3-20-2018

Urban Geocomputation: Two Studies on Urban Form and Its Role in Altering Climate

Jackson Lee Voelkel
Portland State University

Urban Geocomputation: Two Studies on Urban Form and Its Role in Altering Climate

by

Jackson Lee Voelkel

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

Master of Urban Studies

Thesis Committee:
Vivek Shandas, Chair
Liming Wang
David S. Banis

Portland State University
2018

Abstract

Our climate and our cities are changing. Though their changes are not completely dependent upon one another, there is still a coupling effect between them. This study assesses the role of urban form as it pertains to elements of climate change. It is comprised of two essays intended for publication. The first of these essays addresses the feedbacks between urban form, energy consumption, and rising global temperatures. The second essay looks at one particular factor of urban form – tree type – as it pertains to air pollution and urban heat island mitigation. Both papers use the analytical approaches necessary to answer the questions they pose, not ubiquitous over-generalizing modeling software or methods found often in the literature. As seen in the analyses, this practice – known as geocomputation – allows for a deeper and more accurate description of complex spatial relationships.

Acknowledgements

In regards to Chapter 1, I would like to thank Energy Trust of Oregon for allowing access to their energy consumption data, as well as their guidance and input in all matters energy. Tim Hitchins and Alec Trusty put forward a massive effort to clean and organize the data provided by Energy Trust of Oregon, which proved invaluable for my analysis. In regards to Chapter 2, I would like to thank Drs. Meenakshi Rao, Linda A. George, Todd N. Rosenstiel, Vivek Shandas, and Alexis Dinno for their efforts to provide me with the data and models used in Rao et al. (2014), which were essential to my NO2 analysis. For the many months spent creating the canopy classification dataset I thank Joe Gordon – the long hours we spent working out how to achieve such a complex task were some of the most rewarding I have had in my studies. Portions of the research were funded by the Institute for Sustainable Solutions at Portland State University, the United States Forest Service Urban and Community Forestry program (2011-DG-11062765-016), and the National Science Foundation's Sustainability Research Network (#1444755). Indispensable support for the computational infrastructure was provided by Portland State University's Office of Information Technologies - Research Computing (OIT-RC), and in particular William Garrick.

In addition, I would like to thank all of those who have inspired me through their own work, creativity, and passion over the last few years. Though some may not be aware of it, the interactions I have had with them were critical to my academic development. These people include, but are not limited to:

Amber Ayers	Micah Babinski	David Banis
Joe Broach	Angie DiSalvo	Geoffrey Duh
William Garrick	Joe Gordon	Jamaal Green
Samantha Hamlin	Tim Hitchins	Changyu Hong
Jennifer Karps	Dillon Mahmoudi	Yasuyo Makido
Kevin Martin	Derek Miller	Randy Morris
Al Mowbry	Philip Orlando	Meenakshi Rao
Vivek Shandas	Alec Trusty	Liming Wang
Michael Weisdorf		

Analyses performed during the following research were entirely dependent on the developers who work tirelessly to continually enhance free and open-source software. I am forever grateful to the core and package developers of the R statistical language and environment, through which a majority of my work was performed.

I would like to acknowledge my friends and family for their unrelenting support, and especially Diliana for helping me get through this in one piece.

Table of Contents

List of Tables

List of Figures

Introduction

Cities are dynamic processes – they are a reflection of a constant adaptation by humans to fit our past, present, and future needs. Changes occurring within our cities can be viewed as a change in needs of the people who live within them. The needs of the powerful often shape cities more so than the needs of the less powerful, yet the form of our urban spaces (be it size, density, greenspace, etc.) conforms to certain human requirements nonetheless. Presently, humanity faces an issue of such magnitude that our greatest collective effort should be addressing it: climate change is altering our planet. Generations in the not-so-distant future will face rising temperatures, melting ice caps, and increased storm frequency. Thusly, our cities must begin to adapt.

Rising global temperatures – one of the most commonly discussed symptoms of our changing climate – produce amplified heat within cities (Oke, 1982). This affect is non-uniform, as the urban heat island effect creates pockets of high temperatures that effect certain areas of the city more than others (Voelkel and Shandas, 2017). One noted side effect of these increases is an increase in energy consumption – as temperatures rise, so does the amount of energy required to cool buildings to comfortable levels (Hassid et al., 2000). The demand for energy production is increased due to this, and the emissions resulting from the burning of fossil fuels at power plants exacerbate climate change (Creutzig et al., 2015). In order to mitigate this effect, we first need to understand it in detail.

In response to the need for better understanding of our urban landscapes, many methodologies have emerged. Often involving the generalization of models to different

geographic regions (see section 1.1.1.1. "Existing Studies" for detailed examples), these methods fail to assess spatial phenomena with the detail and geographic extent necessary. Too often are studies performed at the scales convenient for analysis, rather than *required* for analysis. The analytical methodologies of geographic information systems (GIS) have played a role in this over-simplification of issues by constraining which questions researchers are able to ask. These constraints arise from a limited set of data types and analytical processes applicable to these data types. This oversimplification is the inspiration for geocomputation. The field of geocomputation "represents a conscious attempt to move the research agenda back to geographic analysis" and attempts to break free of the constraints of GIS software (Gahegan, 2017). The underlying goal of geocomputation is to leverage powerful computing systems to perform the spatial analysis necessary to answer complex questions, not to work 'inside the box' of analytical tools and software that are created for mass accessibility. Because of this, geocomputational processes must utilize an array of computing environments (e.g. operating systems, muti-core processing, and parallel computing) and programming languages. I use the term "geocomputation" henceforth to refer to the use of statistical and computational methods which exist outside of standard GIS software and are leveraged in the methods herein.

The following two chapters are comprised of two essays. The analyses performed in each paper were born from the theories of geocomputation: they attempt to answer questions that *need* answering, not those that are convenient in terms of analysis, software, or processing time. The first essay, "Urban Form and Residential Energy

Expenditures: A Potential for Climate Change Intervention", begins by building a logic framework that can help us better understand how even minute changes in our built landscapes can have global and self-reinforcing ramifications. Within this framework, I assess the current state of literature on the connections between urban form and energy consumption (which I operationalize through energy expenditures as one driver of climate change). The methodology employed a spatial energy consumption dataset (the size of which has no equal in the literature) to assess the factors of urban form which drive energy use.

The second essay, "The Role of Broad Tree Functional Types in Urban Heat Island and Nitrogen Dioxide Exposure Models", builds upon the findings of chapter 1. In it, a high-resolution tree canopy classification dataset is introduced to a previous study of the urban heat island effect (Voelkel and Shandas, 2017) and a previous study of nitrogen dioxide (Rao et al., 2017) in the City of Portland. The study is situated in urban studies literature and theory, drawing on historic examples of pioneering tree planting within growing cities. The importance of trees in human health and heat mitigation leads to an important – and previously unanswered – question: what type of trees should we plant? The analysis uses machine learning to assess this question, and answers are produced in terms of air quality improvement and heat mitigation. Chapter 2 especially embraces geocomputation, creating not only a novel methodology for spatial modeling, but doing so with unique data types. These chapters – though they fit into the same field – were written with the intent to submit to academic journals and are structured as such. Both essays aim to contribute to the geocomputation literature in general by addressing

pertinent issues within literature on urban form. The primary goal of this is to see if applying complex geocomputational solutions to questions in the field (in this case the effect of form on energy consumption and environmental detriments) can increase the understanding of problems by providing high-accuracy answers and models. In each essay, analyses required custom-tailored computation in order to account for the high-granularity and richly detailed available data.

Chapter 1. Urban Form and Residential Energy Expenditures: A Potential for Climate Change Intervention

1.1. Introduction

We are living in an unprecedented era of human migration and settlement patterns. Now, for the first time, a majority of Earth's population lives in urban environments - by 2050 the United Nations expects an additional 2.5 billion people to live in cities (United Nations, 2015). Simultaneously, another historic change is occurring in the planet's climate. Human activity is driving global temperatures higher at an increasing rate, and mitigation "will require an urgent and fundamental departure from business as usual" (Pachauri et al., 2015). It is estimated that a single degree Celsius rise in the annual temperature average could have a 448 billion dollar (USD) global economic impact based on changes to agriculture, forestation levels, sea levels, ecosystems, human health, and energy consumption (Tol, 2002). At the city-scale, the effects of climate change are being felt the most: flooding and extreme heat events in urban areas are expected to occur with higher frequency than in rural areas (Gill et al., 2007).

For urban scholars and planners, this poses an opportunity to address sustainability. One of the solutions gaining traction is the inclusion of green infrastructure (the use of natural systems to supplement or replace traditional infrastructure) in the planning process. Green infrastructure implementations such as bioswales and targeted tree plantings have been shown to reduce some of the negative products of accelerated urban climate change (Foster et al. 2011; Gill et al., 2007; Baldinelli and Bonafoni, 2015). Often these green infrastructure initiatives serve to reduce a physical occurrence such as urban flooding/runoff and extreme heat events; however, there exists a potential

to mitigate ancillary effects of climate change as they relate to human vulnerability. When considering extreme urban heat, it is quite often the case that populations residing in the hottest portions of the city also have lower incomes and consist primarily of people of color (Huang et al., 2011). Often the discussion of this inequitable exposure is centered in human health (Mattern et al., 2000; Poumadère et al., 2005; Voelkel et al., 2016); however, the effects of climate change on energy consumption in urban settings may prove a more pertinent route for research due to energy consumption's own role as a driver of climate change.

1.1.1. Urban Form, Energy Consumption, and Climate Change

A key consideration of the aforementioned urban population growth is urban energy consumption. According to a recent Intergovernmental Panel on Climate Change (IPCC) report, urban areas are responsible for 67%-76% of the planet's energy consumption in addition to approximately 66% of global carbon emissions (Seto et al., 2014). Additional studies have found that the increased energy use in cities is a significant contributor to climate change (Creutzig et al., 2015). Previous literature has addressed the connection between urban form and energy consumption. In her 2013 study "Urban Form and Residential Energy Use: A Review of Design Principles and Research Findings" (Ko, 2013), Ko found that 22% of all urban energy consumption occurred in residential households. The primary drivers of this consumption are heating and cooling costs (Brack, 2002). Factors that contribute to the variation in residential energy consumption are housing size, type (single-family versus multi-family), density (e.g. concentration of housing units), vegetation (such as tree plantings), and impervious

surface coverage. In Ko's logical framework (figure 1.1), urban form dictates intra-urban microclimates (e.g. urban heat islands). Microclimates, in turn, drive residences to alter their heating and cooling patterns. Importantly, some portions of the city will see energy consumption increased dramatically by microclimates, whereas others may see positive benefits (such as reduced heating costs in the winter).

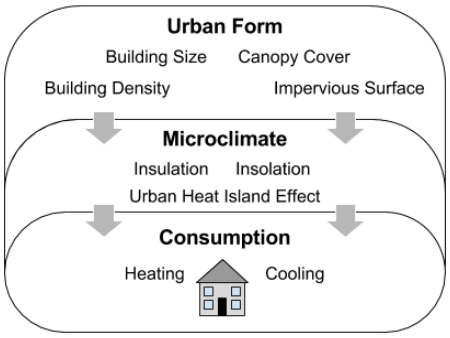

Figure 1.1. Urban form and its effect on residential energy consumption. Adapted from Ko (2013).

However, a critical component is missing in Ko's logic model: the positive feedback loop created by increased consumption. By increasing energy consumption for some residents, urban form is indirectly increasing carbon emissions created during energy production. This, in turn, can lead to higher global temperatures and an exacerbation of the urban heat island effect (Meehl and Tebaldi, 2004).

Several studies have linked individual components of urban form directly to urban energy consumption. One such study (Akbari, 2002) is able to expand upon Ko's work by beginning to consider the climatic feedback loop that occurs while simultaneously studying the effect of vegetation (trees and grasses) on energy consumption. Akbari's logic model considers the effect of trees on energy consumption directly as a result of shading buildings; however, it also acknowledges the city-wide and global effect of

increased canopy by considering that world temperature increases may be mitigated via a reduction in energy production-related emissions.

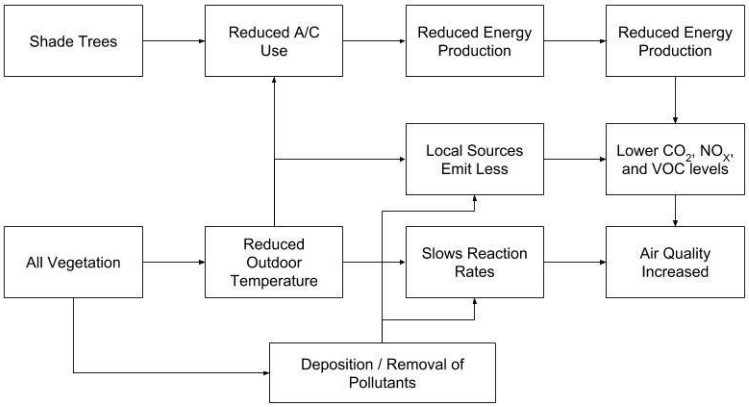

Figure 1.2. Urban vegetation and its effect on detrimental aerosols. Adapted from Akbari (2002).

Akbari's and Ko's logical frameworks for the influence of urban form on energy consumption both have their merits. Where Ko does not consider the impact of household energy consumption on global climate, Akbari does; where Akbari does not consider a multitude of factors contributing to urban form, Ko does. A key component shared between them is that urban form drives urban microclimates, which are the key influencer of energy consumption. The increases in emissions from higher energy consumption/production result in changes to global climate patterns (Kalnay and Cai, 2003) which, in turn, alter urban microclimates in a positive feedback loop. By combining the logic models of Akbari and Ko, this feedback loop is clear (figure 1.3). What is also clear is that changes to urban form are a promising intervention in both local and global temperature increases.

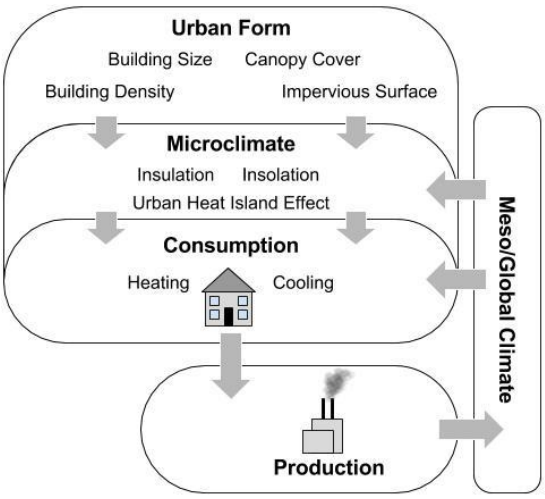

Figure 1.3. Urban form's effect on micro/global climate change and the positive feedback loop therein.

1.1.1.1. Existing Studies

Previous research into urban form and energy consumption exists on a gradient of scale upon which analysis is performed. In most cases, studies of these types have a low number of observations (Simpson and McPherson, 1998; Tso and Yau, 2007; Donovan and Butry, 2009), or are highly generalized simulations (Taha et al., 1988; Akbari et al., 2001; McPherson and Simpson, 2003; Ewing and Rong, 2008). On the opposite end of the spectrum, studies focus on precise building metrics and microclimates with such detail and precision that they cannot practically generalize the effects of urban form on energy consumption beyond individual buildings (Fahmy et al., 2009; Yang et al., 2012). For these precision studies, authors often use computationally intensive computer simulation software such as ENVI-met (for microclimate and computational fluid dynamics modeling) or EnergyPlus (for inter-building material heat transfer modeling). Another common type of study is situated between the aforementioned paradigms on the gradient of "too-precise" and "too-general". Often leveraging software such as CITYgreen and iTree software, these studies require highly detailed information on urban

forests that can only be obtained through rigorous in-situ sampling (i.e. tree trunk diameter); however, once this information is obtained, models resample this detailed information to US Census geometries to make final estimates of urban for influences (Carver et al., 2004; Solecki et al., 2005; Nowak et al., 2009; Hirabayashi et al., 2012; Bodnaruk et al., 2017). In essence, these studies must put forward the effort of a high-resolution study, but final predictions and estimates are limited to a smaller sample of Census Block Groups.

The methodologies employed in the literature on urban form and energy consumption fall into three categories. Though these methodologies are applied to a wide variety of research questions, I refer to only those studies which seek to understand energy consumption as it is driven by urban form. Donovan and Butry (2009) begin to touch on these categories, though their analysis does not provide a sufficient number of household-level energy observations to properly remove itself from the shortcomings of the previous literature (n = 460). Table 1.1 lists these methodological paradigms along with their deficiencies:

Table 1.1. Common energy consumption evaluation methodologies.

Method	**Description**	**Common Software**	**Weaknesses**
Generalized / Global	Comparisons between multiple cities or regions, often leveraging downscaled global climate models.	URBMET, DOE-2.1C	Assumes uniformity in the spatial distribution of urban features (e.g. "Total Canopy Cover" as a metric for determining city-wide energy consumption).
Detailed / Semi-Localized	Inter-building comparisons, usually factoring in trees surrounding said buildings	iTree, CITYgreen	Requires exact metrics of trees (e.g. species, DBH), but generalizes to a smaller number of geometries.
Precise / Localized	Compares intra-building construction materials.	ENVI-met, EnergyPlus	Requires many input variables and is thusly limited to energy profiles of small geographic areas.

According to Santamouris et al. (2015), regional building- and landform-based analyses are needed to better understand the relationship between energy consumption and urban form. This finding – based on a meta-analysis of 15 studies – calls for an analysis that is situated between "Generalized / Global" and "Detailed / Semi-Localized" in order to perform the analysis in a far more rigorous way. In order to accurately assess the role of urban form in altering residential energy consumption, this new "Hybrid" method must a) be conducted at either a city- or regional-scale; b) contain highly-resolved descriptions of landforms such as trees and buildings; c) contain a large sample size of building energy use observations; and d) account for variations in building stock metrics such as square footage and vintage. Using this hybrid method – described in full detail in the "Methods" section – I propose that the configuration of trees around urban residences (i.e. the amount of trees within certain distances in certain directions of residences) lowers total annual energy expenditures. Though this has been studied in the existing literature, it has not been tested for an entire metropolitan area at the building-level with a satisfactory number of energy consumption observations to determine the outcomes of urban form-related climate change interventions. I believe that application of this large-scale fine-granularity will create a highly accurate model for residential energy consumption, specifically tailored for the study area. Additionally, I hypothesize that residents who are situated in areas of extreme localized urban heat islands will exhibit higher annual energy expenditures. This is important to assess, as heat not only increases the demand for energy used to cool residences, but also indicates particular types of urban form (e.g. building density, lack of vegetation).

1.2. Methods

1.2.1. Study Area

The study area for this analysis is the Portland Metropolitan Area (PMA), Oregon, USA. The PMA contains 24 cities and municipal areas constrained by a regional urban growth boundary. The PMA extends over 1218km^2 and has an estimated population of 1.46 million as of the 2010 Census (Metro Data Resource Center, 2017). Though known popularly as an area of high precipitation, the PMA experiences hot and dry summer months.

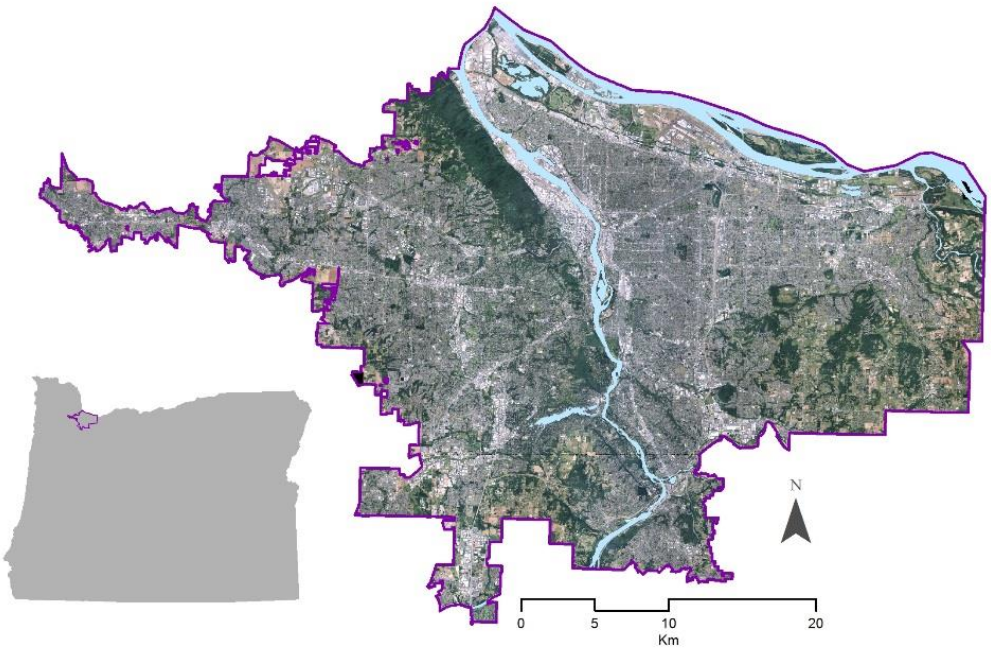

Figure 1.4. Aerial Imagery of the PMA.

1.2.2. Data

1.2.2.1. Energy Consumption

Energy consumption data was obtained at the sub-building level for the PMA through Energy Trust of Oregon. Included in this data were site addresses, yearly electricity consumption in kilowatt hours (kWh), and yearly natural gas consumption in therms (thm). Extensive cleaning and processing by the Sustaining Urban Places Research Lab at Portland State University allowed for joining of the energy data with other regional datasets such as building and tax parcel variables based on geographic location.

1.2.2.2. Canopy Cover

In this analysis, extra attention will be given to tree canopy configuration – this is due to the relative ease and speed in which the amount of trees can be altered within a city compared to other factors such as building density. Information on tree canopy was made available by Oregon Metro's Regional Land Information System (RLIS; Metro Data Resource Center, 2017). The information is provided as a $1m^2$ resolution geographic information system- (GIS) compatible raster. The raster covers an area of 6537 km^2, with pixels containing values of 'null' (i.e. no tree cover at the specific location) or an integer representing the maximum height of trees within the respective pixel. This data was created by combining high-resolution LiDAR-derived elevation data with high-resolution aerial photograph-based vegetation indices. In total, there are 982,859,913 pixels containing tree height information. In order to assess the canopy cover as a percentage of area within multiple distances and directions from each building, the data is first converted to a binary representation (i.e. 'tree' pixels with a value of 1, all others with a value of 0) using the following logic:

$$R_o = (R_{Ci} > 0 \rightarrow R_{Oi} = 1) \wedge (\neg R_{Ci} = \text{NULL} \rightarrow R_{Oi} = 0) \quad (1)$$

where:

R_c = Input canopy data

R_o = Output canopy data

i = Individual raster pixel

The output of this function is a new raster dataset, in which the mean value of pixels within any area represents the percent canopy cover. Using this logic, 20 moving window analyses were performed to create new raster variables in which each pixel represented the canopy cover within 10ft, 20ft, 30ft, 40ft, and 50ft to the North, South, East, and West (figure 1.5). This method allows all 20 values to be easily appended onto the energy consumption data for further analysis.

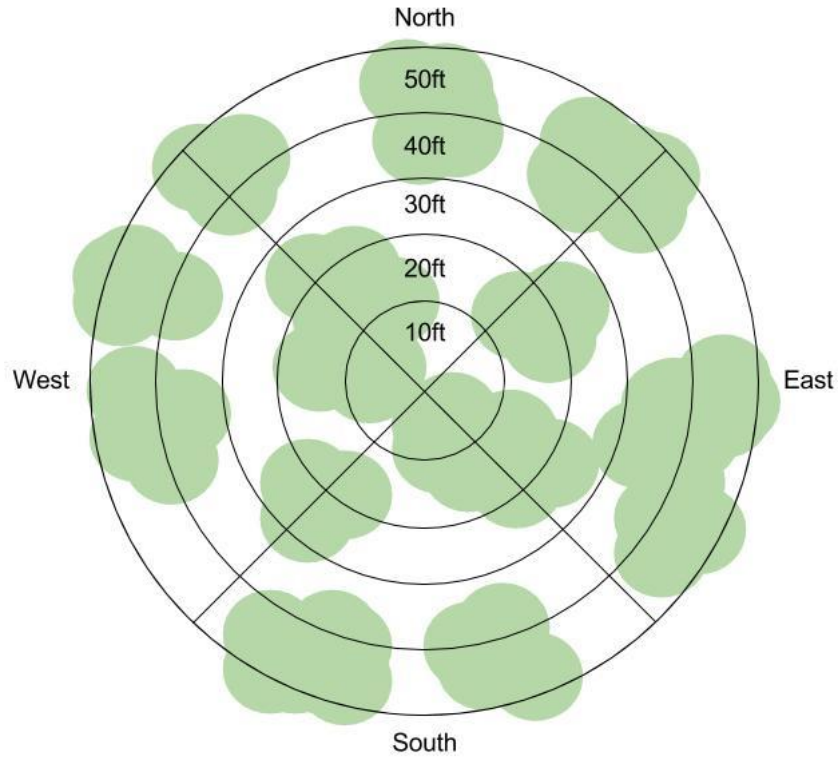

Figure 1.5. Multi-distance and multi-directional assessment of canopy configuration, with building being located at the center of the diagram.

1.2.2.3. Urban Heat

Urban heat island (UHI) data sets were obtained from the Sustaining Urban Places Research Lab at Portland State University. The rasters cover the study area at a $1m^2$ resolution and have modeled temperatures for the region during a heat event. These models were created using the same methods and datasets as Voelkel and Shandas (2017), with the extent of raw observation data being the only major differences. This process requires the measurement of different factors of urban form within many distances across the study area – the result of this measurement is a $1m^2$ staked raster dataset (or, datacube) wherein each pixel contains information on the amount of specific features within a given distance (e.g. "canopy cover within 50m", or "building volume within 1000m"). This data cube is coupled with field measurements of heat taken at specific 1-hour windows during a heat event. These field measurements are collected by attaching a thermocouple to multiple vehicles, which drive in specific portions of the study area and sample temperatures at 1-second intervals. GPS devices are used in conjunction with the thermocouples, allowing values from the datacube to be joined to temperature readings across the study area based on precise spatial location. Random Forest machine learning is used to train a model which predicts temperatures based on these field measurements' temperatures and datacube values. The result of this analysis is the prediction of an entire 'surface' of heat across the study area, realized as a $1m^2$ resolution spatial raster. Details on the final UHI predictive surfaces rasters used in this study can be found in table 1.2:

Table 1.2. Urban Heat Island Model Details

UHI Variable	Time Period	R^2	RMSE	Collection Date	Resolution
Morning	6am-7am	0.9872	0.1661°C	7/29/2016	$1m^2$
Afternoon	3pm-4pm	0.9053	0.3454°C		
Evening	7pm-8pm	0.9665	0.2048°C		

It should be noted that the meaningfulness of the included UHI models extends beyond their temperature information. This is quite important, as the UHIs serve as a snapshot of one hour during one day of a summer heat event - at face value they pose an issue when predicting yearly energy expenditures. Each of these datasets is created from a high-resolution, multi-distance, and multi-dimensional representation of the region's urban form. This combination of measurements and variables which make up the physical city is referred to by Ratti et al. (2005) as "urban texture". While the temperature data predicted by the UHI models is merely an extrapolation of urban texture, these three models are being introduced as quantifiable measurements of urban form – UHI is an entry point for characterizing complex urban forms with particular ecological and/or environmental features. Most notably, they are representative of canopy/biomass cover, vegetative vs. impervious ground cover, building coverage, building heights (including the variation of heights), and building volumes within both close proximity and regionally for any given $1m^2$ pixel in the study area. Each UHI model represents a different combination of urban textural elements that drives micro-climate events in a spatially explicit manner throughout the day; therefore, I have incorporated all three UHI models as variables in the analysis.

1.2.2.4. Unit of Analysis

For this analysis the energy data was joined to spatially explicit residential (both single- and multi-family) building polygons obtained through RLIS, with a sum of kWh and thm calculated in instances where multiple meters existed. Costs per kWh and thm were obtained from Northwest Natural and Portland General Electric. The building-level kWh and thm values were converted to total expenditure (in USD) by multiplying by 0.114 and 0.90723 respectively and calculating the sum. Residential status was obtained through a spatial join to RLIS tax lot data. A count of the number of electricity meters was included in during this building-level aggregation process. Next, the 20 variables representing multi-distance and multi-directional canopy cover were added based on an extraction of the canopy rasters at each building polygon. This process of directional buffers was repeated to calculate total canopy volume configuration in addition to canopy cover configuration. Finally, the UHI raster values were extracted at each building. The output result of this process was a GIS-compatible polygon data set with the following variables (Table 1.3):

Table 1.3. Variables compiled for building-level analysis.

Variable	Description	Units	Source
Total kWh	Total yearly electricity consumption.	kWh	ETO / SUPR Lab
Total thm	Total yearly natural gas consumption.	thm	
Meters	A count of all electricity meters in a building.	N/A	ETO / SUPR Lab / RLIS
Building Size	Floor area of the building.	ft^2	RLIS
Land Value	Value of the land within the parcel where the building is sited.	USD	
Land Area	Area of land within the parcel where the building is sited.	Acres	
Building Value	Value of the structure itself.	USD	
SFR	A dummy variable representing whether a building was likely single-family.	Boolean (1 or 0)	

Building Age	Construction date of the building subtracted from current year (2017).	Years	
UHI - Morning	Morning urban heat model value.	°C	Voelkel and Shandas (2017)
UHI - Afternoon	Afternoon urban heat model value.	°C	
UHI - Evening	Evening urban heat model value.	°C	
Canopy Cover Configuration	A number between 0 and 1 for each distance (10ft, 20ft, 30ft, and 40ft) in each direction (North, East, South, and West) representing canopy cover surrounding each building.	Percent	RLIS
Canopy Volume Configuration	An integer for each distance (10ft, 20ft, 30ft, and 40ft) in each direction (North, East, South, and West) representing the sum of canopy heights in each $1m^2$ pixel surrounding each building.	Percent	RLIS

1.2.3. Regression

1.2.3.1. OLS

The analysis employed Ordinary Least Square Regression (OLS). OLS is a common method for predicting a dependent variable based on a function of one or more independent variables. It is often employed in the literature, regardless of geographic scale (Larivière and Lafrance, 1999; Tso and Yau, 2007; Donovan and Butry, 2009). When model assumptions – such as homoscedasticity or limited multicollinearity in the independent variables – are met, OLS is often the best linear unbiased estimator of a dependent variable (Chumney et al., 2006).

1.2.3.2. Analysis of Spatial Autocorrelation

A fundamental assumption in linear regression is that the independent variables are not excessively autocorrelated (Chumney et al., 2006). When a spatial component is introduced into the analysis, steps must be taken to ensure that these variables are not dependent upon each other in terms of where they occur spatially. If it is found that a spatial relationship – or, spatial autocorrelation – exists between observations (e.g.

clustering of certain values or magnitudes of values), "it impairs our ability to perform standard statistical tests of hypotheses" (Legendre, 1993). In order to evaluate whether any spatial pattern was observable in my final model, I performed a Monte Carlo permutation test for Moran's I (a common indicator of spatial autocorrelation) on the model residuals. This test randomly assigns residual values to the observation points in space 999 times, building a distribution of expected Moran's I values in a hypothetical non-spatially autocorrelated case (i.e. the null-hypothesis). This null distribution is next compared to the single Moran's I value for the model residuals, and significance of the spatial autocorrelation is determined (Cliff et al, 1981).

1.3. Results

1.3.1. OLS

During the OLS modeling process, not all independent variables tested had a significant effect at $\alpha = 0.05$. Those variables that did not meet a significance level of $p < 0.05$ were dropped from the regression, and another iteration took place. More information on these insignificant variables can be found in sections 1.4.1. and 1.4.2. The final model is comprised of 14 variables, which are described in table 1.4:

Table 1.4. Selected OLS independent variables and their descriptions.

Variable	**Description**
Number of Electricity Meters (Count)	A total count of all electricity meters within the building in question. As an accurate count of units does not exist at the building-level for the study area, this is the closest measure of housing units available.
Building Value (USD)	The assessed value of the building only (i.e. excluding land values) in USD.
Parcel Area (acres)	The total land area of the parcel upon which the building sits, measured in acres.
*Is Single-Family * Bldg. Square Footage*	An interaction term between the binary "Is Single-Family" dummy variable and the square footage.
Building Age (Years Old)	The age in years of the building, calculated by subtracting the year built from 2017.

Land Value (USD)	The assessed value of the land, excluding buildings, upon which the building sits.
UHI, Evening (°C)	Evening UHI model temperatures at the location of the building, measured in degrees Celsius.
Canopy Cover, N, 30ft (%)	The area to the North of the building within 30 feet that is covered by trees
Canopy Cover, E, 30ft (%)	The area to the East of the building within 30 feet that is covered by trees
Canopy Cover, S, 40ft (%)	The area to the South of the building within 40 feet that is covered by trees
Canopy Cover, W, 40ft (%)	The area to the West of the building within 40 feet that is covered by trees
Is Single-Family (Binary)	A dummy variable (i.e. '1' or '0') representing whether or not the building is listed as 'Single-Family'. Values of '0' represent 'Multi-Family' households
Building Square Footage (ft^2)	The total floor area of the building in square feet, as determined by the county assessor (i.e. not a GIS calculation from ground area covered by the building footprint)
UHI, Morning (°C)	Evening UHI model temperatures at the location of the building, measured in degrees Celsius.

Overall, the model supports most of the assumptions made between the selected independent variables and total yearly energy expenditure. After fine-tuning some of the specifications (see the "Patterns in Outlier Observations" in the "Discussion" section for details), the model is able to explain 84.36% of the variation in expenditure with 14 variables. Diagnostics of the OLS model are found in table 1.5:

Table 1.5. OLS Regression Results

Variable	**Estimate**	**Beta-Est.***	**Std. Error**	**t Value**	**p-value**
(Intercept)	1114	N/A	218.1	5.107	< 0.00001
Number of Electricity Meters (Count)	771.2	0.7803	1.242	621.1	< 0.00001
Building Value (USD)	0.000795	0.1198	0.00001	81.15	< 0.00001
Parcel Area (acres)	928.6	0.08347	12.24	75.89	< 0.00001
*Is Single-Family * Bldg. Square Footage*	0.3263	0.0689	0.004777	68.32	< 0.00001
Building Age (Years Old)	2.511	0.05958	0.03897	64.43	< 0.00001
Land Value (USD)	0.000107	0.003655	0.000039	2.77	0.005601
UHI, Evening (°C)	18.67	0.002803	6.018	3.102	0.001923
Canopy Cover, N, 30ft (%)	-61.06	-0.00353	19.82	-3.08	0.002069
Canopy Cover, E, 30ft (%)	-61.09	-0.0036	19.47	-3.138	0.0017
Canopy Cover, S, 40ft (%)	-65.32	-0.00385	18.51	-3.53	0.000416
Canopy Cover, W, 40ft (%)	-91.25	-0.0054	18.5	-4.933	< 0.00001
Is Single-Family (Binary)	-306.9	-0.0062	47.03	-6.526	< 0.00001

Building Square Footage (ft^2)	-0.01535	-0.0111	0.001753	-8.754	< 0.00001
UHI, Morning (°C)	-94.58	-0.0158	5.496	-17.21	< 0.00001

Model statistics: n = 219619; Adjusted R^2 = 0.8436; RMSE: 1717; AIC = 3895097
*Standardized relative influence on the model.

Multicollinearity of the independent variables was measured using the variance inflation factor (VIF). It is vital to assess multicollinearity as it is a fundamental assumption of OLS that the independent variables are not excessively dependent upon one another (Chumney et al., 2006). As no variables have a calculated VIF greater than four (see table 1.6), it can be determined that the model has an acceptable and minimal degree of multicollinearity (Montgomery et al., 2015).

Table 1.6. Variance Inflation Factor for OLS independent variables.

Variable	**Estimate**
Building Value (USD)	3.060496
Land Value (USD)	2.444349
Building Square Footage (ft^2)	2.272573
Number of Electricity Meters (Count)	2.216073
Canopy Cover, N, 30ft (%)	1.848899
Canopy Cover, E, 30ft (%)	1.834478
Parcel Area (acres)	1.698599
Canopy Cover, W, 40ft (%)	1.675404
Canopy Cover, S, 40ft (%)	1.672986
*Is Single-Family * Bldg. Square Footage*	1.428508
Is Single-Family (Binary)	1.269184
Building Age (Years Old)	1.200972
UHI, Morning (°C)	1.182058
UHI, Evening (°C)	1.146304

1.3.2. Spatial Autocorrelation

The results of the Monte-Carlo simulation of Moran's I determined that there is enough evidence to reject the null hypothesis and conclude that spatial autocorrelation is present in the OLS model residuals; however, with an I of 0.0305, the test shows that

there is only a slight degree of spatial autocorrelation as a value of 0 represents perfect spatial randomness and a value of 1 represents perfect spatial clustering. Due to this, the advantage of computing a spatial error model is minimal and has been forgone in this study.

1.4. Discussion

1.4.1. Urban Heat

Most notable among the dropped variables (which were not significant at α = 0.05) is the afternoon UHI model, which predicts the highest temperatures observed in the three UHI models. There is a potential that, due to the model's overall lower accuracy (RMSE = 0.3454°C; see table 1.2) relative to the other time periods, any strong correlation between afternoon temperatures and energy expenditure is lost. An alternative explanation, however, is that there is in fact no measurable influence on expenditure because residents are often not home during the time period represented in the afternoon UHI (3pm-4pm). Of the two UHI variables that remain in the OLS model, opposite effects are observed. For residents who live in areas experiencing higher morning UHI temperatures, a reduction in annual energy expenditures is observed in that every additional degree Celsius results in a savings of $94.58 per year. Residents situated in areas corresponding to higher evening UHI temperatures experience an additional annual energy expenditure of $18.65 per year. Notably, morning temperatures are lower than evening temperatures thus the disparity between the two coefficient estimates is not as extreme as it seems: by assessing each variable's Beta-weights it is revealed that both morning and evening UHI have similar effect sizes; however, even the lowest evening

temperatures exceed the highest morning temperatures. The results of this model are still able to show that those residents being exposed to the most extreme temperatures are also facing an additional burden of increased energy expenditures due in part to this localized heat and the particular urban form in which they are situated.

1.4.2. Canopy Configuration

The other variables with surprising results are the directional canopy metrics. Of all 20 canopy volume configuration variables (in which the sum of canopy volume was calculated), a significant relationship cannot be determined. The canopy cover configuration variables reflect discoveries made in previous literature, with the exception of trees to the North of a building. Donovan and Butry (2009) found a significant positive relationship between canopy within 10ft to the North of a building and energy consumption. My analysis reveals that canopy cover within 10ft to the North is insignificant, and that at a distance in which it becomes significant (30ft) a similar cooling pattern and amount as those to the East and South is observed. The OLS model results align with Donovan and Butry (2009) in the finding that canopy cover to the West of homes within 40ft is significantly better at reducing annual energy expenditures, with the OLS model determining a reduction in annual energy expenditure of $9.25 per year for every 10% increase in canopy cover.

1.4.3. Patterns in Outlier Observations

During the model-fitting exercise, outliers were observed which had excessive error. The removal of the 12 highest influence of these outliers (approximately 0.005% of the total observations) results in an R^2 increase of 0.0874. Due to the relatively large

increase in energy expenditure variation explanation resulting from the removal of small number of observations, the final model does not include these twelve. Upon closer inspection of the removed observations, five are retirement homes, two are inpatient drug and alcohol detoxification centers, and the remaining five appeared to be misclassifications in land use (e.g. they contained commercial operations). A major factor contributing to the error introduced by retirement homes is the variable "Number of Electricity Meters" - though the retirement homes removed appeared to contain at least a few dozen housing units, the buildings themselves never contain more than 8 electricity meters. It is highly likely that electricity is included in the housing costs of these residents, thus there is no need for utility services to differentiate between each unit. In addition to the limited electricity meters, it is very likely that the electricity-expensive medical equipment in the retirement and detoxification homes are responsible for relatively abnormal total energy expenditures. The issue of misclassification could also have an influence on multi-family residential observations due to a lack of data identifying mixed use tax lots: there are likely meters included in the data that belong to commercial spaces located within the same buildings as multifamily.

1.4.4. Applications of results

When increasing the canopy cover 10% to the North, South, East, and West within 30ft, 30ft, 40ft, and 40ft respectively we expect a reduction in annual energy expenditure of $27.87 (with a 95% confidence interval between $12.92 and $42.83). Though the effect size of trees on annual energy expenditure appears low when considering a single residence, the impacts at analysis scales above single-building are

significant. When savings for a single home are extrapolated to 300 homes (a sub-neighborhood area commonly consisting of approximately 10-12 residential blocks), the total annual tree-based savings increase to $8,361.60 (with a 95% confidence interval between $3,875.62 and $12,847.63). By increasing the scale to 1,980 homes (the average number of residential buildings per neighborhood in the PMA) savings increase to $55,186.56 (with a 95% confidence interval between $25,579.10 and $84,794.33). Finally, this 10% increase – when considering all 465,368 residential buildings in the study area – results in a savings of $12,970,737 (with a 95% confidence interval between $6,011,967 and $19,929,579). All of these estimates increase based on the total increase in canopy cover: Figure 1.6 shows these percent increases in canopy cover along the x-axis, with the expected savings (in USD) along the left and right y-axes for a sub-neighborhood and the entire study area respectively.

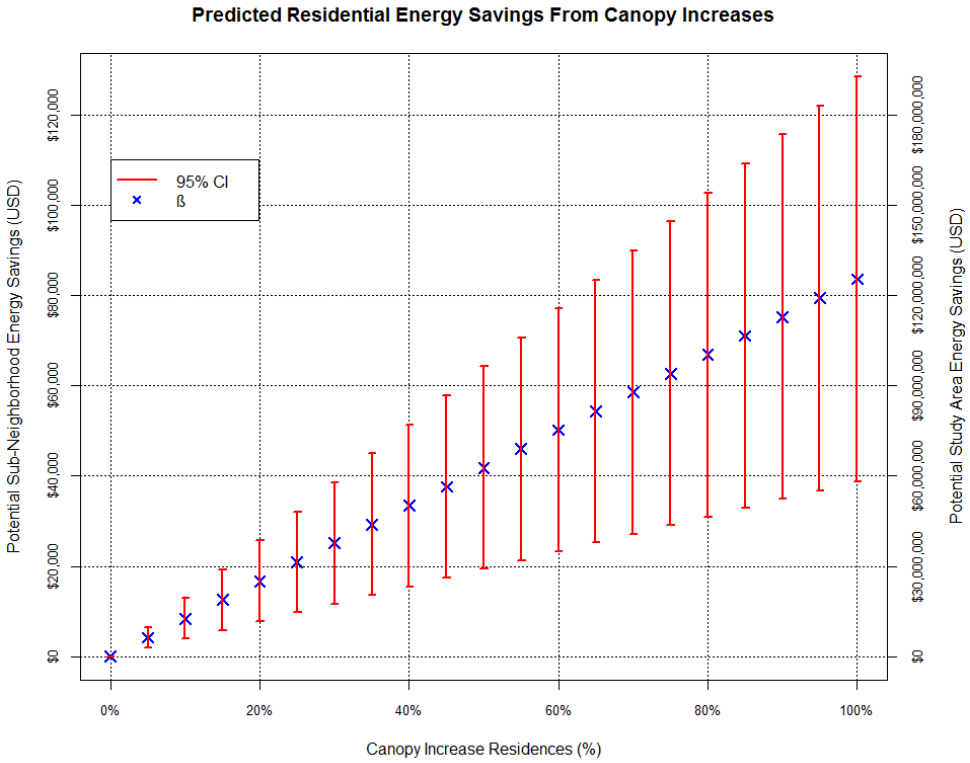

Figure 1.6. Predicted residential energy savings from canopy increases to the North, East, South, and West of a residence within 30ft, 30ft, 40ft, and 40ft respectively.

1.4.5. Temporal Resolution

The greatest drawback to the UHI data in this study revolves around the temporal resolution of energy consumption data. This analysis assesses annual expenditures, yet uses UHI models representing a single-hour snapshot in time. To better understand the influence of temperatures on residential energy expenditures, higher temporal resolution data must be obtained. Studies such as Donovan and Butry (2009) or Simpson and McPherson (1998) combat this issue by using summer-time energy measurements; however, this increase in resolution is coupled with incredibly low numbers of observations (460 and 254 observations, respectively). Even though the analysis in this study assesses annual expenditures, two of the UHI variables still show significant

influence on the model. This reaffirms the use of these UHI variables as metrics for urban form and texture crucial to understanding residential energy use patterns.

1.4.6. The Cost of Trees

A major consideration of targeted tree plantings is the cost, both in terms of initial investment and continued upkeep. Though this analysis has shown an increase of canopy cover around residential households will lower overall energy expenditures, the costs associated with purchasing, planting, watering, and maintenance (e.g. trimming branches or raking fallen leaves) may outweigh the benefits. Further analysis into the costs of trees by McPherson and Simpson (2003) shows that the initial costs of planting and maintaining trees is approximately $50/tree, when purchasing in bulk for a mass-planting campaign. They note that the energy consumption benefits of these trees may take up to 13 years to be realized, meaning that targeted tree plantings are likely cost effective when considering long-term energy consumption goals. Notably, existing and established tree canopy in urban areas are already creating benefits without the immediate planting costs – this means that the preservation of existing trees needs to be considered alongside additional plantings.

1.5. Conclusion

The analysis employed in this study used OLS regression to determine the effects that two metrics of urban form/texture (canopy configuration and UHI) had on residential energy expenditures. After the removal of outliers and non-significant variables (at α = 0.05) the final OLS model specification had an adjusted R^2 of 0.8436 and an RMSE of approximately $1,717 based on 219,619 observations. As hypothesized, the configuration

of canopy around a residential building (specifically canopy cover) at different distances and in different directions has a significant negative relationship with annual energy expenditures.

In addition, the hypothesis that residents located in areas of higher temperatures – inherently driven by complex relationships of urban form – experience an additional burden in terms of energy expenditures is confirmed. Those residents who are located in areas where evening temperatures are relatively high see an increase of $18.67 per year for every additional degree Celsius increase. By lowering this additional cost with interventions in urban form it may be possible to reduce regional and global temperature increases, thus slowing down the processes of climate change. Though residences located in areas of high morning temperatures experience a reduction of $94.58 per year per degree Celsius, the actual temperature ranges seen in the morning are far lower than those of the evening. It is also important to note that morning and evening UHI patterns have dramatic differences, and concentrations of high temperatures occur in different geographic locations.

Urban form can be used as an intervention in energy consumption; however, it isn't the greatest influencer. Personal behavior far exceeds any other variable when determining the amount of residential energy consumed, and knowledge of this behavior is likely to be more important that urban form and microclimates (Pettersen, 1994; Ratti et al., 2005). In this study (and the others mentioned) urban form is used to operationalize energy expenditure patterns because it is one of the only climate change intervention factors that can currently be measured. In order to properly intervene in global climate change, the results found herein must also be coupled with urban planning and education

outreach – failure to do so in a timely fashion will likely result in future global climate patterns that have changed too drastically for intervention.

Chapter 2. The Role of Broad Tree Functional Types in Urban Heat Island and Nitrogen Dioxide Exposure Models

2.1. Introduction

Since critical theory on urban studies began in the early to mid-1800's, scholars have considered the detrimental effects that the urban environment has on the populations that reside in cities. Early industrial cities were rife with pollution — firsthand accounts from Fredrich Engels from *Condition of the Working Class in England* (Engels, 1845) describe the streets of London as cesspools of runoff, the roadways "a long string of stagnant puddles". Engels notes the deplorable air that loomed over England, affecting all of its residents negatively. The direct physical health impacts were noticed by the urban populations as well as the mental 'hardening' that the industrial city imposed on those who lived within it (Simmel, 1903). To escape the blight of the 'urban', many began to look outside of the city: towards nature. In this early literature, and permeating through urban studies scholarship well into the 1900's, nature has been considered the antithesis of the city. To urban philosopher Georg Simmel, 'urban' was — at its core — the absence of nature; to sociologist Louis Wirth, cities were the "removal of the organic" (Wirth, 1938). The dichotomy of the city and nature brought about theories that cities could be made better by allowing some form of nature to reside within them. Some forward-thinking planners went so far as to design theoretical cityscapes that focused on integrating cities with nature, such as Ebenezer Howard's "Garden Cities" (Howard, 1902). Others – such as famed Central Park (New York, New York) and Portland Park Plan (Portland, Oregon) designer Frederick Law Olmsted – viewed nature as impossible to integrate uniformly within cities. This led to a focus on the development of large

central-city parks, where residents would be able to briefly ‘escape’ the city (Olmsted, 1870).

With today’s modern push for sustainable and ‘green’ cities, we see the vision of Olmstead’s centralized nature begin to blur with the ubiquitous nature urged by Howard. With this physical blurring of nature and cities comes a change in the concept of nature within cities: the concept of nature within urban environments now focuses more on cities as an extension of nature, thusly shifting research away from the dichotomous relationship of the past and placing ‘urban’ and ‘nature’ within the same symbiotic unit of analysis (Gandy, 2006). The widespread greening of many post-industrial cities around the world is welcomed, as the pollution-ameliorating powers of urban forests (i.e. widespread canopy cover within a city) are being linked to positive health benefits for urban populations. In one of the earliest quantifications of ‘green’ benefits, Roger S. Ulrich discovered something interesting about trees: after surgery, patients in hospitals had a faster recovery time when their window looked out onto trees (Ulrich, 1984). This study is considered a catalyst for research on trees and health since 1984, and as of the time of this writing (11/20/2017) it has been cited 4129 times (according to Google Scholar).

2.1.1. Trees and Exposure

The body of literature studying the effects of trees on urban environments and human populations is now large, spanning disciplines from environmental sciences to urban planning and assessing the minutia and specificity of the role of urban tree canopy. In Portland, Oregon alone, many studies have assessed the role of trees in pollution and

exposure mitigation. Donovan et al. (2011) found that by increasing tree-canopy by 10% "within 50m of a house reduced the number of small for gestational age births by 1.42 per 1000 births" (Donovan et al., 2011). This increase in birth weights could be in part due to the tree-reduced air pollution in Portland (Rao et al., 2014), which is noted in studies world-wide (Bealey et al, 2007; Clougherty et al., 2013; Henderson et al., 2007; Nowak et al., 2006). In addition to reducing human exposure to air pollution, trees have a documented cooling effect on urban environments (Akbari, 2002; Baldinelli and Bonafoni, 2015; Bolund and Hunhammar, 1999; Cao et al.,2010; Hart and Sailor, 2008), which likely reduces heat-related illnesses in summer months (Borden and Cutter, 2008; Poumadère et al., 2005; Sullivan, 1995). In previous attempts to model the drivers of the urban heat island effect, it has been found that midday temperature models are the poorest performers (Voelkel et al., 2016; Voelkel and Shandas, 2017). It has been suggested that possible drivers of this difficult-to-explain afternoon temperature variation could be related to either the spatial configuration of trees or the functional type (i.e. evergreen or deciduous) of those trees (Henry and Dicks, 1987; Lin and Lin, 2010). This speculation has not been explored thoroughly in urban heat island literature, potentially due to a lack of high resolution data describing the urban forest at a broad functional type level of detail.

The benefits of trees in urban settings are known. Though they have always been considered to positively impact the health of urban populations, modern quantitative techniques have allowed planners to potentially target areas in most need of environmental exposure reduction. Though there has been previous research on the optimization of tree plantings for exposure reduction (Li et al., 2005; Wu et al., 2008), no

such studies exist on defining which types of trees should be planted for optimal air pollution or heat reduction. In the following sections, the role of broad tree functional type (BTFT) — in this case meaning evergreen and deciduous tree canopy — will be assessed in two environmental exposure models. For predicting NO_2 in the City of Portland, I hypothesize that evergreen trees will have a greater effect on aerosol concentrations than deciduous trees due to the more complex brush-like structure of most evergreen trees. As a mitigator of urban heat, I hypothesize that evergreen trees will – again – be a more powerful predictor than deciduous trees due to their denser structure, though the introduction of both will improve model performance.

2.2. Methods

2.2.1. Study Area

The study area for this analysis is the City of Portland, Oregon, USA. Portland is a mid-sized city with approximately 620,000 residents. It has a mild climate with rainy winters, and warm summers. Though data on BTFT exists for the entire Metropolitan area, the methods employed herein require an additional 1000m 'buffer' of data around the study area — because of this — and the limited geographic extent of the input data — the scope of the analysis is limited to Portland alone.

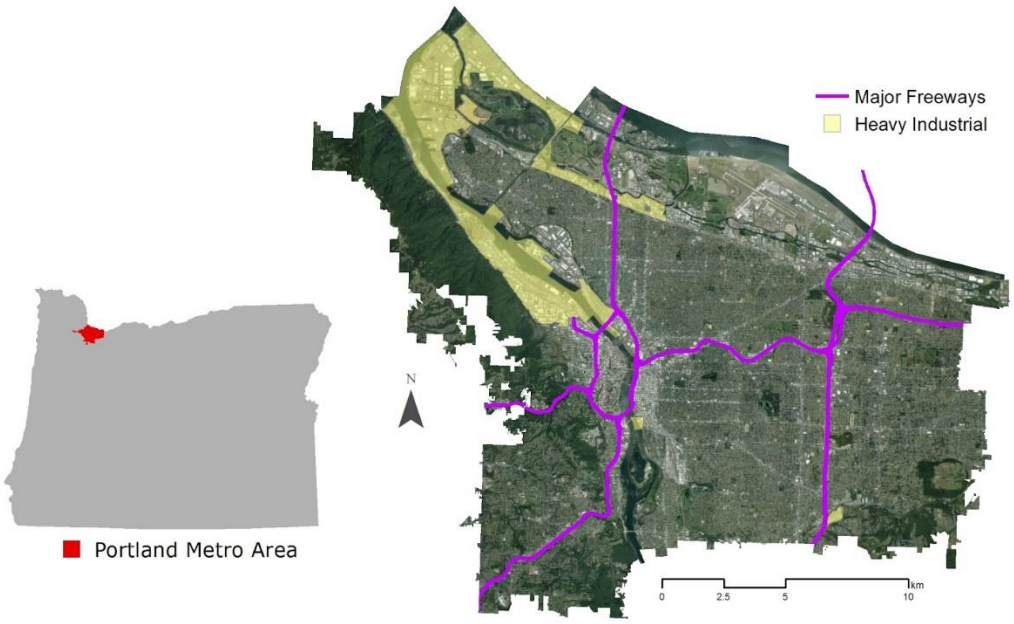

Figure 2.1. Left: Oregon, with the Portland Metropolitan Area in red; Right: the City of Portland, with two common drivers of extreme urban heat and air pollution (heavy industrial zones and major freeways) noted.

2.2.2. Data

Datasets used in this analysis are listed below in table 2.1, with further detailed descriptions in the following sections.

Table 2.1. Data sources.

Data	**Format**	**Year**	**Parent Data**	**Resolution (meters^2)**	**Source**
Building Height	Raster	2014	OLC LiDAR	1	SUPR Lab[†]
Canopy Cover	Raster	2014	OLC LiDAR / Orthophotos	1	SUPR Lab
CDM	Raster	2014	OLC LiDAR	1	SUPR Lab
Low-lying Vegetation	Raster	2014	OLC LiDAR / Orthophotos	1	SUPR Lab
Broad Tree Functional Type (BTFT)	Raster	2014	OLC LiDAR / Orthophotos	1	SUPR Lab / Oregon Metro
Evergreen Canopy Cover	Raster	2017	Canopy Cover / BTFT	1	SUPR Lab

Deciduous Canopy Cover	Raster	2017	Canopy Cover / BTFT	1	SUPR Lab
Evergreen CDM	Raster	2017	CDM / BTFT	1	SUPR Lab
Deciduous CDM	Raster	2017	CDM / BTFT	1	SUPR Lab
In-situ heat	Vector (point)	2014	N/A	N/A	SUPR Lab
In-situ NO_2	Vector (point)	2013 / 2014	N/A	N/A	Rao et al. (2014)

[+]Sustaining Urban Places Research Lab, Portland State University, Portland, Oregon. *http://www.suprlab.org*.

2.2.2.1. Land Use / Land Cover (LULC) Data

LULC data is primarily derived from two data sets. The first data source is airborne LiDAR collected by the Oregon Lidar Consortium (OLC) in the summer of 2014. Due to the high data collection rate of LiDAR systems, the dataset collected over the 3,200km^2 greater Portland Metropolitan area has an average point density of ~12 points/m^2 on a flat surface and over 60 billion individual points. From this 'pointcloud', a highly accurate and detailed 1m^2 resolution Digital Surface Model (DSM) can be created in which each pixel in the geographically positioned raster data set represents the elevation. A Digital Elevation Model (DEM) representing elevations of the surface of the earth with all features removed (e.g. trees and buildings) can also be created with the same parameters as the DSM. LiDAR elevation/height data is coupled with 15cm^2 resolution 4-band orthoimagery (i.e. including an infrared band in addition to red, green, and blue bands). Together, these data were used to classify buildings (with information about height and volume), tree canopy cover, low-lying vegetation (under 10ft), and canopy density metric (CDM). CDM is a measure of tree *amount* rather than tree cover, and is calculated by dividing the points classified as 'tree' in the LiDAR pointcloud by all

points within a $1m^2$ cell. Voelkel and Shandas (2017) found that CDM within multiple distances has a major effect on UHI.

2.2.2.2. Broad Tree Functional Type Data (BTFT)

A geospatial raster dataset representing deciduous and evergreen trees throughout the greater Portland Metropolitan Area was obtained through Oregon Metro's Regional Land Information System (RLIS; Metro Data Resource Center, 2016). The classification was performed leveraging the aforementioned LiDAR and orthoimagery and an ensemble of machine learning algorithms by analysts from Metro and the Sustaining Urban Places Research Lab (Portland State University). The data has an overall classification accuracy of 88%, with a Kappa of ~0.75. As with the LiDAR-derived raster data, the BTFT data is $1m^2$ resolution. Using the BTFT raster, both the tree canopy and CDM rasters are bifurcated into new variables representing evergreen and deciduous canopy and CDM.

2.2.2.3. Urban Heat and Air Pollution Measurements and Models

Heat measurements are borrowed from Voelkel and Shandas (2017). These vector point data were collected during an extreme heat event on August 25, 2014 within the City of Portland. Collection was performed with 6 vehicle-mounted thermocouples over three separate time periods (6am, 3pm, and 7pm) and temperature data was appended to GPS positional data tracking each individual vehicle. The models created by Voelkel and Shandas (2017) are used to compare the effects of evergreen and deciduous trees. Mobile-source Air Pollution – using Nitrogen Dioxide (NO_2) as a cursor for other combustion related aerosols and particulates – was obtained and used with permission from the authors of Rao et al. (2014) (eq. 1). This Portland-based analysis used 144 in-

situ sampling devices around the Portland Metropolitan Area to analyze the distribution of NO_2 through the region. The authors sampled data for the summer of 2013, with a total of 88 sampling sites within Portland proper. Included with this data are the results of a multi-distance analysis — this includes information for each point on characteristics such as annual average daily traffic (AADT), total freeway length, and total railyard area within multiple distances. The authors also included a measure of spatial distribution as a longitudinal distance from Portland's Central Business District.

$$
\begin{aligned}
NO_2 = {} & 7.7 + 1.1e^{-8} * FWY\_AADT_{1200} \\
& + 6.5e^{-4} * MAJ\_ART_{500} \\
& + 1.7e^{-3} * ARTERIES_{350} \\
& + 1.8e^{-8} * STREETS(POP)_{800} \\
& + 1.0e^{-3} * RAILS_{250} \\
& - 1.0e^{-2} * ELEVATION \\
& + 1.4e^{-5} * ELEVATION^2 \\
& - 5.73e^{-6} * TREES_{400} \\
& + 1.1e^{-4} * X\_DIST
\end{aligned}
\qquad (2)
$$

Equation from Rao et al. (2014). Used with permission. Adj R^2 = 0.80; RMSE = 2.2 ppb.

where:

NO2 = NO2 ppb
FWY_AADT$_{1200}$ = freeway (m) in 1200 m, weighted with AADT
MAJ_ART$_{500}$ = major arteries (m) in 500 m
ARTERIES$_{350}$ = arteries (m) in 350 m
STREETS(POP)$_{800}$ = streets (m) in 800 m, weighted by the population
RAILS$_{250}$ = railroads (m) in 250 m
ELEVATION = elevation (ft)
TREES$_{400}$ = tree cover (m^2) in 400 m
X_DIST = distance from center of city (in m), along E–W axis

2.2.3. Continuous Surface Land Use / Cover Regression (CS-LUR)

Standard land use regression (LUR) models employ a series of circular buffers around in-situ measurement sites in order to account for the multi-distance effects of land use/cover on the dependent variable in question (Clougherty et al., 2013; Ghimire et al., 2010; Henderson et al., 2007; Rao et al., 2014). These buffers are next used to calculated

statistics on different land use/cover variables that fall within them. A model is created from these multi-distance statistics to predict the sampled phenomena in question. Upon creating a LUR model, it is common to create dispersed points across the study area and, once again, calculate land use/cover within the same multi-distance buffers and predict the dependent variable – these predicted values can be interpolated to create a continuous surface of values. The method employed here follows a similar logical methodology, but forgoes the process of using vectorized buffers around each point – and, thusly, eliminates the error introduced by a final interpolation of dispersed points. CS-LUR uses a common GIS raster-based technique (a moving-window analysis) to create a new raster dataset for each land use/cover data set and buffer distance that is in the form of a continuous raster surface. This moving-window analysis result will henceforth be referred to as a focal buffer. As an example, take the creation of a 50m canopy mean focal buffer: a circular moving window (with a radius of 50m) will move through every pixel of the original canopy raster calculating the mean. Because the original canopy raster has values 1 and 0 for 'canopy' and 'not canopy' respectively, the resulting focal buffer will have cell values that represent the percent canopy cover within 50m for every pixel in the study area.

These focal buffers are created at 15 distances commonly found in LUR studies: 50m, 100m, 150m, 200m, 250m, 300m, 350m, 400m, 450m, 500m, 600m, 700m, 800m, 900m, and 1000m. By analyzing surrounding areas at multiple distances, it is possible to ameliorate the effects of spatial autocorrelation in the model (Rodriguez-Galiano, 2012). Altogether, 150 new raster datasets are created. Using the R software environment and the "raster" package (Hijmans, 2015), these rasters can be 'stacked' on top of each other

into a single datacube. Once stacked, geographic points with in-situ measurements (the dependent variables in question) are used to extract pixel values from all layers in our datacube. The result of this extraction in a table with a row for each in-situ measurement, a column for the dependent variable (taken from the points), and 150 additional columns representing the value of each focal buffer value. Using this table, a model is trained using the in-situ recorded variable as the dependent variable and the additional land use/cover focal buffer values as the independent variables. After an acceptable model is formed, the true power of this method comes into play. Due to the fact that the datacube represents a detailed characterization of each variable (in a multi-distance context) in the study area, dispersed points and interpolation are not required. The model is used to predict values using the entire datacube as a new dataset, thus creating a predictive surface covering the entire study area. This continuous surface land use regression method (CS-LUR) will be employed for assessing the role of evergreen and deciduous trees as mitigators of the urban heat island effect and vehicle-based pollution.

2.2.4. Urban Heat Island Effect Modeling

The initial step required for modeling is to combine the spatially-located temperature measurement points and focal buffer rasters into tabular form. This process assesses the value for each pixel in each raster in the datacube for every individual point. The resulting tabular data contains a row for each temperature observation, a column containing observed temperatures, and a column for pixel value at the respective observation for each of the 150 rasters in the datacube (figure 2.2).

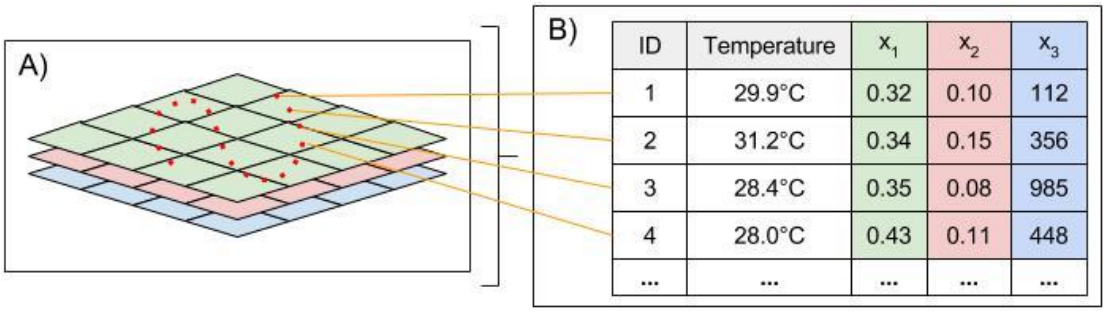

Figure 2.2. A: Temperature observation points are geographically situated with each layer on a simplified datacube; B: Extracted values are used to populate a table with an appropriate format for modeling.

For consistency with the previous UHI models, this analysis predicts temperatures using Random Forest machine learning. Random Forest is a non-parametric bootstrapped technique that 'grows' a large amount of classification and regression (CART) trees. The algorithm compares CART trees created with a subset of variables with a randomized tree (consisting of the same variables) — this process allows for a recording of variable importance based on increasing MSE values (Breiman, 1996; Liaw and Wiener, 2002). Random Forest has been found to ameliorate issues of spatial autocorrelation better than parametric linear regression models (Oliveira et al., 2012) and produces models which are more robust to noise (Buyantuyev and Wu, 2010; Dietterich, 2000). Though multicollinearity does not violate the assumptions of a Random Forest model (as it would a standard OLS regression), overfitting of the model is a risk. To assess potential overfitting, an additional cross-validation is performed on the model using a holdout method (70% of data for modeling, 30% for validation). Though the model itself performs internal cross-validation (noted as "out of bag error", or "OOB"), overfitting can be illuminated by comparing the OOB predictive power to that of the holdout method (Dormann et al., 2013). The number of CART trees is set to 500, with the number of variables randomly selected for each tree set to the default $k/3$ (where k is the number of

variables in the model). This limit of 500 trees is placed on the models for a few reasons: 1) to reduce computational burdens; 2) to balance computation time and model improvement; and 3) to reduce the chance of overfitting a model (thus leading to spurious results). After modeling is completed, spatial autocorrelation of model is assessed by comparing predicted and observed temperatures for *all* observations. In order to assess spatial autocorrelation, a neighborhood of 10 observations is determined for each observation using k-nearest neighbors and $\alpha = 0.01$.

2.2.5. Surface Prediction and BTFT Variable Effect

After models are trained and evaluated for each time period, a UHI raster surface is predicted. This process uses the 'column' of datacube pixels as new inputs into the model, which in turn calculates a new pixel with the predicted temperature value (figure 2.3). Temperature predictions are repeated for every pixel column in the study area — this results in the model being applied to over 3.35 billion pixel columns, each consisting of 150 individual pixels (equating to over 503.1 billion total pixels considered in the modeling).

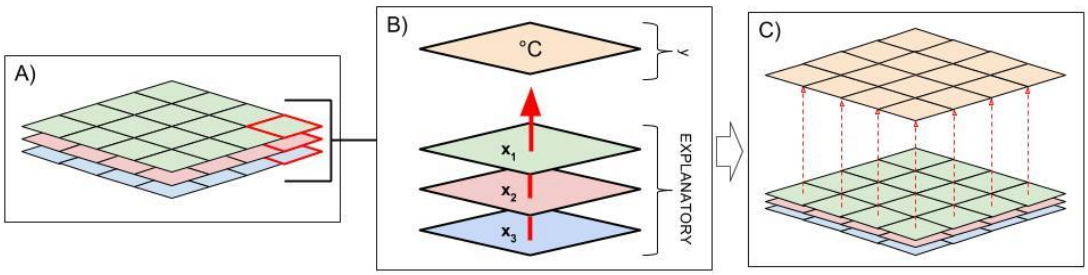

Figure 2.3. A: A simplified example of the datacube, with one column of cells on the far right highlighted in red; B: Each cell, representing a value of each explanatory variable, is run through the fitted random forest model (signified with a red arrow) to create a predicted temperature; C: All cell columns are calculated, and their predictions form a continuous raster surface.

In order to determine the total influence of BTFT on temperatures, extra steps must be taken to convert the predicted surface temperature to the reduction/increase in temperature as a result of BTFT. This is done by calculating a second UHI surface on an altered datacube: all canopy-related rasters have their cell values replaced with 0. By applying the original random forest model to this altered datacube, temperature predictions are changed based on the predicted effect of canopy in that specific column. Next, the original UHI surface is subtracted from the altered UHI, resulting in a third and final raster. This "variable effect" raster's values represent the effect (measured in degrees Celsius) that trees have on temperatures on a cell-by-cell basis across the study area. Two different versions of the variable effect raster are created for each time period to assess the role of 1) canopy cover and CDM for evergreen trees only; and 2) canopy cover and CDM for deciduous trees only. Pixel values for these variable effect rasters are assessed to understand the role of BTFT as a mitigator of the urban heat island effect.

2.2.6. Mobile-source Air Pollution Modeling

The published model in Rao et al. (2014) used simple ordinary least squares regression (OLS) to predict NO_2 values at each sampling site. For this analysis, OLS is used to assess changes in model performance as a result of the addition of the aforementioned BTFT variables (replacing the original "canopy cover" independent variable). Input measurements are restricted to those that fall within in the City of Portland (n=88). NO_2 observation points are combined with the datacube in a similar manner as in 2.4; however, only BTFT variables are added to the tabular output in order to maintain a high level of consistency with the original NO_2 model. Additionally, all

variables in the original study from Rao et al. (2014) were provided by the authors and included in the tabular output in order to hold all variables except for BTFT constant between studies. A combination of BTFT variables are used in the model to assess significance and effect of their inclusion. Though only two variables are addressed ("evergreen" and "deciduous" canopy), 30 total variables are examined when factoring in the effective distances. Because the effect of distance is known to be highly variable in land use regressions (Voelkel et al., 2016; Voelkel and Shandas, 2017), this modeling exercise will select an effective distance with the highest contribution to the overall model base on the p-values of coefficient estimates.

2.3. Results

2.3.1. Urban Heat Island Effect CS-LUR

Table 2.2. Random forest modeling results for morning, afternoon, and evening observations.

Model	**Pseudo $R^{2\dagger}$**	**CV $R^{2\dagger\dagger}$**	**RMSE (°C)**	**Voelkel and Shandas (2017) CV $R^{2\dagger\dagger\dagger}$**	**Top 5 Important Variables**	**Distance (m)**	**%IncMSE**
6am	0.9766	0.9794	0.153	0.9793	Vegetation Cover	50	34.03
					Vegetation Cover	1000	27.90
					Vegetation Cover	900	22.91
					Sum of Deciduous CDM	1000	22.38
					Total Building Volume	1000	21.82
3pm	0.8377	0.8177	0.480	0.8199	Standard Dev. of Building Heights	100	26.80
					Deciduous Canopy Cover	100	23.86
					Standard Dev. of Building Heights	1000	23.35
					Standard Dev. of Building Heights	450	23.09
					Vegetation Cover	50	22.52

7pm	0.9536	0.9587	0.202	0.9715	Standard Dev. of Building Heights	1000	23.17
					Standard Dev. of Building Heights	900	20.51
					Vegetation Cover	100	20.10
					Sum of Evergreen CDM	1000	19.25
					Sum of Evergreen CDM	900	19.01

[†] Directly measured by the random forest algorithm.
[††] Observed from the 70/30 holdout cross-validation. This is the R^2 used in discussion of model fit.
[†††] It is important to note that Voelkel and Shandas (2017) used 1000 trees in their modeling, thus the likelihood of these values being artificially inflated is higher than those of the current study.

The CS-LUR models produced results with high explanatory power for each of the three temperature observation periods (table 2.2). In table 2.2, "%IncMSE" represents the average model MSE change when forcing randomization during the individual tree growth process versus allowing the stated variable to naturally partition data. This is a direct measure of a variable's influence in the model, similar to a beta weight in a linear regression model (Hastie et al., 2009). For the sake of parsimony, I have only included the top 5 most influential/important variables in table 2.2, though it should be noted that all variables were used to determine the model fit. The morning model explained 97.94% of the variation in temperatures across the study area (RMSE = 0.153°C), with the sum of deciduous CDM within 1000m (interpreted best as an index for total biomass of deciduous canopy within broad region) as the fourth most important variable in the model. The afternoon model was able to explain 81.77% of the variation in temperatures (RMSE = 0.480°C), and noted percent deciduous canopy cover within 100m as the second most important variable. Finally, the evening model explained 95.87% of the variation in temperature across the study area (RMSE = 0.202°C), and placed the sum of

evergreen CDM within both 1000m and 900m as the fourth and fifth most important variables respectively.

2.3.1.1. Spatial Autocorrelation of UHI Model Error

For all UHI models, there was a failure to reject the null hypothesis in a test of spatial autocorrelation of the residuals of the full-observation predicted/observed regression (table 2.3). Because of this, it is concluded that there is likely no detrimental spatial autocorrelation in the model.

Table 2.3. Spatial Autocorrelation (Moran's I) results for each UHI model.

UHI Model	**Moran's I**	**p-value**[†]
6am	0.0026375	0.213
3pm	0.0078062	0.023*
7pm	0.0038708	0.143

[†] $\alpha = 0.01$

2.3.1.2. BTFT Variable Effect

As the six resulting rasters from this analysis contained approximately 837 million pixels each, a sample of one million values was taken at random (without substitution) from each. These values were next tested against each other in pairs according to model time to determine a difference in means using a Welch two-sample t-test. In all cases, the models in which evergreen canopy metrics were removed resulted in higher mean values. This indicates that, within the City of Portland, evergreen trees are reducing temperatures more than deciduous trees.

2.3.2. Mobile-source Air Pollution Modeling

The NO_2 OLS model found that the most significant BTFT variable was percent deciduous cover within 150m of a location (table 2.4). Performance of the model increased compared to Rao et al. (2014)'s model with an R^2 of 0.8305 and an RMSE of 1.98ppb.

Table 2.4. Regression results from the addition of a BTFT variable.

Coefficient	Estimate	Beta Weight	VIF	Std. Error	t	p-value
Intercept	1.070e+01	--	--	5.782e-01	18.503	< 2e-16
FWY_AADT_{1200}	1.052e-08	0.367	1.544	1.366e-09	7.703	7.45e-12
MAJ_ART_{500}	5.777e-04	0.143	1.723	2.032e-04	2.843	0.00537
$ARTERIES_{350}$	1.568e-03	0.169	1.114	3.766e-04	4.163	6.42e-05
$STREETS(POP)_{300}$	1.830e-08	0.253	1.307	3.179e-09	5.757	8.43e-08
$RAILS_{250}$	9.315e-04	0.110	1.560	4.060e-04	2.295	0.02373
ELEVATION	-1.159e-02	-0.521	3.287	1.548e-03	-7.490	2.16e-11
$ELEVATION^2$	1.426e-05	0.306	2.737	2.956e-06	4.825	4.71e-06
X_DIST	1.091e-04	0.296	1.065	1.463e-05	7.460	2.50e-11
$DECIDUOUS_{150}$†	-4.429e-05	-0.098	1.192	1.892e-05	-2.341	0.02110

Adjusted R^2: 0.8305; RMSE: 1.98ppb

†Where $DECIDUOUS_{150}$ is the area (in m^2) of deciduous canopy cover within 150m. This is calculated by multiplying the original data (percent deciduous canopy cover within 150m) by $(150)^2$. The translation to an areal measurement of tree cover is performed in order to keep consistency with the methods of Rao et al. (2014) and is otherwise unnecessary.

2.3.2.1. Spatial Autocorrelation of NO_2 Model Error

The Moran's I test for spatial autocorrelation resulted in a failure to reject the null hypothesis and a conclusion that there is no spatial autocorrelation of the NO_2 model residuals (table 2.5; figure 2.4).

Table 2.5. Spatial Autocorrelation (Moran's I) results for each UHI model.

Model	Moran's I	p-value†
Summer NO_2	0.046	0.071

†$\alpha = 0.01$

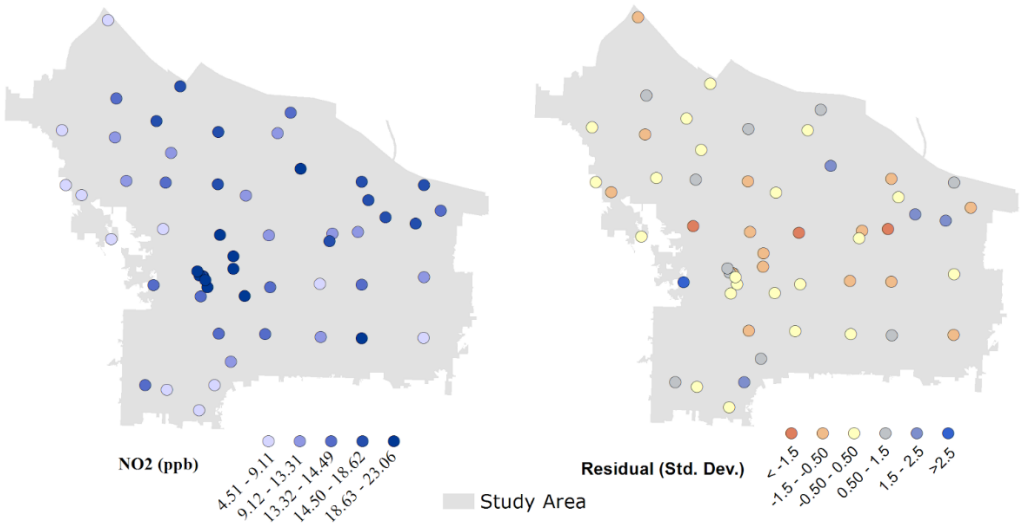

Figure 2.4. Left: Observed NO2 values in the study area; Right: Standard deviation of model residuals.

2.4. Discussion

2.4.1. BTFT and the Urban Heat Island Effect

The first hypothesis for the effect of BTFT on the urban heat island effect was that the introduction of such variables would provide models with better performance than overall canopy metrics. An adjustment was made to this current study in the form of 'pruning' — the number of generated trees was reduced in order to mitigate any inflated metrics stemming from an over-fitted model. This pruning was performed by using only 500 trees to build the random forest model — half the amount used in Voelkel and Shandas (2017). Even with this reduction of trees, similar cross-validation R^2 values are observed. Because of the high similarity in these values, it is concluded that the replacement of overall canopy metrics with BTFT variables results in a more accurate and better performing model. Future work should re-compute the models in Voelkel and Shandas (2017) using 500 training trees, making a comparison more practical.

The second hypothesis predicted that evergreen tree canopy would have a greater effect on temperatures than deciduous trees. The variable effect analysis confirms this hypothesis for all three model time periods. Additionally, evergreen trees account for only 35.55% of canopy cover and 43.48% of total CDM within the City of Portland. This further indicates that, as evergreen trees are cooling the study area more than deciduous – and with less representation in total canopy – evergreen trees are more effective at mitigating the urban heat island effect.

2.4.1.1. Reflections on the Use of CS-LUR

The use of the CS-LUR method allows for the creation of UHI surface rasters without the need for interpolation. This method – though convenient for surface creation and capable of high-explanatory power results – has drawbacks which must be addressed. First, the actual size of the data can impose massive limitations: due to the 1m resolution and large study area size, the focal buffer rasters comprising the datacube used in this study totaled 2.056TB. This large size leads to a second limitation: processing power. The CS-LUR models in this study were calculated on a UNIX computation server with 16 cores and 757GB of available RAM. Even when taking advantage of all 16 processor cores and leveraging as much RAM as possible, the CS-LUR process can take several days to run. This is a major limitation for cities or other local governments wishing to run such models, as computational resources are likely to be limited in comparison to a university.

2.4.2. BTFT and Mobile-Source Air Pollution

Unlike the UHI CS-LUR results, a rejection of the hypothesis is determined for the NO_2 model: instead of the hypothesized large influence of evergreen tree presence in reducing NO_2, it is observed that deciduous canopy cover within 150m from a location has the greatest effect. The replacement of a non-BTFT canopy metric in the model resulted in an increase in R^2 of 0.0305. Based on this updated model, an increase in deciduous tree cover of $1m^2$ within 150m of a location will result in an NO_2 reduction of 4.429e-5 ppb. Though this influence is small, it is greater than that of non-BTFT from Rao et al. (2014)'s model, which determined that an increase in canopy cover for any tree type of $1m^2$ within 400m of a location will result in a reduction of 5.73e-6 ppb. In addition, RMSE improved by 0.22ppb over the original model.

2.5. Conclusion

The era of centralized 'nature' in cities is beginning to fade from view. With a multitude of studies on the specific role that distributed urban tree canopy plays in reducing harm to human populations, cities must focus on distributing canopy throughout their jurisdictions. As urban foresters and planners target areas to increase canopy cover they must consider more than a simple metric such as 'number of trees' or 'canopy cover' alone: they must consider that the type of tree they are planting will have a specific effect that likely differs from another type. This study has shown that deciduous trees — at a localized scale — are likely better at reducing mobile-source air pollutant (namely NO_2), yet evergreen trees are better at mitigating the urban heat island effect. In order to apply this research to any recommendations on planting or policy, data on temperatures and NO_2 must be collected for multiple seasons. This is primarily due to the fact that

deciduous trees lose their leaves for a large portion of the year, and the effect of this on mitigating air pollution will likely change.

Final Conclusions

The preceding analyses provide examples of the application of geocomputation. Chapter 1 combined several data sources to create a spatial dataset describing energy consumption and urban form that is unparalleled in similar studies. Chapter 2 applied a common environmental science analytical approach – Land Use Regression – with multiple terabytes of land use/cover data, creating high-performing machine learning models capable of describing the effects of tree functional types as they pertain to urban heat island mitigation. Though components of each study mirror those found in the literature, they rely on custom analyses that could only be made possible through the leveraging of a large computing infrastructure. In doing so, these analyses we forced to move outside of standard GIS software and rely on statistical computing languages in order to process in a timely fashion (or, rather, at all). The ability of these analyses to provide both high predictive and explanatory power highlights a key component of geocomputation: by incorporating large volumes of difficult-to-integrate datasets – and not being pigeon-holed into asking only those questions which can be answered with available software – higher accuracy answers to research questions can be gleaned.

Geocomputation is not an answer to all spatially-based questions, however. Firstly, the size of the datasets employed in these analyses are difficult to work with. The need for large disk spaces and large RAM is a roadblock to many researchers and practitioners, as the costs associated with such computers is far higher than with a standard desktop computer. Secondly, geocomputational approaches to problems are likely inappropriate for answering questions with a time-sensitive nature. Due to the noted size of the datasets used in the preceding analyses, the time devoted to data

cleaning, normalization, and compilation totaled many months of dedicated time. In an emergency (i.e. analyses in response to a natural disaster, where human lives are at stake), analyses like those performed here would be inappropriate. Lastly, data availability is a major limiting factor. Chapter 1 relied on records of energy consumption which are often unavailable to researchers – in fact, it contained almost 200 times as many energy observations than most other studies. With a limited number of energy consumption observations, this analysis does not require a geocomputational approach. Chapter 2 leveraged highly-detailed $1m^2$ resolution information on land uses and land covers. Again, this level of detail is often unavailable to researchers and is highly dependent on the study area.

Though the drawbacks are clear for wide-scale adoptions of geocomputation in research, the studies performed here show that – when appropriate – geocomputation allows highly accurate and informed answers to be gained from research questions. For cities, where landscapes are dynamically changing due in part to planning practices, geocomputation is highly applicable. By utilizing a wide range of data within cities, it is possible to produce accurate models of spatial phenomenon which can be used to inform decisions which alter the built landscape. Most importantly, the analyses in this study show that it is important to conduct city- and region-wide analyses at highly granular resolutions to produce predictions and descriptions.

References

Akbari, H. (2002). Shade trees reduce building energy use and CO2 emissions from power plants. *Environmental Pollution*, *116, Supplement 1*, S119–S126. https://doi.org/10.1016/S0269-7491(01)00264-0

Akbari, H., Pomerantz, M., & Taha, H. (2001). Cool surfaces and shade trees to reduce energy use and improve air quality in urban areas. *Solar Energy*, *70*(3), 295–310. https://doi.org/10.1016/S0038-092X(00)00089-X

Armson, D., Stringer, P., & Ennos, A. R. (2012). The effect of tree shade and grass on surface and globe temperatures in an urban area. *Urban Forestry & Urban Greening*, *11*(3), 245–255. https://doi.org/10.1016/j.ufug.2012.05.002

Baldinelli, G., & Bonafoni, S. (2015). Analysis of Albedo Influence on Surface Urban Heat Island by Spaceborne Detection and Airborne Thermography. In V. Murino, E. Puppo, D. Sona, M. Cristani, & C. Sansone (Eds.), *New Trends in Image Analysis and Processing -- ICIAP 2015 Workshops* (pp. 95–102). Springer International Publishing. Retrieved from http://link.springer.com/chapter/10.1007/978-3-319-23222-5_12

Bealey, W. J., McDonald, A. G., Nemitz, E., Donovan, R., Dragosits, U., Duffy, T. R., & Fowler, D. (2007). Estimating the reduction of urban PM10 concentrations by trees within an environmental information system for planners. *Journal of Environmental Management*, *85*(1), 44–58. https://doi.org/10.1016/j.jenvman.2006.07.007

Bodnaruk, E. W., Kroll, C. N., Yang, Y., Hirabayashi, S., Nowak, D. J., & Endreny, T. A. (2017). Where to plant urban trees? A spatially explicit methodology to explore ecosystem service tradeoffs. *Landscape and Urban Planning*, *157*, 457–467. https://doi.org/10.1016/j.landurbplan.2016.08.016

Bolund, P., & Hunhammar, S. (1999). Ecosystem services in urban areas. *Ecological Economics*, *29*(2), 293–301. https://doi.org/10.1016/S0921-8009(99)00013-0

Borden, K. A., & Cutter, S. L. (2008). Spatial patterns of natural hazards mortality in the United States. *International Journal of Health Geographics*, *7*, 64. https://doi.org/10.1186/1476-072X-7-64

Brack, C. L. (2002). Pollution mitigation and carbon sequestration by an urban forest. *Environmental Pollution*, *116*(Supplement 1), S195–S200. https://doi.org/10.1016/S0269-7491(01)00251-2

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

Buyantuyev, A., & Wu, J. (2010). Urban heat islands and landscape heterogeneity: linking spatiotemporal variations in surface temperatures to land-cover and socioeconomic patterns. *Landscape Ecology*, *25*(1), 17–33. https://doi.org/10.1007/s10980-009-9402-4

Cao, X., Onishi, A., Chen, J., & Imura, H. (2010). Quantifying the cool island intensity of urban parks using ASTER and IKONOS data. *Landscape and Urban Planning*, *96*(4), 224–231. https://doi.org/10.1016/j.landurbplan.2010.03.008

Carver, A. D., Unger, D. R., & Parks, C. L. (2004). Modeling Energy Savings from Urban Shade Trees: An Assessment of the CITYgreen® Energy Conservation

Module. *Environmental Management*, *34*(5), 650–655. https://doi.org/10.1007/s00267-002-7003-y

Chumney, E. C. G., Simpson, K. N., & American Society of Health-System Pharmacists (Eds.). (2006). *Methods and designs for outcomes research*. Bethesda, MD: American Society of Health-System Pharmacists.

Cliff, A. D., Ord, J. K., & Cliff, A. D. (1981). *Spatial processes: models & applications*. London: Pion.

Clougherty, J. E., Kheirbek, I., Eisl, H. M., Ross, Z., Pezeshki, G., Gorczynski, J. E., … Matte, T. (2013). Intra-urban spatial variability in wintertime street-level concentrations of multiple combustion-related air pollutants: The New York City Community Air Survey (NYCCAS). *Journal of Exposure Science and Environmental Epidemiology*, *23*(3), 232–240. https://doi.org/10.1038/jes.2012.125

Creutzig, F., Baiocchi, G., Bierkandt, R., Pichler, P.-P., & Seto, K. C. (2015). Global typology of urban energy use and potentials for an urbanization mitigation wedge. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(20), 6283–6288. https://doi.org/10.1073/pnas.1315545112

Dietterich, T. G. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Machine Learning*, *40*(2), 139–157. https://doi.org/10.1023/A:1007607513941

Donovan, G. H., & Butry, D. T. (2009). The value of shade: Estimating the effect of urban trees on summertime electricity use. *Energy and Buildings*, *41*(6), 662–668. https://doi.org/10.1016/j.enbuild.2009.01.002

Donovan, G. H., Michael, Y. L., Butry, D. T., Sullivan, A. D., & Chase, J. M. (2011). Urban trees and the risk of poor birth outcomes. *Health & Place*, *17*(1), 390–393. https://doi.org/10.1016/j.healthplace.2010.11.004

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., … Lautenbach, S. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x

Ellison, D., Morris, C. E., Locatelli, B., Sheil, D., Cohen, J., Murdiyarso, D., … Sullivan, C. A. (2017). Trees, forests and water: Cool insights for a hot world. *Global Environmental Change*, *43*, 51–61. https://doi.org/10.1016/j.gloenvcha.2017.01.002

Engels, F. (1845). Introduction; The Great Towns,; Results. In *The Condition of the Working Class in England* (Panther Edition, 1969). Moscow: Otto Wigand. Retrieved from https://www.marxists.org/archive/marx/works/1845/condition-working-class/

Ewing, R., & Rong, F. (2008). The impact of urban form on U.S. residential energy use. *Housing Policy Debate*, *19*(1), 1–30. https://doi.org/10.1080/10511482.2008.9521624

Fahmy, M., Sharples, S., & Eltrapolsi, A. (2009). Dual Stage Simulations to Study the Microclimatic Effects of Trees on Thermal Comfort in a Residential Building,

Cairo, Egypt. Presented at the Eleventh International IBPSA Conference, Glasgow, Scotland.
Foster, J., Lowe, A., & Winkelman, S. (2011). *The value of green infrastructure for urban climate adaptation.* (No. 750). Center for Clean Air Policy.
Gahegan, M. (2017). GeoComputation. Retrieved November 20, 2017, from http://www.geocomputation.org/what.html
Gandy, M. (2006). Urban nature and the ecological imaginary. In *In the nature of cities: Urban political ecology and the politics of urban metabolism* (pp. 63–74).
Ghimire, B., Rogan, J., & Miller, J. (2010). Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic. *Remote Sensing Letters*, *1*(1), 45–54. https://doi.org/10.1080/01431160903252327
Gill, S. ., Handley, J. ., Ennos, A. ., & Pauleit, S. (2007). Adapting Cities for Climate Change: The Role of the Green Infrastructure. *Built Environment*, *33*(1), 115–133. https://doi.org/10.2148/benv.33.1.115
Hart, M. A., & Sailor, D. J. (2008). Quantifying the influence of land-use and surface characteristics on spatial variability in the urban heat island. *Theoretical and Applied Climatology*, *95*(3–4), 397–406. https://doi.org/10.1007/s00704-008-0017-5
Hassid, S., Santamouris, M., Papanikolaou, N., Linardi, A., Klitsikas, N., Georgakis, C., & Assimakopoulos, D. N. (2000). The effect of the Athens heat island on air conditioning load. *Energy and Buildings*, *32*(2), 131–141. https://doi.org/10.1016/S0378-7788(99)00045-6
Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (2nd ed). New York, NY: Springer.
Henderson, S. B., Beckerman, B., Jerrett, M., & Brauer, M. (2007). Application of Land Use Regression to Estimate Long-Term Concentrations of Traffic-Related Nitrogen Oxides and Fine Particulate Matter. *Environmental Science & Technology*, *41*(7), 2422–2428. https://doi.org/10.1021/es0606780
Henry, J. A., & Dicks, S. E. (1987). Association of urban temperatures with land use and surface materials. *Landscape and Urban Planning*, *14*, 21–29. https://doi.org/10.1016/0169-2046(87)90003-X
Hijmans, R. J. (2015). *raster: Geographic Data Analysis and Modeling*. Retrieved from http://CRAN.R-project.org/package=raster
Hirabayashi, S., Kroll, C. N., & Nowak, D. J. (2012). *i-Tree eco dry deposition model descriptions*. itreetools. Retrieved from http://www.itreetools.org/eco/resources/iTree_Eco_Dry_Deposition_Model_Descriptions_v1.0.pdf
Howard, E. (1902). *Garden Cities of Tomorrow*.
Huang, G., Zhou, W., & Cadenasso, M. L. (2011). Is everyone hot in the city? Spatial pattern of land surface temperatures, land cover and neighborhood socioeconomic characteristics in Baltimore, MD. *Journal of Environmental Management*, *92*(7), 1753–1759. https://doi.org/10.1016/j.jenvman.2011.02.006
Kalnay, E., & Cai, M. (2003). Impact of urbanization and land-use change on climate. *Nature*, *423*(6939), 528–531. https://doi.org/10.1038/nature01675

Ko, Y. (2013). Urban Form and Residential Energy Use: A Review of Design Principles and Research Findings. *Journal of Planning Literature*, *28*(4), 327–351. https://doi.org

Legendre, P. (1993). Spatial Autocorrelation: Trouble or New Paradigm? *Ecology*, *74*(6), 1659–1673. https://doi.org/10.2307/1939924

Li, F., Wang, R., Paulussen, J., & Liu, X. (2005). Comprehensive concept planning of urban greening based on ecological principles: a case study in Beijing, China. *Landscape and Urban Planning*, *72*(4), 325–336. https://doi.org/10.1016/j.landurbplan.2004.04.002

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22.

Lin, B.-S., & Lin, Y.-J. (2010). Cooling Effect of Shade Trees with Different Characteristics in a Subtropical Urban Park. *HortScience*, *45*(1), 83–86.

Mattern, J., Garrigan, S., & Kennedy IV, S. B. (2000). A Community-Based Assessment of Heat-Related Morbidity in North Philadelphia. *Environmental Research*, *83*(3), 338–342. https://doi.org/10.1006/enrs.2000.4067

McPherson, E. G., & Simpson, J. R. (2003). Potential energy savings in buildings by an urban tree planting programme in California. *Urban Forestry & Urban Greening*, *2*(2), 73–86. https://doi.org/10.1078/1618-8667-00025

Meehl, G. A., & Tebaldi, C. (2004). More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century. *Science*, *305*(5686), 994–997. https://doi.org/10.1126/science.1098704

Metro Data Resource Center. (2016, October 30). Regional Land Information System (RLIS). Retrieved from 2015-10-30

Metro Data Resource Center. (2017, October 30). Regional Land Information System (RLIS). Retrieved from 2015-10-30

Montgomery, D. C., Peck, E. A., & Vining, G. G. (2015). *Introduction to Linear Regression Analysis*. John Wiley & Sons.

Nowak, D. J., Crane, D. E., & Stevens, J. C. (2006). Air pollution removal by urban trees and shrubs in the United States. *Urban Forestry & Urban Greening*, *4*(3–4), 115–123. https://doi.org/10.1016/j.ufug.2006.01.007

Nowak, D. J., Hoehn III, R. E., Crane, D. E., Stevens, J. C., & Fisher, C. L. (2009). *Chicago's Urban Forest* (Assessing Urban Forest Effects and Values). USDA Forest Service.

Oke, T. R. (1982). The energetic basis of the urban heat island. *Quarterly Journal of the Royal Meteorological Society*, *108*(455), 1–24. https://doi.org/10.1002/qj.49710845502

Oliveira, S., Oehler, F., San-Miguel-Ayanz, J., Camia, A., & Pereira, J. M. C. (2012). Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest. *Forest Ecology and Management*, *275*, 117–129. https://doi.org/10.1016/j.foreco.2012.03.003

Olmsted, F. L. (2011). Public Parks and the Enlargement of Towns. In R. T. LeGates & F. Stout (Eds.), *The City Reader* (5th ed, pp. 321–327). London ; New York: Routledge.

Pachauri, R. K., Mayer, L., & Intergovernmental Panel on Climate Change (Eds.). (2015). *Climate change 2014: synthesis report*. Geneva, Switzerland: Intergovernmental Panel on Climate Change.

Pettersen, T. D. (1994). Variation of energy consumption in dwellings due to climate, building and inhabitants. *Energy and Buildings*, *21*(3), 209–218. https://doi.org/10.1016/0378-7788(94)90036-1

Posada, D., Buckley, T. R., & Thorne, J. (2004). Model Selection and Model Averaging in Phylogenetics: Advantages of Akaike Information Criterion and Bayesian Approaches Over Likelihood Ratio Tests. *Systematic Biology*, *53*(5), 793–808. https://doi.org/10.1080/10635150490522304

Poumadère, M., Mays, C., Le Mer, S., & Blong, R. (2005). The 2003 Heat Wave in France: Dangerous Climate Change Here and Now. *Risk Analysis*, *25*(6), 1483–1494. https://doi.org/10.1111/j.1539-6924.2005.00694.x

Rao, M., George, L. A., Rosenstiel, T. N., Shandas, V., & Dinno, A. (2014). Assessing the relationship among urban trees, nitrogen dioxide, and respiratory health. *Environmental Pollution*, *194*, 96–104. https://doi.org/10.1016/j.envpol.2014.07.011

Ratti, C., Baker, N., & Steemers, K. (2005). Energy consumption and urban texture. *Energy and Buildings*, *37*(7), 762–776. https://doi.org/10.1016/j.enbuild.2004.10.010

Rodriguez-Galiano, V. F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P. M., & Jeganathan, C. (2012). Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, *121*, 93–107. https://doi.org/10.1016/j.rse.2011.12.003

Santamouris, M., Cartalis, C., Synnefa, A., & Kolokotsa, D. (2015). On the impact of urban heat island and global warming on the power demand and electricity consumption of buildings—A review. *Energy and Buildings*, *98*(Supplement C), 119–124. https://doi.org/10.1016/j.enbuild.2014.09.052

Seto, K. C., Dhakal, S., Bigio, A., Blanco, H., Delgado, G. C., Dewar, D., … Ramaswami, A. (2014). Chapter 12 - Human settlements, infrastructure and spatial planning. In *Climate Change 2014: Mitigation of Climate Change. IPCC Working Group III Contribution to AR5*. Cambridge University Press. Retrieved from http://www.ipcc.ch/pdf/assessment-report/ar5/wg3/ipcc_wg3_ar5_chapter12.pdf

Simmel, G. (2010). The Metropolis and Mental Life. In G. Bridge & S. Watson (Eds.), *The Blackwell city reader* (2nd ed, pp. 11–19). Chichester, West Sussex, U.K. ; Malden, MA: Wiley-Blackwell.

Simpson, J. R., & McPherson, E. G. (1998). Simulation of tree shade impacts on residential energy use for space conditioning in Sacramento. *Atmospheric Environment*, *32*(1), 69–74. https://doi.org/10.1016/S1352-2310(97)00181-7

Solecki, W. D., Rosenzweig, C., Parshall, L., Pope, G., Clark, M., Cox, J., & Wiencke, M. (2005). Mitigation of the heat island effect in urban New Jersey. *Global*

Environmental Change Part B: Environmental Hazards, *6*(1), 39–49. https://doi.org/10.1016/j.hazards.2004.12.002

Sullivan, K. D. (1995). *July 1995 Heat Wave* (Natural Disaster Survey Report). Silver Spring, Maryland: U.S. Department of Commerce - National Oceanic and Atmospheric Administration. Retrieved from http://www.nws.noaa.gov/os/assessments/pdfs/heat95.pdf

Taha, H., Akbari, H., Rosenfeld, A., & Huang, J. (1988). Residential cooling loads and the urban heat island—the effects of albedo. *Building and Environment*, *23*(4), 271–283. https://doi.org/10.1016/0360-1323(88)90033-9

Tol, R. S. J. (2002). Estimates of the Damage Costs of Climate Change. Part 1: Benchmark Estimates. *Environmental and Resource Economics*, *21*(1), 47–73. https://doi.org/10.1023/A:1014500930521

Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, *32*(9), 1761–1768. https://doi.org/10.1016/j.energy.2006.11.010

Ulrich, R. S. (1984). View through a window may influence recovery from surgery. *Science (New York, N.Y.)*, *224*(4647), 420–421.

United Nations. (2015). *World Urbanization Prospects: The 2014 Revision* (No. (ST/ESA/SER.A/366)). Department of Economic and Social Affairs, Population Division. Retrieved from http://esa.un.org/unpd/wup/Publications/Files/WUP2014-Report.pdf

Voelkel, J., & Shandas, V. (2017). Towards Systematic Prediction of Urban Heat Islands: Grounding Measurements, Assessing Modeling Techniques. *Climate*, *5*(2), 41. https://doi.org/10.3390/cli5020041

Voelkel, J., Shandas, V., & Haggerty, B. (2016). Developing High-Resolution Descriptions of Urban Heat Islands: A Public Health Imperative. *Preventing Chronic Disease*, *13*. https://doi.org/10.5888/pcd13.160099

Wirth, L. (2011). Urbanism as a Way of Life. In R. T. LeGates & F. Stout (Eds.), *The City Reader* (5th ed, pp. 96–104). London ; New York: Routledge.

Wu, C., Xiao, Q., & McPherson, E. G. (2008). A method for locating potential tree-planting sites in urban areas: A case study of Los Angeles, USA. *Urban Forestry & Urban Greening*, *7*(2), 65–76. https://doi.org/10.1016/j.ufug.2008.01.002

Yang, X., Zhao, L., Bruse, M., & Meng, Q. (2012). An integrated simulation method for building energy performance assessment in urban environments. *Energy and Buildings*, *54*(Supplement C), 243–251. https://doi.org/10.1016/j.enbuild.2012.07.042

/10.1177/0885412213491499