AVOIDING A NEVER-ENDING HEAT WAVE IN OUR CITIES: A SPATIO-TEMPORAL ANALYSIS OF VEGETATION AND TEMPERATURE IN PHILADELPHIA

by

Kelly Wang

An Undergraduate Thesis Submitted as Part of the

WHARTON RESEARCH SCHOLARS

Faculty Advisor:

Shane T. Jensen

Professor in the Department of Statistics and Data Science

THE WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA

May 2024

TABLE OF CONTENTS

Avoiding A Never-Ending Heat Wave in Our Cities: A Spatio-Temporal Analysis of Vegetation and Temperature in Philadelphia

Kelly Wang

Abstract

As Earth continues to warm, urban centers are struggling to disperse the additional heat that they are experiencing. Cities are not prepared for record-breaking temperatures in part due to their lack of natural vegetation that contribute to the urban heat island (UHI) effect. This study analyzes the spatial and temporal trends in land surface temperatures (LST) and vegetation across Philadelphia, Pennsylvania. This city has greening initiatives that plan to add green spaces to the built environment, intended to help mitigate the increasing heat. However, only a few studies have examined how changes in vegetation in cities, whether increasing or decreasing, have impacted the surface temperatures in its surrounding area over time. This paper uses data from Landsat 8's satellite imagery program to calculate a normalized vegetation index (NDVI) and surface temperatures aggregated at the Census block group level to analyze their relationship in over 1000 neighborhoods of Philadelphia over the years 2013–2022. The results of this paper indicate that there exists a strong negative relationship between LST and NDVI spatially and temporally.[†]

Keywords:

Land Surface Temperature, Urban Heat Island, Green Space, Spatio-temporal, Philadelphia

†. Code Repository: https://github.com/k-wang-10/phl-temp-ndvi-thesis

1. Introduction

1.1. Motivation

During the Northern Hemisphere's meteorological summer (June to August) in 2023, people experienced some of the most intense heat waves on Earth. NASA reported record-breaking temperatures that were 2.1 degrees Fahrenheit (1.2° C) warmer than any average summer between 1951 and 1980. These unprecedented temperatures have exacerbated the frequency and severity of natural disasters like wild fires and hurricanes. The heat itself has led to deadly consequences, especially in larger cities. As society industrialized, urban centers were designed to be more compact and built vertically to maximize the number of people that could live in a set area. However, as time has passed, these cities' constituents are now facing significant ramifications from the foundations that have been constructed, one of which is the urban heat island (UHI) effect (Hibbard et al. 2017). This phenomenon occurs when land surface temperatures (LST) increase in an area due to its infrastructures containing concrete, asphalt, or other materials that trap in heat and release it to their immediate surrounding environment. Therefore, the lack of vegetation in an area filled with urban-built structures can intensify the UHI effect because water, which is high in heat retention, can not be evaporated away through the natural greenery.

In the city of Philadelphia, there has been a focus on initiatives that improve the city's vegetation conditions, maintain their quality, and grow the amount of green spaces in the city. The local government's Office of Sustainability partners with groups around the area to help plan projects that can improve the quality of life in every neighborhood, like preparing for the extreme heat.[1] For example, the Pennsylvania Horticultural Society's (PHS) Philadelphia LandCare project aims to develop the vacant lots in the city into green spaces.[2] Vacant lots are undeveloped property that usually contain empty urban structures, rather than vegetation.

1. City of Philadelphia - Office of Sustainability
2. PHS - Transforming Vacant Land

These initiatives have become more prevalent, but their effectiveness in mitigating UHI effects within the city have not been quantified. This paper provides models to quantify their impact by analyzing the association between the changes in land surface temperatures with changes in amounts of green space over several years in Philadelphia. Satellite imagery is used to calculate surface temperatures and a vegetation index. Their relationship has been extensively explored in different countries, especially in places with more tropical climates (Andronis et al. 2022). More recently, this technique of using satellite imagery to map out temperatures and vegetation has been used on more urban environments to quantify UHI effects in a city (Almeida, Teodoro, and Gonçalves 2021).

Specifically in Philadelphia, this study's focal city, Pearsall (2017) analyzed the spatial variation in green spaces with surface temperature on August 7, 2014 and mapped out which areas in the city were experiencing more extreme effects from the UHI effect on that one day. Pearsall (2017) demonstrated that while vegetation is associated with cooling an area, vacant lots contributed more in magnitude to an increase in land surface temperatures, suggesting the city to convert these vacant lots into green spaces rather than development that would generate tax revenues.

1.2. Contribution

The purpose of this paper is to extend Pearsall (2017)'s analysis by incorporating the temporal trends of land surface temperatures and green spaces into the spatial analysis. Therefore, the first contribution of this paper is the compilation of a dataset, using satellite remote sensing images, that quantifies changes in vegetation density and temperatures at a neighborhood level for Philadelphia's summers over the years 2013 to 2022. Adding a time component to the analysis is important because the development of green spaces are not expected to immediately mitigate the UHI effect. This paper explores these spatial-temporal trends by performing a cross-sectional and longitudinal study on NDVI and LST to discern the extent to which vegetation in an urban setting is effective in reducing increases in surface temperatures in the area.

Satellite imagery was processed to create summary statistics of Philadelphia's regions. This study defines the areas of its analysis to be at the Census block group level that was established in 2010, which allows this paper to compare changes between these $n = 1336$ block group neighborhoods. These statistics will be aggregated to map out the city-wide trends in vegetation and temperature during the summer of each year. Using the results from the exploratory data analysis, this paper will examine the strength and direction of the association between changes in vegetation and temperature over a span of years through a time series model.

The objective is to demonstrate that the results of this paper can help green space program planners in Philadelphia to more optimally choose their greening project locations such that they better mitigate UHI effects in the city over time. Philadelphia is an interesting location to analyze because it is home to over 1.6 million residents, making it the sixth most populous city in the United States, who come from different socioeconomic backgrounds. It is also a city with various neighborhoods that contain a variety of commercial, industrial, and residential land uses, so this paper could assist other urban centers in implementing similar models to locate regions in their areas that are vulnerable to UHI effects.

1.3. Literature Review

1.3.1. Environmental Impact on Humans

Studies about the effects of temperature on the physical human condition have shown that there are numerous health-related risks that arise from a heated environment (Luber and McGeehin 2008). Heat exhaustion is the most common heat related illness, and it could lead to heat strokes if it goes untreated. Another health consequence is the deterioration of the cardiovascular system. Kondo et al. (2018) found that the odds of death from heart failure was higher for people living in areas with low amounts of green space, which was linked to more extreme heat events. Their paper's main focus was on the effects of exposure to natural greenery in urban environments. Kondo et al. (2018) found improvements in mental health and reductions in stress when people were exposed to urban green spaces outdoors.

There is also extensive research about the economic implications of a heating world. Heat waves are disrupting human capital, from students not learning as effectively in classrooms (Park, Behrer, and Goodman 2021) to workers becoming less productive (Opper, Park, and Husted 2023). This impacts future economic mobility and growth as less people are going into post-secondary education, along side the growing importance in a bachelor's degree to doing well in the job market. Furthermore, the increased frequency of extreme heat events means people are going to turn on their air conditioners for longer periods of time. However, lower income households cannot always bear the high costs of paying their electricity bill, so they may choose to not turn on their air conditioners. This economic burden feeds back into health consequences since many of the most vulnerable are not able to protect themselves from heat-related illnesses (Luber and McGeehin 2008).

However, more people are beginning to focus on the upward trend of land surface temperatures. Notably, the emphasis has been around the increased temperatures in urban spaces, which has accelerated many of the negative impacts that people are facing from these heat waves. The significantly higher temperatures that are recorded in city environments as opposed to more rural areas has been termed as the urban heat island effect. With the world becoming more urbanized, the UHI effect intensifies because there is not enough natural vegetation in the surrounding areas to disperse the heat (Hibbard et al. 2017).

1.3.2. Technology for Measuring Temperature and Vegetation

Most research on UHI effects have to first decide on what methods and datasets are used to measure land surface temperatures. A widely-used dataset is the US Geological Survey's (USGS) Landsat Satellite Imagery.[3] With satellites that send back images available for analyses, it has helped improve measurements of land surface temperatures. Landsat is not the only option to calculate LST, but it has been shown to be more accurate and at a higher

3. USGS.gov

resolution than the Moderate Resolution Imaging Spectroradiometer (MODIS) when Wang et al. (2016) ran their LST calculation algorithms on Landsat data in an arid region of Northwest China that is composed of sandy deserts and a natural oasis with mixed forests. Even within the Landsat dataset, there are different tiers of satellite images that have been processed differently for distinct use cases. As of now, Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) Level 1 Collection 2 images have the best quality and accuracy for analysis, as opposed to the older Collection 1 or the Level 2-processed images, which is explained in Section 2.1.

LIDAR is a different type of instrument that instead uses laser light to map out the land. LIDAR has a higher resolution of 1 meter compared to Landsat's 30–100 meter resolution. Studies show that land surface temperatures calculated with LIDAR show more detail in an area and can more accurately identify different infrastructure types (Li 2021). LIDAR can also account for the problem of cloud cover, but with its high resolution, there is a computational cost to using LIDAR as opposed to Landsat. Studies have shown that even though LIDAR can be more accurate, Landsat's calculations of LST are still close approximations to the in situ data measurements (Chen et al. 2016). This paper aggregates the data into larger regions, so the finer spatial resolution of LIDAR is not necessary for the data compilation portion of this study. Because this paper is also running a longitudinal study, the remote sensing data's temporal resolutions are also important. However, LIDAR is inconsistent in the time it takes to capture images in the same location, often taking more than two years. Landsat, instead, captures about two images every month in the same area.

1.3.3. Cross-Sectional Studies

Research has shown that one of the highest predictors of land surface temperature is the presence of green vegetation. The majority of these studies have come from places in China and India (Almeida, Teodoro, and Gonçalves 2021), where there are higher population densities within very condensed urban regions. Almeida, Teodoro, and Gonçalves (2021) also show that most research on temperatures and vegetation have focused on more tropical or arid climates. The normalized difference vegetation index (NDVI) is a commonly-used in those studies to measure the health and density of green spaces, which can be calculated from Landsat data. NDVI has been found to be especially useful in more urban areas because it can account for areas covered by infrastructure or cement (Townshend et al. 1991). Most studies use NDVI because the index can not only detect the presence of vegetation, but also areas of water and human residency.

Incorporating vegetation into a on study heat is important in analyzing urban environments. Compared to more tropical or more arid locations, cities are usually between the two climates in terms of vegetation amounts. However, the UHI effect emerges because the majority composition of a city's landscape is human-made. A study measured soil temperatures to identify the UHI effect, and it found that there was significant decrease in soil temperatures the further they moved out from the Leicester city's center (Edmondson et al. 2016). A study of over 100 prefectures in China showed with day and nighttime data that the intensity of UHIs decreased rapidly at night, particularly along the coastal zones and riverside buffers (Ming et al. 2023). Their results also noted that some of the effects that influence the intensity of urban heat islands are possibly nonlinear, like the level of compactness in the city.

1.3.4. Longitudinal Studies

The studies mentioned in Section 1.3.3 only analyzed the relationship between temperature and vegetation at a single point in time or a span of one year. However, now that the number of studies about increasing temperatures have grown, there are more papers that aim to model land surface temperatures over time (Sanecharoen et al. 2019). Creating a time series is valuable for future predictions, as shown through the spatial and temporal associations discovered between deforestation and heat islands in Mexico (Carrillo-Niquete et al. 2022). Carrillo-Niquete et al. (2022) found that 14% of deforestation was found in newly developed residential areas of a city, and they discovered statistically significant thermal difference between areas of deforestation versus without deforestation. Not only has temperature changed over the years, but so has vegetation. For example, a negative trend between NDVI and LST was reported from 2013–2018 in Cyprus' Paphos Forest, which covers an area of about 700 square kilometers (Andronis et al. 2022).

The techniques that have been implemented to model a time series include artificial neural networks (Sekertekin, Arslan, and Bilgili 2020) and types of regression models (Carrillo-Niquete et al. 2022). There is more to study in this area because as noted by Ming et al. (2023), there are some variables that have nonlinear effects on temperature, so a simple linear regression may not have the most predictive power. However, using AI powered models may make the models too complex and not interpretable for its use cases.

A study that monitored the dynamics of urban vegetation created an algorithm to detect if the changes in NDVI over the years were gradual or more abrupt jumps (Cortinovis, Haase, and Geneletti 2023). They found that abrupt changes in vegetation were captured in areas where land cover changed due to urban projects like the removal of vegetation or implementing new greening initiatives. Gradual changes were instead associated with natural vegetation growth or decline. Their findings indicate that policies targeting vegetation in urban areas can impact the NDVI, which consequently could affect the land surface temperatures residents experience in their city.

1.3.5. Studies in Philadelphia

Philadelphia has been the subject location of a few UHI/LST studies because it has highly urbanized regions along with stretches of parks, rivers, and more suburban life along the edges (Li, Chakraborty, and Wang 2023). Another aspect of Philadelphia that makes it an interesting city to study is the diversity of its residents within the city. However, Philadelphia's vast economic and demographic background is not evenly dispersed around the city, with people with similar backgrounds concentrated in the same areas. Pearsall (2017) notes that there are significantly higher demographic and economic disparities in the city's temperature hot spots. This correlation is not specific to Philadelphia because another study found that in general, workers from areas that are the most heavily impacted by the heat come from lower-income communities (Behrer et al. 2021).

Looking at Philadelphia from 1970 to 2010 through five Landsat images, urban land cover changes have been mostly stable, meaning areas that consist of impervious surfaces have mostly stayed impervious and areas of vegetation continue to have vegetation (Locke et al. 2023). However, Locke et al. (2023) found that it was more common for areas with trees and shrubbery to be converted to impervious cover with buildings than the other way around. The United States Department of Agriculture came out with a report in 2016 that analyzed tree canopy cover changes in Philadelphia from 2008 to 2012 (Nowak et al. 2016). The report also found little changes in land cover types over the years. In addition to analyzing changing land cover, Nowak et al. (2016) compared the environmental benefits of different tree characteristics (e.g., species, trunk diameters, etc.) to find that large diameter trees have a strong correlation with environmental benefits (e.g., pollution removal, carbon storage, etc.). However, the report notes that most of the tree canopy with these types of characteristics are located in Pennypack and Wissahickon Park (over 80% tree cover in each park), while other parts of the city, like North and South Philadelphia, only consist of less than 10% tree cover.

In Philadelphia, there are also approximately 40,000 vacant lots that have not been developed. It would be ideal to convert the vacant lots into green spaces, but commercial development is also competing for these pieces of land. The additional benefits of developing these vacant lots into green spaces include decreases in crime rates compared to before the implementation of a greening project (Cui, Jensen, and MacDonald 2022) and an increase in property values through the Philadelphia Landcare program that cleans vacant lots to convert them into green spaces (Heckert and Mennis 2012). However, there is a subtlety to the increase in house property values that was discovered in a later study (Lin, Jensen, and Wachter 2021). The median household income level and the percent of vacant lots within an area influence the effect of greening vacant lots have on house prices. The impact in increasing property value is most significant when a neighborhood's income is around

20-40% above the median and the vacant lot share is above 5%.

Even though there are green conversion initiatives that have been developed for the city, there has not been an analysis on the long term efficacy of such projects in mitigating rising temperatures. Also, as shown through the published literature about Philadelphia, socioeconomic factors determine locations where projects are placed, or more importantly, not placed. For these reasons, it is necessary to model land surface temperatures and green space changes over time to identify areas that can provide long term benefits to its constituents, mainly in reducing the urban heat island effect.

2. Data & Methodology

2.1. Landsat Data

This paper uses Landsat satellite imagery, which is a remote sensing dataset that is jointly managed by NASA and the US Geological Survey. Its two most recent satellites, Landsat 8 and 9, were launched in February 11, 2013 and September 27, 2021, respectively. These two satellites contain the Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) instruments that capture 11 different bandwidths of the electromagnetic spectrum emitted by the Earth's surface.[4] Bands 1–9 have a 30 meter resolution and bands 10–11 have a 100 meter resolution. USGS regularly updates its images to have better information, but its biggest update was when they converted all their images from Collection 1 to Collection 2, which is better calibrated. The next section details how certain bandwidths are processed into a temperature measure and the normalized difference vegetation index, and how LST and NDVI get aggregated at a Census block group level.

Landsat's satellites pass every point on Earth once every 16 days, which allows for about 2 images of Philadelphia per month. This analysis uses Landsat 8's summer 2013 to 2022 images as its sampled dataset. Note that this timeline follows the meteorological summers of the Northern Hemisphere (June to August) so that the most extreme heat events in Philadelphia can be captured every year. The focus is on the summer months because during Philadelphia's winters, all the vegetation will have withered. Therefore, any effect of vegetation on temperature would not be captured by the models. Satellite instruments can only measure the green hue from vegetation, which makes the winter images an ineffective dataset to analyze and distinguish areas of natural greenery.

Landsat categorizes its data into two levels of processing. Level 1 data are georectified to align the satellite image to a physical map of the location and calibrated in their radiances for all the detectors that might have drifted off from their tuned mark. Level 2 data are actually inferences made from the Level 1 data to estimate the geophysical state. Level 1's

4. USGS - Landsat Missions

dataset will be used for this paper's purposes because it allows this study to make its own spatio-temporal estimates by working as closely as possible to the raw images that were collected by the satellite.

2.2. Processing Landsat Variables

The Level 1 Collection 2 Landsat images need to be transformed into interpretable numerical values. The normalized vegetation index is calculated using Landsat's Band 4 and Band 5. The images are converted into pixel arrays where the following equation is calculated for each pixel at the same location.

$$\text{NDVI} = \frac{NIR - R}{NIR + R} = \frac{\text{Band 5} - \text{Band 4}}{\text{Band 5} + \text{Band 4}} \tag{2.2.1}$$

Temperatures can be calculated using Band 10 or 11, but it has been shown that Band 10 yields more accurate results (Wang et al. 2016). Note that the following formulas can be generalized to Band 11. Define the values directly from the Band 10's satellite image as T_{Band}, which can be transformed by a band-specific multiplicative factor M and additive factor A to get the top-of-atmosphere (TOA) radiance.

$$T_{\text{TOA-Rad}} = M \times T_{\text{Band}} + A \tag{2.2.2}$$

The TOA radiance subsequently needs to be converted into a top-of-atmosphere brightness temperature using the band specific thermal conversion values K_1 and K_2.

$$T_{\text{Bright}(^\circ\text{K})} = \frac{K_2}{\ln\left(\frac{K_1}{T_{\text{TOA-Rad}}} + 1\right)} \tag{2.2.3}$$

Land surface temperatures are calculated using the following formula.

$$\text{LST}_{(^\circ\text{K})} = \frac{T_{\text{Bright}(^\circ\text{K})}}{1 + \left(\lambda \times T_{\text{Bright}(^\circ\text{K})}/\rho\right) \times \ln(\varepsilon)} \tag{2.2.4}$$

The formula uses the TOA brightness temperature in degrees Kelvin with the constant $\rho = \frac{h \times c}{s}$ (where h = Planck's constant, c = velocity of light, s = Boltzmann constant), λ = wavelength of the emitted radiance from a given image, ε = land surface emissivity that depends on λ (calculated in Section 2.5). This paper will use LST in degrees Fahrenheit because it allows for a wider range of values than degrees Celsius (and Kelvin).

$$\mathrm{LST}_{(^\circ\mathrm{F})} = \frac{9}{5} \times (\mathrm{LST}_{(^\circ\mathrm{K})} - 273.15) + 32 \tag{2.2.5}$$

After getting the NDVI and LST values, the summary statistics of the data were computed at the Census block level in Philadelphia by aggregating the pixels within their respective block group region. This paper performs a preliminary analysis of the data to see if any noticeable trends can be identified that may need further examination (e.g., the distribution of summer NDVI and temperatures in Philadelphia). As mentioned before, neighborhood-level statistics provide an overview of the city-wide trends and its changes over the years. The summary statistics also assisted in a sensitivity analysis that was performed on the temperature changes when there is a change in vegetation values.

2.3. Incorporating Geographic Data

The defined geographic breakdown of Philadelphia used in this paper can be found on OpenDataPhilly or through the US Census' website. Census block group levels were chosen because they are small enough to separate Philadelphia into 1,336 distinguishable regions while also being a more manageable number to work with compared with over 10,000 Census blocks that combine into block groups. Using these block group regions, a spatial correlation factor was computed based on a neighborhood's borders and distance to other block groups. A queen-contiguity spatial weights matrix was created using the block groups shape file. This matrix determines an area's neighbors by labeling any region that touches the current area's border a 1, else the region is labeled as 0. This spatial breakdown allows for neighboring block groups to have some interdependence between one another.

2.4. Models

To analyze the various spatial-temporal relationships for NDVI and land surface temperatures, cross-sectional and longitudinal regression analyses were performed on the data. This paper initially analyzes NDVI alone by running a longitudinal Ordinary Least Squares (OLS) regression on the city-wide average NDVI with the years 2013 to 2022 (encoded with $t = 0, ..., 9$) as the explanatory variable to estimate the intercept (α) and slope (β) for time.

$$\text{NDVI}_t = \alpha + \beta \cdot \text{Year}_t \tag{2.4.1}$$

Then the averages were broken up into $n = 1336$ block group level mean NDVI values (calculated individually for each year) and a linear regression was modeled on all the values together against time to get intercepts and slopes for each neighborhood $i = 1, ..., n$.

$$\text{NDVI}_{it} = \alpha_i + \beta_i \cdot \text{Year}_{it} \tag{2.4.2}$$

After only considering NDVI trends in Philadelphia, temperature is added into our analysis. The first model that incorporates LST is a cross-sectional OLS regression that ignores temporal trends by using the 1,336 block-group mean temperature across the 10-year span as the response variable, with their respective mean NDVI as the covariate. Spatial versions of the cross-sectional model, like a Geographically Weighted Regression and Spatial Durbin Error Model, will also be tested on the neighborhoods to capture some of the spatial correlations between block groups that are near one another. The Geographically Weighted Regression (GWR) creates neighborhood-specific NDVI slopes by relating its estimates to its neighbors' slopes. An adaptive bandwidth kernel used the 46 nearest neighbors for its calculations. The Spatial Durbin Error Model (SDEM) incorporates the spatially weighted NDVI values (using the queen contiguity weight matrix W_{PHL} for Philadelphia) and spatial autocorrelations in the residuals as additional explanatory variables to the model to predict temperatures. The SDEM model is outlined in Equation 2.4.3.

$$\text{Temp}_i = \alpha + \beta_{\text{NDVI}} \cdot \text{NDVI}_i + \beta_{\text{W\_NDVI}} \cdot W_{\text{PHL}} \cdot \text{NDVI}_i + \lambda \cdot W_{\text{PHL}} \cdot \text{Residual}_i \tag{2.4.3}$$

Adding back time (in years) as a variable, the change in vegetation is the main explanatory variable to predict changes in the response variable, land surface temperatures, for a given year.

$$\text{Temp}_{it} = \alpha_i + \beta_i \cdot \text{Year}_{it} + \beta_{\text{NDVI}} \cdot \text{NDVI}_{it} \tag{2.4.4}$$

This time series model uses Philadelphia's neighborhood-level average values for NDVI and LST in neighborhood i in time t. The model returns estimates for neighborhood-level intercepts and time dependent slopes, while also returning a city-wide NDVI slope.

The models are expected to output a negative values for β_{NDVI}, which would suggest an inverse relationship between vegetation density and land surface temperatures. This is based on the idea that the urban heat island effect is amplified by the lack of vegetation. As mentioned in the literature review, Philadelphia has many greening projects that aim to improve the environment and increase the vegetation where their plans are implemented. The goal is to demonstrate and explain how vegetation trends and temperature trends are changing in relation to one another over time.

2.5. Limitations

There are constraints on what images can be used in the sampled dataset. For example, some days have significant cloud cover, which can lead to inaccurate calculations of LST and NDVI using the formulas from Section 2.2. Even if days with similar lower percentages of cloud cover are chosen, there could still be slight differences in how the clouds are formed and their shape such that it alters the calculations in differing ways. This paper identified 22 images of Philadelphia from 2013 to 2022 that had low enough cloud cover to not skew any of the calculations of NDVI and LST at the neighborhood-level.

Another limitation is the emissivity calculation. There is extensive literature on contrasting methods to calculate ε, but there is no definitive procedure that results in the best accuracy. Both of these concerns can be addressed by reviewing more extensively through the literature to see what the majority of remote sensing papers use. This paper uses

$$P_v = \left(\frac{\text{NDVI} - \text{NDVI}_{\text{min}}}{\text{NDVI}_{\text{max}} - \text{NDVI}_{\text{min}}} \right)^2 \quad \rightarrow \quad \varepsilon = 0.004 \times P_v + 0.986 \qquad (2.5.1)$$

as an approximation for emissivity, where P_v is the proportion of vegetation using NDVI.

3. Results

3.1. NDVI Analysis

From 2013 to 2022, Philadelphia's city-wide average summer NDVI shown in Figure 3.1 appears to demonstrate a slight decreasing trend with a slope of -6.87×10^{-4} NDVI/Year, though this yearly result was not significant. Over these ten years, Philadelphia's average NDVI was around 0.224 (including areas near water).

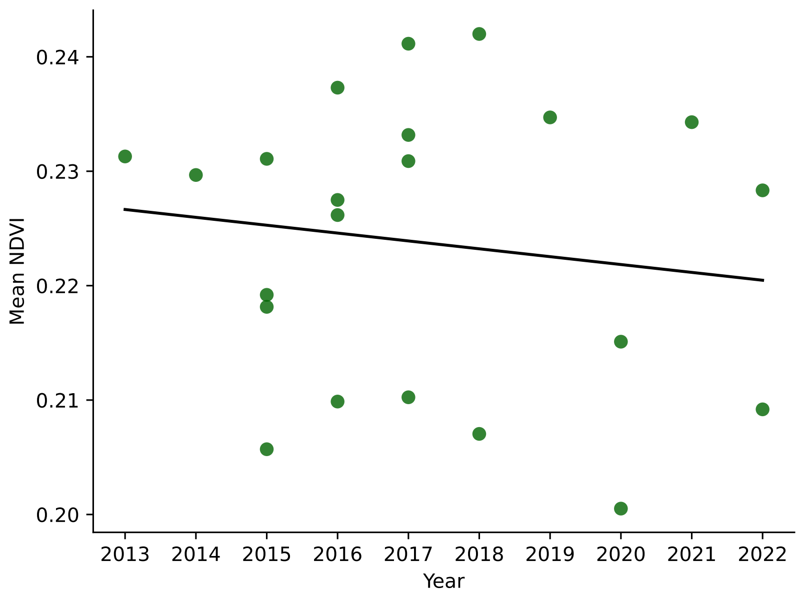


Figure 3.1: Each point represents Philadelphia's city-wide NDVI average for a summer month in the years 2013–2022.

When broken down at the neighborhood level, the Ordinary Least Squares model of NDVI in each neighborhood explained by time results in 136 statistically significant slopes. Figure 3.3 shows 129 block groups experiencing a decrease in vegetation compared with 7 positively sloped neighborhoods, which indicates that Philadelphia's vegetation is reducing in more areas than it has gained over the years.

Without including any spatial correlations, the OLS model shows that most of the decreasing trends in green spaces are located in the Northern Philadelphia region. The intercept estimations in Figure 3.2 visualize which areas naturally have more or less greenery.

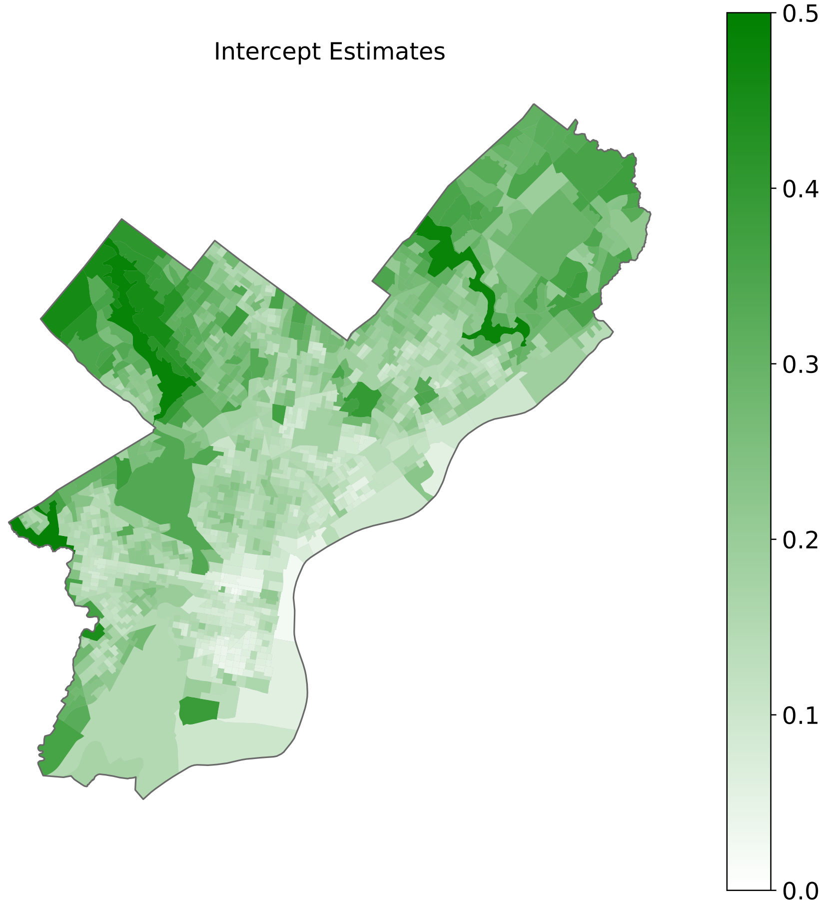


Figure 3.2: Intercept coefficient estimates for NDVI vs Time (in years).

The regions with the highest NDVI in Philadelphia are unsurprisingly located in its parks. The average NDVI around Center City is close to zero, which is a reasonable value because it is located in Philadelphia's downtown area where commercial buildings and pavement are densely packed together, leaving little space for vegetation.

The linear models at the block group level with the most extreme slopes, visualized in Figure 3.3, appear to have a better fit for neighborhoods that have positive slopes. The most extreme negatively sloped neighborhoods in Figure 3.4a can be characterized by one abrupt decrease in their NDVI over the span of ten years. The nonlinear decreasing trends of NDVI for neighborhoods with negative slopes suggest that a different phenomenon may be occurring in these regions that lead to their sudden changes in their NDVI. Note that the areas experiencing the most extreme decreases in vegetation are also the ones located

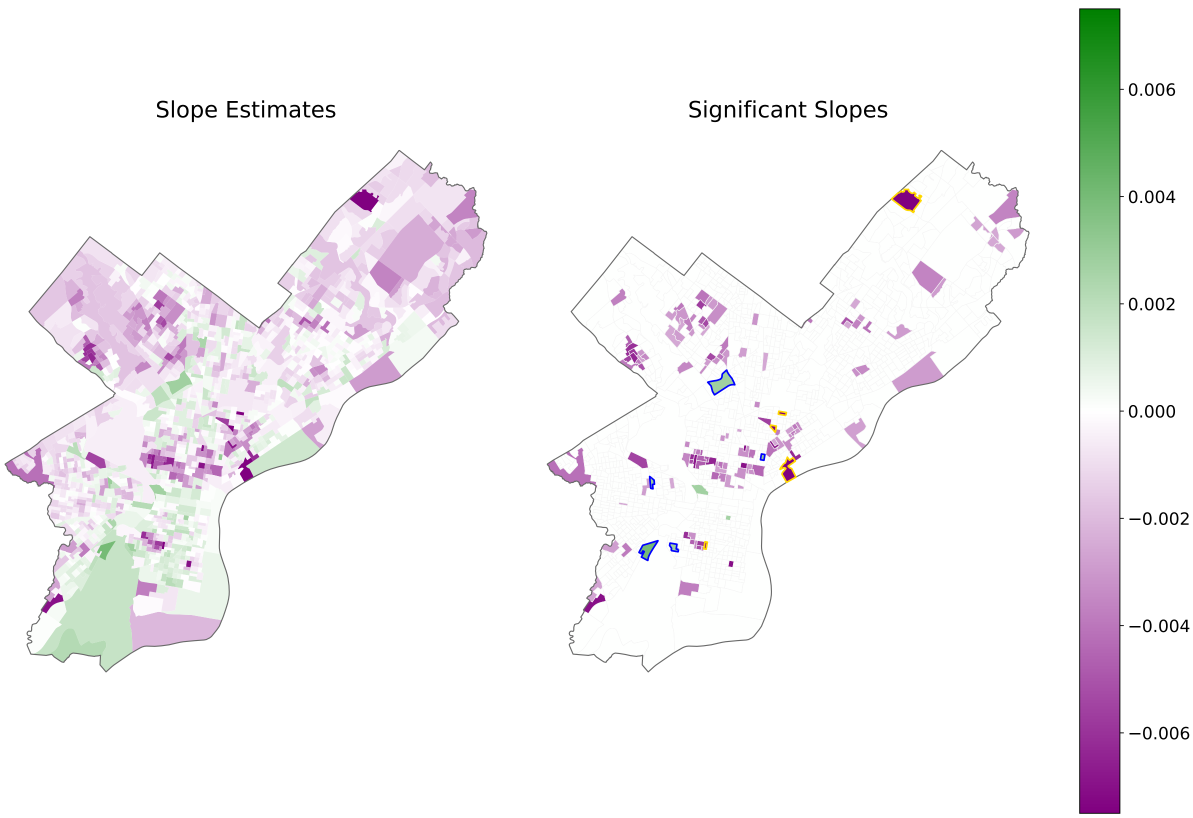
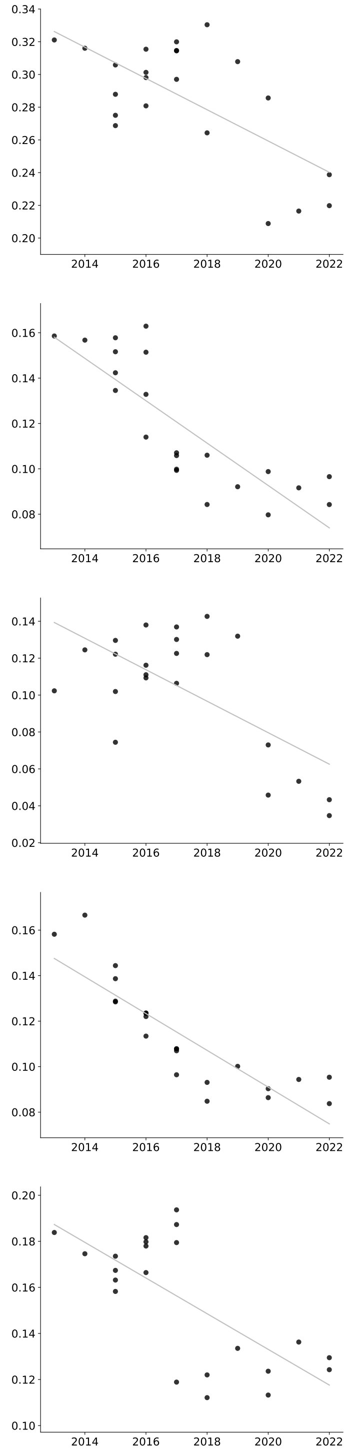


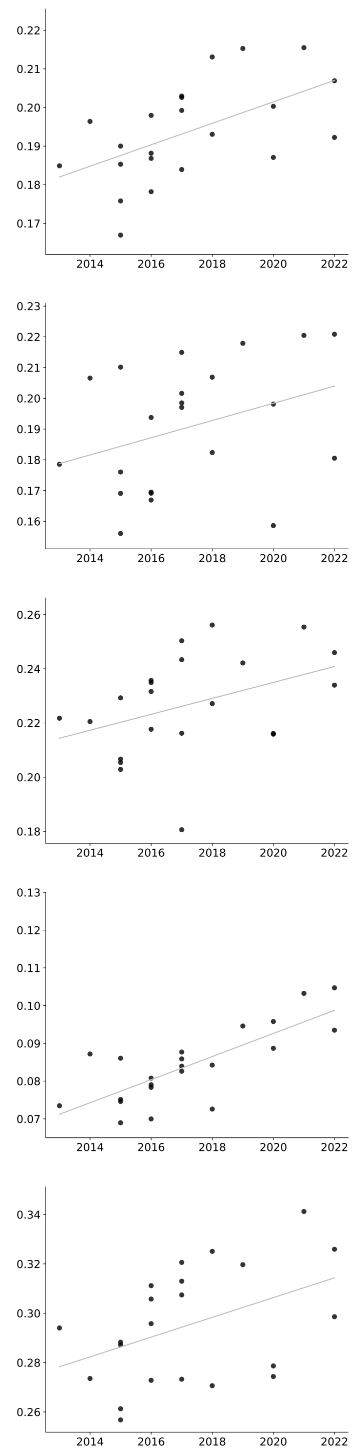
Figure 3.3: Slope coefficient estimates for Time (in years). Neighborhoods outlined in blue are the five most positive slopes and outlined in yellow are the five most negative slopes.

in the North Philadelphia area. Of the five neighborhoods shown in Figure 3.4a, four of them started with NDVI of less than 0.2 in 2013, which is below the city-wide average, and has continued to decrease over time. In contrast, the neighborhoods with the most positive slopes, as shown in Figure 3.4b, have observed NDVI that gradually increase over the years while following a more linear pattern. It is possible that human interference in a region can lead to rapid decreases in vegetation whereas landscapes that are left alone will naturally grow in abundance.

Taking the slope and intercept coefficients from the Ordinary Least Squares model for NDVI at the neighborhood level and graphing them together in Figure 3.5, it is shown that there is a significant negative trend between the two, where the slope of the regression line is -6×10^{-3} NDVI Intercept/NDVI Slope. This result suggests that on average, areas with higher NDVI tend to lose their vegetation more over the years. However, in Figure



(a) Five most negatively sloped neighborhoods ranked with the most negative on the top.



(b) Five most positively sloped neighborhoods ranked with the most positive on the bottom.

Figure 3.4: Ten neighborhoods with the most extreme changes of mean NDVI (y-axis) in Philadelphia from 2013–2022 (x-axis).

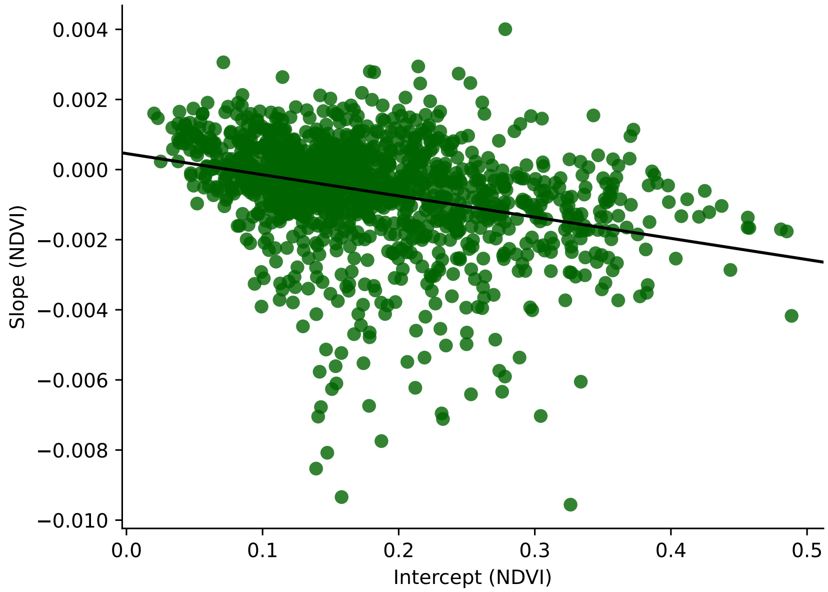


Figure 3.5: Graphical representation of slope and intercept coefficients for all 1336 neighborhoods from the OLS model for NDVI at the neighborhood level.

3.5, there are also many outlier neighborhoods that are far away from the regression line. These points have intercepts that are around or below the average NDVI intercept across all Philadelphia neighborhoods, but their slopes are a lot more negative than what the regression line suggests. These locations require further analyses, but these initial findings illustrate the possibility that there are some neighborhoods with lower vegetation that are losing their greenery at a faster rate over the years than neighborhoods with high amounts of vegetation.

3.2. Temperature Analysis

Average NDVI and Temperature values are shown in Figure 3.6. A spatial hot spot analysis found areas of in Philadelphia that had high LST values surrounded by neighborhoods that also experienced high temperatures. The temperature hot spots are located in North, South, and West Philadelphia, which are areas of the city that already have low NDVI. Philadelphia's parks, on the other hand, yield the lowest land surface temperatures.

Before accounting for the temporal trends, aggregating the average temperatures and

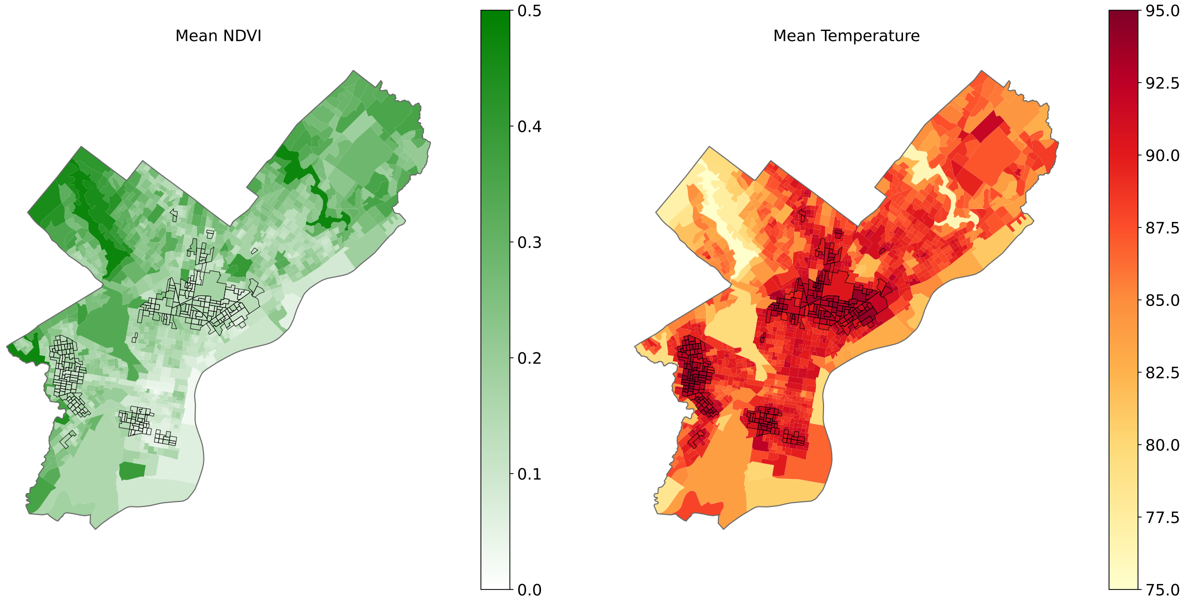


Figure 3.6: Comparing average NDVI values in each neighborhood over 10 years with average land surface temperatures. The block groups outlined in black (on both maps) represent temperature hot spots in Philadelphia.

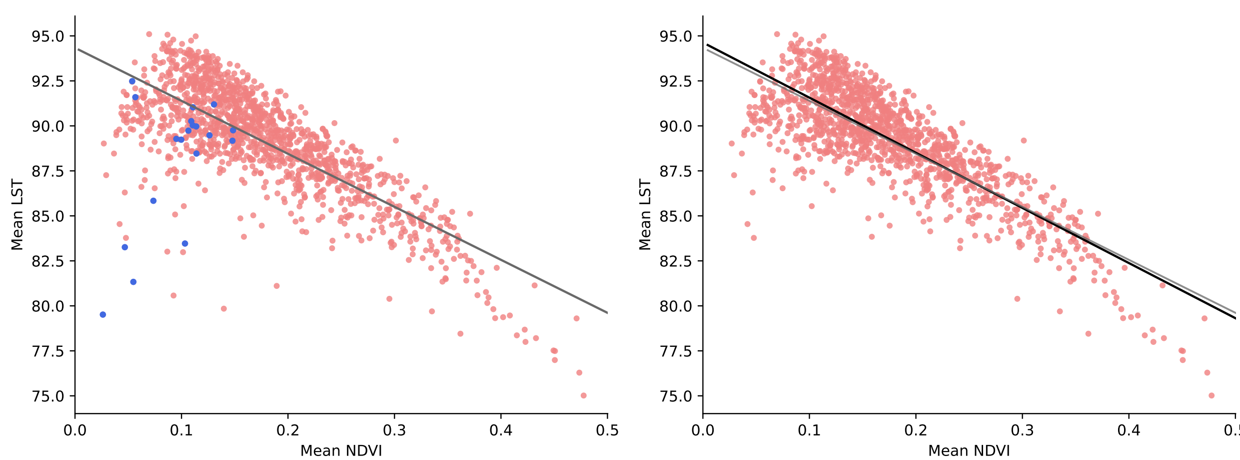
NDVI for each neighborhood across the years shows that there is a strong negative relationship between NDVI versus land surface temperatures. This result is shown in Figure 3.7, which aligns with the findings from studies that were discussed in the literature review (Section 1.3) because it suggests that neighborhoods with more green vegetation have cooler temperatures. The parameter estimates in Table 3.1 show that if NDVI were to increase by

Data	Variable	Coefficient	St. Err.	Moran's I	Moran's I (z-score)	RMSE	AIC
All Block Groups	Intercept	94.3362	0.1304	0.6830	43.331	1.9459	5574
	NDVI	−29.4464	0.6739				
Removed Points	Intercept	94.6309	0.1176	0.7548	47.311	1.7363	5218
	NDVI	−30.6316	0.6055				

Table 3.1: Estimates from the Ordinary Least Squares Model.

0.1 in a neighborhood, it is expected to reduce the surface temperatures of the area by about 3° F. However, in Figure 3.7a, there are some neighborhoods that appear to be outliers due to their large negative residuals. These points have low NDVI values, but their land surface temperatures are cooler than neighborhoods with similar values of NDVI. The linear model was re-run with all block groups that border the Delaware River removed from the model, which resulted in Figure 3.7b where there are no more points that have an average NDVI below 0.2 as well as mean surface temperatures below 82.5° F. The NDVI slope is shown in Table 3.1 to have decreased when the water neighborhoods were removed.

With a significant positive spatial correlation reported in Table 3.1, the following results are from models that incorporate the spatial weighting between neighborhoods. Figure 3.8 illustrates the reasoning behind the large outliers found in the Ordinary Least Squares model between the mean land surface temperature and mean NDVI in each block group. The Geographically Weighted Regression model results in positive slopes in certain neighborhoods (highlighted in Figure 3.7a) that border the Delaware River (on the east side of Philadelphia). These estimates are consistent with Pearsall (2017)'s GWR model. The positive slopes occur in areas close to water because NDVI is pulled towards zero due to the negative NDVI values that occur in areas of water. However, temperatures in these locations are also cooler than inland neighborhoods with the same NDVI due to the evaporative cooling effect from the presence of water.



(a) Regression run on all $n = 1336$ block groups. Neighborhoods with positive GWR slopes marked in blue.

(b) Regression run on $n = 1323$ block groups by removing those bordering the Delaware River. New estimated line in black.

Figure 3.7: Each point represents a census block group plotted by their respective average average NDVI and average land surface temperature (°F) from 2013–2022.

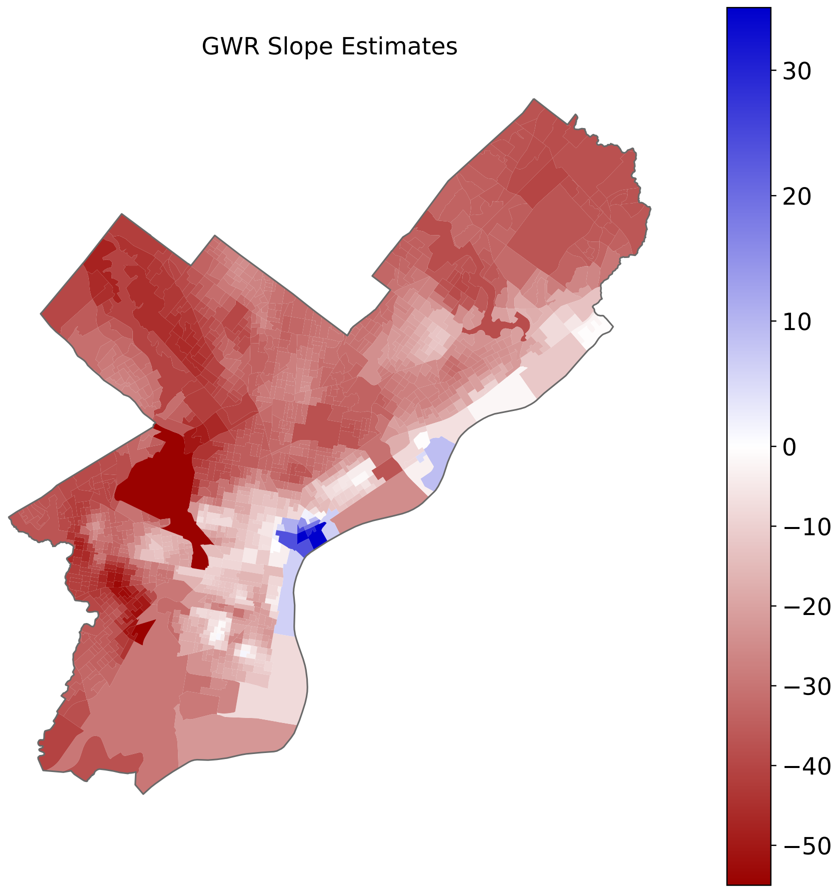


Figure 3.8: GWR slope coefficient estimates for NDVI.

The Spatial Durbin Error Model outputs similar negative slopes for NDVI, seen in Table 3.2. Though its predictive power is slightly worse than the OLS model (by comparing the Root Mean Square Errors (RMSE) with Table 3.1), the SDEM demonstrates that a block group surrounded by neighborhoods with higher NDVI could also reduce temperatures in its own area. The estimate $\lambda > 0$ indicates residuals in the model are positively correlated with its neighbors.

Variable	Coefficient	St. Err.	RMSE	AIC
Intercept	95.1274	0.3751	1.9696	4330
NDVI	-27.7617	0.6744		
W_NDVI	-4.4966	1.5369		
λ	0.8465	0.0171		

Table 3.2: Estimates from the Spatial Durbin Error Model.

Adding the time component back into the OLS model between NDVI and LST, NDVI unexpectedly results in a positive slope. Because the OLS model treats all the variables as fixed effects, which resulted in only six significant Year slopes, a mixed linear model was also run on the same data set by treating the intercept and Year slopes as random effects grouped by neighborhoods. A mixed linear model's random effects pulls its estimates towards the data's expected values. The random effects model results in a negative NDVI slope across the city, shown in Table 3.3.

Model	Variable	Coefficient	Random Effect Variance	RMSE
OLS (Fixed)	Intercept*	78.0315	–	5.839
	Year*	-0.3254	–	
	NDVI	70.6329	–	
Mixed (Random)	Intercept*	92.6309	2.405×10^{-1}	5.979
	Year*	-0.3768	6.348×10^{-3}	
	NDVI	-11.2626	–	

Table 3.3: Estimates for the longitudinal model from a fixed effect OLS and a mixed model with random intercepts and random Year slopes. Variables with * are the average of their respective coefficients across all neighborhoods.[5]

The negative slope estimates for the Year variable appear to be more of an artifact of the data that was used rather than the true relationship between surface temperatures and time. Figure 3.9 shows that most of the estimates are close to zero, with most of the slopes in the fixed OLS model not being significantly different from zero. The city-wide temperature trend over the years results in a slope of $-0.294°\,\mathrm{F/Year}$, but it is also not a significant result. Over the years 2013–2022, Philadelphia's land surface temperature has consistently varied around $86°\,\mathrm{F}$, the average temperature across the city.

The change in sign for the NDVI slope can be attributed to the different intercept estimates between the fixed and random models. Though the intercepts in Figure 3.10 appear to follow similar distributions in Philadelphia, the fixed intercepts range from $40°\,\mathrm{F}$

5. Note the mixed random effects model did not converge, so the significance for the NDVI variable's coefficient in the model could not be tested.

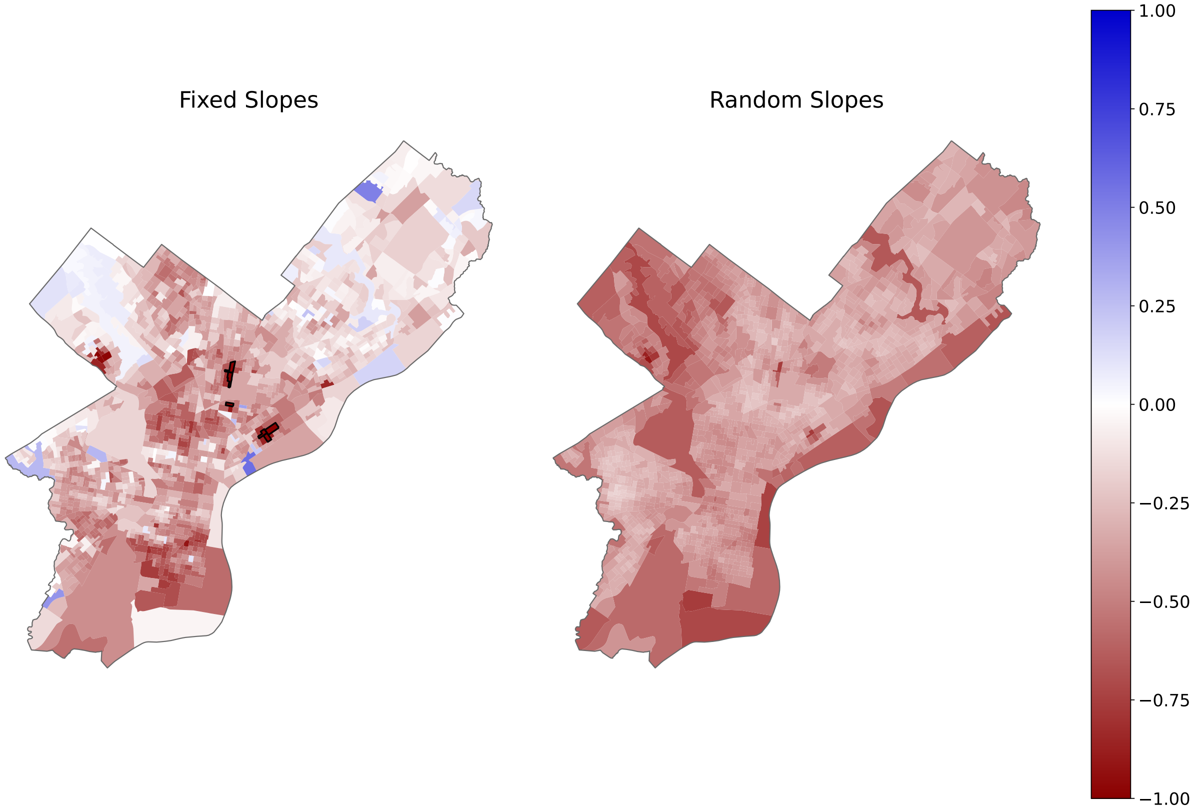


Figure 3.9: Block group level slopes for the Year variable in the models in Table 3.3. Six neighborhoods were outlined in black on the Fixed Slopes map because they had significant Year slopes in context of the fixed OLS model.

to 95° F, compared to the smaller range of 84° F to 96° F for the random intercepts. The random effects intercepts are more interpretable and better align with the distribution of mean temperatures in Philadelphia (shown in Figure 3.6). Interpreting the negative slope on NDVI in the random effects model indicates that an increase in NDVI from one year to the next, after controlling for a yearly effect, is expected to decrease land surface temperatures in that area. Therefore, both the cross-sectional and longitudinal models on NDVI and LST show that increasing vegetation in an area could potentially mitigate extreme temperatures from the urban heat island effect.

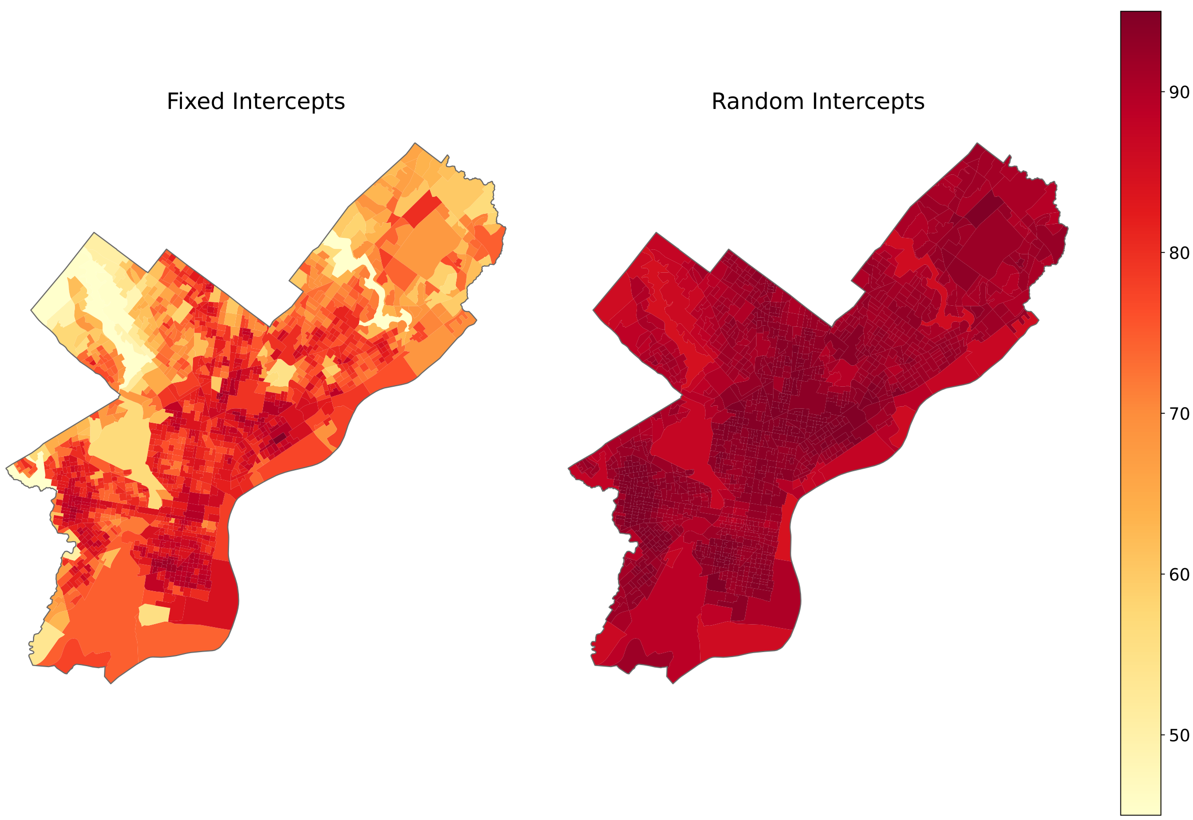


Figure 3.10: Block group level intercepts for the models in Table 3.3. Note the difference in ranges between the fixed and random intercepts. The random intercepts fall between the 80° F to 100° F temperature range.

4. Discussion and Conclusion

4.1. Implications of the Results

This paper identified areas in Philadelphia that lost the most vegetation over the years 2013 to 2022. The analysis showed that neighborhoods with low NDVI also experienced the highest temperatures in the city over the summer months. The models in this study demonstrated a strong negative correlation between amounts of green vegetation and land surface temperatures spatially and over time. This result aligns with the conclusions made by the studies mentioned in Section 1.3. The notable finding of this paper is that the negative trend between NDVI and LST persists over time, which supports the idea of increasing implementation of green conversion project in the city.

As mentioned in Section 1.2, the results of this study would also help identify the neighborhoods in Philadelphia that would most benefit from a greening project. For example, the temperature hot spots identified in Figure 3.6 show that these neighborhoods also suffer from the lack of vegetation when compared to the rest of the city. When looking at the block groups that were identified to have lost the most vegetation (Figure 3.4a), it would be helpful for greening projects to identify why these areas have a sudden drop in NDVI. As shown in the positively sloped neighborhoods, vegetation tends to gradually increase NDVI once it has been planted, but urban development in the commercial and industrial space could drastically decrease an area's NDVI.

With a strong spatial correlation between neighborhoods, greening efforts should consider targeting areas with the lowest vegetation first instead of expanding out from regions that already have some vegetation. This decision would also prevent non-greening redevelopment projects from obtaining the space first. Overall, this study has demonstrated that the lack of vegetation contributes to the urban heat island effect, which can be addressed by greening initiatives so that residents do not suffer from heat-related health problems.

4.2. Future Work

Though this paper has spatially and temporally analyzed the relationship between vegetation and surface temperatures in Philadelphia, additional data could be incorporated into this study to add more nuance to the relationship between vegetation and temperature around the city. For example, green spaces can be characterized by different types of vegetation (e.g., grass roofs versus tree canopy parks), so the next steps to this paper would be to investigate if these differences in vegetation types can be detected along with their own effects on temperature. Further studies would also examine if it could distinguish a difference in effect between green spaces that are placed all together in one space or dispersed between urban infrastructures. These analyses would require data at finer resolutions, so it would useful to have LIDAR data to label the different types of vegetation and to distinguish areas of vegetation that are aggregated together versus more dispersed.

As demonstrated in Pearsall (2017)'s and in Section 1.3, economic and demographic data should also be incorporated into the study to analyze who lives in the areas that are most affected by the changes in vegetation and surface temperatures in Philadelphia over the years. The purpose of adding these factors is to see if there are additional variations in temperature changes that can be explained by a neighborhood's economic and demographic layout. The results from adding socioeconomic characteristics could help policymakers identify areas for greening projects that would benefit the environment as well as under-served communities.

Furthermore, the models that were run in this paper could be updated to better represent the observations. A Bayesian Additive Regression Trees (BART) model could be tested on the neighborhood-level models in Section 3 in comparison with the linear regressions. BART does not assume anything about the shape of the regression function. Instead, it approximates the function with a step function and decision tree that requires a prior assumption. Therefore, a BART model is more complicated than an OLS regression, but the benefit is that it may characterize non-linear relationships that cannot be considered in the linear models. For example, the neighborhoods that were identified with the most negative slopes have an abrupt decrease in NDVI over the years, so a BART model could capture this change by estimating two separate slopes before and after the change point. However, with more complexity, BART model's predictions could become harder to interpret. Spatial correlations should be considered in the longitudinal models from Section 3 as well. A spatial conditional autoregressive (CAR) model would incorporate the spatial dependence between neighborhoods while maintaining the temporal trends.

Finally, the most meaningful future work would be the successful implementation of green space conversions in the city of Philadelphia using the findings from this study. The effects of these greening projects in an area would better inform the models that were presented in this paper and provide other cities with a framework on how to mitigate increasing temperatures with vegetation, which would consequently benefit the well-being of its residents and the environment.

References

Almeida, Cátia Rodrigues de, Ana Cláudia Teodoro, and Artur Gonçalves. 2021. "Study of the Urban Heat Island (UHI) Using Remote Sensing Data/Techniques: A Systematic Review." *Environments* 8, no. 10 (October 9, 2021): 105.

Andronis, Vassilis, Vassilia Karathanassi, Victoria Tsalapati, Polychronis Kolokoussis, Milto Miltiadou, and Chistos Danezis. 2022. "Time Series Analysis of Landsat Data for Investigating the Relationship between Land Surface Temperature and Forest Changes in Paphos Forest, Cyprus." *Remote Sensing* 14, no. 4 (February 18, 2022): 1010.

Behrer, A. P., R. J. Park, G. Wagner, C. M. Golja, and D. W. Keith. 2021. "Heat has larger impacts on labor in poorer areas*." *Environmental Research Communications* 3, no. 9 (September): 095001.

Carrillo-Niquete, Gerardo A., José Luis Andrade, José René Valdez-Lazalde, Casandra Reyes-García, and José Luis Hernández-Stefanoni. 2022. "Characterizing spatial and temporal deforestation and its effects on surface urban heat islands in a tropical city using Landsat time series." *Landscape and Urban Planning* 217 (January 1, 2022): 104280.

Chen, Yu-Cheng, Chih-Yu Chen, Andreas Matzarakis, Jin-King Liu, and Tzu-Ping Lin. 2016. "Modeling of mean radiant temperature based on comparison of airborne remote sensing data with surface measured data." *Atmospheric Research* 174-175 (June): 151–159.

Cortinovis, Chiara, Dagmar Haase, and Davide Geneletti. 2023. "Gradual or abrupt? An algorithm to monitor urban vegetation dynamics in support of greening policies." *Urban Forestry & Urban Greening* 86 (August): 128030.

Cui, Jesse, Shane T Jensen, and John MacDonald. 2022. "The effects of vacant lot greening and the impact of land use and business presence on crime." *Environment and Planning B: Urban Analytics and City Science* 49, no. 3 (March): 1147–1158.

Edmondson, J. L., I. Stott, Z. G. Davies, K. J. Gaston, and J. R. Leake. 2016. "Soil surface temperatures reveal moderation of the urban heat island effect by trees and shrubs." *Scientific Reports* 6, no. 1 (September 19, 2016): 33708.

Heckert, Megan, and Jeremy Mennis. 2012. "The Economic Impact of Greening Urban Vacant Land: A Spatial Difference-In-Differences Analysis." *Environment and Planning A: Economy and Space* 44, no. 12 (April 23, 2012): 3010–3027.

Hibbard, Kathy A., Forrest M. Hoffman, Deborah Huntzinger, and Tristram O. West. 2017. "Changes in Land Cover and Terrestrial Biogeochemistry." *Climate Science Special Report* 1:277–302.

Kondo, Michelle, Jaime Fluehr, Thomas McKeon, and Charles Branas. 2018. “Urban Green Space and Its Impact on Human Health.” *International Journal of Environmental Research and Public Health* 15, no. 3 (March 3, 2018): 445.

Li, Xiaojiang. 2021. “Investigating the spatial distribution of resident’s outdoor heat exposure across neighborhoods of Philadelphia, Pennsylvania using urban microclimate modeling.” *Sustainable Cities and Society* 72 (September): 103066.

Li, Xiaojiang, TC Chakraborty, and Guoqing Wang. 2023. “Comparing land surface temperature and mean radiant temperature for urban heat mapping in Philadelphia.” *Urban Climate* 51 (September 1, 2023): 101615.

Lin, Desen, Shane Jensen, and Susan M. Wachter. 2021. “Amenity Value of Green Space: Evidence from Philadelphia.” *SSRN Electronic Journal.*

Locke, Dexter Henry, Lara A. Roman, Jason G. Henning, and Marc Healy. 2023. “Four decades of urban land cover change in Philadelphia.” *Landscape and Urban Planning* 236 (August): 104764.

Luber, George, and Michael McGeehin. 2008. “Climate Change and Extreme Heat Events.” *American Journal of Preventive Medicine* 35, no. 5 (November 1, 2008): 429–435.

Ming, Yujia, Yong Liu, Jianping Gu, Jinzhu Wang, and Xue Liu. 2023. “Nonlinear effects of urban and industrial forms on surface urban heat island: Evidence from 162 Chinese prefecture-level cities.” *Sustainable Cities and Society* 89 (February 1, 2023): 104350.

Nowak, David J., Allison R. Bodine, Robert Hoehn, Alexis Ellis, Sarah C. Low, Lara A. Roman, Jason G. Henning, Emily Stephan, Tom Taggert, and Ted Endreny. 2016. *The Urban Forests of Philadelphia.* NRS-RB-106. Newtown Square, PA: U.S. Department of Agriculture, Forest Service, Northern Research Station, November.

Opper, Isaac M., R. Jisung Park, and Lucas Husted. 2023. “The effect of natural disasters on human capital in the United States.” *Nature Human Behaviour* 7, no. 9 (September): 1442–1453.

Park, R. Jisung, A. Patrick Behrer, and Joshua Goodman. 2021. “Learning is inhibited by heat exposure, both internationally and within the United States.” *Nature Human Behaviour* 5, no. 1 (January): 19–27.

Pearsall, Hamil. 2017. “Staying cool in the compact city: Vacant land and urban heating in Philadelphia, Pennsylvania.” *Applied Geography* 79 (February 1, 2017): 84–92.

Sanecharoen, Warisara, Kanchana Nakhapakorn, Aeumphorn Mutchimwong, Supet Jirakajohnkool, and Rattapon Onchang. 2019. “Assessment of Urban Heat Island Patterns in Bangkok Metropolitan Area Using Time-Series of LANDSAT Thermal Infrared Data.” *Environment and Natural Resources Journal* 17, no. 4 (September 1, 2019): 87–102.

Sekertekin, Aliihsan, Niyazi Arslan, and Mehmet Bilgili. 2020. "Modeling Diurnal Land Surface Temperature on a Local Scale of an Arid Environment Using Artificial Neural Network (ANN) and Time Series of Landsat-8 Derived Spectral Indexes." *Journal of Atmospheric and Solar-Terrestrial Physics* 206 (September 15, 2020): 105328.

Townshend, John, Christopher Justice, Wei Li, Charlotte Gurney, and Jim McManus. 1991. "Global land cover classification by remote sensing: present capabilities and future possibilities." *Remote Sensing of Environment* 35, no. 2 (February 1, 1991): 243–255.

Wang, Yujia, Ji Zhou, Mingsong Li, and Xiaodong Zhang. 2016. "Validation of Landsat-8 TIRS LAND surface temperature retrieved from multiple algorithms in an extremely arid region." IEEE International Symposium on Geoscience and Remote Sensing (IGARSS). November 3, 2016.