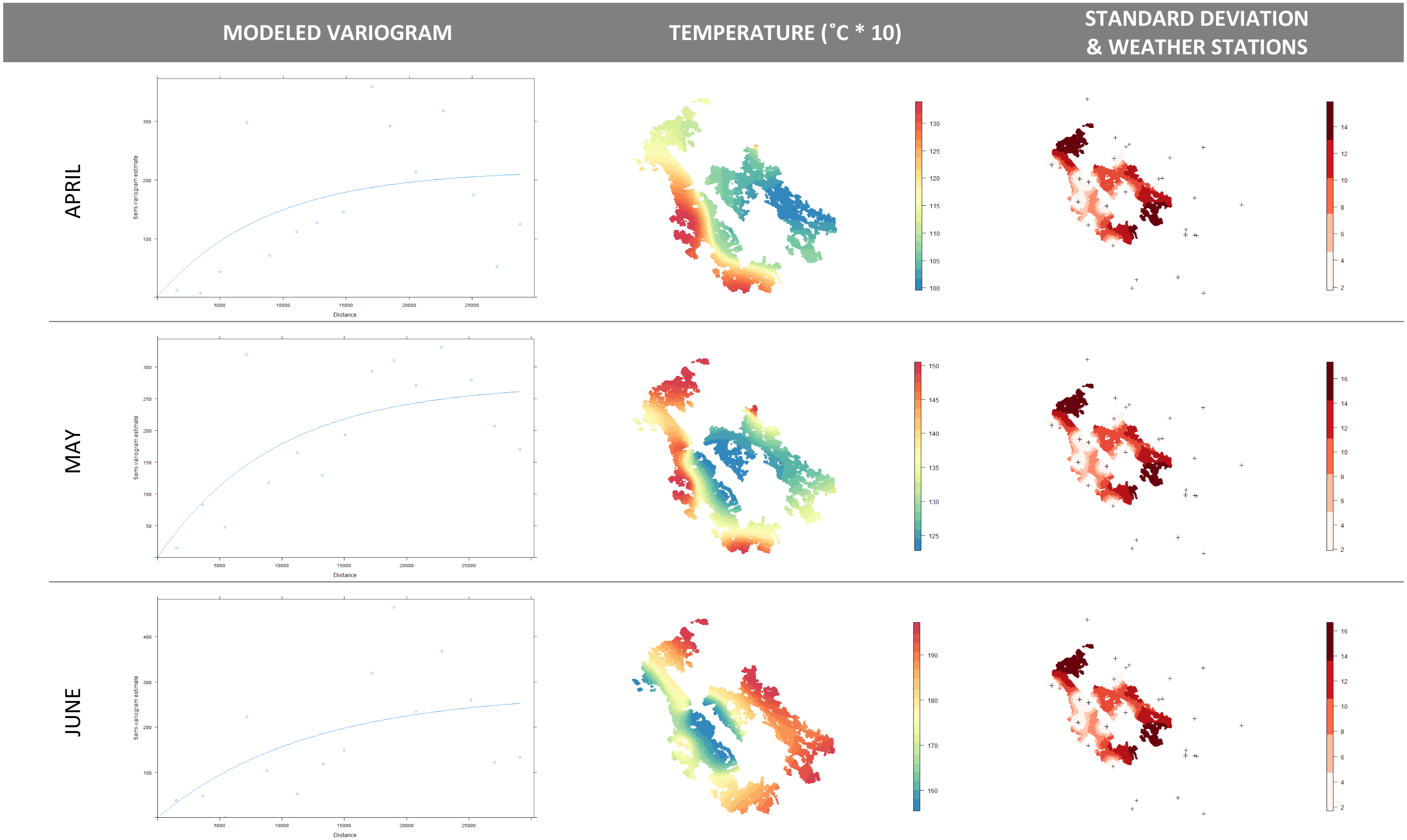


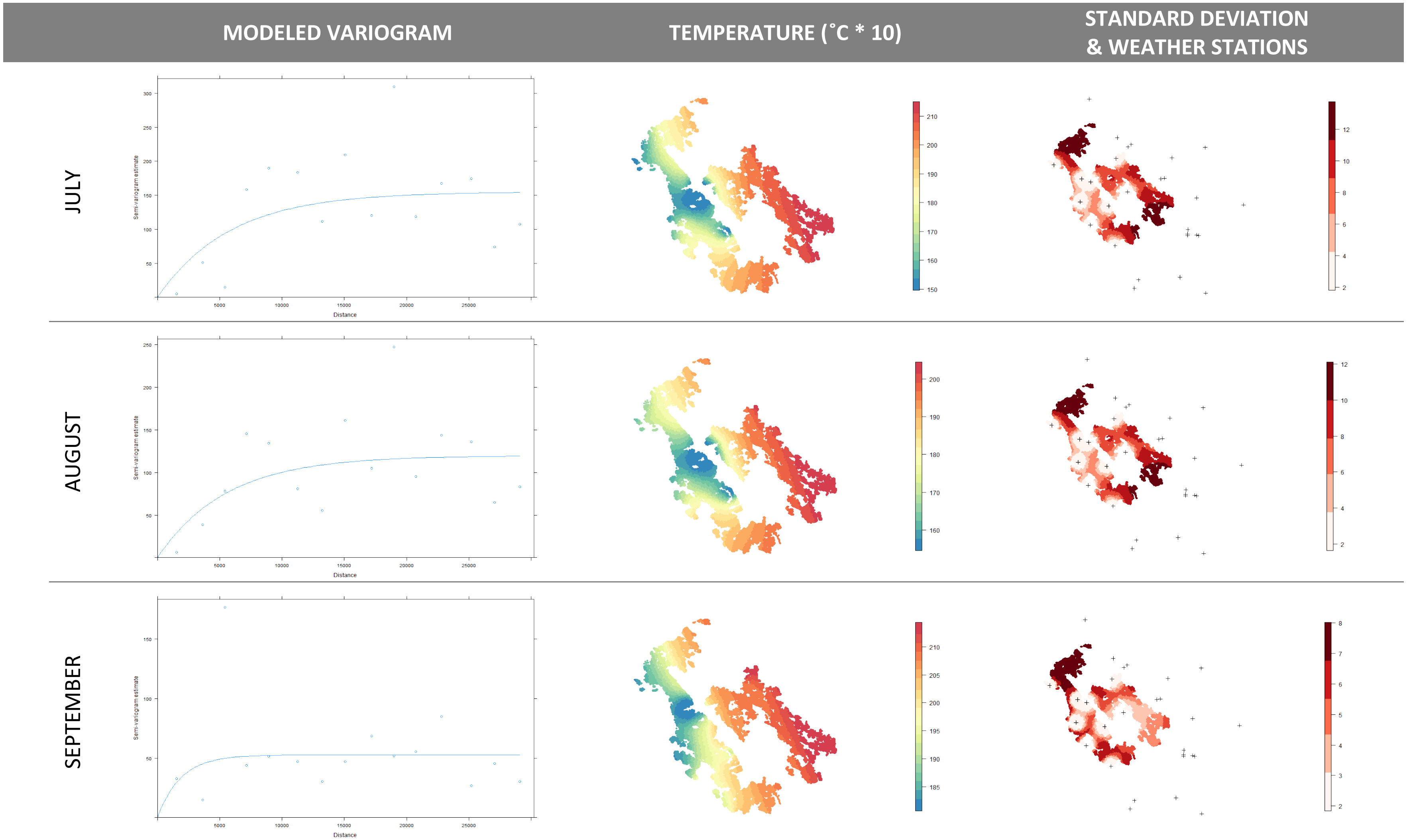

MODELED VARIOGRAM
TEMPERATURE (˚C * 10)
STANDARD DEVIATION & WEATHER STATIONS
APRIL
MAY
JUNE
Semi-variogram estimate
Distance

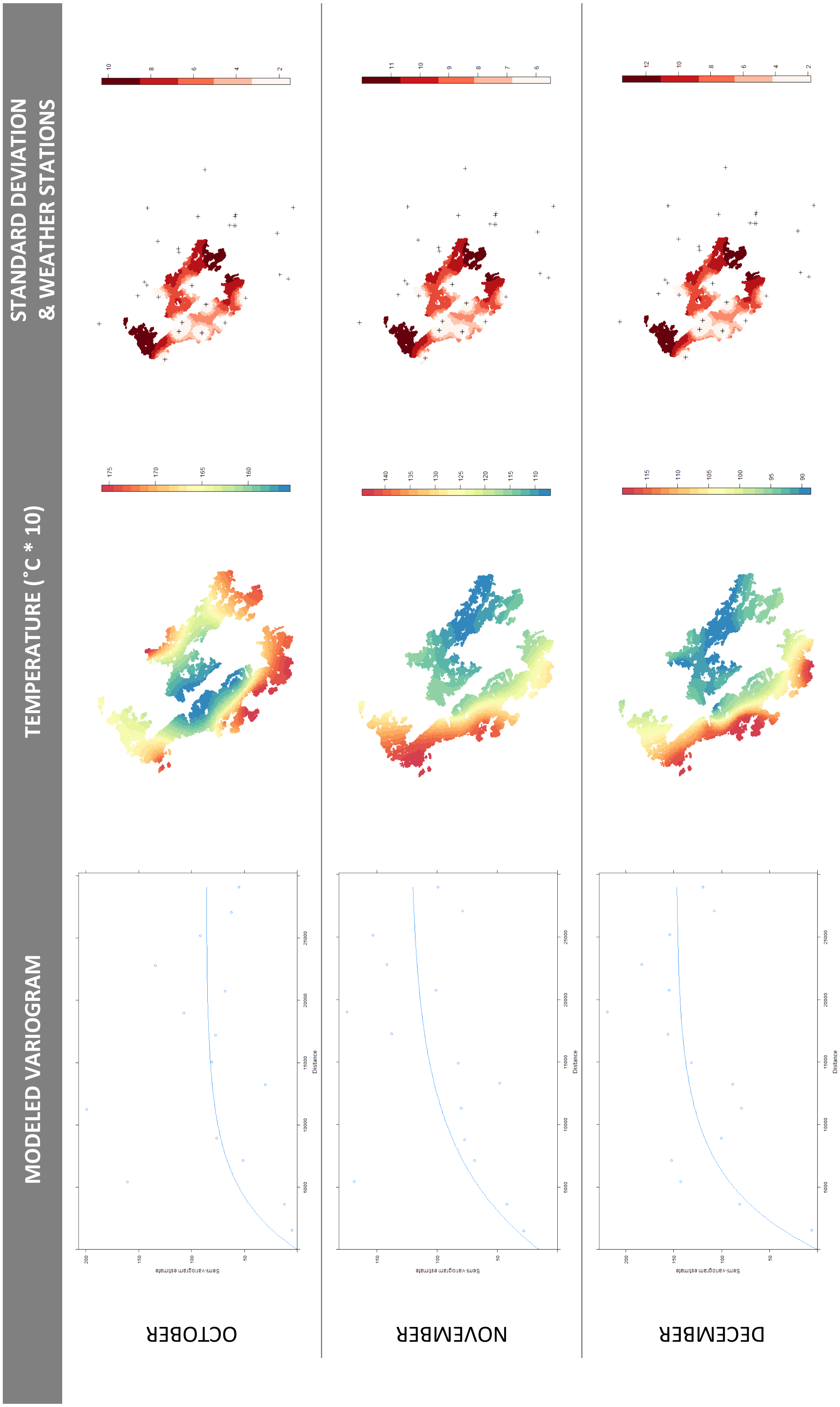
	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
JULY			
AUGUST			
SEPTEMBER			


MODELED VARIOGRAM
TEMPERATURE (˚C * 10)
STANDARD DEVIATION
& WEATHER STATIONS
OCTOBER
NOVEMBER
DECEMBER
Semi-variogram estimate
Distance

2010	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
JANUARY			
FEBRUARY			
MARCH			


MODELED VARIOGRAM
TEMPERATURE (˚C * 10)
STANDARD DEVIATION
& WEATHER STATIONS
APRIL
MAY
JUNE
Semi-variogram estimate
Distance

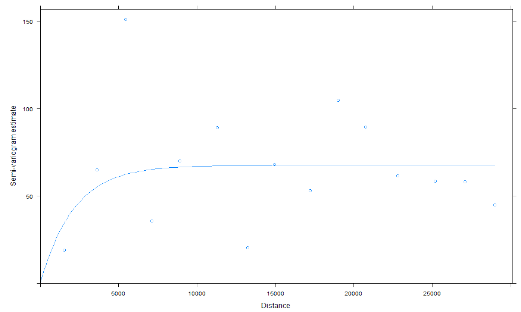

MODELED VARIOGRAM
TEMPERATURE (˚C * 10)
STANDARD DEVIATION & WEATHER STATIONS
JULY
AUGUST
SEPTEMBER
Semi-variogram estimate
Distance

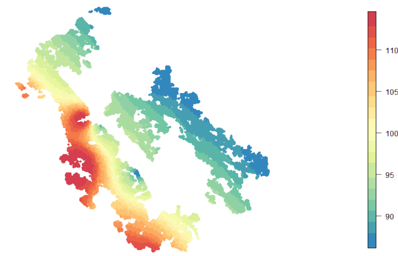
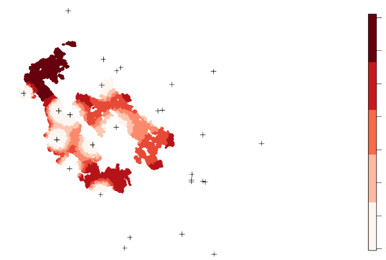

MODELED VARIOGRAM
TEMPERATURE (°C * 10)
STANDARD DEVIATION
& WEATHER STATIONS
OCTOBER
NOVEMBER
DECEMBER
Semi-variogram estimate
Distance

	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS

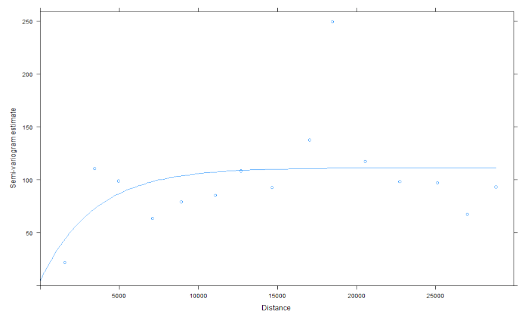
2011

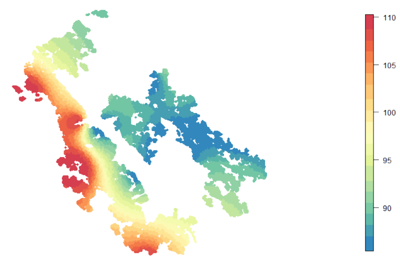
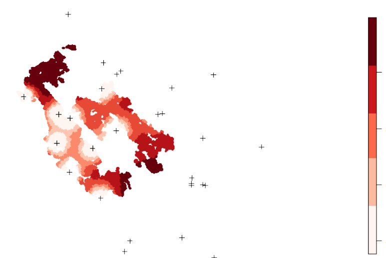
JANUARY



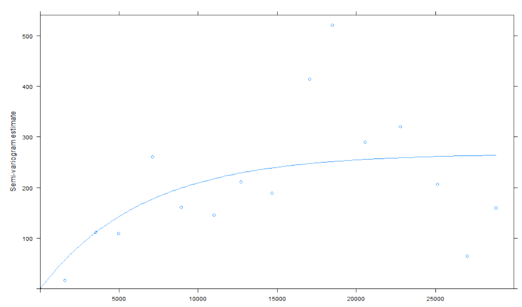



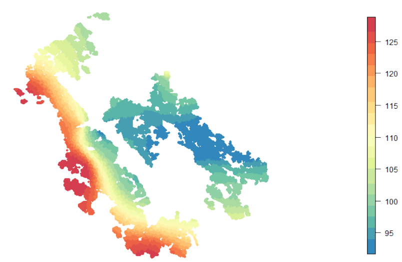
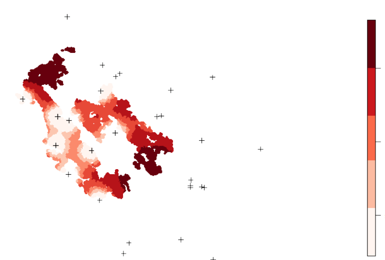
FEBRUARY



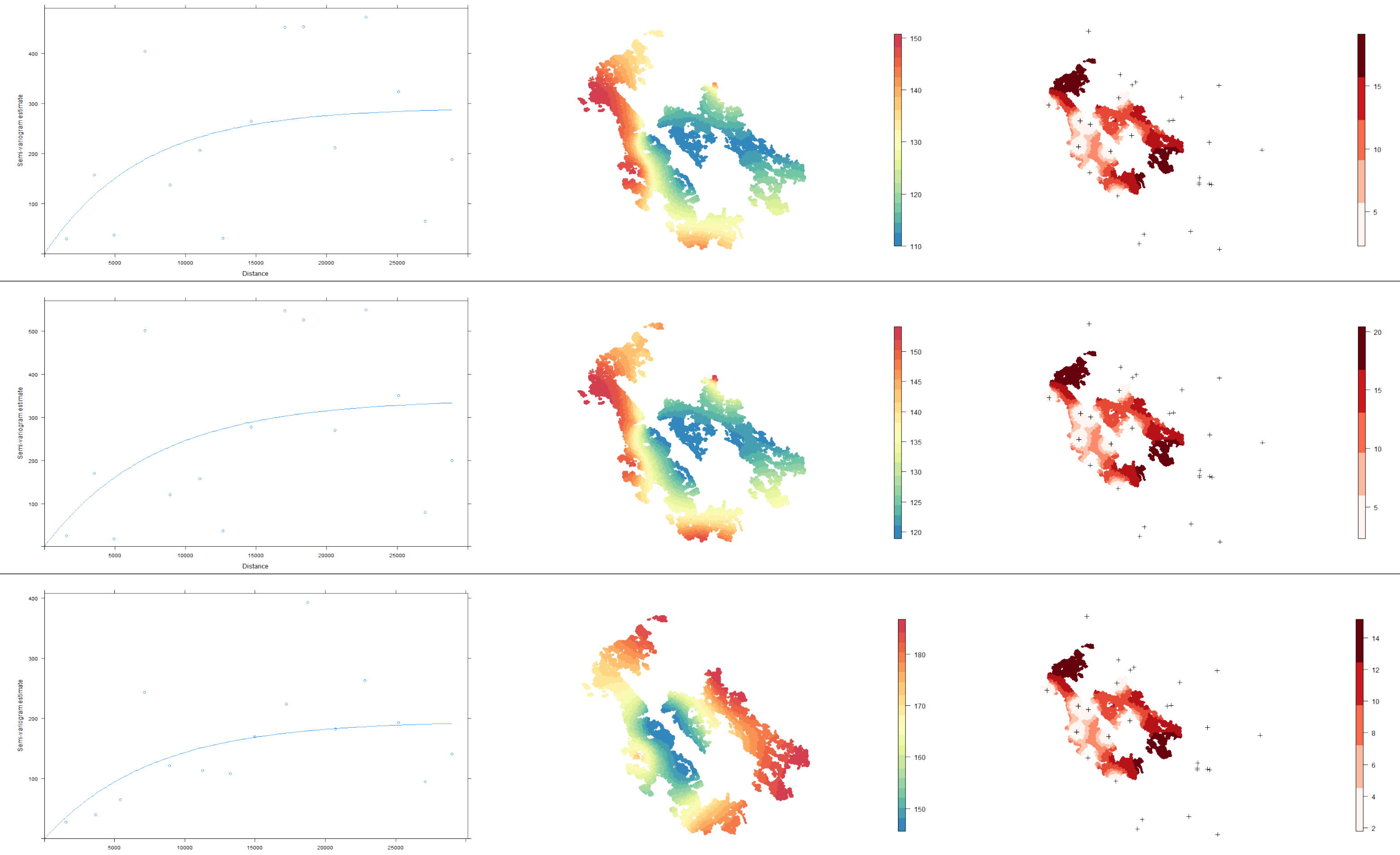



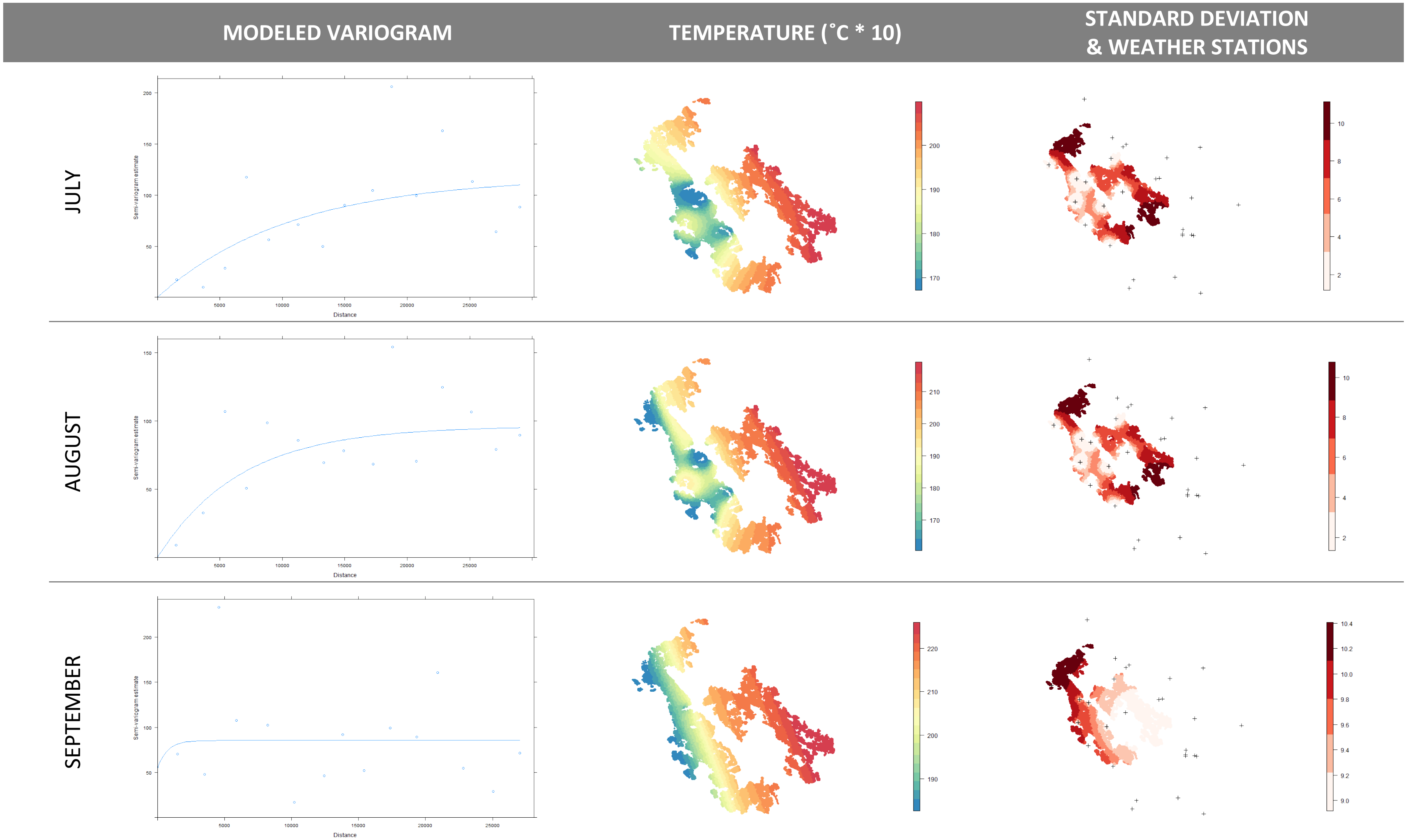
MARCH

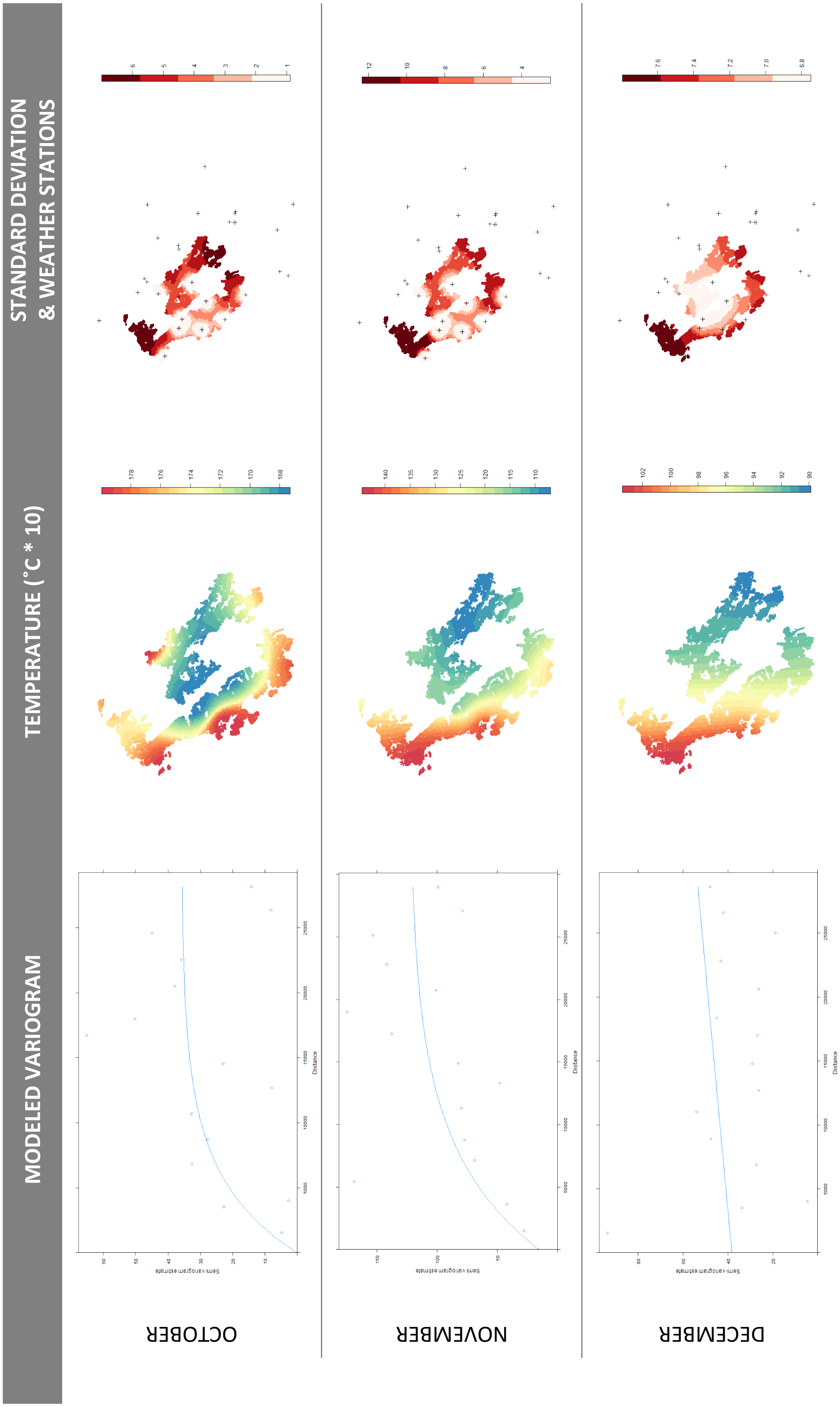


	MODELED VARIOGRAM	TEMPERATURE (˚C * 10)	STANDARD DEVIATION & WEATHER STATIONS
APRIL			
MAY			
JUNE			




MODELED VARIOGRAM
TEMPERATURE (˚C * 10)
STANDARD DEVIATION
& WEATHER STATIONS
JULY
AUGUST
SEPTEMBER
Semi-variogram estimate
Distance
5000
10000
15000
20000
25000
50
100
150
200
170
180
190
200
210
220
2
4
6
8
10
9.0
9.2
9.4
9.6
9.8
10.0
10.2
10.4


MODELED VARIOGRAM
TEMPERATURE (˚C * 10)
STANDARD DEVIATION & WEATHER STATIONS
OCTOBER
NOVEMBER
DECEMBER
Semi-variogram estimate
Distance

Appendix 3. Modeling intraurban precipitation

While climatic temperature may vary smoothly over space, precipitation has been shown to be patchy in the short term (Jensen and Pedersen 2005). For this reason, it can be difficult to model precipitation volumes at a fine spatiotemporal scale with the available data and modeling techniques. This appendix summarizes precipitation data sources and interpolation methods in the context of monthly water demand modeling, develops a "Rainy Days" variable using Ordinary Kriging, and then discusses directions for future research on intraurban precipitation.

Data sources

There are two major sources of precipitation data. The first is radar, collected by the National Weather Service's Next Generation Radar III (NEXRAD III). The second is rain gauge, collected at NWS's network of weather stations. Radar data is attractive because it is the most accurate method of capturing the spatial distribution of the intensity of a storm, even though it cannot directly measure precipitation volumes (precipitation volumes are determined by calibrating against rain gauge data). Rain gauge data is attractive because it is the most accurate volumetric measurement of precipitation.

NEXRAD III

The study area is covered by four NEXRAD III Doppler radar stations (see Figure 1). These stations measure several different precipitation variables, including reflectivity, which indicates precipitation intensity (National Climatic Data Center n.d,). Doppler radars will detect most precipitation within 145 km (90 miles) and heavy precipitation within as much as 250 km (155 miles) (Weather Underground n.d.). Variables are collected at six to ten minute intervals (see Figure 2).

It is possible to download radar data from NWS. The maximum temporal aggregation available is a daily summary. Since this study takes place over seven years, this amounts to 2,555 days of weather data. Accessing the data requires a proprietary viewer, such as NOAA's Weather and Climate Toolkit.[3] The viewer can export to more common data formats. Due to the necessity of opening in the viewer and then exporting, it is difficult process to automate and makes radar data challenging to work with over a long study period

[3] http://www.climate.gov/news-features/decision-makers-toolbox/weather-and-climate-toolkit

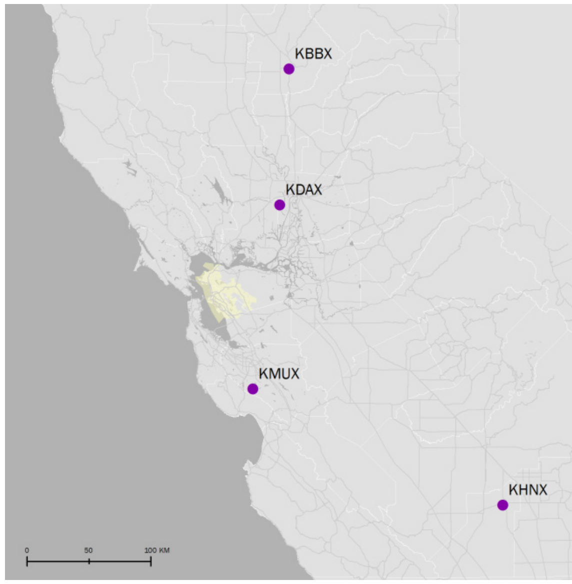


Figure 43. Bay Area radar stations (KBBX, KDAX, KMUX, KHNX) are located from approximately 90 to 130 miles apart. Study area highlighted in yellow.

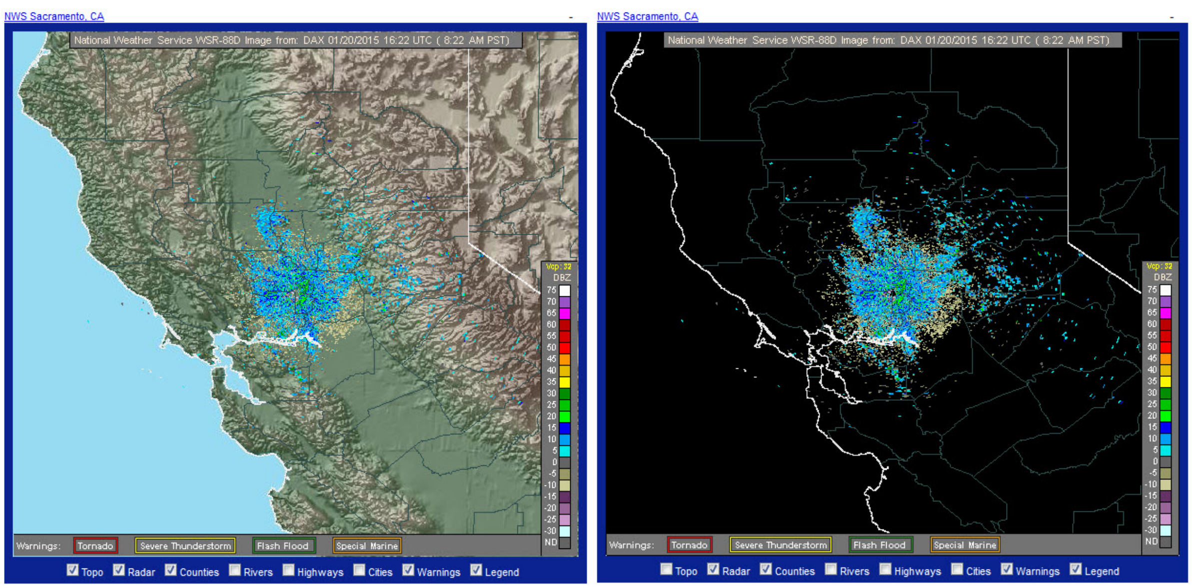


Figure 44. A sample radar image from station KDAX, visualized two ways (topographic relief visible on the left, none on the right). A new radar scan will produce a new image every six to ten minutes. Source: http://radar.weather.gov/radar.php?rid=dax

Radar is also challenging to work with because reflectivity data products (the variable directly measured by radar) do not immediately provide a useful input to a household water demand model. The spatial resolution is relatively coarse in the context of a parcel-level study (precise resolution depends on distance from radar station). The precipitation volumes that NWS provides are best approximations based on calibration from available rain gauges. Calibrating against rain gauges may not always be reliable, however, since a single radar pixel may contain a high degree of variability (Wood, Jones, and Moore 2000). Jensen and Pederson (2005) find that within a single radar pixel, neighboring rain gauges may vary up to 100%. Increasing the resolution of precipitation data is an active area of research by urban hydrologists. Because radar data is computationally intensive and does not provide direct volumetric data, it is not used in this study.

Rain Gauge

Rain gauge data are point measurements of precipitation volumes, as opposed to characterizations of the spatial pattern and intensity of a storm. From NWS's National Climactic Data Center, it is possible to download monthly rain gauge summaries, the most appropriate temporal scale for this study.

There are fifteen to thirty-seven weather stations reporting precipitation data near the study area. Increasingly more weather stations opened through the course of the study period, more than doubling the number of reporting stations from start to finish (see Figure 3). There are two precipitation variables of interest: (1) Total precipitation amount for the month; (2) Number of days in month with greater than or equal to 0.1 inch of precipitation.

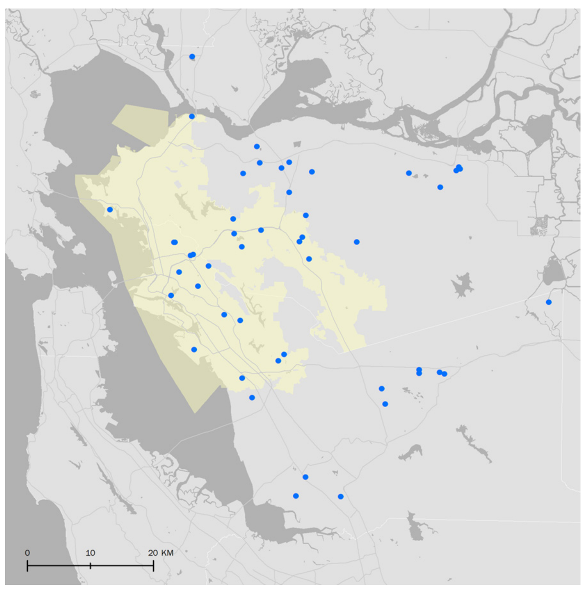


Figure 45. National Weather Service stations recording precipitation (blue) in the EBMUD area (yellow) between January 2005 and December 2011. Distributed by the National Climactic Data Center.

Most studies that account for precipitation focus on precipitation volumes. There are a smaller number of studies that examine counts of the number of rainy days. Studies may evaluate rainy days because of the psychological impact of the presence of any amount of rain, or because the only available data is the number of days above a precipitation threshold (Schleich and Hillenbrand 2009).

Rainy days varies more smoothly over space than patchy precipitation volumes. Because most interpolation methods model smoothly varying surfaces, this analysis focuses on rainy days (the number of days in a month with greater than or equal to 0.1 inch). Given the parameters of the study, rainy days is a more robust variable than precipitation volumes.

Data cleaning

Temperature modeling (see Appendix 1) indicates that there are errors in the reported weather data—both gross errors and subtle errors. Yet, given that the rainy days data will likely be more patchy than temperature, it is more difficult to determine where errors lie. One challenge is that rain gauges are most likely to malfunction by ceasing to collect data, which would produce a false zero. Yet, zero values are sufficiently common that detecting false zeros is only possible on the most rainy months.

As a result, the data was cleaned exclusively to remove gross outliers. All points four standard deviations outside mean of the mean were dropped from the analysis. There is not a strong reason to believe that the data should be normally distributed, so this method was used only to act as a coarse filter. All small reporting errors remain.

Model selection

Interpolating precipitation data has received some attention in the literature, though there is not a consensus on the best methods. Some researchers find that Kriging offers the most accurate interpolations (Vicente-Serrano, Saz-Sánchex, and Cuadrat 2003), while others assert that Kriging provides no additional benefit over computationally easier, deterministic methods (Dirks et al. 1998). It is likely that the most appropriate method is context dependent and relates to the quality of the measured data, the size of the study area, and local precipitation patterns. This study begins by examining Kriging, Inverse Distance Weighting and spline as possible approaches to interpolation.

Comparing Kriging, IDW and Spline

The most significant advantages of Kriging, a stochastic method, is that it accommodates possible measurement error in the data. Deterministic algorithms, such as Inverse Distance Weighting (IDW) and spline, assume that the measured values are accurate. When there may be error in the data, Kriging should lead to more reliable predictions.

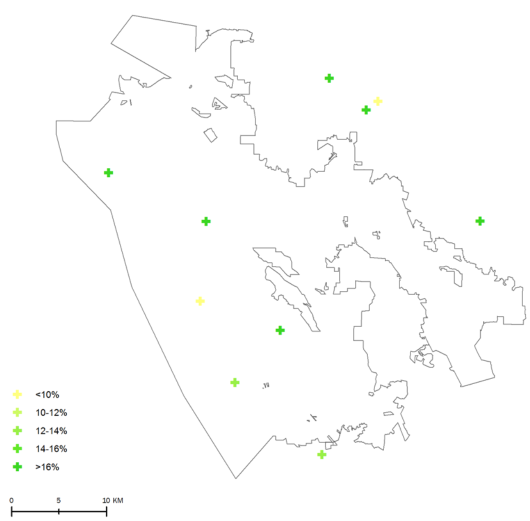
Unlike IDW or spline, however, Kriging is most suited to data with clear spatial relationships. To derive the spatial parameters required for Kriging from the variogram (nugget, sill, range), R's gstat package optimizes model fit from a set of starting parameters (see Table 2). In some cases, there may be multiple mathematical solutions. This is most frequently the case when there are not strong spatial relationships among the data or few available data points. In these instances, it is necessary to choose the set of parameters that visually appears to best fit the data, as it is displayed in the variogram. Visually fitting a variogram cannot be automated, which makes Kriging difficult to scale over a longer study period.

The precipitation data, in contract to temperature, often exhibited weak spatial relationships. One particularly problematic month is May 2005, which showed no spatial trend (see Table 2). Because no trend was indicated, all interpolated points in the study area took on the same value as the global mean. May 2005 is a valuable month to examine more closely. Figure 4 compares the measured data points and the results from Kriging, Inverse Distance Weighting, and spline.

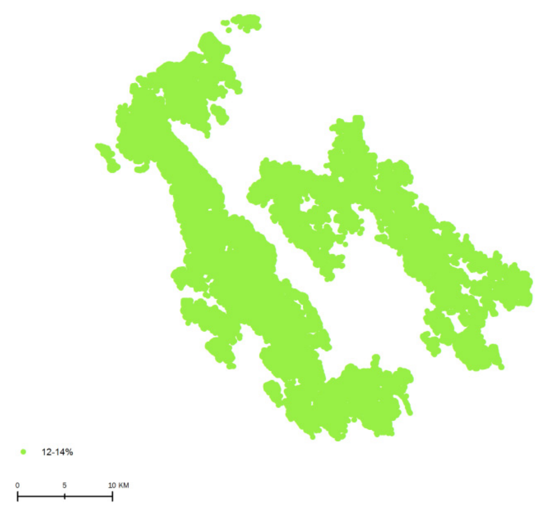
Each method results in a different pattern of interpolated data. It is impossible to be sure which is the most correct, given the sparsely available data. By comparing against the measured values, however, spline gives a range of values far wider than the values observed in the measured data (see Table 1). This indicates that spline is not the best method. [4] The choice between Kriging and IDW is more subtle.

Table 13. Range of rainy day values in the measured data and interpolated results for May 2005.

Method	Minimum	Maximum	Mean
Measured	6.4%	16.1%	13.2%
Kriging	13.2%	13.2%	13.2%
Inverse Distance Weighting (IDW)	6.4%	16.1%	13.7%
Splines	2.9%	31.8%	17.6%



Measured



Interpolated: Kriging

[4] It is possible to continue to refine the spline interpolation by manipulating the number of data points included in local approximation, but there is not a generalizeable approach. Spline does not offer an efficient, semi-automated interpolation method, as is necessary to process the 84 months of precipitation data used in this study.

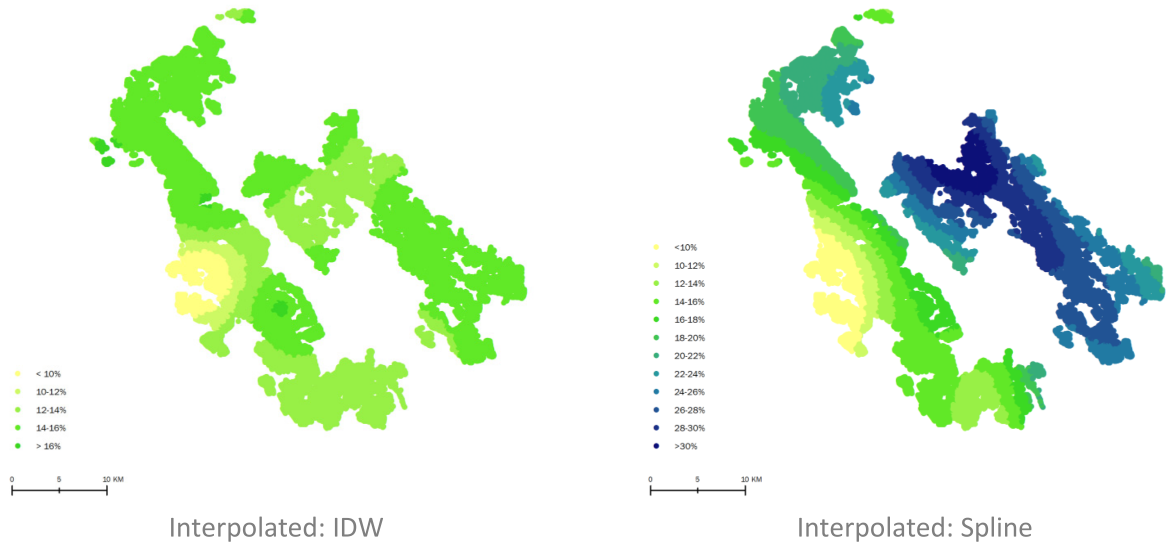


Figure 46. Rainy days in May 2005. Measured data and interpolated results.

To evaluate Kriging versus IDW, Figure 5 maps the measured values again. Inset 1 highlights three proximate stations. Two stations reported 16.1% rainy days, while the third reports 9.7% rainy days. If this difference is accurate or an error is not possible to discern, given the available information. Inset 2 highlights a reported value of 6.5%, much lower than any surrounding points. Again, it is not possible to determine if this area is truly experiencing fewer rainy days, or if the data is in error.

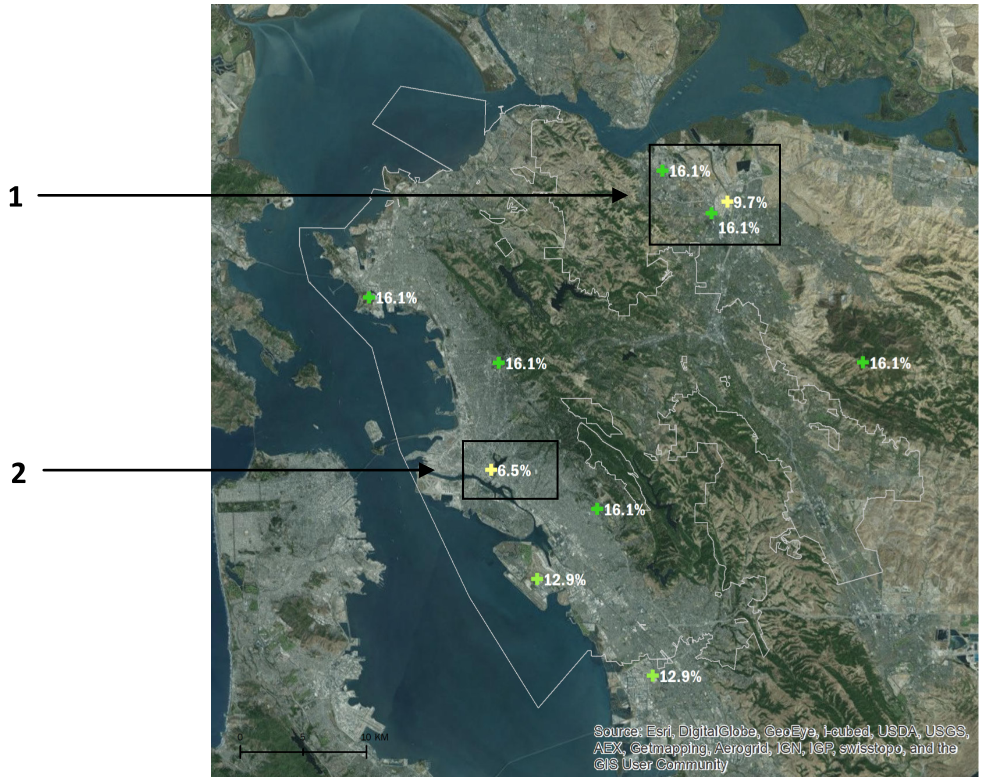


Figure 47. Data from May 2005. Inset 1 shows three neighboring readings, where one is different. Inset 2 shows an isolated reading that is lower than any of the nearest data.

The choice between Kriging and IDW is primarily dependent on whether the measured data is reliable. Because of observed error with temperature data (recorded at the same weather stations, see Appendix 1) and well documented challenges with rain gauges, this dissertation assumes that there are errors in the data. Though Kriging may, at times, exhibit less variation in interpolated values than captured in the measured data, it is a more cautious approach.

Method: Ordinary Kriging

Universal Kriging was appropriate for temperature, given the spatial trends in the data (see Appendix 1), but precipitation shows little evidence of stationary trends (see Figure 6). Ordinary Kriging is appropriate when there is a constant, non-zero mean through the study area.

This precipitation analysis implements Ordinary Kriging by following five steps:

1. Normalizing the rainy days variable by the number of days in the month.
2. Dropping data point s greater than four standard deviations outside of the mean.
3. Fitting the variogram using R's gstat package.
4. Interpolating values at each house within the study area with gstat.
5. Taking the square root of the variance to find one standard deviation from the predicted value.

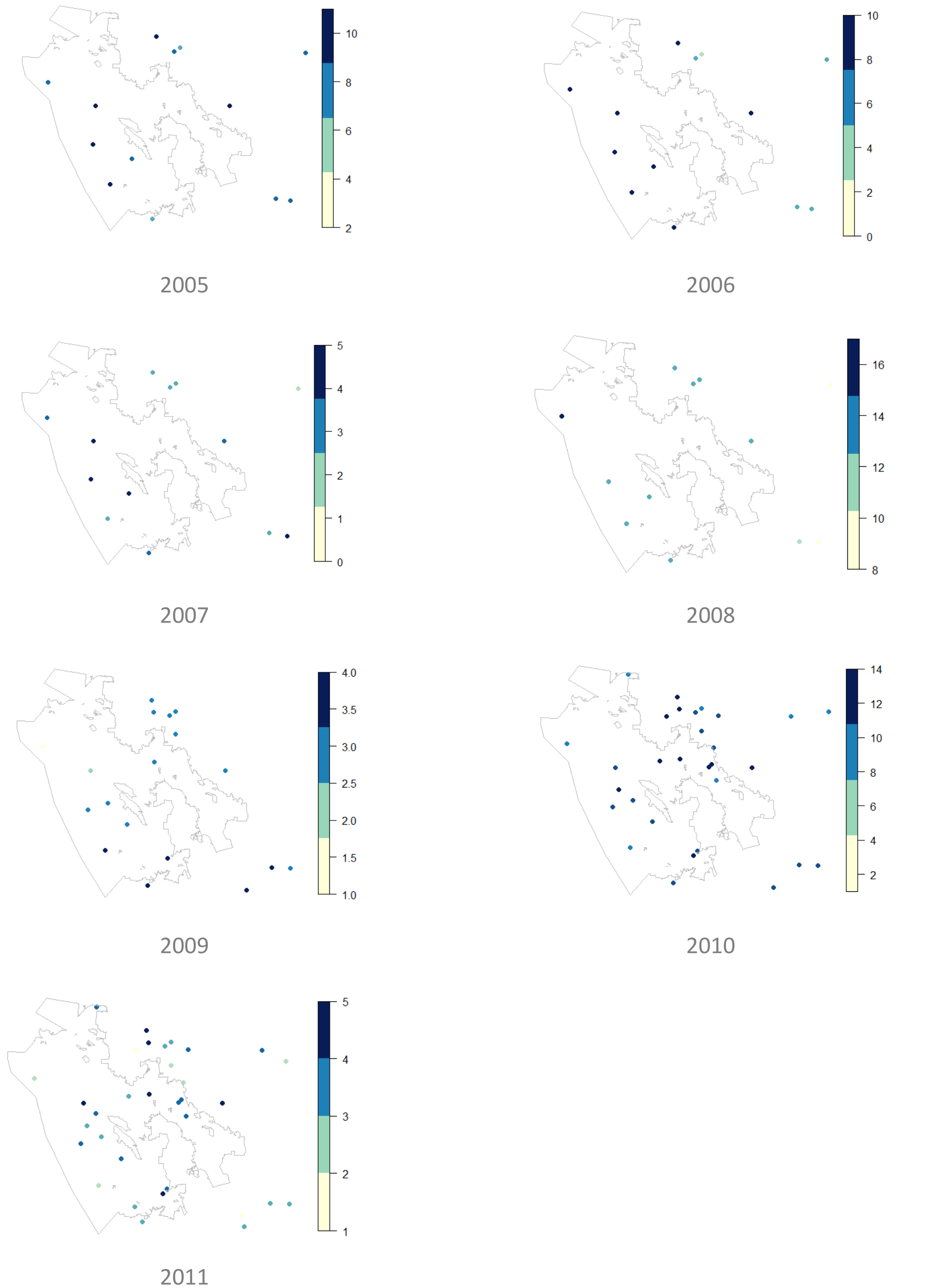


Figure 48. Rainy Days in January. There were 15 reporting stations in Jan 2008, increasing to 37 in Jan 2011. In this figure, the count of rainy days is less important than the pattern. For this reason, all months are displayed with the same range of colors. Each month shows a distinct point pattern, with no evidence of location-based trend.

Future research: finding patterns in precipitation

Our current precipitation data products have primarily been motivated by questions of short-term meteorological forecasts, long-range climate patterns, and basin-level hydrology. Though research agendas are shifting, to date there has been less interest in urban rainfall dynamics at a high spatial resolution (parcel) and medium temporal resolution (monthly). While this dissertation examines the parcel/month spatiotemporal scale due to the available water demand data, transactional data is often available at monthly aggregations, so this may be a useful unit when relating human behavioral patterns to precipitation.

Like temperature, high resolution rain gauge data is necessary for model adjudication. Though precipitation does not exhibit strong spatial patterns within the study area, it is possible that there are relationships with variables such as elevation, aspect, wind direction, or the presence of air pollution–all of which are significant predictors in other studies (Shepherd 2005)—that would imply the model should be approached with Universal Kriging or co-Kriging. With the available data it is difficult to test alternative model formulations: weather stations in the area are located at low elevations on flat aspects; wind data is sparse (collected primarily on NOAA weather buoys along the water) and subject to large errors; air pollution data is not collected at a high resolution. With a suite of higher resolution environmental data, it may be possible to identify stronger spatial patterns.

Table 14. Modeled variograms, interpolated precipitation, and standard deviation in the East Bay Municipal Utility District from January 2005 to December 2011. Temperature given in degrees Celsius * 10. The map of standard deviations includes the location of reporting weather stations, denoted with a cross.

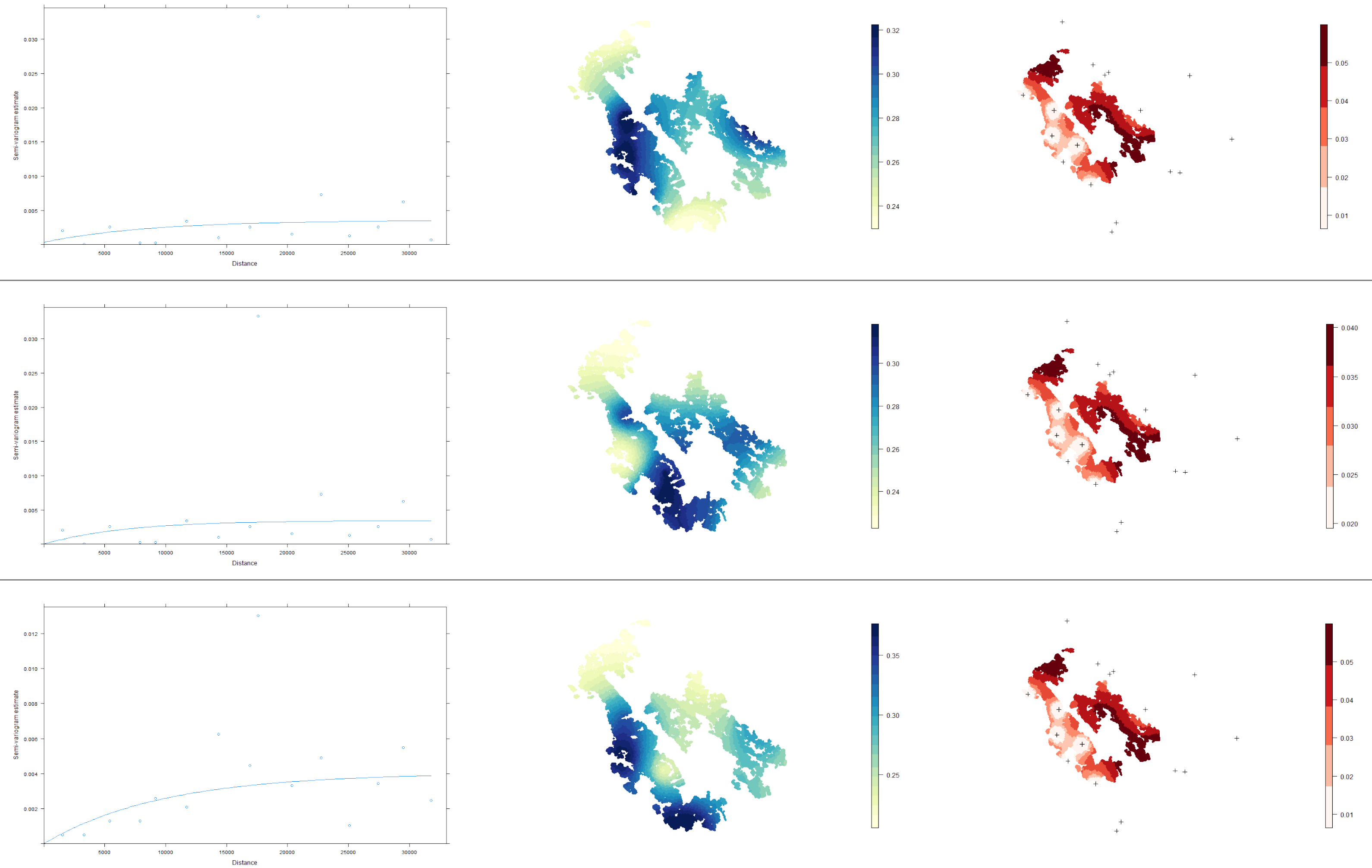
	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS

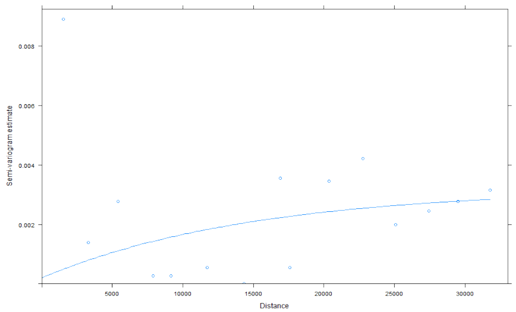
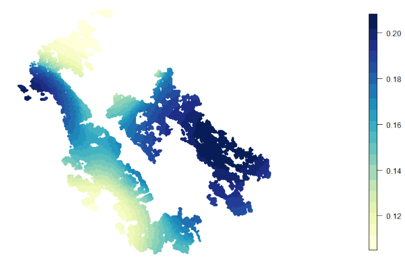
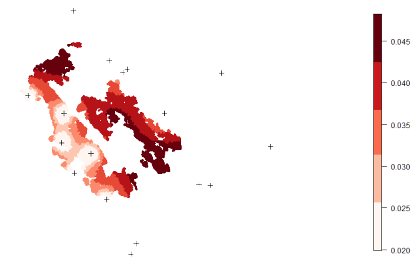
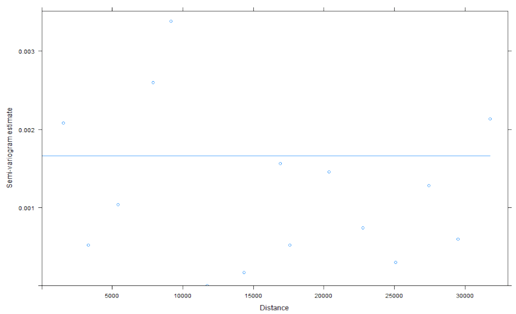
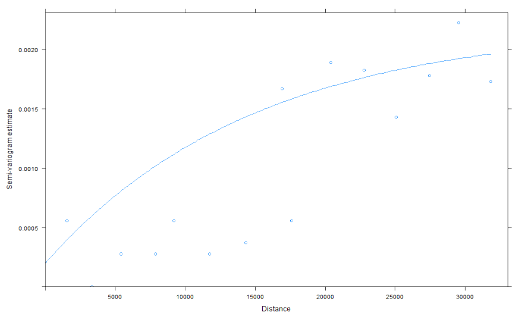
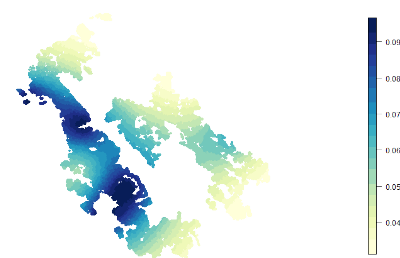
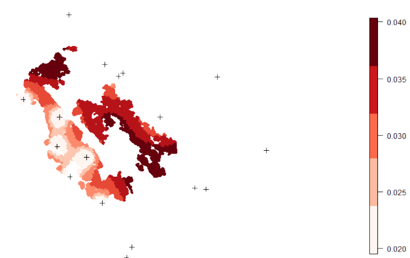
2005

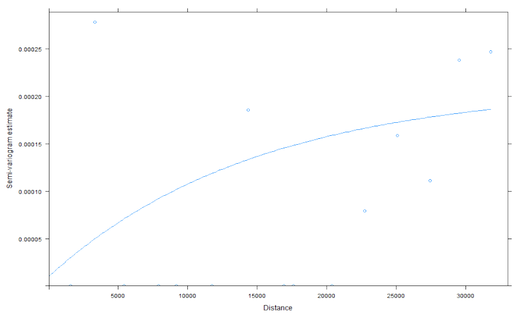
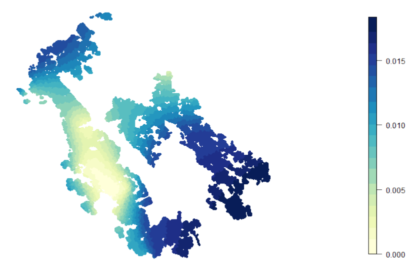
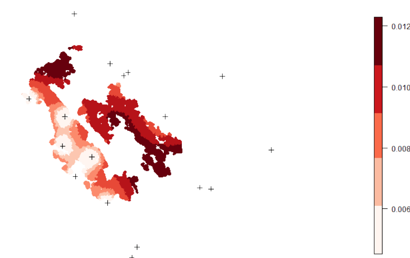
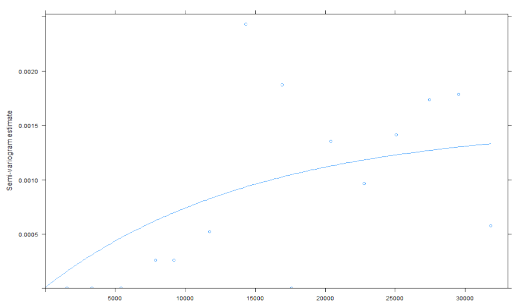
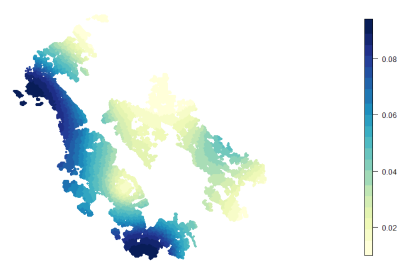
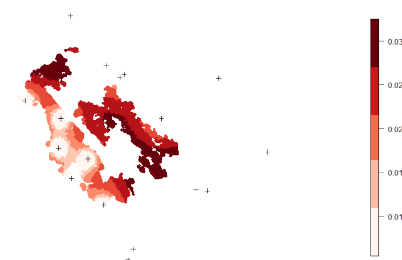
JANUARY

FEBRUARY

MARCH



	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
APRIL			
MAY			0.042
JUNE			

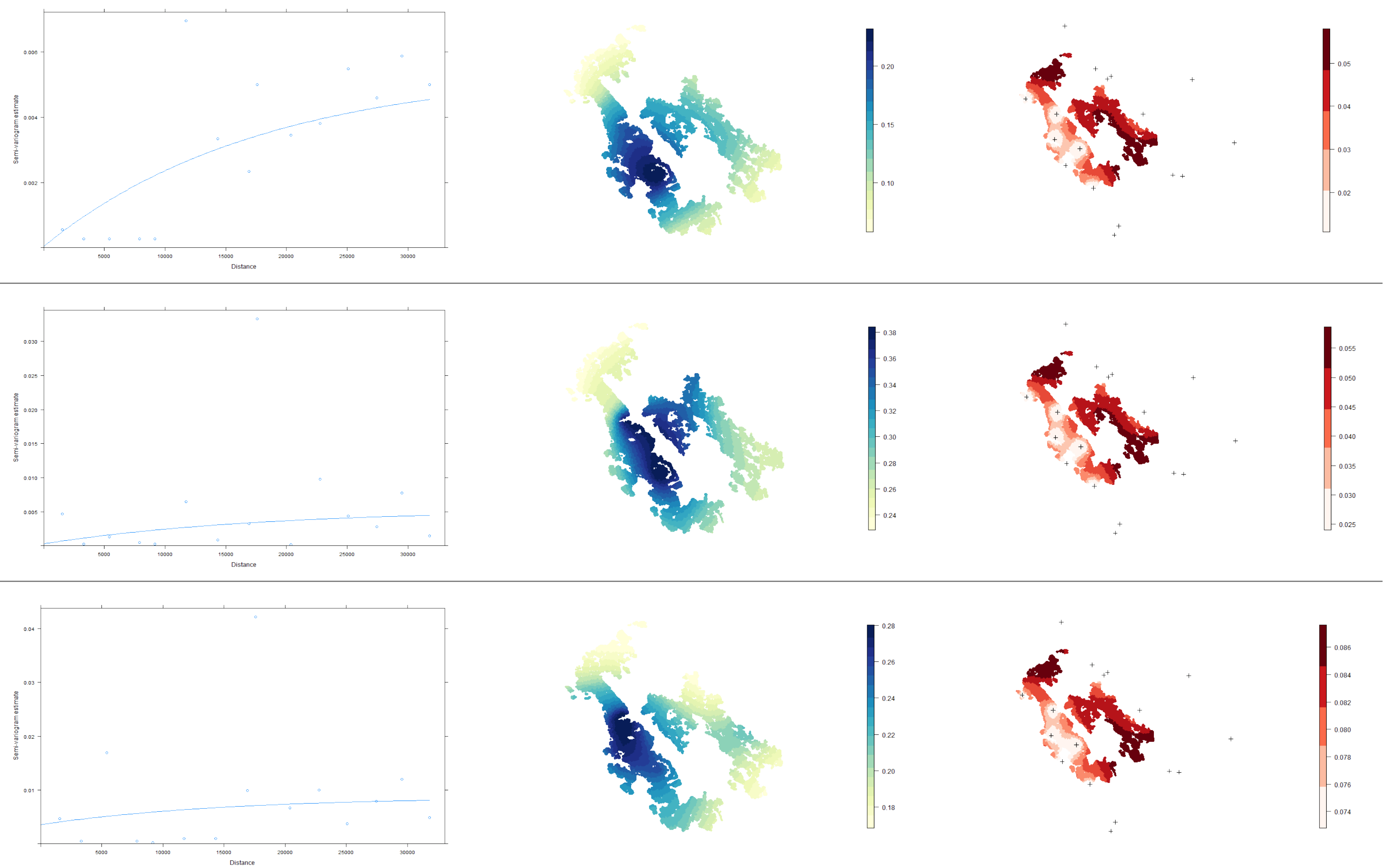
	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
JULY	n/a	n/a	n/a
AUGUST	n/a	n/a	n/a
SEPTEMBER			
OCTOBER			

MODELED VARIOGRAM | RAINY DAYS (% PER MONTH) | STANDARD DEVIATION & WEATHER STATIONS

NOVEMBER

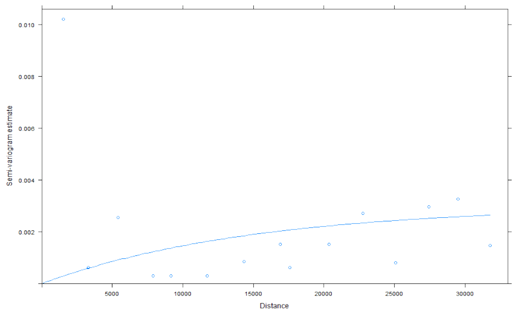
DECEMBER

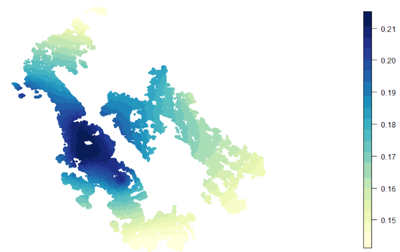
2006 JANUARY

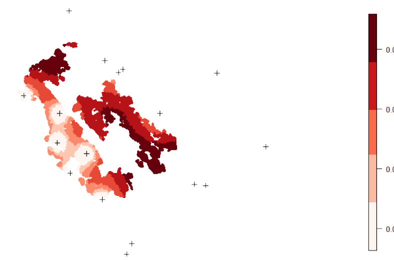


MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS

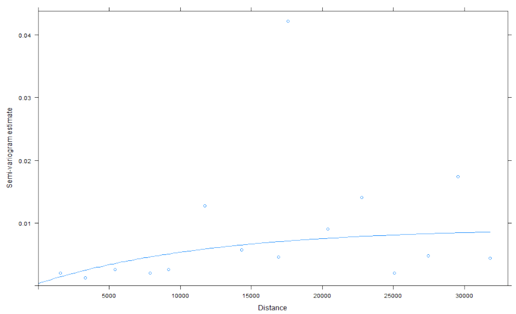
FEBRUARY

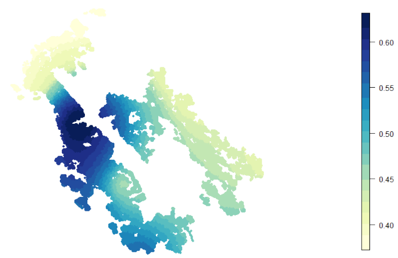


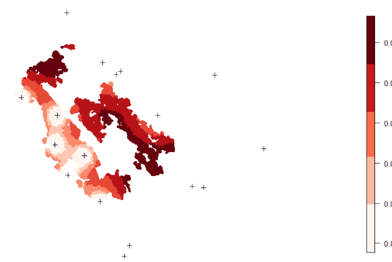




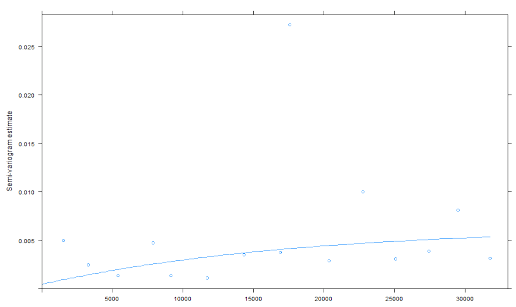
MARCH

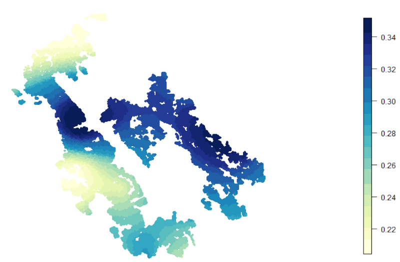


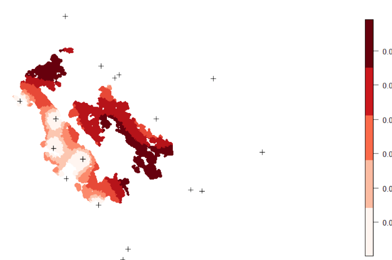


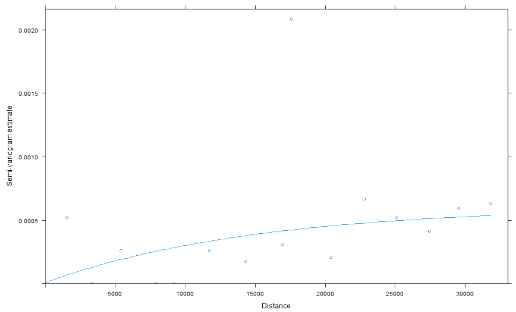
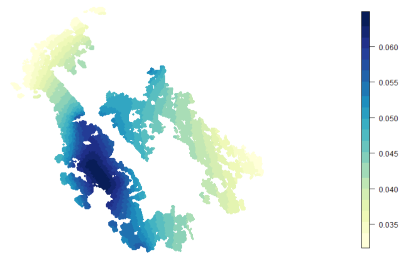
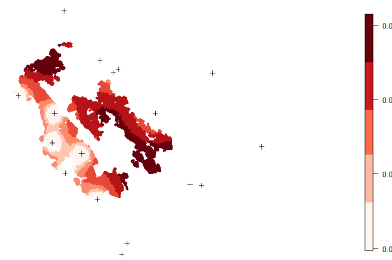


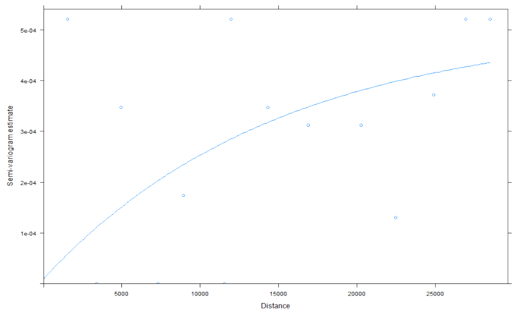
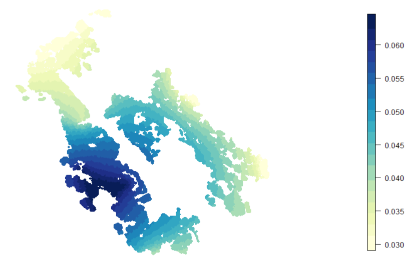
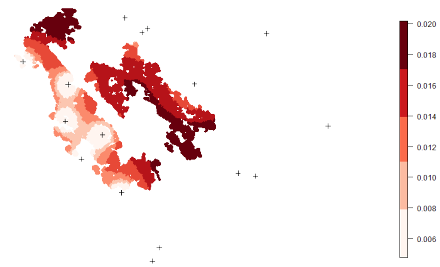
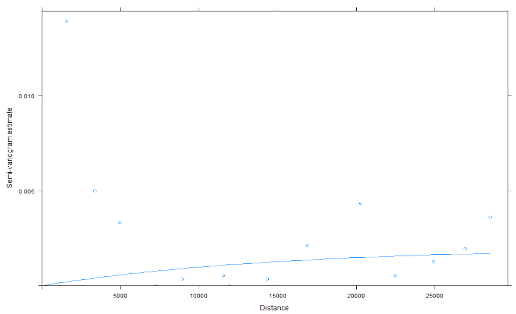
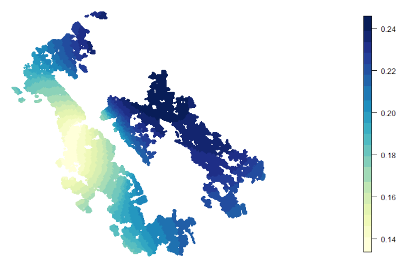
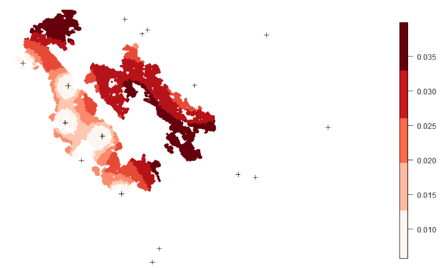
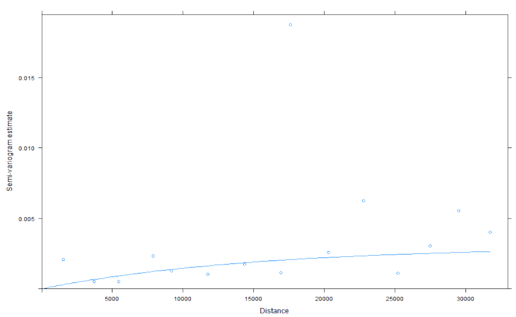
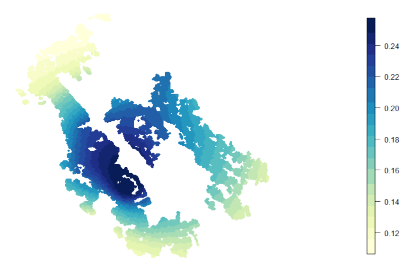
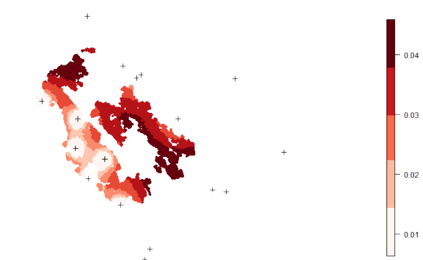
APRIL

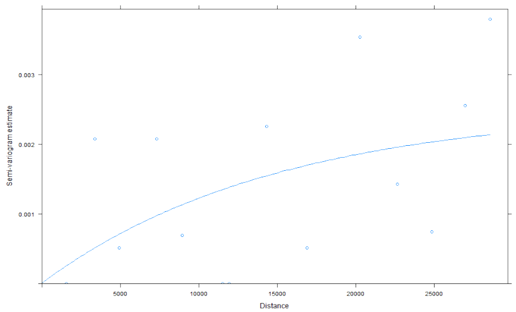
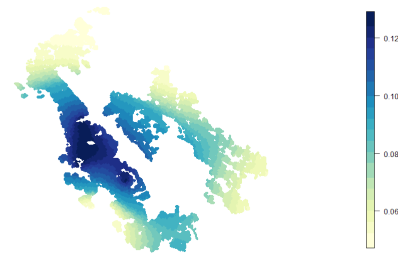
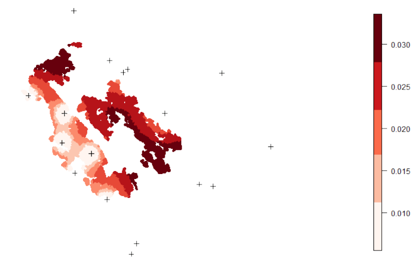
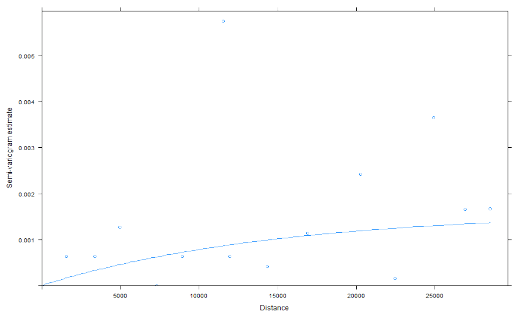
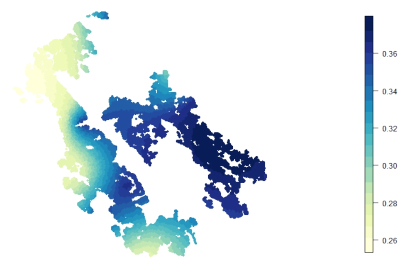
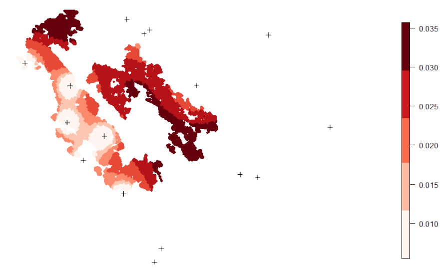
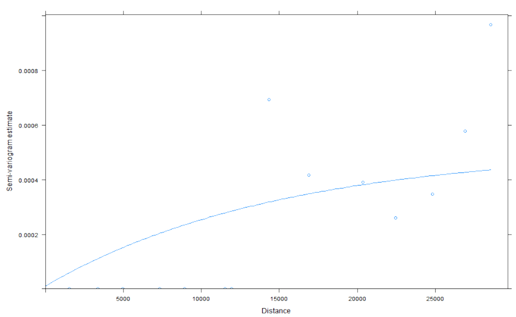
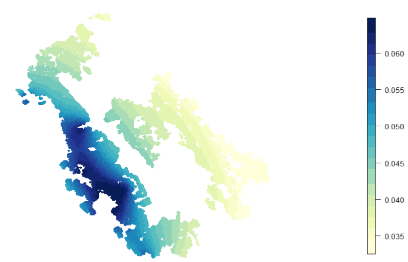
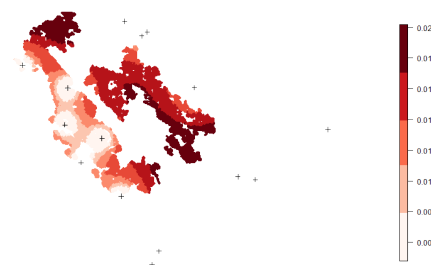




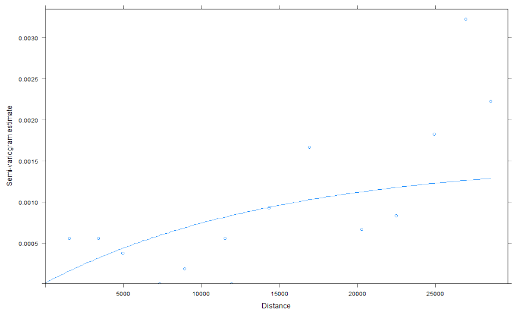
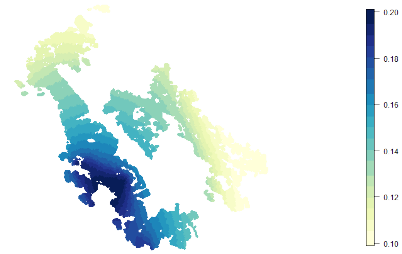
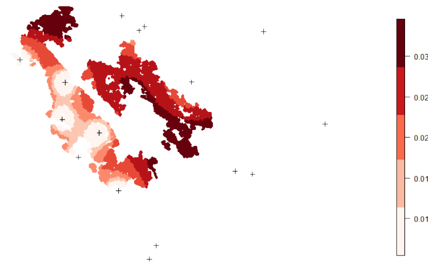
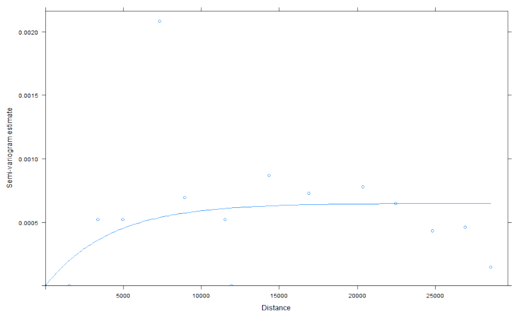
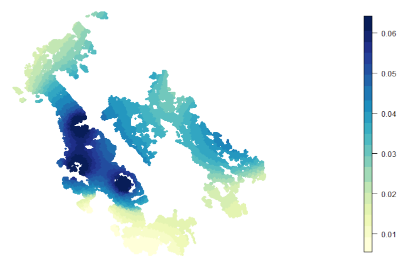
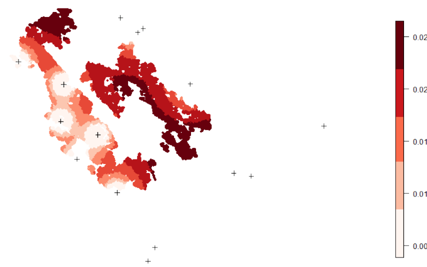


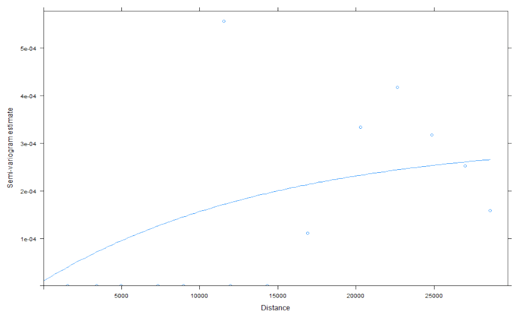
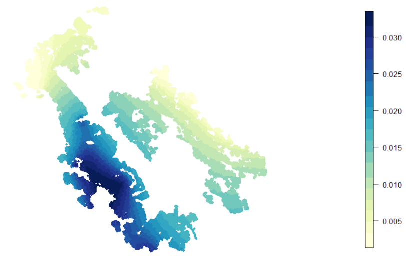
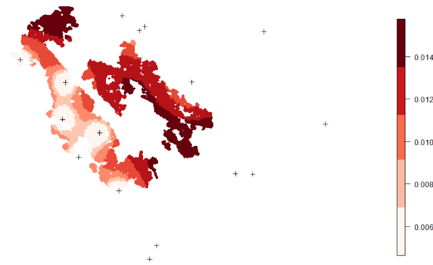
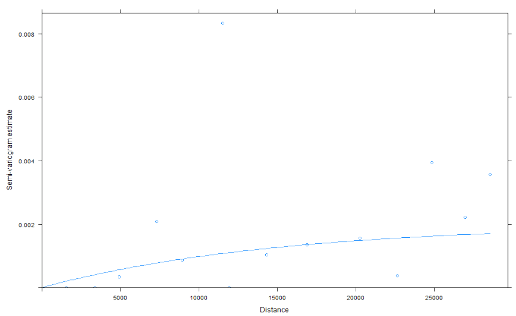
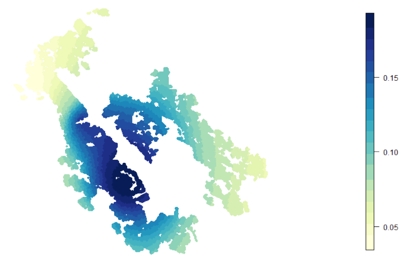
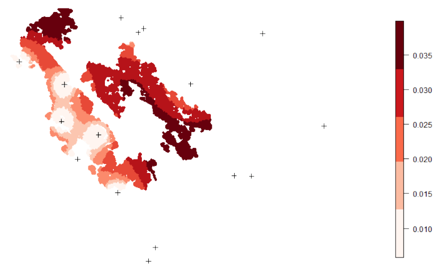
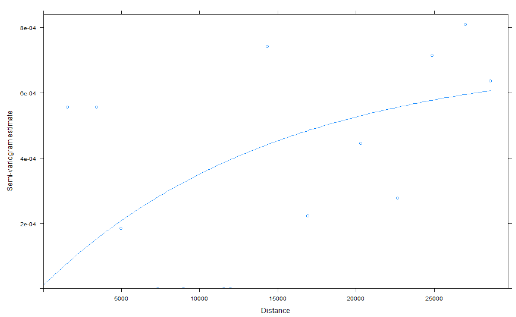
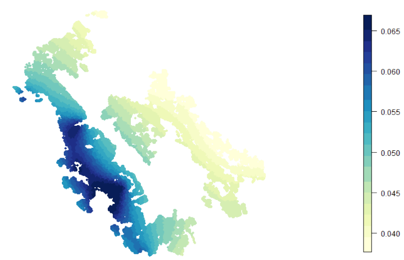
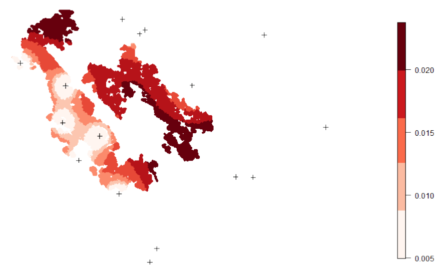
	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
MAY			
JUNE	n/a	n/a	n/a
JULY	n/a	n/a	n/a
AUGUST	n/a	n/a	n/a
SEPT	n/a	n/a	n/a

	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
OCTOBER			
NOVEMBER			
DECEMBER			

	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
JANUARY			
FEBRUARY			
MARCH			

2007

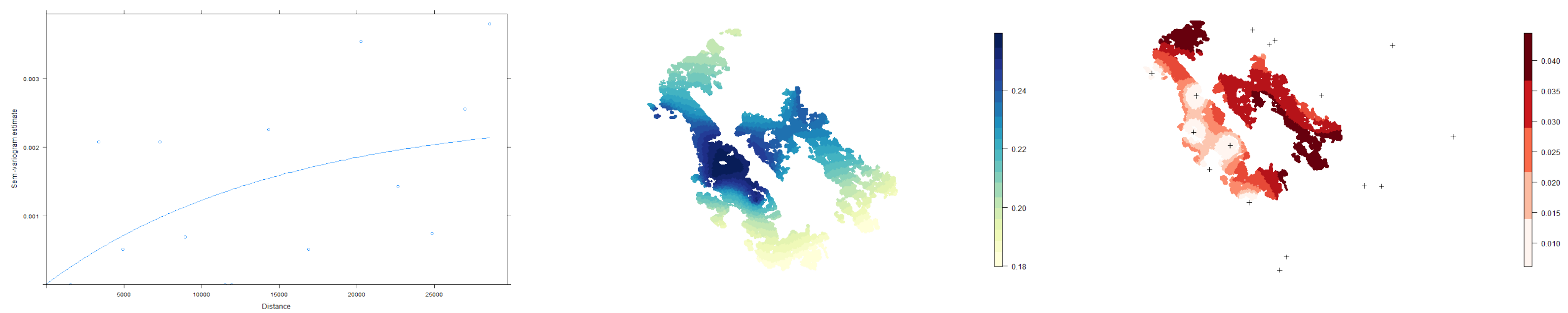
	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
APRIL			
MAY			
JUNE	n/a	n/a	n/a
JULY	n/a	n/a	n/a
AUGUST	n/a	n/a	n/a

	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
SEPTEMBER			
OCTOBER			
NOVEMBER			

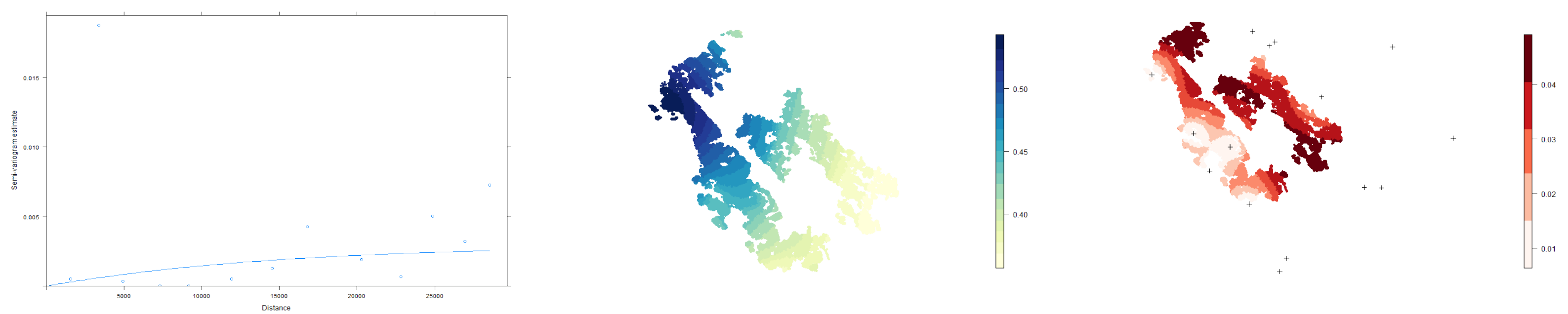
	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS

2008

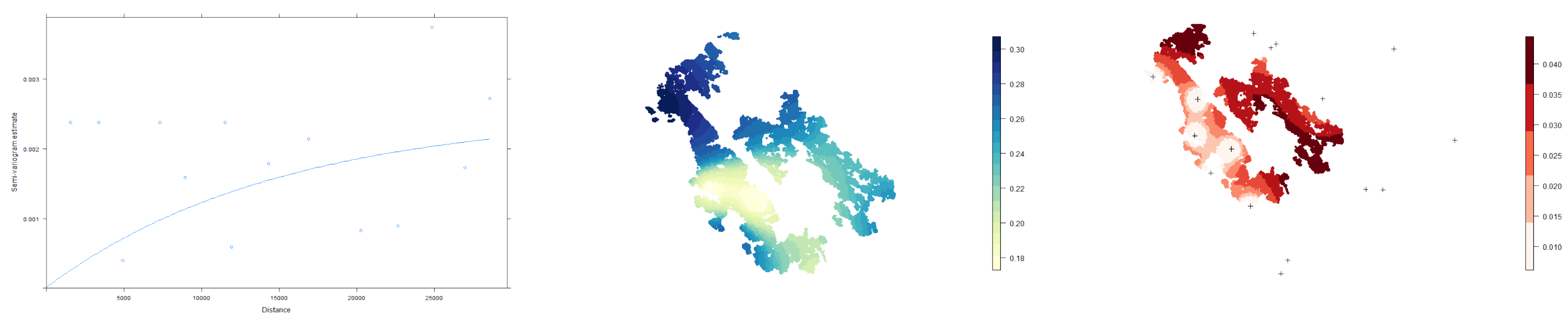
DECEMBER

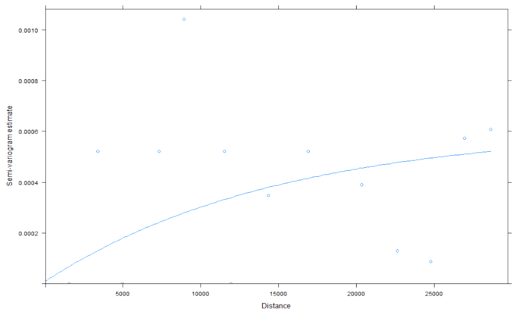
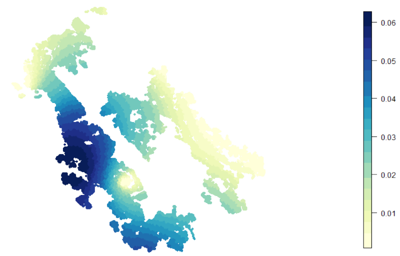
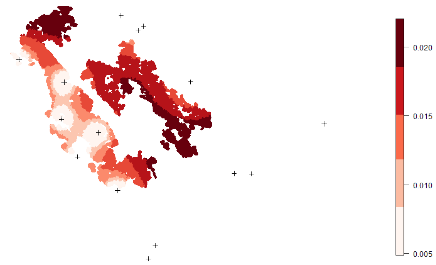
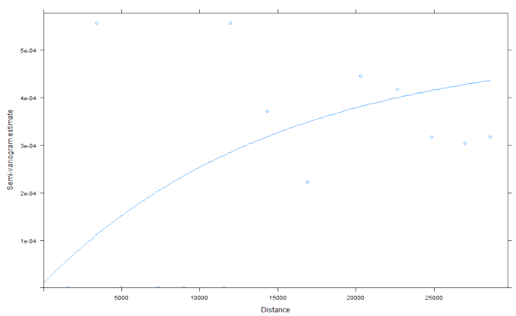
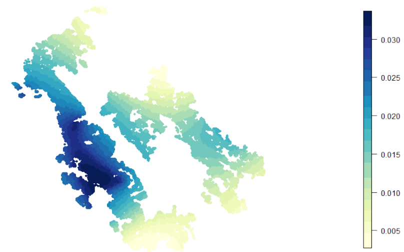
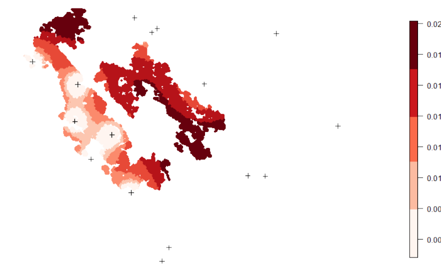


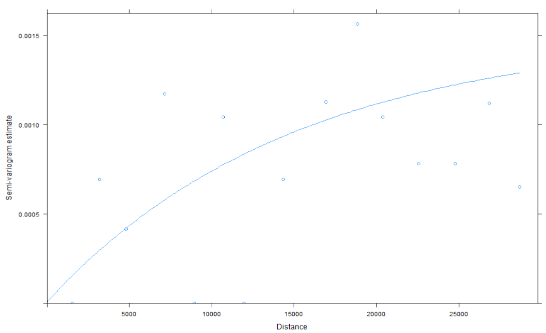
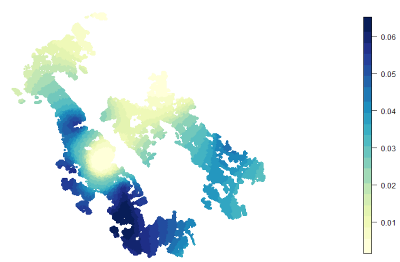
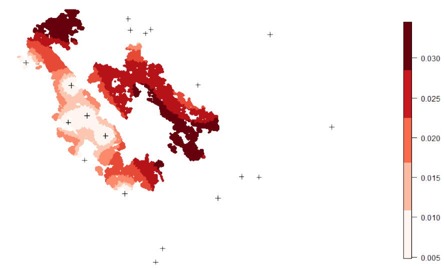
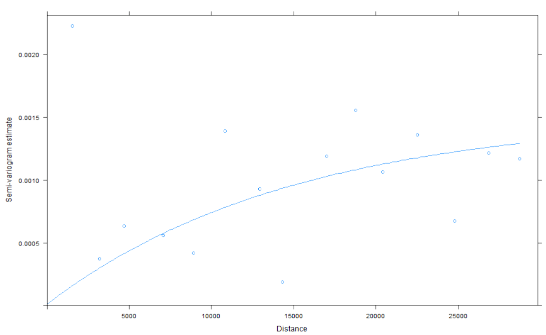
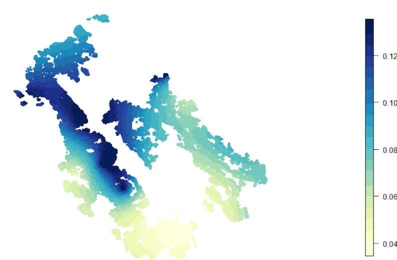
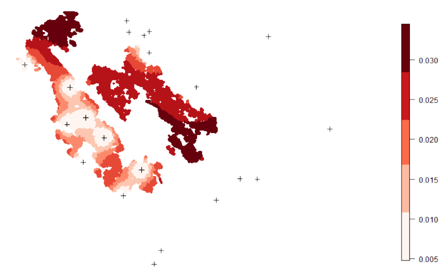
JANUARY

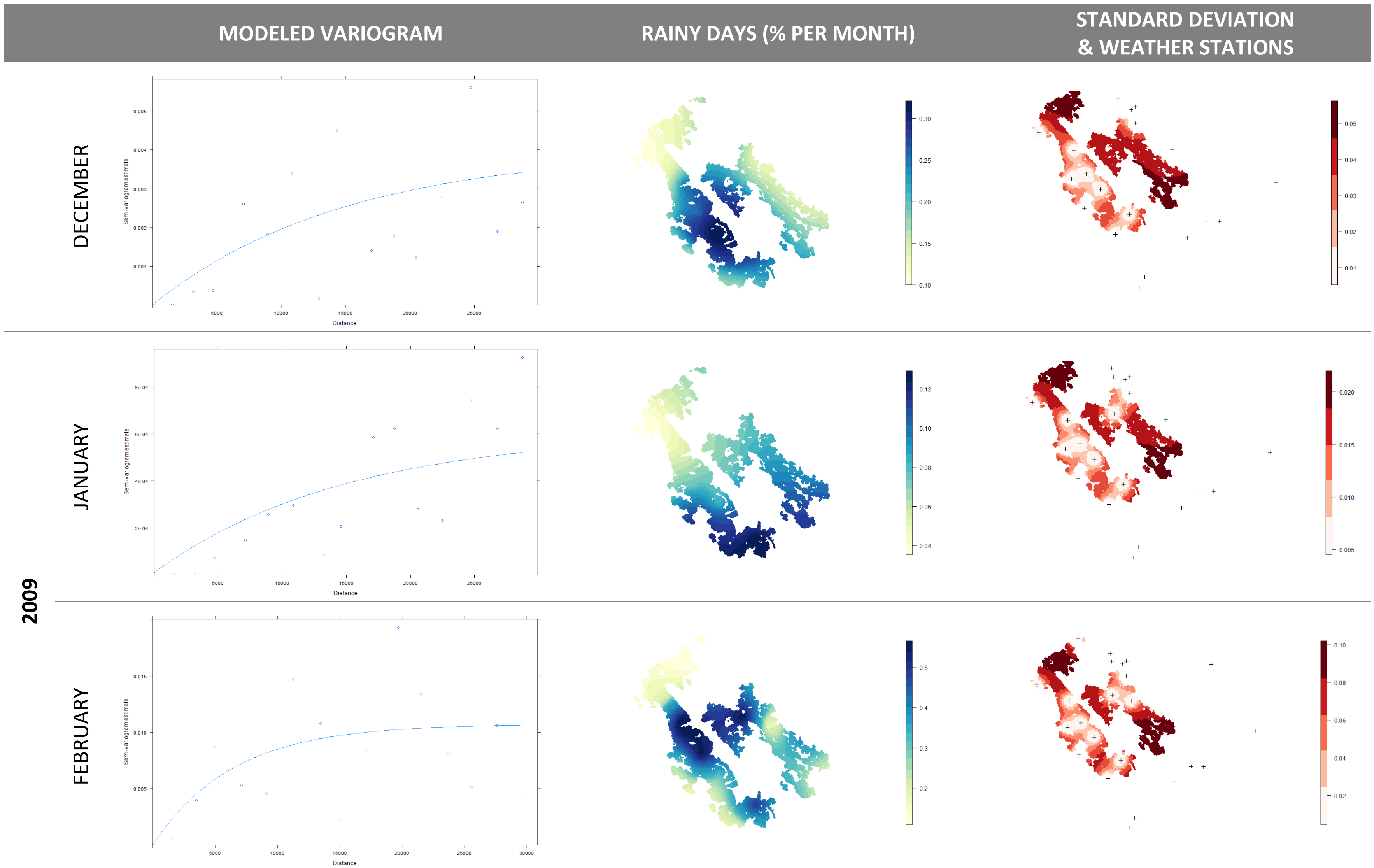


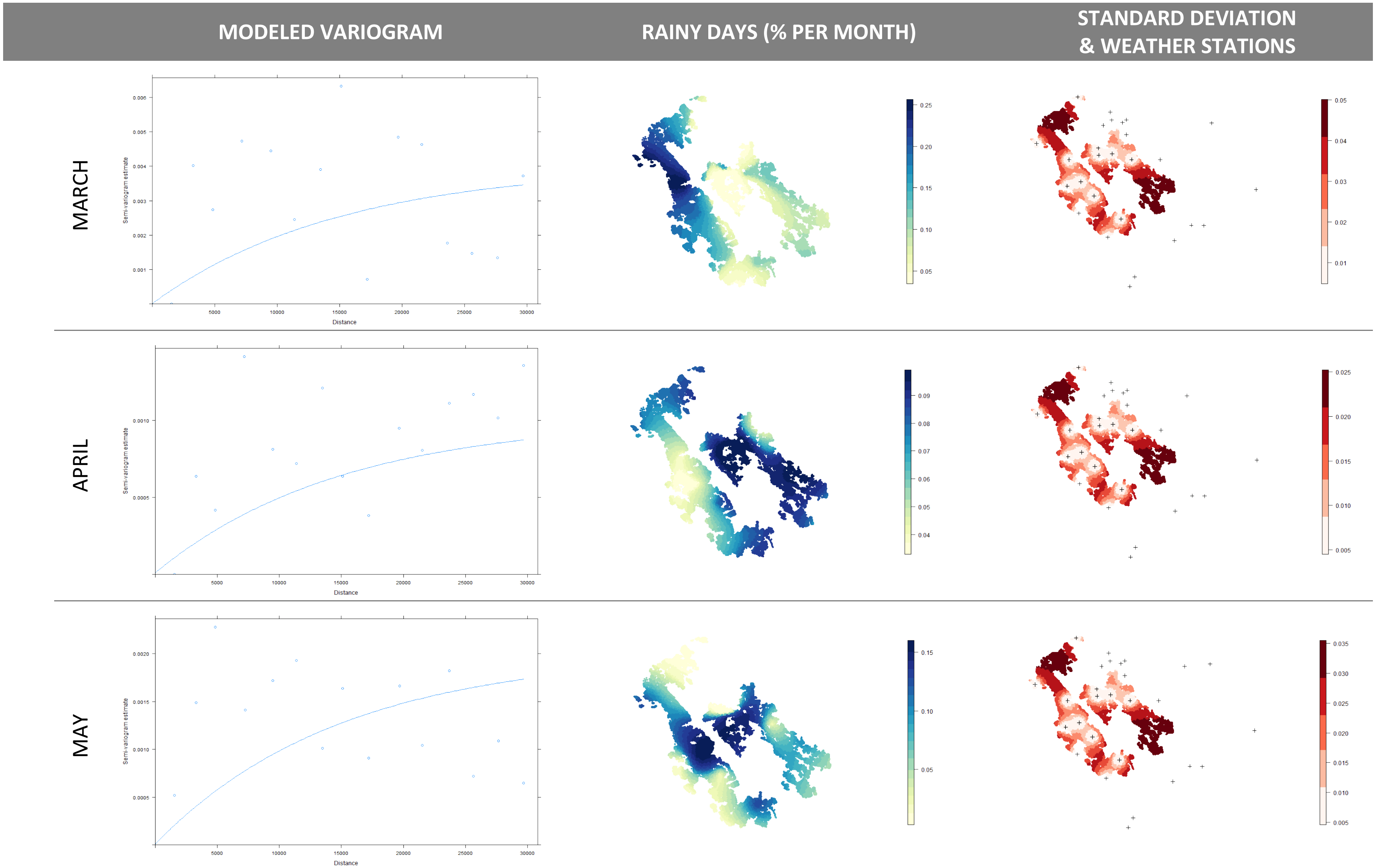
FEBRUARY

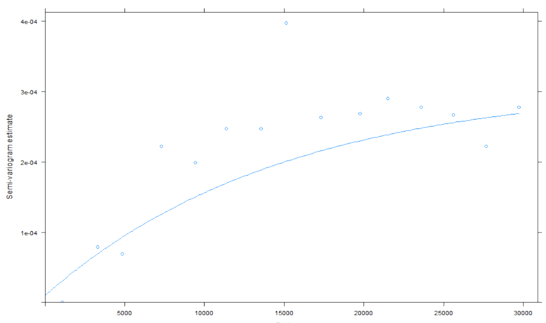
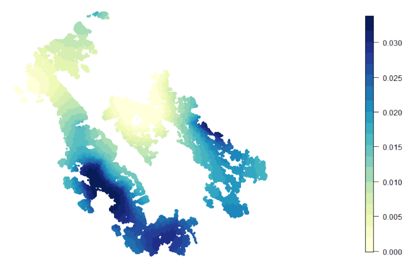
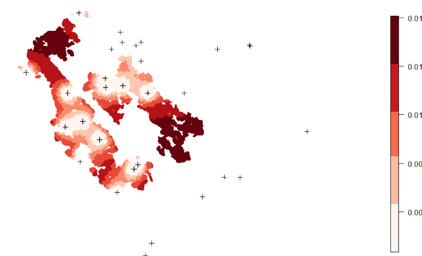
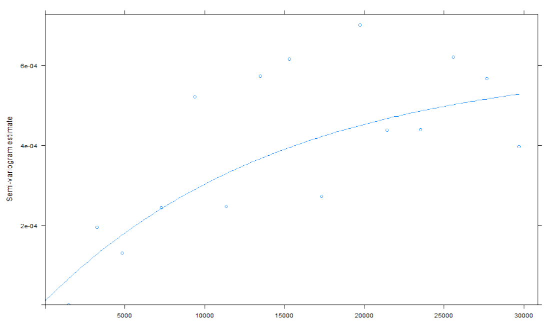
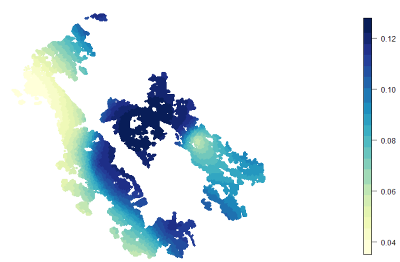
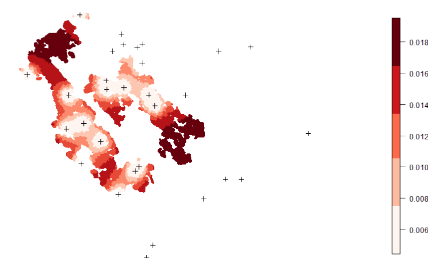


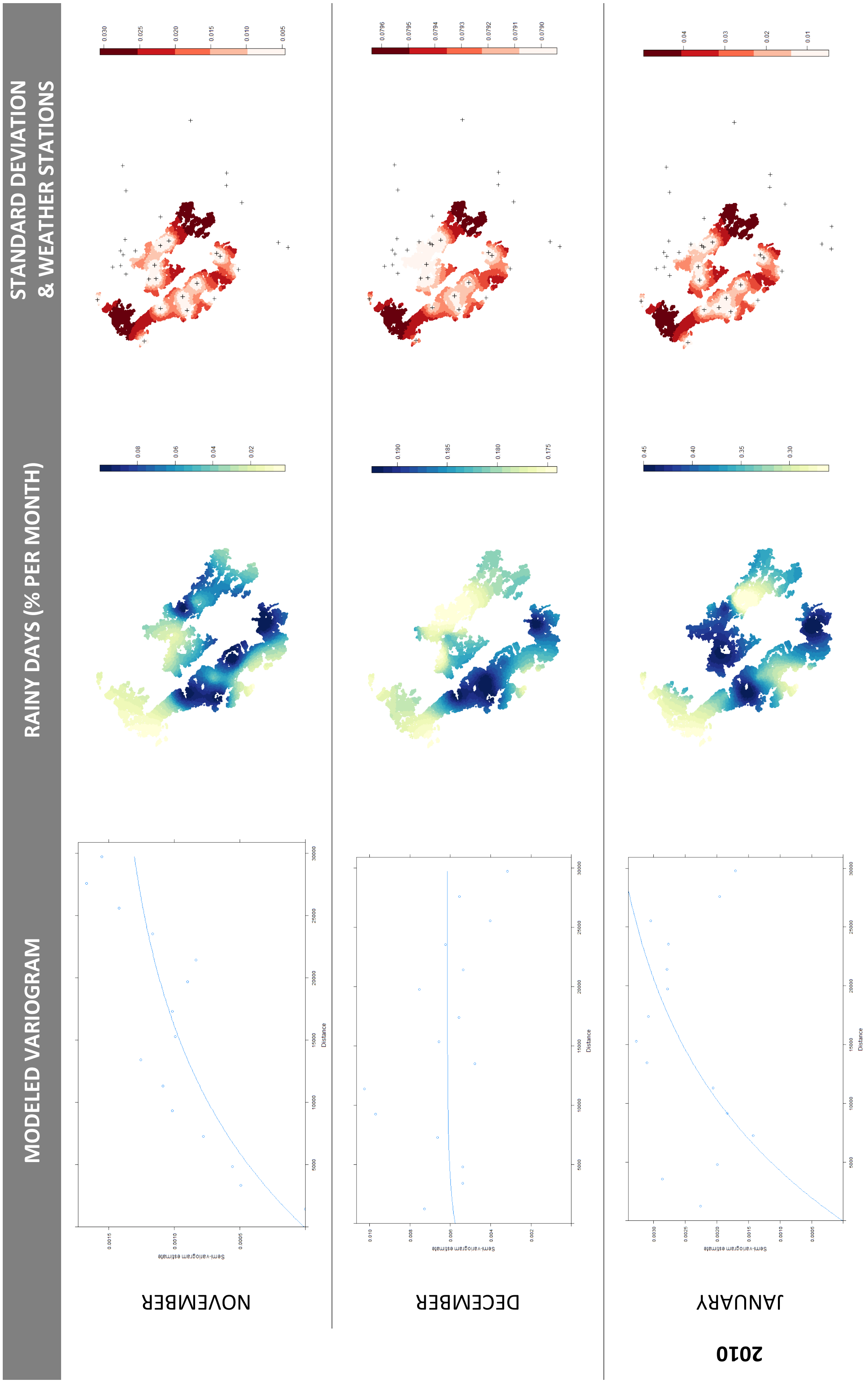
	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
MARCH			
APRIL			
MAY	n/a	n/a	n/a
JUNE	n/a	n/a	n/a
JULY	n/a	n/a	n/a

	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
AUGUST	n/a	n/a	n/a
SEPT	n/a	n/a	n/a
OCTOBER			
NOVEMBER			

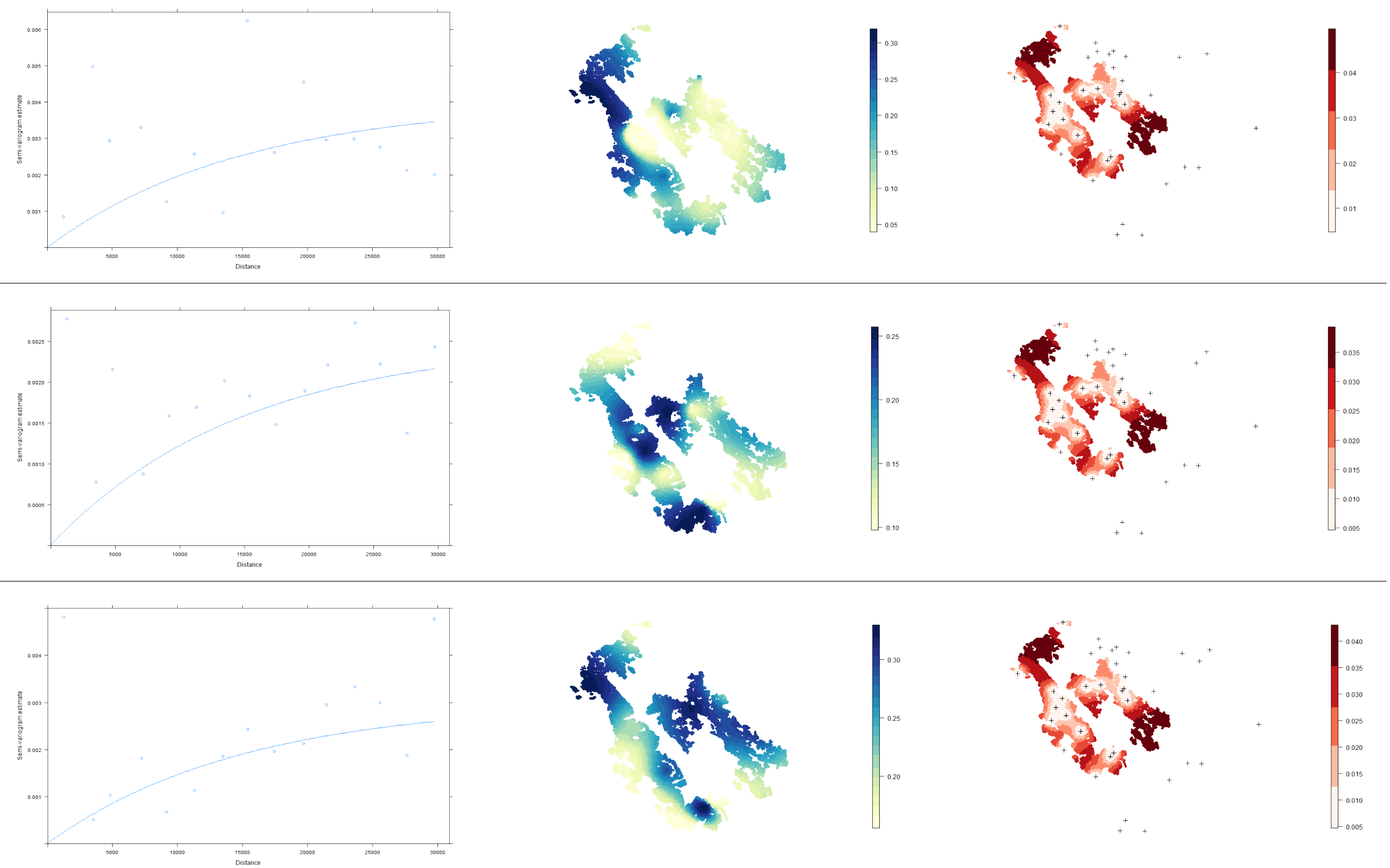

MODELED VARIOGRAM
RAINY DAYS (% PER MONTH)
STANDARD DEVIATION & WEATHER STATIONS
DECEMBER
JANUARY
FEBRUARY
2009
Semi-variogram estimate
Distance

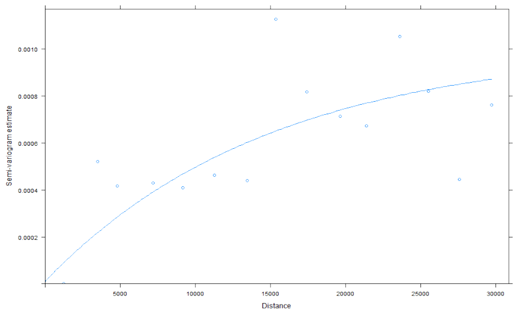
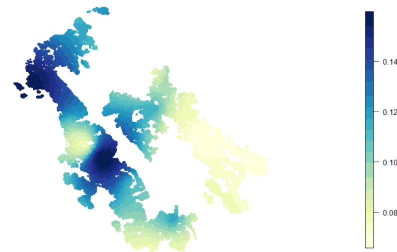
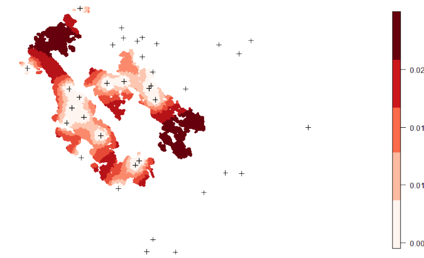

MODELED VARIOGRAM
RAINY DAYS (% PER MONTH)
STANDARD DEVIATION & WEATHER STATIONS
MARCH
APRIL
MAY
Semi-variogram estimate
Distance

	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
JUNE	n/a	n/a	n/a
JULY	n/a	n/a	n/a
AUGUST	n/a	n/a	n/a
SEPTEMBER			
OCTOBER			


MODELED VARIOGRAM
RAINY DAYS (% PER MONTH)
STANDARD DEVIATION & WEATHER STATIONS
NOVEMBER
DECEMBER
JANUARY
2010
Semi-variogram estimate
Distance

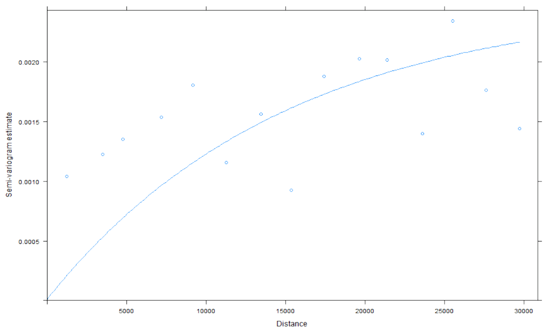
	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
FEBRUARY			
MARCH			
APRIL			

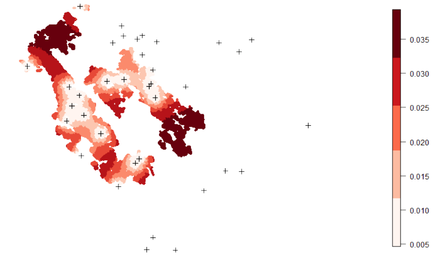


	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
MAY			
JUNE	n/a	n/a	n/a
JULY	n/a	n/a	n/a
AUGUST	n/a	n/a	n/a
SEPT	n/a	n/a	n/a

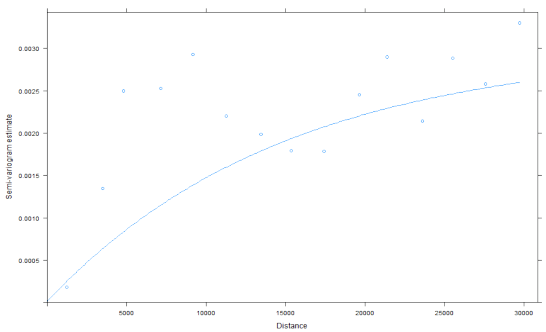
MODELED VARIOGRAM | RAINY DAYS (% PER MONTH) | STANDARD DEVIATION & WEATHER STATIONS

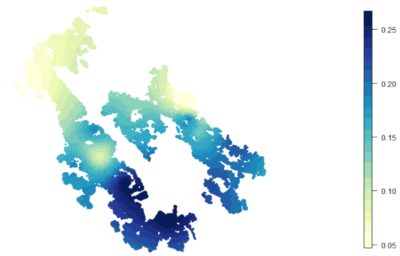
OCTOBER

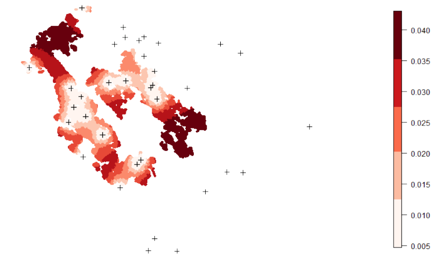




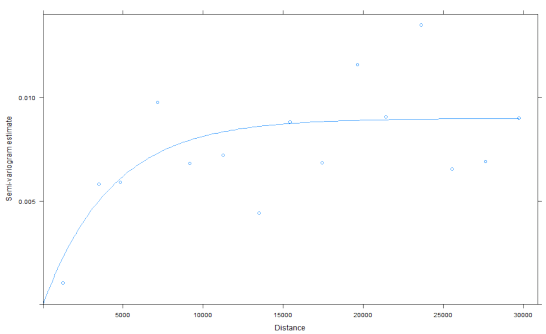
NOVEMBER

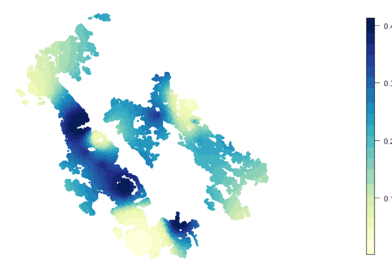


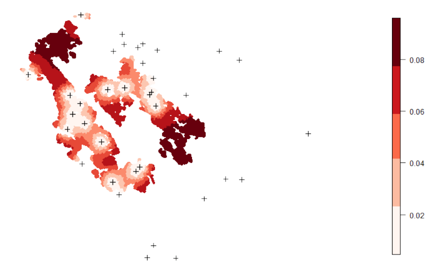




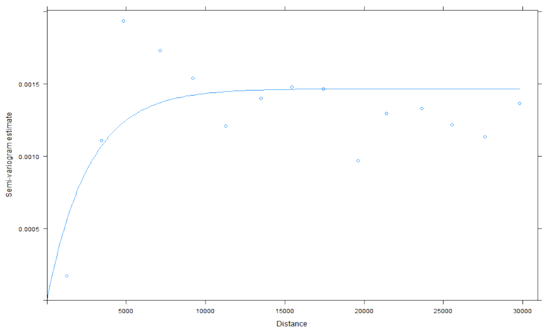
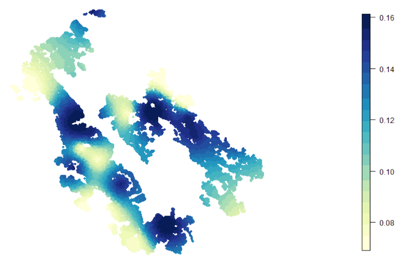
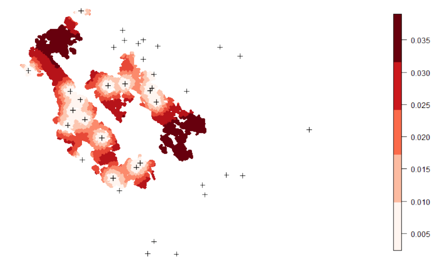
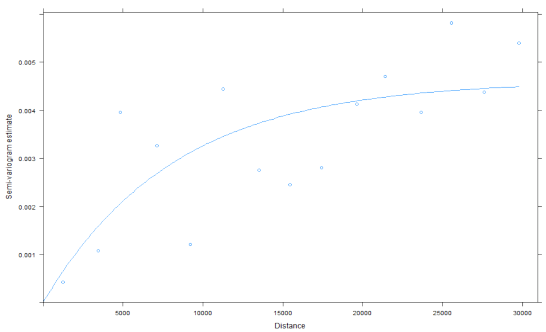
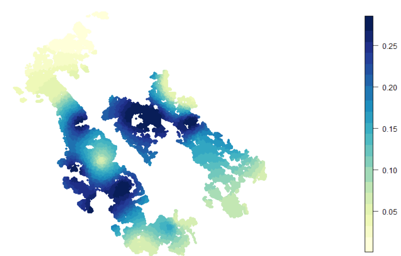
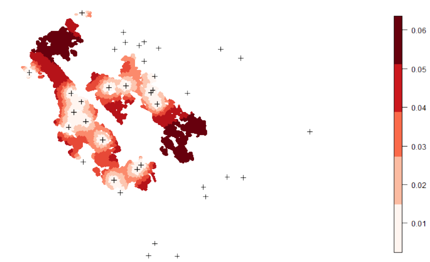
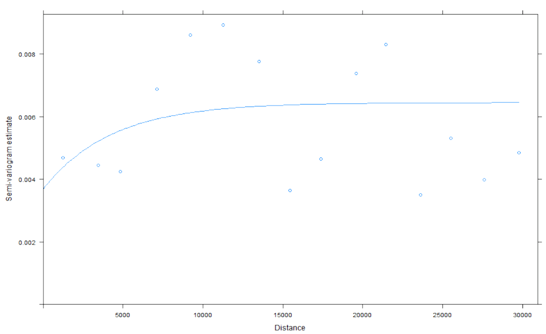
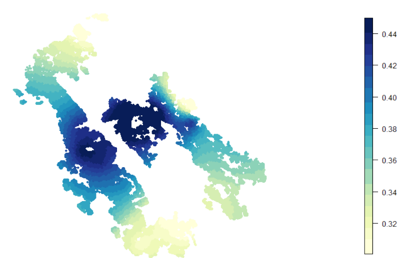
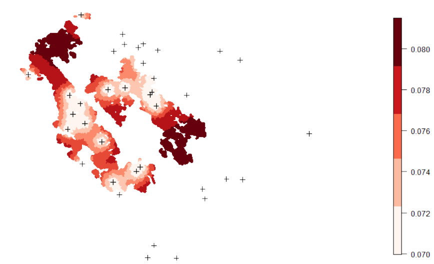
DECEMBER

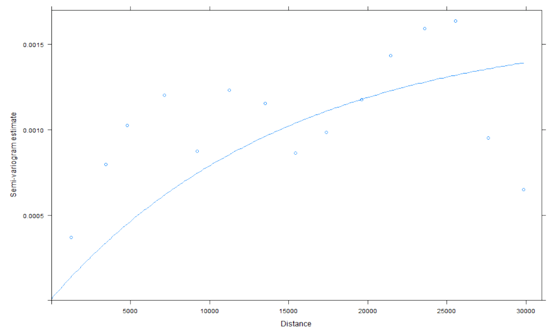
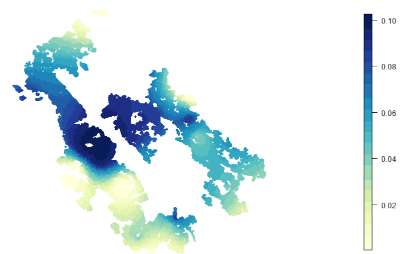
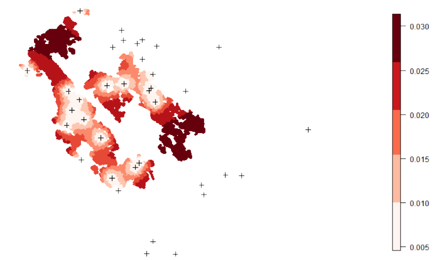
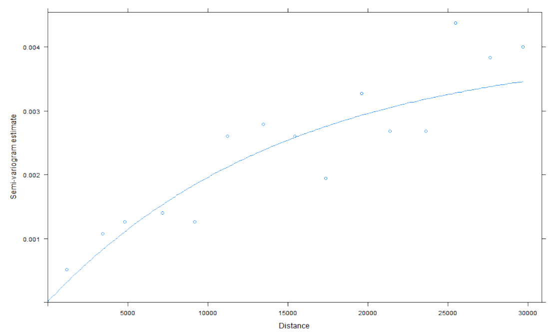
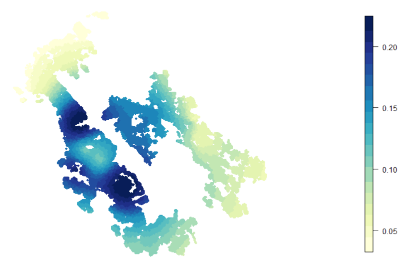
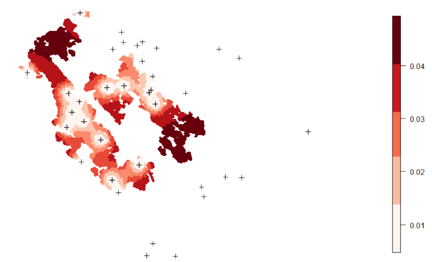
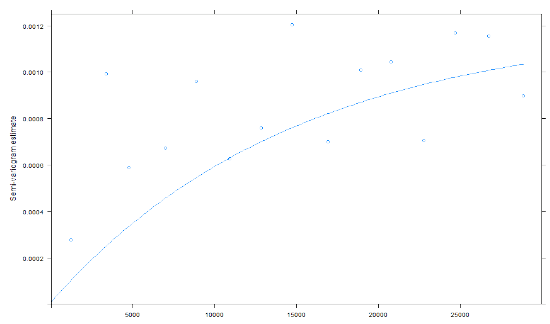
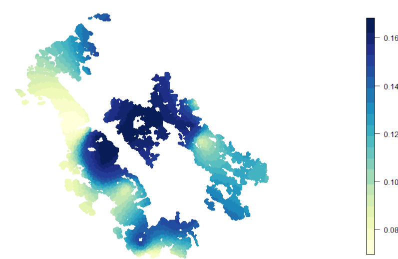
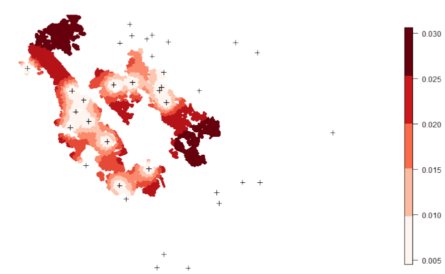


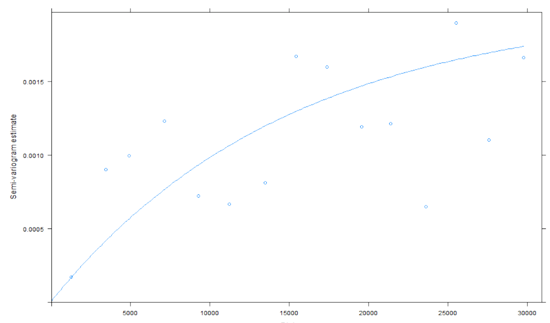
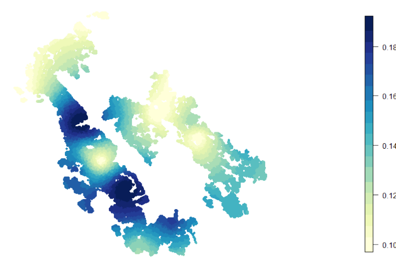
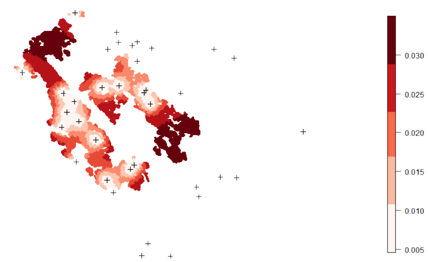
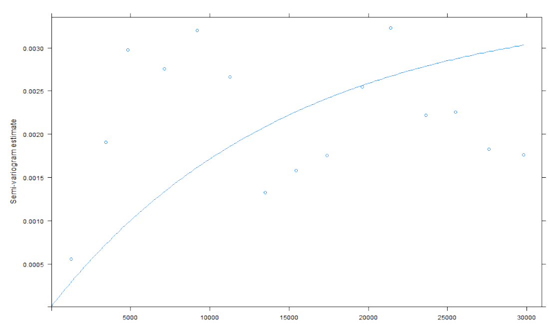
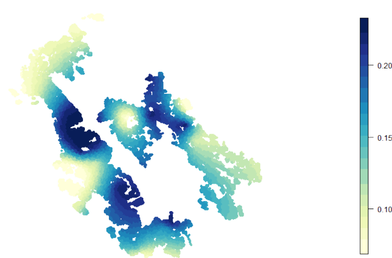
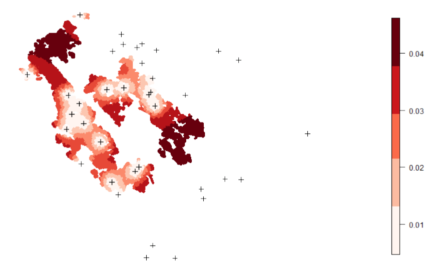


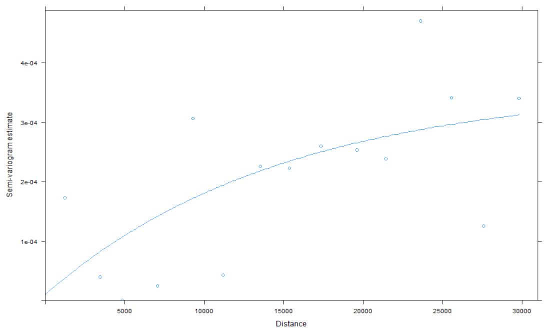
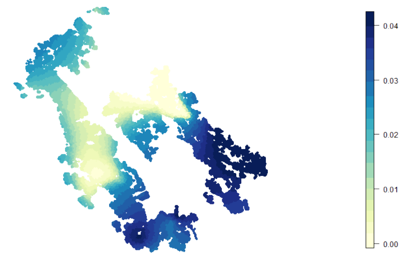
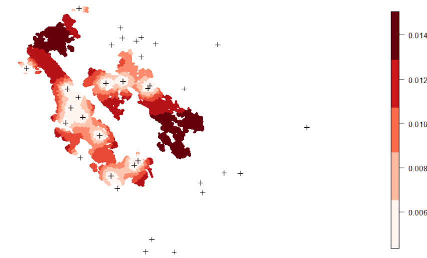


2011

	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
JANUARY			
FEBRUARY			
MARCH			

	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
APRIL			
MAY			
JUNE			

	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
JULY	n/a	n/a	n/a
AUGUST	n/a	n/a	n/a
SEPT	n/a	n/a	n/a
OCTOBER			
NOVEMBER			

	MODELED VARIOGRAM	RAINY DAYS (% PER MONTH)	STANDARD DEVIATION & WEATHER STATIONS
DECEMBER			

Appendix 4. Identifying classes of water users

The quantile regression executed in chapter 4 requires all of the households in the study area to be divided into groups based on water use. Most commonly, quantile regressions recalculate the quantile for each time period, which means that households may shift among quantiles from period to period. In this study, such an approach may introduce problems of simultaneity. As a result, this appendix seeks to identify consistent classes of water users.

First, the appendix evaluates potential quantile classification schemes, determining that terciles are a relatively stable categorization (low, moderate, and high water users). Next, it examines characteristics of water users by tercile. Finally, it evaluates how low and high water users differ from moderate users through a multinomial logit model. Ultimately, it suggests that terciles are a valuable lens, but there are many different possible classification schemes that may help tease apart the observed heterogeneity in water users.

Identifying quantiles

Within the seven year study period, do households maintain consistent water use relative to peers? Or does household use jump around? We tested several quantile divisions to find a scheme that resulted in the most stable classification over time.

To evaluate stability, we calculated the quantile of water use for each household for every month in the study period. Then, we identified the median quantile value over time for each household. Next, we calculated the average number of quantile jumps per month. This is represented by the following equation:

$$average\ monthly\ change_i = \frac{\sum_1^t |\widetilde{quantile}_i - \ quantile_{it}|}{t}$$

where i the number of households and t is the number of months in the study period. We repeated the analysis for several different quantiles: 20, 10, 5, 4, 3 (divisions of 5, 10, 20, 25, and 33.3 percentiles, respectively).

Figure 1 plots the change score by household, showing the distribution of quantile jumps per month among households. For both 20 and 10 quantiles, the average change was greater than 1. This indicates these quantile separates are too fine—households would, on average, change group membership every month. Classifying by 5 quantiles may still be too fine, with a mean stability score of 0.6 this indicates that in over half of the months of the study period, the household changed user class. Four or 3 quantiles are the most steady divisions. Three quantiles shows a heavily left-skewed distribution.

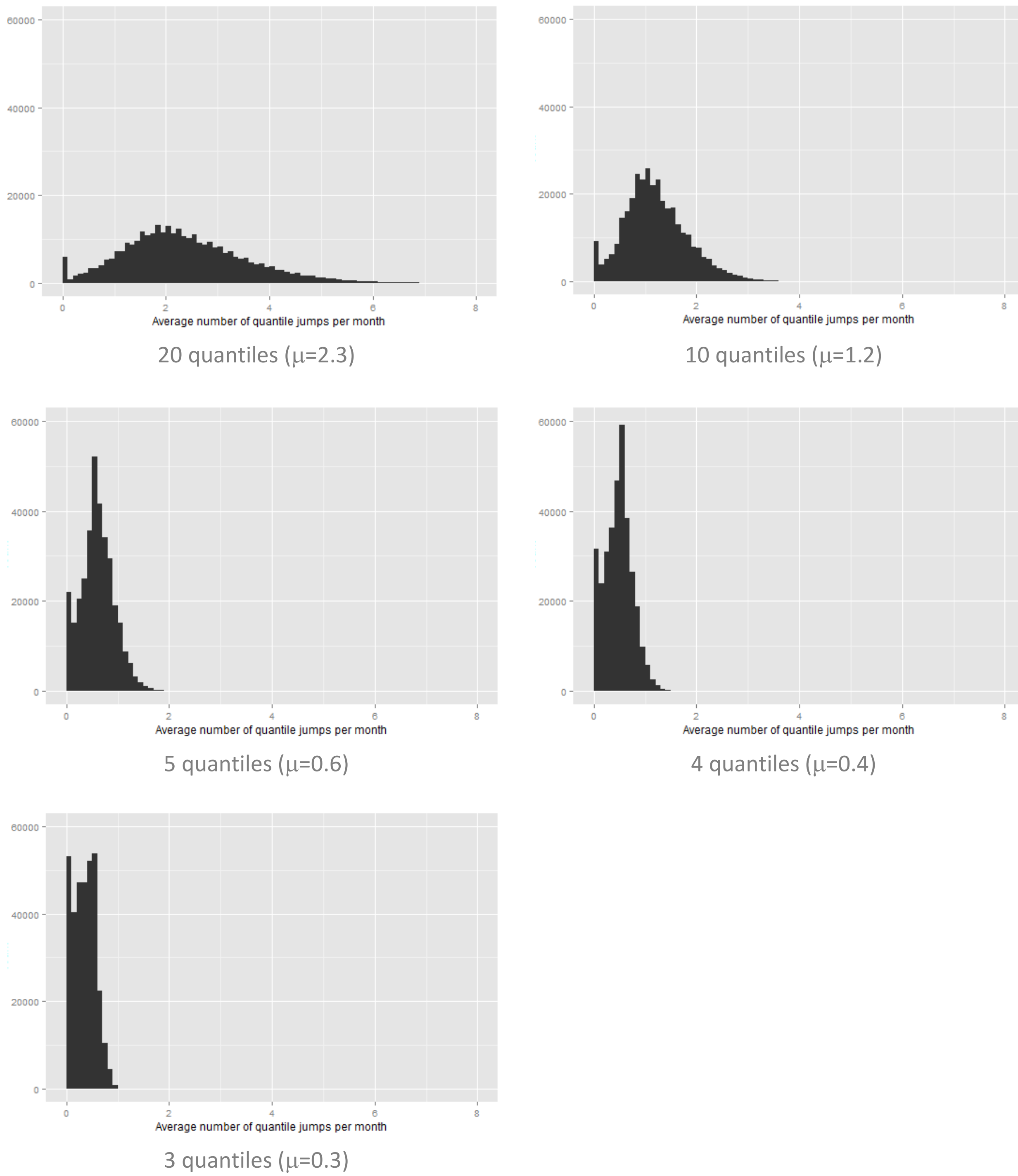


Figure 49. Distribution across all households of the average number of quantile jumps per month during the study period.

Tercile characteristics

We chose to evaluate user groups by 3 quantile divisions—terciles—because of their stability. This classification represents low, moderate, and high water users. The following reviews the characteristics of households by tercile.

Distributions of monthly water use

Within each tercile, there is variation in use. All terciles exhibit the same range of water use, from 25 to 2917 gpd. This minimum and maximum is defined by the data cleaning process (described in chapter 4). The median use varies by group, however. Low water using households consume an average of 110 gpd. Moderate households consume a median of 218 gpd. High households consume a median of 502 gpd. Distributions are shown in Figure 2. Median consumption is 100, 200, and 400 gpd, respectively.

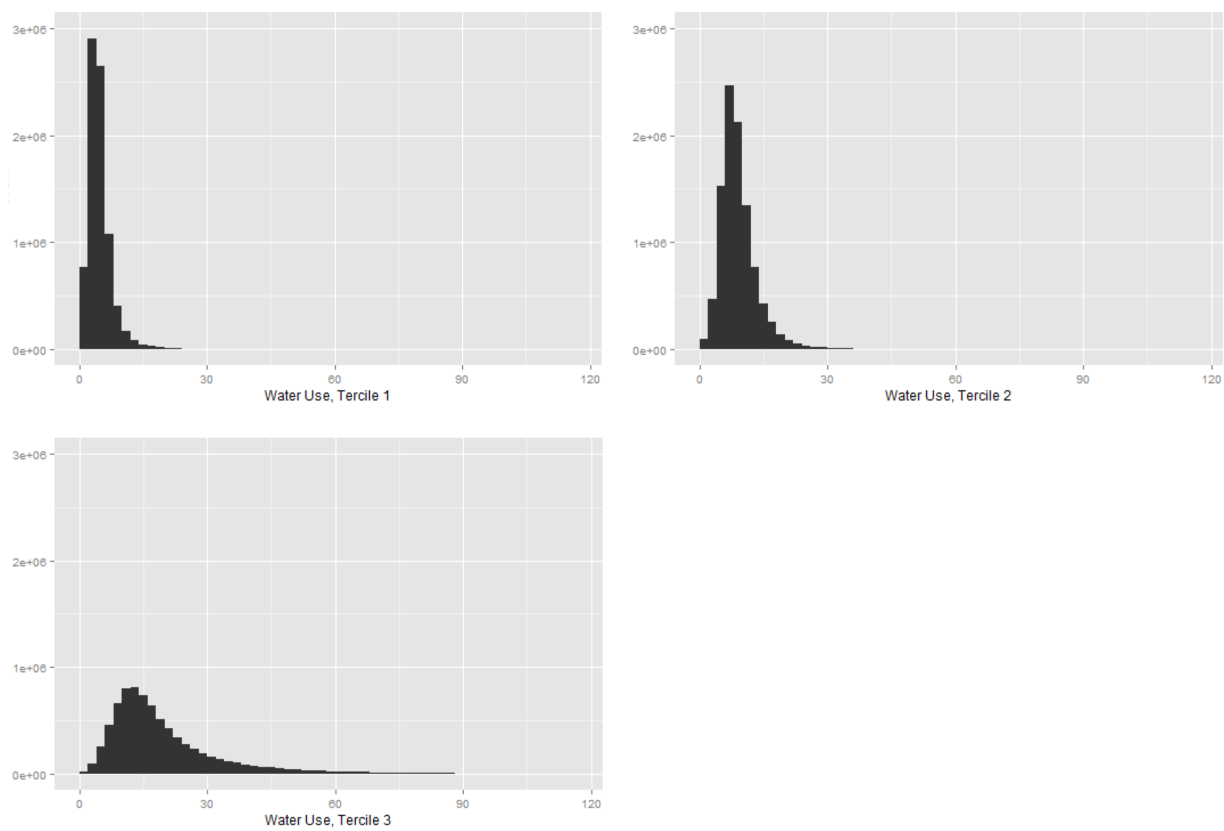


Figure 50. Water use by tercile (units in ccf). Tercile 1, low water users, consume a median of 100 gpd. Tercile 2, moderate water users, consume a median of 200 gpd. Tercile 3, high water users, consume a median of 400 pgd. The long right tail of the full water use distribution is evident in the high water users, which have a much wider range than the other terciles.

Average water use over time

Each tercile exhibits the expected sinusoidal variability in consumption over time (see Figure 3), though the amplitude of the change in seasonal consumption is different for each group. Low water users consume roughly 60 additional gpd in summer months. Moderate waters consume an additional 140 gpd. High water users consume an additional 400 gpd.

Because seasonal use is associated with outdoor water, this indicates that all water users are consuming water outdoors. High water users are consuming substantially greater volumes outside than low water or moderate water users, however.

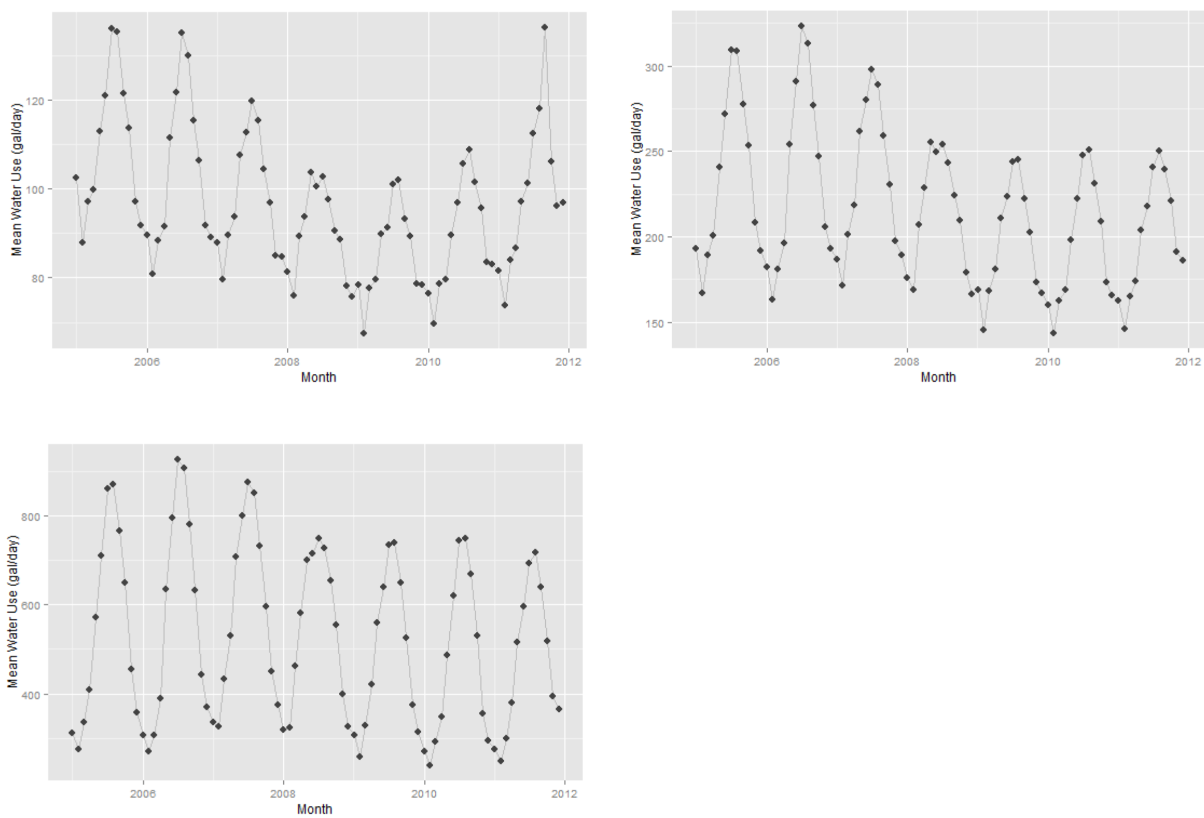


Figure 51. Water use by tercile over time. Low and moderate are in the top row. High water users are in the bottom row. The change in use from peak to trough varies across low, moderate, and high users: approximately 60, 140, and 400 gpd, respectively. Note that the lowest, winter water use for high water users is larger than the highest summer use for moderate water users. Similarly, the lowest monthly use for moderate water users is larger than the highest monthly use for low water users.

Household locations

Low, moderate, and high water users are located throughout the study area, though there are a greater percentage of low water users on the western side of the study area and a greater percentage of high water users on the eastern side of the study area. The spatial distribution of the groups is shown in Figure 4.

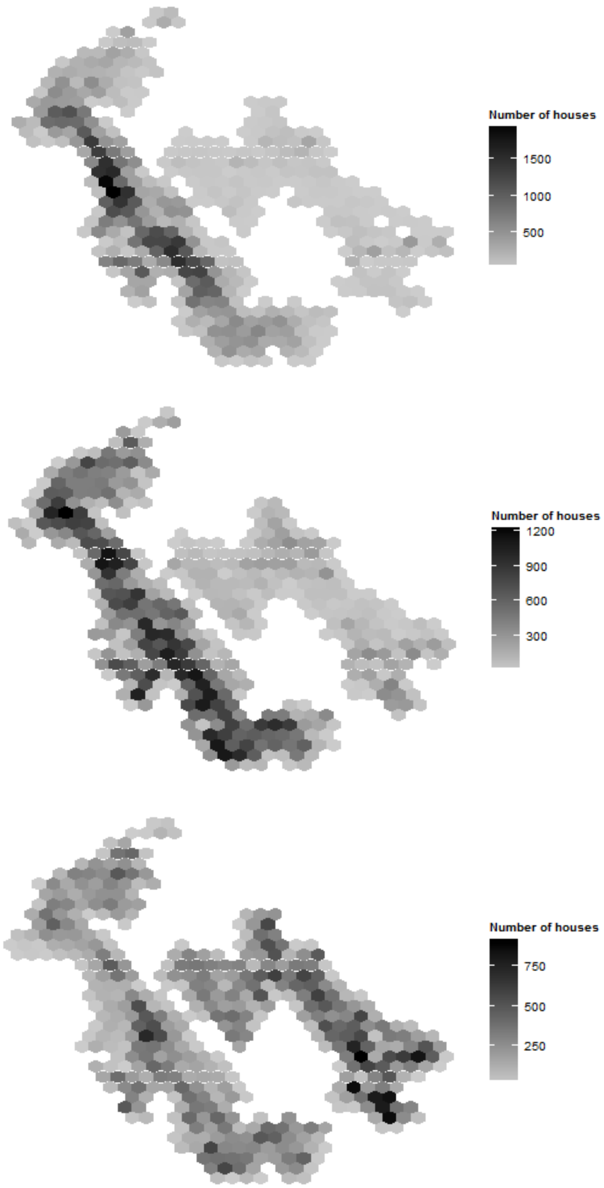


Figure 52. Tercile classification by location. Households are summarized on a 1 km hexagonal grid to preserve household anonymity. Each hexagon counts represents the total number of households by tercile. Low, moderate, and high water users are shown from top to bottom.

Descriptive statistics

In addition to different water consumption profiles, terciles exhibit different parcel attributes. The descriptive statistics for low water using households is available in Table 1, moderate in Table 2, and high in Table 3.

Table 15. Low water using households.

Variable	Units	N	Mean	Std dev	Min	Max
lotsize	sq ft	110,866	24,191.56	425,464.70	49	15,160,622
bldgsize	sq ft	104,900	2,170.38	5,462.63	1	147,048
year		99,097	1,943.13	26.18	1,854	2,011
elev	Meters	113,109	58.69	71.53	-0.57	454.60
slope	Degrees	113,109	4.80	5.53	0.18	40.38
ndvi		113,109	0.02	0.05	-0.19	0.25
temp	Celsius	113,109	14.47	0.17	13.68	15.43
rainyday	%/month	113,109	0.12	0.01	0.09	0.14
hhsize		113,109	12.09	0.76	9.04	13.54
hhinc	$/year	113,109	73,793.28	38,150.70	0	250,001

Table 16. Moderate water using households.

Variable	Units	N	Mean	Std dev	Min	Max
lotsize	sq ft	118,445	11,031.72	127,456.80	77	40,209,670
bldgsize	sq ft	115,248	1,865.21	3,681.41	1	147,048
year		113,393	1,949.52	25.88	1,850	2,011
elev	Meters	120,213	68.46	75.68	0.05	454.83
slope	Degrees	120,213	5.05	5.52	0.14	38.80
ndvi		120,213	0.02	0.05	-0.15	0.25
temp	Celsius	120,213	14.50	0.18	13.72	15.45
rainyday	%/month	120,213	11.91	0.77	9.04	13.54
hhsize		120,213	2.76	0.58	0.00	5.55
hhinc	$/year	120,213	84,420.16	41,121.61	0	250,001

Table 17. High water using households.

Variable	Units	N	Mean	Std dev	Min	Max
lotsize	sq ft	92,890	17,217.28	213,445.00	200	41,973,458
bldgsize	sq ft	91,487	2,333.88	1,541.19	275	147,048
year		90,939	1,962.85	25.26	1,850	2,011
elev	Meters	94,412	110.07	79.31	0.66	454.81
slope	Degrees	94,412	5.79	5.41	0.19	37.99
ndvi		94,412	0.06	0.06	-0.17	0.25
temp	Celsius	94,412	14.55	0.21	13.67	15.45
rainyday	%/month	94,412	11.77	0.67	9.04	13.54
hhsize		94,412	2.80	0.48	1.26	5.55
hhinc	$/year	94,412	114,089.40	47,248.77	6,125	250,001

Evaluating tercile differences

We tested the relationships among terciles through a multinomial logit analysis. Moderate water users define the reference level. The characteristics of low and high water users are evaluated against moderate users. The results of the analysis are available in Table 4.

The model indicates that, relative to average users, low water users are in cooler and more rainy areas, they have a smaller household size, they have fewer pools, and they have drier landscape styles. Relative to moderate water users, high water users are in hotter areas, have a slightly larger household size, have more pools, and have more lush landscaping styles.

The multinomial model only explains 13% of the variance among users, however. This indicates that there are many factors that are not adequately captured by the model that determines water user class. The tercile classifications are tested further in chapter 4.

Table 18. Multinomial logit analysis results, demonstrating how low and high water using households relate to moderate water users.

Variable	Low Water Users		High Water Users	
	β	(SE)	β	(SE)
Temperature	-0.67^{***}	(0.04)	1.12^{***}	(0.04)
Rainy Days	0.20^{***}	(0.01)	0.03^{***}	(0.01)
Elevation	-0.001^{***}	(0.0001)	0.002^{***}	(0.0001)
Slope	0.02^{***}	(0.001)	-0.05^{***}	(0.001)
Income	-0.0000^{***}	(0.0000)	0.0000^{***}	(0.0000)
Household size	-0.26^{***}	(0.01)	0.19^{***}	(0.01)
Building size	0.0000^{***}	(0.0000)	0.0000^{***}	(0.0000)
Year built	-0.002^{***}	(0.0002)	0.01^{***}	(0.0003)
Pool	-0.53^{***}	(0.04)	1.35^{***}	(0.02)
NDVI	-1.21^{***}	(0.12)	8.45^{***}	(0.12)
Lot size	-0.0000^{**}	(0.0000)	0.0000^{***}	(0.0000)
Intercept	11.65^{***}	(0.80)	-36.50^{***}	(0.80)
Observations	302,598			
R^2	0.13			
Log Likelihood	-289,355.10			
LR Test	$83,647.95^{***}$ (df = 24)			

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Future research: Alternate specifications of user groups

This appendix is part of a larger inquiry into classifying user groups. As discussed in chapter 3, there are many different possible mechanisms of subsetting a population, like households, that can be described along a continuous, multivariate spectrum. Future studies may refine these classes. One possible extension will be to evaluate high water users by season: what are the characteristics of the households that have the highest summer use versus winter use?

An alternative formulation of this study could focus on change, rather than durable behaviors. Which households change quantiles over time? What contributes to change? Is there a behavioral shift within the same household? If so, why?

References

Alley, William M. 1984. "The Palmer Drought Severity Index: Limitations and Assumptions." *Journal of Climate and Applied Meteorology* 23 (7): 1100-1109.

Andrew, John T. 2015. "Adapting California's Water Sector to a Changing Climate." Chap. 1, In *Sustainable Water: Challenges and Solutions from California*, edited by Allison Lassiter. Oakland, California: University of California Press.

Arbués, F., I. Villanua, and R. Barberan. 2010. "Household Size and Residential Water Demand: An Empirical Approach." *Australian Journal of Agricultural and Resource Economics* 54 (1): 61-80.

Arbués, Fernando, Ramón Barberán, and Inmaculada Villanúa. 2004. "Price Impact on Urban Residential Water Demand: A Dynamic Panel Data Approach." *Water Resources Research* 40 (11).

Arbués, Fernando, María Ángeles García-Valiñas, and Roberto Martínez-Espiñeira. 2003. "Estimation of Residential Water Demand: A State-of-the-Art Review." *The Journal of Socio-Economics the Journal of Socio-Economics* 32 (1): 81-102.

Arbués, Fernando, Inmaculada Villanúa, and Ramón Barberán. 2010. "Household Size and Residential Water Demand: An Empirical Approach." *Australian Journal of Agricultural & Resource Economics* 54 (1): 61-80.

Ault, Toby R., Julia E. Cole, Jonathan T. Overpeck, Gregory T. Pederson, and David M. Meko. 2014. "Assessing the Risk of Persistent Drought using Climate Model Simulations and Paleoclimate Data." *Journal of Climate* 27: 7529-7549.

Balling, Jr, 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).

Bramley, Glen and Sinéad Power. 2009. "Urban Form and Social Sustainability: The Role of Density and Housing Type." *Environment & Planning B: Planning & Design* 36 (1).

Brent, Daniel Aaron. 2013. "Essays on Water Resource Economics." PhD, University of Washington.

Breusch, Trevor, Michael B. Ward, Hoa Thi Minh Nguyen, and Tom Kompas. 2011. "FEVD: Just IV or Just Mistaken?" *Political Analysis* 19 (2): 165-169.

Breyer, Betsy, Heejun Chang, and G. Hossein Parandvash. 2012. "Land-use, Temperature, and Single-Family Residential Water use Patterns in Portland, Oregon and Phoenix, Arizona." *Applied Geography* 35 (1–2): 142-151.

Executive Department, State of California. 2015. *Executive Order B-29-15*, no. http://gov.ca.gov/docs/4.1.15_Executive_Order.pdf (April 1, 2015).

Bruegmann, Robert. 2008. "Learning from Sprawl." In *Worlds Away : New Suburban Landscapes*, edited by Andrew Blauvelt, 257-288. Minneapolis: Walker Art Center.

Bruegmann, Robert. 2005. *Sprawl : A Compact History*. Chicago: University of Chicago Press.

Burchell, Robert W., George Lowenstein, William R. Dolphin, Catherine C. Galley, Anthony Downs, Samuel Seskin, Katherine Gray Still, and Terry Moore. 2002. *Costs of Sprawl—2000*. Washington, DC: National Academy Press.

California Building Standards Commission. 2013. *California Green Building Standards Code*. Sacramento, California.

California Department of Finance. "Report E-1 Population Estimates for Cities, Counties and the State with Annual Percent Change — January 1, 2013 and 2014.", accessed 4/30/2015, http://www.dof.ca.gov/research/demographic/reports/estimates/e-1/view.php.

———. "Report P-1 (Total Population): State and County Population Projections, 2010-2060.", accessed 4/30/2015, http://www.dof.ca.gov/research/demographic/reports/projections/P-1/.

California Department of Water Resources. 2013. *California Water Plan Update 2013: Investing in Innovation and Infrastructure*. Sacramento, California.

———. "Sierra Nevada Snowpack is Virtually Gone; Water Content is Now Only 5 Percent of Historic Average, Lowest since 1950." News for Immediate Release., last modified April 1, accessed 5/1/2015, http://www.water.ca.gov/news/newsreleases/2015/040115snowsurvey.pdf.

———. "Updated Model Water Efficient Landscape Ordinance AB 1881.", accessed 5/1/2015, http://www.water.ca.gov/wateruseefficiency/landscapeordinance/.

Cervero, Robert and Jin Murakami. 2010. "Effects of Built Environments on Vehicle Miles Traveled: Evidence from 370 US Urbanized Areas." *Environment & Planning A* 42 (2).

Chang, Heejun, G. Parandvash, and Vivek Shandas. 2010. "Spatial Variations of Single-Family Residential Water Consumption in Portland, Oregon." *Urban Geography* 31 (7): 953-972.

Christian-Smith, Juliet and Matthew Heberger. 2015. "Urban Water Demand and Pricing in a Changing Climate." Chap. 5, In *Sustainable Water: Challenges and Solutions from California*, edited by Allison Lassiter. Oakland, California: University of California Press.

Congress for the New Urbanism, Natural Resources Defense Council, and U.S. Green Building Council. 2011*. LEED for Neighborhood Development*. Washington, DC. http://www.usgbc.org/ShowFile.aspx?DocumentID=8880.

Cook, B. I., T. R. Ault, and J. E. Smerdon. 2015. "Unprecedented 21st Century Drought Risk in the American Southwest and Central Plains." *Science Advances* 1 (1).

Costello, L. R. and K. S. Jones. 2014*. Water use Classification of Landscape Species (WUCOLS IV)*. Davis, California: University of California Division of Agriculture and Natural Resources.

Dalhuisen, J. M., Florax, R. J. G. M., de Groot, H. L. F., and P. Nijkamp. 2003. "Price and Income Elasticities of Residential Water Demand." *Land Economics* 79: 292-308.

Daly, Christopher, Michael Halbleib, Joseph I. Smith, Wayne P. Gibson, Matthew K. Doggett, George H. Taylor, Jan Curtis, and Phillip P. Pasteris. 2008. "Physiographically Sensitive Mapping of Climatological Temperature and Precipitation Across the Conterminous United States." *International Journal of Climatology* 28 (15): 2031-2064.

Department of Water Resources. 1993*. California Water Plan Update*.

Ding, Chris and Xiaofeng He. 2004. "K-Means Clustering Via Principal Component Analysis."ACM, .

Dirks, K. N., J. E. Hay, C. D. Stow, and D. Harris. 1998. "High-Resolution Studies of Rainfall on Norfolk Island: Part II: Interpolation of Rainfall Data." *Journal of Hydrology* 208 (3–4): 187-193.

Division of Water Resources. 1930*. Report to the Legislature of 1931 on State Water Plan*.

Domene, Elena and David Saurí. 2006. "Urbanisation and Water Consumption: Influencing Factors in the Metropolitan Region of Barcelona." *Urban Studies* 43 (9): 1605-1623.

Domene, E., D. Sauri, and M. Pares. 2005. "Urbanization and Sustainable Resource use: The Case of Garden Watering in the Metropolitan Region of Barcelona." *Urban Geography* 26 (6): 520-535.

Duany, Andres, Elizabeth Plater-Zyberk, and Jeff Speck. 2000. *Suburban Nation : The Rise of Sprawl and the Decline of the American Dream*. New York: North Point Press.

East Bay Municipal Utility District. "Service Area Map." About., accessed 5/7/15, https://www.ebmud.com/about/service-area-map.

———. . 2008. *Water use during Water Shortage Emergency Condition*.

Echenique, M. H., A. J. Hargreaves, G. Mitchell, and A. Namdeo. 2012. "Growing Cities Sustainably." *Journal of the American Planning Association* 78 (2): 121-137.

Ewing, Reid and Fang Rong. 2008. "The Impact of Urban Form on U.S. Residential Energy Use." *Housing Policy Debate* 19 (1): 1-30.

Fox, C., B. S. McIntosh, and P. Jeffrey. 2009. "Classifying Households for Water Demand Forecasting using Physical Property Characteristics." *Land use Policy* 26 (3): 558-568.

Franczyk, Jon and Heejun Chang. 2009. "Spatial Analysis of Water use in Oregon, USA, 1985-2005." *Water Resources Management* 23 (4): 755-774.

Gober, Patricia, Kelli L. Larson, Ray Quay, Colin Polsky, Heejun Chang, and Vivek Shandas. 2013. "Why Land Planners and Water Managers Don'T Talk to One another and Why they should!" *Society & Natural Resources* 26 (3): 356-364.

Golden Gate Weather Services. "Climate of San Francisco: Monthly Rainfall (1849-Present).", accessed 5/1/2015, http://ggweather.com/sf/monthly.html.

Greene, William. 2011. "Fixed Effects Vector Decomposition: A Magical Solution to the Problem of Time-Invariant Variables in Fixed Effects Models?" *Political Analysis* 19 (2): 135-146.

Guhathakurta, Subhrajit and Patricia Gober. 2007. "The Impact of the Phoenix Urban Heat Island on Residential Water Use." *Journal of the American Planning Association* 73 (3): 317-329.

Häb, Kathrin, Ariane Middel, and Hans Hagen. 2013. "Using K-Means Clustering for a Spatial Analysis of Multivariate and Time-Varying Microcliatme Data." In , edited by O. Kolditz, K. Rink and G. Scheuermann. Vol. Workshop on Visualization in Environmental Sciences.

Hall, William Hammond. 1889. "Irrigation in California." *National Geographic Magazine* 1 (4).

Halper, Eve Brook. 2011. "Residential Outdoor Water use in Tucson, Arizona: Geospatial, Demographic, and Temporal Perspectives." PhD, The University of Arizona.

Hargrove, William W. and Forrest M. Hoffman. 2004. "Potential of Multivariate Quantitative Methods for Delineation and Visualization of Ecoregions." *Environmental Management Environmental Management : An International Journal for Decision Makers, Scientists and Environmental Auditors* 34 (1): S39-S60.

Harlan, Sharon L., Scott T. Yabiku, Larissa Larsen, and Anthony J. Brazel. 2009. "Household Water Consumption in an Arid City: Affluence, Affordance, and Attitudes." *Society & Natural Resources* 22 (8): 691-709.

Hayden, Dolores and Jim Wark. 2004. *A Field Guide to Sprawl*. New York: W.W. Norton.

Hewitt, Julie A. and W. Michael Hanemann. 1995. "A Discrete/Continuous Choice Approach to Residential Water Demand Under Block Rate Pricing." *Land Economics* 71 (2): 173-192.

House-Peters, Lily A. and Heejun Chang. 2011a. "Urban Water Demand Modeling: Review of Concepts, Methods, and Organizing Principles." *Water Resources Research* 47 (5).

House-Peters, Lily A. and Heejun Chang. 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.

House-Peters, Lily, Bethany Pratt, and Heejun Chang. 2010. "Effects of Urban Spatial Structure, Sociodemographics, and Climate on Residential Water Consumption in Hillsboro, Oregon1." *JAWRA Journal of the American Water Resources Association* 46 (3): 461-472.

Howe, Charles 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.

Ingram, B. L. and Malamud-Roam, Frances. "The West without Water: What Past Floods, Droughts, and Other Climatic Clues Tell Us about Tomorrow." University of California Press.

Ito, Koichiro. 2014. "Do Consumers Respond to Marginal Or Average Price? Evidence from Nonlinear Electricity Pricing." *American Economic Review* 104 (2): 537-563.

Jacobi, John, Debra Perrone, Leslie Lyons Duncan, and George Hornberger. 2013. "A Tool for Calculating the Palmer Drought Indices." *Water Resources Research* 49 (9): 6086-6089.

Jensen, N. E. and L. Pedersen. 2005. "Spatial Variability of Rainfall: Variations within a Single Radar Pixel." *Atmospheric Research* 77 (1–4): 269-277.

Jorgensen, B. S., J. F. Martin, M. W. Pearce, and E. M. Willis. 2013. "Predicting Household Water Consumption with Individual-Level Variables." *Environment and Behavior Environment and Behavior*.

Jorgensen, Bradley, Michelle Graymore, and Kevin O'Toole. 2009. "Household Water use Behavior: An Integrated Model." *Journal of Environmental Management* 91 (1): 227-236.

Kenney, Douglas S., Christopher Goemans, Roberta Klein, Jessica Lowrey, and Kevin Reidy. 2008. "Residential Water Demand Management: Lessons from Aurora, Colorado." *Journal of the American Water Resources Association* 44 (1): 192-207.

Klaiber, H. A., V. K. Smith, M. Kaminsky, and A. Strong. 2014. "Measuring Price Elasticities for Residential Water Demand with Limited Information." *Land Economics* 90 (1): 100-113.

Lassiter, Allison, ed. 2015. *Sustainable Water: Challenges and Solutions from California*. Oakland, California: University of California Press.

Lee S.-J. and Wentz E.A. 2008. "Applying Bayesian Maximum Entropy to Extrapolating Local-Scale Water Consumption in Maricopa County, Arizona." *Water Resources Research* 44 (1).

Lenoir. 1951. *California's 500-Mile-Long Central Valley Project : Solution to a Perennial Need to Supply Life-Giving Water to a Thirsty Agricultural Empire*. Not drawn to scale. ed. [Sacramento] : Sacramento Lithograph Co.,.

McHarg, Ian L. 1994. *Design with Nature*. 25th Anniversary Edition ed. New York: J. Wiley.

Mini, C., T. S. Hogue, and S. Pincetl. 2014. "Patterns and Controlling Factors of Residential Water use in Los Angeles, California." *Water Policy* 16 (6): 1054-1069.

Monteith, J. L. 1965. "Evaporation and Environment." *Symposia of the Society for Experimental Biology* 19: 205-34.

National Climatic Data Center. "NEXRAD Products." National Oceanic and Atmospheric Administration, accessed 1/19/2015, http://www.ncdc.noaa.gov/data-access/radar-data/nexrad-products.

———. "State of the Climate: National Overview for December 2014.", last modified January, accessed 5/2/2015, http://www.ncdc.noaa.gov/sotc/summary-info/national/201412.

Neuman, Michael. 2005. "The Compact City Fallacy." *Journal of Planning Education and Research* 25 (1): 11-26.

Olmstead, Sheila M., Michael W. Hanemann, and Robert N. Stavins. 2007. "Water Demand Under Alternative Price Structures." *Journal of Environmental Economics and Management* 54 (2): 181-198.

Olmstead, Sheila M. and Robert N. Stavins. 2009. "Comparing Price and Nonprice Approaches to Urban Water Conservation." *Water Resources Research* 45 (4).

Plumper, T. and V. E. Troeger. 2007. "Efficient Estimation of Time-Invariant and Rarely Changing Variables in Finite Sample Panel Analyses with Unit Fixed Effects." *Political Analysis Political Analysis* 15 (2): 124-139.

Polebitski, Austin S. and Richard N. Palmer. 2010. "Seasonal Residential Water Demand Forecasting for Census Tracts." *Journal of Water Resources Planning and Management* 136 (1): 27-36.

Randolph, John. 2008. "Comment on Reid Ewing and Fang Rong's “The Impact of Urban Form on U.S. Residential Energy use”." *Housing Policy Debate* 19 (1): 45-52.

Renwick, Mary and Richard 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.

Russell, Sally and Kelly Fielding. 2010. "Water Demand Management Research: A Psychological Perspective." *Water Resources Research* 46 (5).

Schleich, J. and T. Hillenbrand. 2009. "Determinants of Residential Water Demand in Germany." *Ecological Economics* 68 (6): 1756-1769.

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.

Shepherd, J. Marshall. 2005. "A Review of Current Investigations of Urban-Induced Rainfall and Recommendations for the Future." *Earth Interactions* 9 (12): 1-27.

Slavikova, L., V. Maly, O. Vojacek, M. Rost, and L. Petruzela. 2013. "Impacts of Climate Variables on Residential Water Consumption in the Czech Republic." *Water Resources Management* 27 (2): 365-379.

Sohn, Jungyul. 2011. "Watering Cities: Spatial Analysis of Urban Water use in the Southeastern United States." *Journal of Environmental Planning and Management* 54 (10): 1351-1371.

State Water Resources Control Board. "Monthly Urban Water Supplier Report Data." Sacramento, California, accessed 5/1/2015, http://www.waterboards.ca.gov/waterrights/water_issues/programs/drought/conservation_reporting_info.shtml.

———. "State Water Board Expands and Extends Emergency Water Conservation." Sacramento, California, last modified March 17, accessed 5/2/2015, http://www.waterboards.ca.gov/press_room/press_releases/2015/pr031715_renewed_emergency_wtr_regs.pdf.

Syme, Geoffrey J., Quanxi Shao, Murni Po, and Eddy Campbell. 2004. "Predicting and Understanding Home Garden Water Use." *Landscape and Urban Planning.* 68 (1): 121.

Troy, Patrick and Darren Holloway. 2004. "The use of Residential Water Consumption as an Urban Planning Tool: A Pilot Study in Adelaide." *Journal of Environmental Planning and Management* 47 (1): 97-114.

US Environmental Protection Agency. 2006*. Growing Toward More Efficient Water use: Linking Development, Infrastructure, and Drinking Water Policies*. Washington, DC.

Vicente-Serrano, Sergio M., M. Angel Saz-Sánchex, and José M. Cuadrat. 2003. "Comparative Analysis of Interpolation Methods in the Middle Ebro Valley (Spain): Application to Annual Precipitation and Temperature." *Climate Research* 24 (2).

Waterman, Ryan. 2004. "Addressing California's Uncertain Water Future by Coordinating Long - Term Land use and Water Planning : Is a Water Element in the General Plan the Next Step ?" *Ecology and Law Quarterly* (31).

Weather Underground. "Radar Frequently Asked Questions.", accessed 1/19/2015, http://www.wunderground.com/radar/help.asp?.

Wentz, Elizabeth and Patricia Gober. 2007. "Determinants of Small-Area Water Consumption for the City of Phoenix, Arizona." *Water Resources Management* 21 (11): 1849-1863.

Wichman, Casey J. 2014. "Perceived Price in Residential Water Demand: Evidence from a Natural Experiment." *Journal of Economic Behavior & Organization* 107 (Part A): 308.

Wilkinson, Robert. 2015. "The Water–Energy–Climate Nexus in California." Chap. 2, In *Sustainable Water: Challenges and Solutions from California*, edited by Allison Lassiter. Oakland, California: University of California Press.

Wood, S. J., D. A. Jones, and R. J. Moore. 2000. "Accuracy of Rainfall Measurement for Scales of Hydrological Interest." *Hydrology and Earth System Sciences* 4 (4): 531-544.

Worster, Donald. 1985. *Rivers of Empire: Water, Aridity, and the Growth of the American West*. New York: Pantheon Books.

Zetland, David. 2011. *The End of Abundance: Economic Solutions to Water Scarcity*. Kindle ed. Netherlands: Aguanomics Press.