Monitoring and Modeling Urban Heat Patterns in the State of Iowa, USA Utilizing Mobile Sensors and Geospatial Data

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of Master of Arts

Clemir Abbeg Coproski

University of Northern Iowa

July 2023

Abstract

With cities experiencing faster warming rates than their surroundings (Stone 2007) and two thirds of the global population projected to be living in urban areas by 2050 (United Nations 2019), studies on temperature patterns in the urban environment have the potential to address concerns regarding cities' livability. While recent progress has been made in understanding urban heat spatial heterogeneity, more research in a variety of urban settings is necessary. While there have been studies conducted in the U.S. Midwest (Kunkel et al. 1996; Rajasekar and Weng 2009; and Gallo et al. 1993), no studies have examined the pattern of temperature in urban areas using high spatial and temporal resolution methods across multiple different sized cities in the state of Iowa. The goal of this research was to examine the spatial pattern of temperature across multiple cities of different sizes in the state of Iowa utilizing mobile temperature sensors along with high spatial and temporal resolution methods that leverage geospatial data on morphometric and natural features.

Air temperature data were collected in ten urban areas in Iowa during afternoon (4-5 p.m.), evening (9-10 p.m.), and night (4-5 a.m.) offering the necessary data to generate predictive temperature models based on natural and man-made features found in the urban fabric. Utilizing a Random Forest algorithm, predicted surface models, which consider Canopy Cover, Canopy Density Metric, Building Height, Building Volume, and NDVI as independent variables showed an R^2 between 0.879 and 0.997, with 20 out of 24 models falling into a range of R^2 higher than 0.95. When examining the relationship between air temperature and income, all values with statistical significance presented a

weak to moderate negative correlation, which is consistent with the literature that suggests that areas with higher income experience less heat than areas with lower income, while the correlation between non-ethnic groups showed a very weak to weak positive correlation, which is interpreted as in some locations, ethnical minorities are more likely to experience heat than white population.

This research, funded by the Iowa Economic Development Authority Iowa Energy Center Grant Program, aims to benefit communities in urban areas across Iowa. The data collected, including temperature measurements and models, will be made publicly available to assist government officials and a diverse group of professionals in mitigating urban heat and improving living conditions. A deeper understanding of urban heat patterns will lead to preventive measures and improved techniques to enhance the overall quality of life for individuals and communities.

This Study by: Clemir Abbeg Coproski

Entitled: Monitoring and Modeling Urban Heat Patterns in the State of Iowa, USA Utilizing Mobile Sensors and Geospatial Data

has been approved as meeting the thesis requirements for the

Degree of Master of Arts

Date	Dr. Binqing Liang, Chair, Thesis Committee
Date	Dr. Alex Oberle, Thesis Committee Member
Date	Dr. Kamyar Enshayan, Thesis Committee Member
Date	MSc. John DeGroote, Thesis Committee Member
Date	Dr. Stephanie Huffman, Dean, Graduate College

Acknowledgements

First of all, I would like to express my deepest gratitude to my mom, Loreni. Without your love and support, life would have never been possible to begin with. You are and will always be the most incredible human being that has ever walked on this planet. I would also like to express my most sincere appreciation to my sister, Franciele. Siblings might not get along all the time, but in every disagreement, there is something to be learned, and oh well, we have learned a lot, haven't we? I love you both, more than you could ever imagine.

To my friends everywhere, thank you for your support throughout this journey. In different ways, you guys could share the best in you. Laughter, tears, and something in between, you will always be regarded as the family I got to choose.

To my advisor, Dr. Bingqing Liang, for motivating me to do my best. Your advice, comments, and constructive criticism shaped me to be a better student and professional.

A big thank you to John DeGroote. Even though this acknowledgment is part of my thesis, my gratitude to you extends to way before and beyond that. You are not only one of the best professors I have ever had but also an inspiring human being. Your help will never be forgotten.

A special thank you to the UNI Geography Department, and staff. Without you, this journey would not have been as motivating as it was. Every single one of you has contributed to an outstanding experience and a quality education.

To the University of Northern Iowa, College of Social and Behavioral Science, Office of International Education, and all departments that, in one way or another, contribute to making education a priority for people from all over the world, my sincerest thank you.

Special acknowledgment goes to the committee members, Dr. Alex Oberle, and Dr. Kamyar Enshayan, whose significant contributions were instrumental in concluding this study.

The successful execution of this study was made possible by the funding provided by the Iowa Economic Development Authority Iowa Energy Center Grant Program (agreement number 21-IEC-012).

Finally, thank you to every single person that I had the pleasure to encounter during the past two years. You guys helped me to create great memories that will be remembered for years to come.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Rising temperatures and extreme heat events have been seen across the globe in recent decades leading to increased concern across the world (Vicedo-Cabrera et al. 2021). Mean annual heatwave season lengths were longer in 2010s as compared to 1980s, with the most pronounced exposure increases being in the lowest-quartile income regions (Alizadeh et al. 2022). Urban areas specifically, home to a majority of the world's population, have experienced higher temperatures caused by the built environment morphology (Environmental Protection Agency 2022). In 2022, headlines or stories such as 'U.S. cities are heat islands boiling under deadly extreme temperatures. It's only expected to get worse' as seen on CBS News (Cohen 2022) were common. The high probability of extreme heat events has been brought up in many studies. Fischer et al. (2021) revealed that week-long heat extremes that break records by three or more standard deviations are predicted to be two to seven times more likely to happen in 2021-2050 and three to 21 times more in 2051-2080 when compared to the last three decades. Dahl et al. (2019) indicated the increase in global temperatures poses a secondary problem: the number of people affected by extreme events. Compared to the last three decades of the 20th century, the annual number of days with heat indices exceeding 100F and 105F are projected to double or triple by the mid-21st century (2036-2065) and affect more than 25% of the US by area compared to only 1% from 1971-2000. The National Oceanic and Atmospheric Administration (NOAA) and the National Aeronautics and

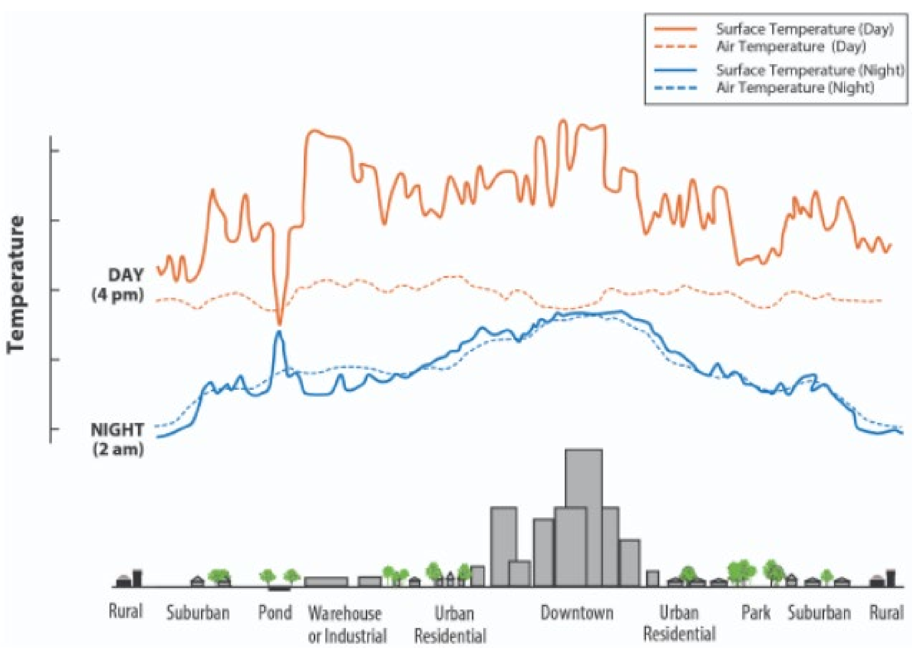
Space Administration (NASA) found that that 2010-2019 was the hottest decade ever recorded and the world's five warmest years have all occurred since 2015 (NOAA 2020).

With cities experiencing faster warming rates than rural areas (Stone 2007), major public health problems (e.g., chronical diseases, death) can arise, affecting a larger portion of the population, including those with higher socioeconomic vulnerability. According to Wong et al. (2013), more casualties have resulted from heat waves than other hazardous weather events such as hurricanes, floods, and tornadoes combined. While mortality rates caused by heat-related illnesses have dropped since 1980 in the USA, in certain places like Paris, an increase of 0.5°C above the average minimum nighttime temperature could double the risk of death in the elderly (Wong et al. 2013). In a study compromising 108 US urban areas, Hoffman et al. (2020) found that 94% of studied areas displayed elevated land surface temperatures in formerly redlined areas (historical discriminatory real estate practices) with approximately 2.6° C warmer temperatures than non-redlined area. It is also observed that socioeconomically vulnerable groups such as Native Hawaiian and Other Pacific Islander (NHPI), Hispanics, African Americans, as well as young children often closely linked to higher temperatures and even heat events (Voelkel et al. 2018; Wong et al. 2013).

The concept of contrasting temperature regimes between urban and surrounding areas goes back historically and has been studied with more modern techniques and data for several decades. The phenomenon of urban heat has been studied for over 150 years (Oke 1982). Indeed, Gartland (2011) cited papers dating back as far as 1833 by Howard (England) and 1855 by Emilien Renou (France). In 1958, the term known as Urban Heat Island (UHI) was coined by the British climatologist Gordan Manley when he compared

London's snowfall pattern to the city's surrounding outlying districts. According to The United States Environmental Protection Agency (2022), an UHI is defined as an urbanized area that experiences higher temperature than outlying areas. The main factors contributing to UHI include reduced natural landscapes in urban areas, urban material properties, urban geometry, heat generated from human activity, and variation in weather (e.g. cloud coverage and wind patterns). Figure 1 illustrates conceptually and figuratively the potential patterns of spatial heterogeneity in air and surface temperature across urban areas. This figure, while being useful, is an oversimplification, illustrating the complicated nature of the urban fabric and the associated temperature regimes.

Figure 1
Heat Island Effect Diagram. United States Environmental Protection Agency 2022



Studies evaluating urban temperatures using remote sensing analysis techniques (Balázs et al. 2009; Saydelles 2005; Yu et al. 2020) have been common over the last couple of decades. According to Zhou et al. (2018), the first known surface urban heat island (sUHI) study using satellite imagery was published by Rao (1972) and subsequent studies have continued often using Landsat and other satellite imagery for measures of land surface temperature (LST). A systematic review of regional and urban heat island conducted by Degefu et al. (2022) found that greater than 50% of studies used Landsat while 36% used MODIS (Degefu et al. 2022). Such studies based on remote sensing imagery are sometimes supplemented using stationary thermometers in selected locations or mobile sensor systems to collect temperature across urban land cover gradients (Alcoforado 2010).

While leveraging satellite imagery provides the most affordable way to collect data covering large areas and across multiple urban areas, these techniques do suffer from limited spatial resolutions (e.g., 30 m for Landsat, 250 m for MODIS). Imageries of such resolution are unable to capture the highly detailed spatial heterogeneity of complicated urban matrices. Another challenge due to the sparse temporal resolution is the inability to describe temperature changes throughout a day, which "is necessary for understanding how fast specific areas of the city heat and cool" (Voelkel and Shandas 2017, 2). To address this shortcoming, recently low-cost sensor technologies and high-resolution geospatial data have been used to provide highly detailed spatial data capturing spatial heterogeneity across and within the urban area, as seen in studies like Voelkel and Shandas (2017), which made use of mobile sensor systems to record detailed temperatures across multiple times on a very hot day in Portland coupled with high resolution Light Detecting and Ranging (LiDAR) data (1m). According to Stewart (2000), the use of mobile thermometers enables very high-resolution temperature data collection, thus allowing for a better understanding of small variations of the temperature found within the urban area (Szymanowski and Kryza 2012).

The improvements in remote sensing technologies, including higher spatial resolutions, is a major benefit to study temperature variation in different areas of the urban environment. However, they are still far from the capabilities and fine resolution of mobile temperature sensors. Temporal resolution of traditional remote sensing platforms also does not provide the flexibility that mobile sensors can provide. The ability of data collection in different times of the day, for consecutive days, without the interference of unstable atmospheric conditions and cloudy skies is considered a great advantage over traditional remote sensing technology.

Research Problem

While recent progress has been made in understanding urban heat spatial heterogeneity, more research in a variety of urban settings would be useful. The majority of previous studies have focused on larger urban areas (e.g., Voelkel and Shandas 2017; Shandas et al. 2019) and often have only considered one study area at a time. While there have been studies conducted in the U.S. Midwest (Kunkel et al. 1996; Rajasekar and Weng 2009; and Gallo et al. 1993), no studies have examined the pattern of temperature in urban areas using high spatial and temporal resolution methods across multiple different sized cities in the state of Iowa.

Research Goals, Objectives

The primary goal of this research is to monitor and to model urban heat pattern in the state of Iowa by using high spatial and temporal resolution mobile sensors and geospatial data. The specific objectives are:

1. To collect very high-resolution temperature data across selected urban areas using mobile sensors in the state of Iowa.
2. To model detailed temperature variation patterns among different Iowa urban areas using selected morphometric and natural features.
3. To examine the temperature variation across urban neighborhoods with varying socio-demographic characteristics.

Chapter 2

Literature Review

Health and Economic Issues Related to Urban Heat

In 1950, only 30% of the world's population lived in urban areas. By 2007 the United Nations estimated that the urban population surpassed the rural population globally, with more than 4 billion people living in the cities. By the year 2050, it is projected that more than two thirds of the global population will be living in urban areas. This population trend can be seen worldwide, but urbanization has been faster in less developed regions compared to historical trends in more developed countries. Regions such as Latin America and the Caribbean reached staggering levels with 80% of populations living in urban areas in 2015 (United Nations 2019).

Average daytime temperatures in urban areas in the USA are up to 4°C higher than average temperatures in outlying areas (U.S. EPA 2022), and in extreme cases can be up to 10°C compared to the rural countryside (Heaviside et al. 2017). In a study conducted by Kunkel et al. (1996), the authors examined one of the worst heat waves that happened in the Midwest in the last century, which caused hundreds of fatalities across the central and eastern United States in mid-July 1995. Fatalities were reported in 19 states, 87% of those occurred in the Midwest, and 65% of all deaths took place in Chicago. In the event that lasted for 4 consecutive days, the daily average temperature exceeded 36°C over some of the areas, with the maximum apparent temperature reaching 48.1°C (118.6°F) in Chicago, the highest temperature since 1916. Several health problems can be attributed to extreme heat ranging from exacerbation of minor existing conditions to increased risk of hospitalization and death (Basu 2009).

Other problems associated with high temperatures include higher electricity demand and the costs related to it. According to Santamouris et al. (2014), although electricity demands can vary according to location, for each additional 1°C electricity usage can increase in 8.5%, causing a financial burden to people with socio-economic vulnerability and possibly overloading the energy grid due to extensive use of thermal controlling devices (e.g., air conditioning).

Studies such as Voelkel et al. (2018), Hattis et al. (2012), and Shandas (2009) explored the connection between climate and social justice. Voelkel et al. (2018), used three factors to evaluate the vulnerability to heat – exposure, sensitivity, and adaptive capacity – and found minority populations were significantly correlated with heat exposure and with low adaptive capacity. They thus suggested it is imperative for governments to address social disparities in heat resilience efforts. Hattis et al. (2012) provided evidence that minority groups such as elderly population (65+ years old), African Americans, and people without a high school diploma were the most vulnerable to heat-related mortality in the state of Massachusetts from 1900 to 2008. Shandas (2009) discussed how certain social groups were disproportionately living in the hottest areas of Portland, Oregon. He also observed not only low-income, but elderly and specific ethnic groups were at higher risk from heat waves.

Heat events can impact different parts of the globe, they are not exclusive to urban areas. However, as seen in several studies, urban areas are more likely to experience higher temperatures than non-urban areas, posing a bigger threat to already marginalized groups.

History and Advances in Urban Heat Studies

Studies on mapping the spatial variation of temperature on the planet can be tracked back as early as more than 2 thousand years when it was first documented by Aristotle and Plutarch on 6 B.C (Bailey 1964). In 1817, Alexander von Humboldt created the first isothermal map, presenting evidence that not only latitudes and the amount of radiation received from the sun could fully explain the variation in temperature in different parts of the planet (SmithsonianMag 2019). Later in 1884, the German-Russian climatologist Wladimir Köppen created the model for climate classification we use to this day, presenting 5 main climate groups, subdivided by the amount of precipitation and heat that occur in each of those areas (Chen and Chen 2013).

Studies for examining temperatures in dense human settlements can be traced back to around 150 years ago according to Oke (1982). Specifically, observations related to the idea of UHI have been well documented throughout history, being first published in 1833 by Luke Howard (Stewart 2010). In his observations, temperatures in the city of London were warmer than its countryside. In France, during the second half of the 19th century, Emilien Renou (1855, 1862) made similar discoveries when measuring the differences in temperature between urban and rural areas in the country.

In the United States, the first studies examining urban heat patterns started in the first half of the 20th century (Gartland 2011), carried out by the American climatologist J. Murray Mitchell (1953, 1961). Although already a widely known concept among scholars, it was only in 1958 that the term urban heat island was first used by the British climatologist Gordan Manley, when studying the effects of snowfall on the temperatures in London and its outskirts. A UHI can be defined as urban and suburban areas presenting a higher air and surface temperature than their surroundings (Gartland 2011). In addition, Tim Richard Oke, one of the most eminent figures on urban climate studies, the effects of high temperatures in urbanized areas are “probably both the clearest and the best documented example of inadvertent climate modification” (Oke 1987, 288).

Stone (2007) examined the temperature trends in the largest 50 urban areas across the USA, comparing the variation in urban and rural temperatures over 50 years. The study found that 58% of the sampled urban areas experienced a mean decadal increase in heat island intensity, while 42% experienced a decrease with decreases found in northeastern US and Ohio River Valley with southern US generally showing increases. When considering only absolute mean temperature for each area, 12 urban and 12 rural areas experienced a decrease in temperature over the period of 50 years. Cities such as Buffalo, Sacramento, Los Angeles, Jacksonville, and Orlando were some of the cities that presented the cooling trends in both, rural and urban areas. Cities such as Fresno, Pittsburgh, Rochester, and Syracuse, experienced cooling trends in their urban areas, while the rural areas of those cities presented warmer temperatures over time (1951 – 2000).

An important part of understanding the intensity of urban heat has to do with the landscape features and processes that influence the magnitude of heat. Paravantis et al. (2015, 4548), as part of a thorough urban heat island intensity review article, summarized as follows: ‘latitude and elevation; climate characteristics such as wind velocity, cloud cover and rainfall; surface morphology and view to the sky; vicinity to bodies of water such as the sea; distance from industrial sites; degree of urbanization (i.e. population density); vegetation (i.e. presence of open and green areas, tree canopy etc.); land usage, surface albedo and presence of impervious surfaces (and their heat balance); urban geometry and presence of urban canyons; characteristics of urban residences including whether dense/sparse, distribution and consumption of heating or cooling energy and ventilation of buildings; as well as road traffic intensity and urban air pollution'. Other characteristics such as roughness (Oke 1987), pavement composition (Gartland 2011) and anthropogenic heat flux generated by cultural activities (Oke 2009) have also been shown to be associated with higher urban temperatures.

Advances in technology and techniques such as increases in spatial resolution from satellite and airborne-based imagery and other geospatial data have allowed researchers to examine the influence that features other than those mentioned above have on the variation of temperature, including the shape of tree canopy (Gkatsopoulus 2017), and tree shade (Yu et al. 2020). According to Chun and Guhathakurta (2016), the type of vegetation also plays a role in regulating urban temperatures. Stone and Norman (2006) pointed out that if the suburban areas of Atlanta/Georgia were to reduce their lawn areas by 25% and replace them with trees, the contribution of latent heat to the increase of temperatures would be reduced by 13%.

Even though studies have proven the mitigating effects of bodies of water on urban heat (Oke 1987; Chun and Guldmann 2012; Paravantis et al. 2015), some authors are cautious in stating that bodies of water can strongly regulate the air temperature due to their high thermal inertia (Chun and Guhathakurta 2016). Hathway and Sharples (2012) highlighted the cooling effects close to rivers, lakes, and seas, and found differences in temperature up to five degrees Celsius above the river and nearly 100m from the riverbanks compared to the rest of the urban areas. They also showed that the level of cooling provided by bodies of water depended on the vegetation and the local urban form, with higher levels of cooling on the bank in areas with high levels of vegetation. They also found the cooling was noticeable up to 30m from the bank, but negligible after 40m from the river.

Mirzaei (2015) found that highly populated areas, which tend to develop either vertically or horizontally, can not only work as a blockage against urban ventilation but also absorb more solar radiation due to the composition of artificial materials, leading to reduced long-wave emission to sky, therefore contributing to warmer temperatures in urban areas. On the other hand, buildings can cast shadows to offer respite from the heat during the day (Chun and Guhathakurta 2016) and to prevent direct radiation from the sun (Hathway and Sharples 2012). Complemented by other elements of urban geometry such as roughness (the ratio between the perimeter of the building and the perimeter of a circle of a similar area) and compactness (a function of density, plot ratio, land-use and travel proximity), sky view factor is also an important factor that correlates negatively with urban heat according to Kusaka and Kimura (2004), since “canyon structure with small sky view factor keeps higher surface temperature during the night” (Kusaka and Kimura 2004, 75).

Atmospheric conditions are also regarded as important factors on the variation of temperature in the urban environment. Stewart (2000) examined those conditions and their effects in the city of Regina in the Canadian province of Saskatchewan. In his study, the author measured wind speed, cloud cover/Boltz factor, atmospheric pressure, and vapor pressure happening right before, during, and after temperature collection. Temperature data were collected between 2 and 3 hours after sunset and acquired through mobile sensors in 31 different locations throughout the city and outskirts. The time was chosen to (1) allow for the seasonal change in the time of sunset, and (2) coincide with the expected time of daily maximum heat island intensity as documented in the literature by Oke (1982). Through the use of linear and non-linear regression techniques, Stewart determined that urban temperature in Regina is highly sensitive to changes in cloud and wind conditions, and relatively insensitive to changes in humidity and atmospheric pressure. Disregarding any weather event or atmospheric condition prior to heat events, wind speeds seem to have a higher influence than any other independent variable for explaining variability in nocturnal UHI, however, if weather conditions preceding heat events are considered, cloud cover supersedes wind speed as the most important control of nocturnal UHI.

While measuring air temperature is commonly used to evaluate variation of temperature within the urban environment, the study of surface temperature is also used to identify the phenomena (Zhou et al. 2018). The evaluation of temperature patterns in urban areas can be done by examining radiative temperature differences between urban and non-urban surfaces or by accessing the variation of air temperature in the urban canopy level.

Investigating further and elaborating on the difference between measuring surface temperature versus air temperature, Gartland (2011) presents a list of the most common methods which are used in measuring temperature in the urban environment: fixed stations, mobile traverses, remote sensing, and vertical sensing. Remote sensing is perhaps the most common for measuring surface temperature, while fixed stations, mobile traverses, and vertical sensing, are usually used to measure air temperature. Oke

(1987) suggested that measuring air temperature using either climate stations or mobile traverses is the most clear-cut way of evaluating urban heat. According to the author, none of the methods to identify variations in temperature in urban areas is mutually exclusive, in fact, different methods and techniques can be used as complimentary tools for measuring temperature patterns or even validating models.

As mentioned earlier, the use of satellite imagery-derived land surface temperature (LST) has been common across a variety of geographic regions. Studies by Saydelles (2005), Yu et al. (2020), and Chen et al. (2020) broadly used LST to estimate the variation of temperature in different urban settings. The variation in surface temperature was identified through the use of thermal satellite images from Landsat 7 ETM+ (Saydelles 2005), Landsat 5 TM (Yu et al. 2020), and ASTER (Chen et al. 2020), with spatial resolutions of 60 meters, 120 meters, and 90 meters respectively. Examining the myriad of satellites/sensors utilized for the purpose of understanding temperature variation across different areas, Zhou et al. (2018) provided a comprehensive review of the main sensors used to this day. In a large sample of studies (N=492), the authors searched for peer-reviewed journals in English language from 1972 to August 2018 using the ISI Web of Science and Google Scholar databases and set a combination of keywords that look for temperature pattern in the urban environment through the use of satellite imagery. The mean number of papers during the latest years (2015 – 2020) related to the search was 75.7 papers per year, while before the year 2005 this number was smaller than 4 papers per year. The Landsat, MODIS, and ASTER satellites were used in 418 papers, showing the importance of those satellites/sensors to urban studies worldwide.

According to Zhou et al. (2018), each constellation of satellites presents different advantages. For Landsat satellites, the main advantages are that imagery can be freely obtained and the swath size of 185km x 185km "is big enough that it allows scientists to process a single image to investigate an entire urban environment" (Zhou et al. 2018, 8). The main benefit of MODIS is its quality-checked data products generated by the MODIS team, on top of the convenient swath dimensions (2330km by 10km), allowing for research of large study areas. Finally, ASTER, the third most frequently used data in the reviewed studies, provides imagery of daytime and nighttime, and since 2015 is available to all users at no cost.

Although imagery derived from satellites is still widely used in different studies and provides invaluable benefits due to its large coverage and possibly low cost for vast areas, their limitations call for other methods to capture the detailed spatial heterogeneity of urban environments. Limited spatial and temporal resolution, being constantly affected by atmospheric conditions, and the high costs when utilized for small areas can be enough limitations that could sway the collection method to other alternatives such as use of mobile traverses.

The use of mobile traverses to measure air temperature in cities and regions of the world in the past decades has grown supplementing measurements in fixed locations. The temperature measuring devices are usually either mounted on cars (Saaroni et al. 2000; Balász et al. 2009; Voelkel and Shandas 2017; and Shandas et al. 2019), motorbikes (Yokobori and Ohta 2009), bicycles (Shandas et al. 2019) or utilizing non-vehicle mobile devices (Stewart 2000; Sakakibara and Matsui 2005; Szymanowski and Kryza 2012). Studies conducted by Stewart (2000), Sakakibara and Matsui (2005), and Yokobori and

Ohta (2009) made use of temperature data collected in pre-defined locations/sites, while those by Saaroni et al. (2000), Balász et al. (2009), Yokobori and Ohta (2009), Voelkel and Shandas (2017), and Shandas et al. (2019) used data measured at points along pre-defined routes at specific planned times throughout the day. The studies mentioned above successfully utilized fine-resolution data from mobile devices that helped to explain temperature patterns across different areas in the urban environment.

For vehicle-mounted devices the position of the device varied in distance to the ground, however, consistency can be seen in elevating the device over at least a meter from the ground in order to not be affected by surface temperature or the heat generated by the car engine. Stewart (2000) took measurements from distances up to 20m from the car to avoid external heat sources including the influence of car engine heat.

With the capability of generating precise, three-dimensional information about the shape of the Earth and its surface characteristics, LiDAR is also widely used to understand the urban landscape and can produce an array of data, ranging from spatial distribution of land cover, tree canopy, and information on building structure such as height or volume. Chun and Guldmann (2012) used LiDAR data paired with land surface temperature (LST) for delineating the urban characteristics and factors that enhance or mitigate urban heat. In the study, the main geospatial data derived from LiDAR were building heights, which combined with GIS building footprints and a digital elevation model (DEM) helped to develop a 3D model of the city concluding that 3D urban characteristics are critical at a neighborhood scale in influencing temperature variations.

Yu et al. (2020) focused on exploring the cooling effect of tree shade in urban landscapes through the use of LiDAR data, land surface temperature data derived from

Landsat TM imagery, and the employment of spatial analysis techniques to understand the spatiotemporal patterns of the temperatures in Tampa and New York City (NYC). Results showed that shade cast by urban trees can lower LST, however, the effects of tree canopy (in % coverage) in Tampa had a higher effect on mitigating temperature than in NYC (r^2 = - 0.64 vs r^2 = - 0.53), while impervious surface presented almost the same magnitude in both cities (r^2 = 0.60 in Tampa, r^2 = 0.61 in New York City).

The role of LiDAR in identifying horizontal and vertical tree canopy structures can also be seen in Chen et al. (2020, 1). They employed 2d grid cells with a 2m spatial resolution in processing the LiDAR point cloud data. LiDAR data and ancillary imagery were used to derive raster metrics for tree canopy which they used in comparison to daytime and nighttime LST data. Ground points derived from LiDAR cloud data were also used to create a Digital Elevation Model (DEM). Their results showed that tree canopy had stronger influences on LST during the day than at night, while the most important independent factors that affected spatial variation of LST were percent cover of tree canopy, and mean tree canopy height during the day, and percent cover of tree canopy, and maximum height of tree canopy at night.

With higher spatial and temporal resolution methods of data collection, studies like the ones conducted by Voelkel and Shandas (2017), and Shandas et al. (2019) were of great influence in the conception of this study. The authors used mobile traverse data collection methods to build extensive datasets (i.e., thousands to hundreds of thousands) of temperature measurements that were subsequently used to study and model temperature variation across different areas of the urban environment, leading to an increased understanding of urban heat patterns in different cities in the United States.

Indeed, the protocols that began in these studies have been further carried on through National Oceanic and Atmospheric Administration-funded projects over several years including in 14 cities across the United States in summer 2022 (NOAA 2022a).

Voelkel and Shandas (2017) used statistical techniques such as multiple linear regression (MLR), classification regression tree (CART) combined with multiple linear regression, and random forest (RF) to model and predict urban heat in the city of Portland based on a wide variety of urban form metrics derived from LiDAR, building footprint vector, and imagery data. They found the random forest to produce the highest predictive power, varying from r^2=0.8199 for temperature data collected in the afternoon (3 p.m.) to r^2=0.9793 for temperature data collected in the morning (6 a.m.). For temperature collections that took place in the evening (7 p.m.), the model presented a predictive power of r^2=0.9715.

Shandas et al. (2019) published a similar study by employing a satellite pixel-based modeling approach for the creation of a continuous surface of predicted temperatures across three different US cities: Richmond, Washington D.C., and Baltimore. They agreed that common patterns could be seen across all three locations: “forested and otherwise vegetated areas are cooler than urbanized areas; lower-density urban areas are often cooler than high-density urban areas; morning high temperatures are always lower than afternoon and evening low temperatures; the greatest relative concentration of heat is in the morning; major arterial roadways are visible in all UHI surfaces, though they are often amplified in the evening” (Shandas et al. 2019, 6). Using a combination of ground-based measurements, spectral data from the Sentinel-2 constellation (LULC), spatial analysis, and machine learning techniques (Random Forest regression) the models showed a high predictive power of no less than $r^2=0.9644$ for all nine models (morning, afternoon, and night for each of the 3 study areas).

In the study conducted by Szymanowski and Kryza (2012), temperature data was collected in 206 different points across Wroclaw, Poland to systematically represent a variety of land-use categories in areas that presented "interesting and geometrically diverse areas in the city center" (2012, 55). The study took into consideration automatic weather station measurements in both urban and rural areas and utilized Landsat ETM+ data to derive information such as albedo, land surface temperature and vegetation indices. Data on buildings roughness length, building density, porosity and Sky View Factor (SVF) were obtained using LiDAR point cloud data which were later converted to a raster dataset of 1m resolution. Although different statistical methods to analyze and improve the spatial interpolation of the urban heat structure were used (multiple linear regression, geographically weighted regression, Akaike information criteria, and Moran's Index), results showed that geographically weighted regression were better suited for spatial modelling of urban heat.

Balázs et al. (2009) discussed how empirical models based on previous datasets were used to analyze the annual mean urban heat intensity for Szeged and Debrecen (Hungary). The authors used single and multiple variable models based on linear and multiple regression techniques and considered independent variables such as surface cover ratio (streets, roofs, parking lots, and pavements), distance from the city boundary/center, and a combination that considers the built-up ratio and its area extent. Results showed a high correlation of $r^2=0.774$ and $r^2=0.816$ between the modeled and observed temperature values for the areas (depending on which independent variables were used) as an average for the three cities (Hajduboszormeny, Hajdudorog, and Hajdunanas) to which the model was applied.

Originating from a project called ICALON (Ilhas de Calor de Londrina) in southern Brazil, Oukawa et al. (2022) compared two different modeling approaches (multiple linear regression and random forest) to analyze and predict the spatiotemporal occurrence of the UHI intensity using air temperature as the dependent variable. To do that, urban temperature was measured in 12 different sites in the city of Londrina using HOBO sensors (U23-001, Onset) with built-in data loggers, and were complemented with data from two more permanent weather stations. A great number of independent spatial variables broken into four categories (land cover, topography, urban geomorphology, and population and traffic) were used to create the models. For example, land cover data was derived from Sentinel-2, elevation, wind speed, and relative humidity were derived directly from data from weather stations. According to the authors, the urban heat island intensity was more pronounced during the night (10 p.m. – 06 a.m.) due to the combination of sustained clear skies, high radiation, and no rainfall caused by lingering high-pressure systems (LHP). From the two statistical techniques used to create the models (multiple linear regression and random forest), random forest performed considerably better than multiple linear regression, with explanatory power of over 96% in both, daytime and nighttime (compared to 64% and 34% respectively). The model was validated using a training dataset (80%) and testing (20%) and reiterated 300 times to ensure transferability.

Table 1 displays the range of selected independent variables by studies that used similar temperature collection methods and or statistic approaches.

Table 1
List of studies that used similar temperature collection method, independent variables and/or similar statistical methods

STUDY	TEMPERATURE COLLECTION METHOD	INDEPENDENT VARIABLE(S) CONSIDERED	ALGORITHM/STATISTICAL METHOD UTILIZED
Chen et.al, 2020	LST	Percent of tree canopy, patch density, edge density, patch shape index, patch cohesion index, tree height	MLR
Chen et.al, 2022	LST + meteorological stations	Anthropogenic parameters, distance from the city center, proportion of LULC area, altitude, longitude, latitude, slope, aspect, proportion of impervious surface (IS) area, albedo, (NDVI), (NDBI), (gNDVI), (SAVI), (NDMI).	RF
Hart & Sailor, 2008	Mobile sensors	Land use, surface cover	MLR, RT
Oukawa et.al, 2022	Fixed sensors + meteorological stations	Land Cover, topography, urban geomorphology, population and traffic, weather data, atmospheric vertical indices	MLR, RF
Shandas et.al, 2019	Mobile sensors	LULC, NDVI, NBAI	RF
Voelkel & Shandas, 2017	Mobile sensors	Land cover, vegetation, building height, building volume, elevation, canopy density metric	CART, MLR, RF
Yokobori & Ohta, 2009	Mobile sensors	Vegetation, cloud cover, wind speed, land cover	LR

Studies focused on understanding the impacts of anthropic activities on urban temperatures as part of their objectives often make use of independent variables such as population density or city size. Studies that focus on the spatial effects of two- and three-dimensional urban features on the formation of UHI (and its variations as sUHI) set their independent variables as for example building geometry and SVF, as seen in Chun and Guhathakurta (2017). Studies that investigate how morphometric features and human-induced activities influence and can be influenced by urban temperature variation often use a broad set of independent variables, going from LULCs, and building volume, to tree canopy, and traffic.

Literature Review Summary and Context of this Study

Elevated urban heat regimes have been demonstrated across the globe and the likelihood of such events is predicted to increase in the future. Better data on the spatial heterogeneity of urban heat can benefit a range of stakeholders. Mirzaei (2015, 200) summarized the negative consequences of elevated urban heat as such: '...elevated air temperature of a city, … increases the heat and pollution-related mortality, reduces the habitats' comfort, and elevates the mean and peak energy demand of buildings'. Urban temperatures have been shown to vary significantly across relatively small distances in urban areas (e.g., Voelkel et al. 2018) and affect different populations disproportionately (e.g., Alizadeh et al. 2022). While there is a long history of urban heat studies, especially using land surface temperature values derived from satellite imagery, recent studies (e.g., Shandas et al. 2019) have been leveraging low-cost mobile air temperature sensors in conjunction with highly detailed urban metric data derived from LiDAR, imagery, and other sources to develop highly detailed (e.g., ~1m spatial resolution) maps or urban temperature. Agencies such as the US National Oceanic and Atmospheric Administration are implementing these types of studies in larger urban areas across the country yet none in Iowa. This study supplements such efforts and contributes by being the first study (as known to the author) to examine urban heat in multiple urban areas ranging from small to medium-sized towns in a single state.

Chapter 3

Methodology

Study Area

Iowa is one of the 12 states in the Midwest region of the United States, located in the center of the region between 40°35’N - 43°30’N latitude and 90°8’W - 96°38’W longitude. Surrounded by the Mississippi (east) and Missouri Rivers (west), it borders six other states: Minnesota (north), Wisconsin (northeast), Illinois (southeast), Missouri (south), Nebraska (southwest), and South Dakota (northwest). With an area of 55,857 square miles, Iowa figures as the 9th largest state by area in the region, its highest elevation is in Osceola County at 1670 feet and the lowest elevation point is in Lee County at 480 feet (USGS 2022).

According to Köppen Climate Types, Iowa is in a hot-summer humid continental climate area (Dfa) and in the past decade (2012 – 2022) had an average temperature of 9.1°C and a precipitation of 35.52 inches per year. The maximum average temperature for the same period was 14.8°C and the minimum average temperature was 3.4°C (NCEI-NOAA 2022).

Iowa is the 19th smallest state population-wise in the United States, with a resident population of 3,190,369 in 2020 (U.S. Bureau Census Data 2022). It has a Gross Domestic Product of approximately 170 billion dollars, and it is positioned as the 29th largest GDP out of the 50 states (U.S. Bureau of Economic Analysis 2022). Iowa’s economy is based primarily on products and manufacturing services related to agriculture. According to an economic report published by Iowa State University (2020), 17.4% of the state's GDP relies exclusively on the agricultural industry and approximately one in every six Iowans work directly within the same industry.

Statistically, in the United States, states that economically rely more on agriculture, tend to present lower levels of urbanization. That is true for the state of Iowa, where the urban population represents 64% of the total state population, as compared to 80.7% nationwide (U.S. Census Quick Facts 2021). Less than one-third of 99 counties in Iowa have an urban population of over 50%, and the population of several counties was classified exclusively as rural.

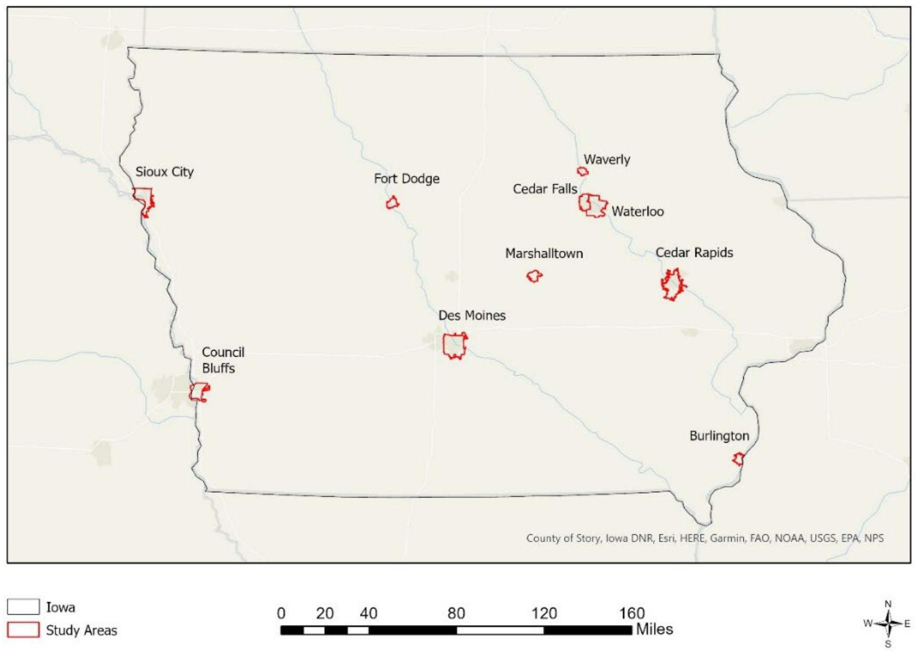
From 2000 to 2020, the population growth in the United States was 17.4% (from 282.2 to 331.45 million) compared to 8.9% in Iowa, going from a population of 2.93 million to 3.19 million (U.S. Census Bureau Data 2022). Considering only the 6 most populated cities (U.S Census Bureau Data 2022) in the state (Des Moines, Cedar Rapids, Davenport, Sioux City, Iowa City, and Ankeny), the average population growth was 7.9%, 9.5% behind the national rate but 1% behind the state rate, indicating that rural areas have the potential to see a more rapid increase in population than urban areas.

While Iowa has no urban area with a population over one million inhabitants (e.g., Des Moines is the largest by population with approximately 210,000 inhabitants) it presents an opportunity to supplement past and recent urban heat pattern research (e.g., Shandas et al. 2019) which has focused on large urban areas. This study attempted to examine urban heat patterns across urbanized areas of various sizes including investigating whether smaller cities with comparable highly local urban structures (e.g. tree canopy, building density) experience urban heat similarly (with the same proportional variation in temperature) to larger cities. This heterogeneity of the urban division in Iowa provides an interesting scenario for this study to be conducted. Based on the lower urban population compared to the national average, on top of a different populational growth rate among urban areas implies that cities would have a more diverse urban setting than large urban centers (e.g., not only tall structures in downtown), enabling a comparative analysis between cities that have a similar area with differing urban configurations, and cities with a similar level of urbanization but different overall areas.

This study collected temperature data from urban areas with populations ranging from 10,000 to 210,000 inhabitants throughout the state (different regions) on very hot days in the summer of 2022. Although specific weather patterns (e.g., rain) and logistical constraints (e.g., distance) played a role in the specific locations, the cities visited included Burlington, Cedar Falls, Cedar Rapids, Council Bluffs, Des Moines, Fort Dodge, Marshalltown, Sioux City, Waterloo, and Waverly.

Figure 2

Cities for which temperature data was collected in 2022



Data Acquisition

Five primary datasets were used throughout this study: air temperature (collected through mobile sensors); morphometric features derived from high-resolution elevation LiDAR data (provided by the state of Iowa); satellite imagery (provided by the United States Department of Agriculture); building footprint vector files (provided by counties, cities, and generated by deep learning techniques originated from Microsoft USA); and socio-economic data adapted from Census Data at the census block group level (provided by the United States Census Bureau). The first four were used in the building of spatial temperature models while the socio-economic data were used to evaluate modeled air temperature surfaces to investigate if there are noticeable differences across neighborhoods with varying socio-economic conditions.

Air Temperature Data Collection

The mobile devices used in this study (Figure 3) were built and tested by Dr. James Thomas Dietrich, who was an Assistant Professor of Geography at the University of Northern Iowa during the investigation of the study. A total of 9 fully functional devices were built to allow multiple collections at the same time in the same or different cities. Dr. Dietrich also developed a Python-based coding that allowed sensor communication and data collection and for post-processing of the raw data collected by the sensors. The main components of the devices, which allow for air temperature collection, were built using Adafruit Sensirion SHT40 Sensors along with an Adafruit Mini GPS Stemma QT PA1010D.

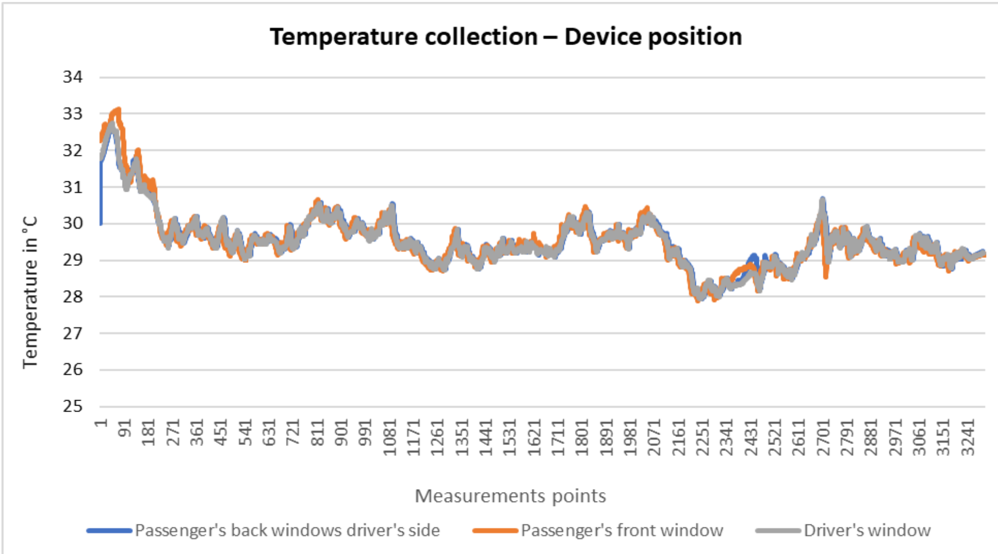
Figure 3

Temperature sensor devices. Source: James Dietrich, and the author

Although air temperature data collection is the focus of this study, sensors also record additional data such as surface temperature (Malexis Infrared Sensor MLX906143V), humidity (Adafruit Sensirion SHT40), and elevation (Adafruit BMP280I2C). Sensors were pre-programmed to collect data every second. The precision of the thermometers is two decimals' Celsius degrees, and the accuracy is up to 0.2°C (Adafruit 2022).

Previous studies that used vehicle-mounted temperature sensors showed the influence of the heat generated by the car engine on the air temperature readings (Stewart 2000; Balász 2009). However, testing with three devices being mounted on three different positions (driver's window, passenger front window, and passenger's back window on the driver's side) showed negligible difference (and within the accuracy of the devices) on air temperature measurements across all devices, as can be seen in Figure 4.

Figure 4
Testing for vehicle-mounted device positions. June 12th, 2022

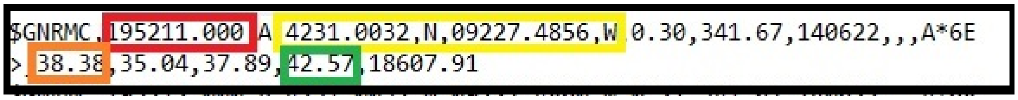


In ideal data collection scenarios, multiple vehicles and sensors could be used on a specific day in a given urban area. However, due to staff, funding, and logistical constraints, individual data collection missions were sometimes necessary.

Since the devices were programmed to store the collected data (e.g., temperature, humidity) to a text file (.txt) format (Figure 5), a post processing phase took place after in-situ temperature collection. Below is a representation of the data collected per point. The information outlined in red shows the time of recording (19:52:11), the information outlined in yellow shows the latitude and longitude (42.310032 N, 92.274856W) of the location that the data was collected, the information outlined in orange (38.38) is the air temperature recorded in Celsius, and the information outlined in green (42.57) is the humidity. A post-processing Python script was used to translate this raw file format into a .csv file which was then imported as an ESRI File Geodatabase feature class.

Figure 5

Representation of the raw data collected by mobile devices

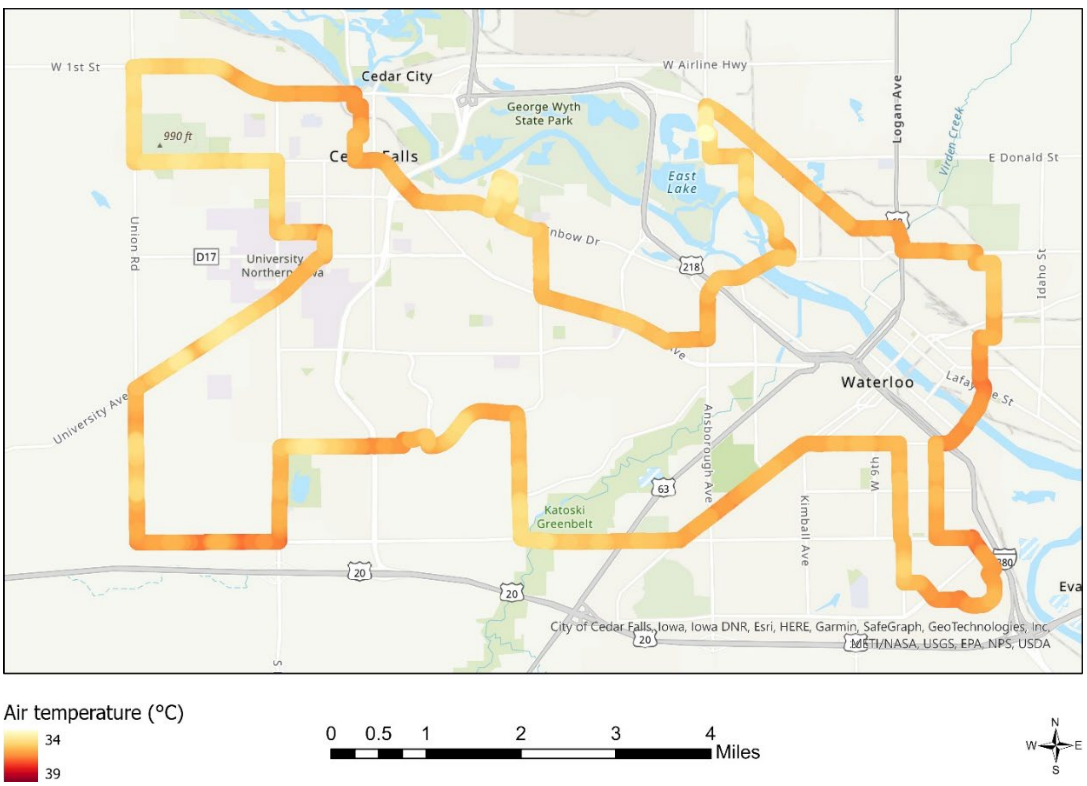


```
$GNRMC,195211.000,A,4231.0032,N,09227.4856,W,0.30,341.67,140622,,,A*6E
>,38.38,35.04,37.89,42.57,18607.91
```

Figure 6 shows an example of the test collection of air temperature on June 12th, 2022, during a hot day (34°C/93°F+) in the cities of Cedar Falls and Waterloo/Iowa.

Figure 6

Processed temperature collection data in Cedar Falls and Waterloo - Iowa on June 12th 2022



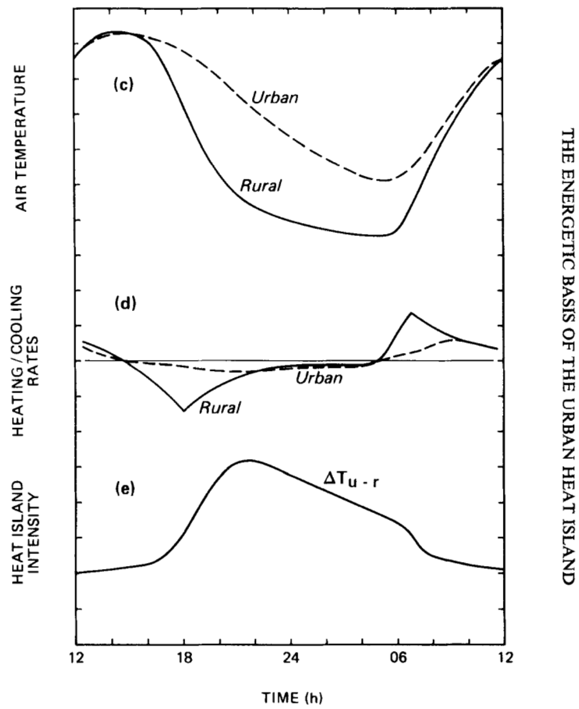
Efforts were made to follow established precedence in data collection methodologies. For example, the ideal speed of the vehicle, on which the thermometer was mounted, was determined to generate a consistent data collection throughout the whole area. Studies such as Voelkel and Shandas (2017) suggested that speeds should be kept under 56 kilometers per hour to prevent any cooling of the sensor due to turbulence at higher speeds, while other studies as Saaroni et al. (2000), and Wong and Yu (2005), worked with speeds not higher than 30 kilometers per hour and 50 kilometers per hour, respectively. In addition, although studies such as Voelkel and Shandas (2017) discarded all data collected while the vehicle was stationary, Balász et al. (2009) utilized data collected while the vehicle was in movement between the speeds of 20 km/h to 30km/h to provide the necessary ventilation of the sensor. Following similar studies utilizing mobile sensors to collect high-resolution temperature data such as Voelkel and Shandas (2017), and Shandas et al. (2019), temperature points collected at speeds that exceed 35 miles per hour were discarded. However, to promote a better understanding of the correlation between vehicle speed and the variation in air temperature, a statistical analysis was used to evaluate and compare air temperature data collected and the speed at which vehicles were moving.

To investigate the variation of temperature across all cities in a consistent manner, the time of data collection was based on the findings suggested by Oke (1982). According to the author, the higher variation of heat between urban areas and non-urban areas can be seen between 8 p.m. and 12 a.m., while the heat is more intense between 2 p.m. to 4 p.m. (Figure 7). Although this study focused on examining the patterns of temperature in urban environments, the presence of areas with low density of built-up structures and with characteristics that resemble non-urban areas are expected to be seen within an overall city or urban area. Days of the collection were chosen based on days that were forecasted to exceed 90°F (32.2°C), which is believed to exacerbate the differences in temperature in the urban environment (Voelkel and Shandas 2017). It is known that wind speed and clear skies (cloudless skies) have a great influence on UHI intensity, especially on summer nights (Oke 1982). Yet, aligning these meteorological conditions on the planning aspect of data collection has proven to be challenging, either for enabling one of the predetermined conditions to be strictly followed (days with temperatures above 90F) or for the logistics involved on the availability of personnel for temperature collection in one or multiple areas. In this study, the time of collection happened three times within the time frame of one day in each study area for a period of one hour: 4 p.m. to 5 p.m., 9 p.m. to 10 p.m., and 4 a.m. to 5 a.m. The period of one hour for each data collection is required to control for the variation of radiation experienced by different areas while working to minimize the variability in temperature and weather conditions that might occur in the city (Saaroni et al. 2000).

Temperature data were collected throughout different urban areas in the state of Iowa, representing small to medium size cities, varying from approximately 10,000 to 210,000 inhabitants and areas varying from 11 to approximately 88 square miles. Although the area covered by temperature collection varied according to traffic, road condition, and speed limit, the time frame of collection was consistent throughout the different study areas for each run (approximately 1 hour).

Figure 7

The energetic basis of the urban heat island. (Oke 1982)



LiDAR Data and Imagery

For this study, different independent variables were generated from a variety of geospatial datasets through automated geoprocessing techniques. LiDAR point cloud data collected by the state of Iowa during the years 2019-2022 was used to estimate urban vegetation and building morphology; building vector files obtained directly from counties and cities were used to improve the accuracy of individual 3D building features present in the study areas; and airborne imagery from the United States Department of Agriculture National Agriculture Imagery Program (NAIP) was used to generate normalized difference vegetation index (NDVI).

Initial efforts to collect high-resolution LiDAR data in the state of Iowa started in 2007 (LiDAR for Iowa 2019) and an updated statewide coverage of detailed LiDAR data was collected from 2019-2021 (Iowa 2022). With a minimum of 2 points per square meter with a vertical and horizontal accuracy of 95%, these data provide highly detailed coverage across all of Iowa and can be used to derive a variety of landscape structure metrics, which can be evaluated in relation to collected urban temperatures similar to efforts made by studies such as Voelkel and Shandas (2017). To generate the input files for the needed variables, various geoprocessing operations were performed, such as converting downloaded LAZ data into LAS data format; classification of lidar point cloud data; production of raster files representing the proportion of specific classes in each study area such as ground density and vegetation density; and the creation of models representing the bare-earth (digital elevation models), and above ground features existent in the study areas (digital surface models). The LiDAR point cloud data, in order to accurately represent the shape and surface area of buildings, were processed along with building vector polygons, to construct building multi-patch feature classes.

The literature suggests a variety of urban morphometric features that, if not directly responsible for the changes in temperature within the urban environment, at least are seen as important components in explaining urban heat patterns. A main feature that seems to have a high influence on urban temperature throughout the literature is vegetation (or the proportion of it compared to other features). For that reason, imagery downloaded from the NAIP was used to create individual NDVI raster files for each urban area. According to the United States Department of Agriculture (USDA), the NAIP program was created to support two main strategic goals both centered on agriculture production: “(1) increase stewardship of America's natural resources while enhancing the environment, and (2); to ensure commodities are procured and distributed effectively and efficiently to increase food security.” (NOAA 2022b). Collected between August 3rd, 2021 and January 20th, 2022, the 2021 Iowa NAIP 4-Band 8-bit Imagery has an accuracy of +-4 meters to the reference image, and a resolution of 60 centimeters.

All three mentioned datasets (LiDAR point cloud data, Building Footprints, and NAIP Imagery) provided the necessary data to generate all 5 independent variables that were used to interpret, analyze, correlate, and modeled all predictive temperature surface raster results during the modeling stage of this project. Each of these independent variables was developed as a raster dataset with a one-meter resolution. The variables and their definitions are as follows:

- Canopy Cover (CC): the proportion of vegetation compared to the amount of bare-earth and built-up structure in a given area (derived from LiDAR point cloud data classified as bare-earth and as vegetation points). The results are expressed in pixel values going from 0 to 1, where 0 represents no vegetation, and 1 represents fully vegetated area.
- Canopy Density Metric (CDM): the result of pixel values of CC multiplied by pixel values found in the normalized digital surface model (nDSM) which represents the relative height of features from the DSM.
- Building Height (BH): the height of buildings expressed in meters in each individual pixel.
- Building Volume (BV): the volume of buildings expressed in cubic meters.

- Normalized Difference Vegetation Index (NDVI): the value per pixel ranging from minus one (-1) to plus one (+1) representing the density, intensity of vegetation as well as vegetation health.

Table 2 provides a compilation of the datasets that were used throughout this study to derive different sets of independent variables including the socio-economic variables (ethnicity and median household income) used in comparison to the modeled temperature surfaces:

Table 2
List of datasets to derive independent variables

DATASET	SOURCE	DERIVED DATA	SPATIAL RESOLUTION	YEAR
LiDAR	USGS - State of Iowa	Vegetation height, vegetation volume, building height, building volume	2 points per sqm	2019 - 2021
NAIP	USDA	NDVI	60cm	2019 - 2021
Building Footprints	Various local government sources	Building polygon	Various	Various
Socio-economic	U.S. Census Bureau	Income, Ethnicity	(Data by block group)	2021

Temperature Modelling

Data from the temperature collection and the urban landscape metrics derived through geoprocessing of primary sources from Table 2 were organized based on location to create comprehensive ESRI File Geodatabases per urban area. The geodatabases allowed the development of spatial models, described below, which were applied on a spatial resolution of one square meter per study area and time periods (afternoon, evening, and night). The models basically took the collected temperature point data as the dependent variable while using the derived urban morphometric data surrounding the points as independent variables to extrapolate spatial models that estimate urban temperatures at those urban locations where no temperatures were recorded. Although several techniques can be used to create such models and to analyze and predict urban heat patterns, the most common methods found in the reviewed literature were MLR, CART, RF, and sometimes a combination of multiple methods in the same study. Even though multiple statistic techniques/algorithms have been used to develop temperature models, recent studies utilizing mobile devices for temperature data collection showed that RF significantly outperformed MLR and CART (Voelkel and Shandas 2017; Shandas, et al. 2019; Oukawa et al. 2022; Chen et al. 2022).

Recent advances in spatial statistical modeling techniques have been built into embedded tools in the software ArcGIS Pro (3.0.1). Thus, for this study, ArcGIS Pro was used for most of the steps in both the pre-processing of dependent and independent variables and also in spatial modeling. The Forest-based Classification and Regression (nomenclature utilized by ArcGIS Pro for RF algorithm) model built into the Spatial Statistics toolbox in ArcGIS Pro was chosen for air temperature modeling. Random forest is a supervised machine learning method developed by Leo Breiman and Adele Cutler (ESRI 2022). Random forests are a non-parametric and nonlinear machine learning technique and are a "combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest" (Breiman 2001, 1). In addition, models generated using a random forest algorithm not only provided the predicted temperature in areas where temperature data was not collected but also indicated the most influential independent variables that explain temperature patterns throughout each individual study area.

Natural and built-up structures (e.g., canopy cover, building height, building volume) surrounding the area in which temperature data was collected vary and thus statistical values for those characteristics were calculated around each site with fixed neighborhood distances of 50, 100, 200, 400, and 800 meters. These values were captured in the form of raster surfaces utilizing the Focal Statistics tool in ArcGIS Pro. This methodology has been utilized previously when using RF for modeling, as surrounding urban conditions provide a better understanding of the spatial relationship between different features on the variation of temperature (Voelkel and Shandas 2017; Chen et al. 2022; Oukawa 2022).

Systematic efforts such as consistent mobile devices check (before temperature data collection), post-temperature collection data filtering (e.g., eliminating points collected when vehicles were stationary), and a LiDAR overlap classification (discarding multiple LiDAR points collected for the same area) were performed to control for potential errors or inconsistencies in both the collected temperature measurements and geospatial data used to derive independent variables. These efforts sought to identify malfunctioning devices, to isolate duplicate data that could be generated throughout data collection, or abnormalities that can be presented in LiDAR raw data point cloud datasets (.las).

A hold-out method (which are part of the settings existent within Forest-based Classification and Regression tool in ArcGIS Pro) for model evaluation (Oukawa et al. 2022; and Voelkel and Shandas 2017), which separates the temperature dataset in two sets was used to assess the accuracy of the model, was used, which according to Berrar

(2019) is one of the simplest data resampling strategies and capable of reliably estimate errors of the model for unseen/new cases.

Independent Variables product development through geoprocessing

A significant effort was necessary to geoprocess a variety of spatial data in order to produce files that served as independent variables for this study. Canopy Cover (CC) and Canopy Density Metric (CDM) were directly derived from lidar point cloud data originating from the United State Geological Survey (USGS). A total of 1,110 Lidar 1x1 km tiles (~36 gb), with a point spacing from 0.233 to 0.649 m stored in LAZ format were downloaded through an automated Python script. A combination of LiDAR data and building footprints provided by counties and cities was used to derive two other independent variables used for the model: Building Volume (BV) and Building Height (BH).

Normalized Difference Vegetation Index, known as NDVI, was derived from the 2021 United States Department of Agriculture National Agricultural Imagery Program (NAIP) aerieal imagery and sourced via the geodata.iowa.gov web application.

Metrics on all of the above independent variables were also calculated based on varying neighborhood distances (similar to buffers) to investigate the influence of the surrounding environments in relation to each measured temperature point similar to other studies such as Voelkel and Shandas (2017). In order to define an overall processing area for each set of temperature measurements, a 1,500 m buffer was calculated in ArcGIS Pro (Buffer tool) around the measured temperature points. The decision to set an area of interest at 1,500 meters, despite the focal statistics tool using distances no larger than an

800-meter radius, was made to accommodate potential future projects requiring a broader neighborhood distance or to address unexpected errors during the data processing phase.

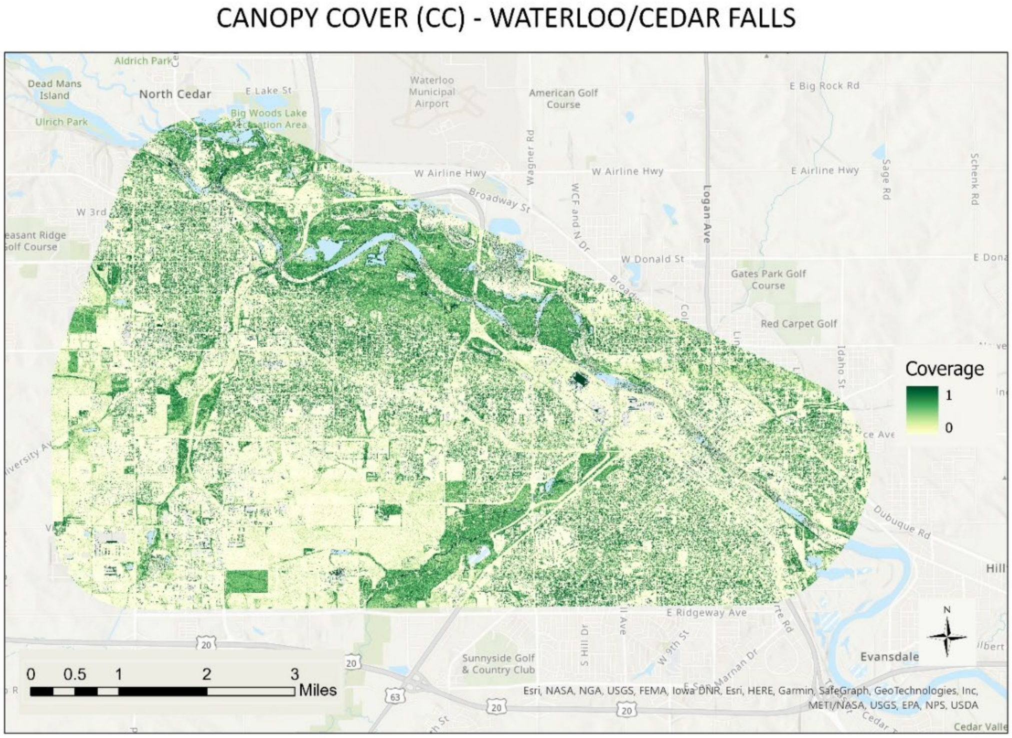
Automated geoprocessing was carrried out using Python scripting utilizing Arcpy; which is a site package that allows all ArcGIS Pro geoprocessing functionality to be accessed via Python. Specifically all LAZ files were converted to LAS files, subsequently creating a LAS Dataset for each urban area and its respective tiles. The LAS Dataset allows for the raw data to be stored in LAS files but through an indexing scheme, allowing for faster processing time as well as allowing for contiguous areas to be analyzed by different tools individually. In addition, a backup of all LAS files was created to prevent unnecessary reprocessing due to possible errors that could occur during the LAS classification stage.

The first step to classifying aerial lidar points was to filter and eliminate points that are present within a distance shorter than the nominal point spacing due to collections from different flight lines. After filtering overlapped points, LAS files were then classified into 4 main classes. The main classes are ground points, building points, noise points, and vegetation points. Vegetation points, classified through the use of the tool named Classify Las by Height in ArcGIS Pro, was subdivided into 3 more classes: low vegetation (up to 5 meters tall), medium vegetation (up to 25 meters tall), and high vegetation (over 25 meters tall). Even though height classification as well as the number of classes can vary according to the researcher, for this project, the vegetation height was not an influence on the model since vegetation type was not used as an independent variable.

With the exception of LAS Noise, all other classes were used to derive subsequential geospatial raster datasets. Ground points were used as base to generate Digital Elevation Models (DEMs), which in conjunction with building and vegetation points were used to derive a Digital Surface Model (DSM). Subtracting the DSM raster values from DEM generated a Normalized Digital Surface Model (NDSM), which is the difference between the built-up and vegetated surface and the bare ground.

To generate an estimate of vegetation density, which is used to create a measure of Canopy Cover (CC), one of the independent variables of the model, points classified as vegetation were used as well as points classified as ground. The ratio of vegetation points to ground points in a specific area yields the Canopy Cover (Figure 8), expressed as a percentage ranging from zero (non-vegetated area) to one (completely vegetated area). The maps of independent variable spatial data displayed below would be similar for each city but an example from the Waterloo/Cedar Falls area is provided.

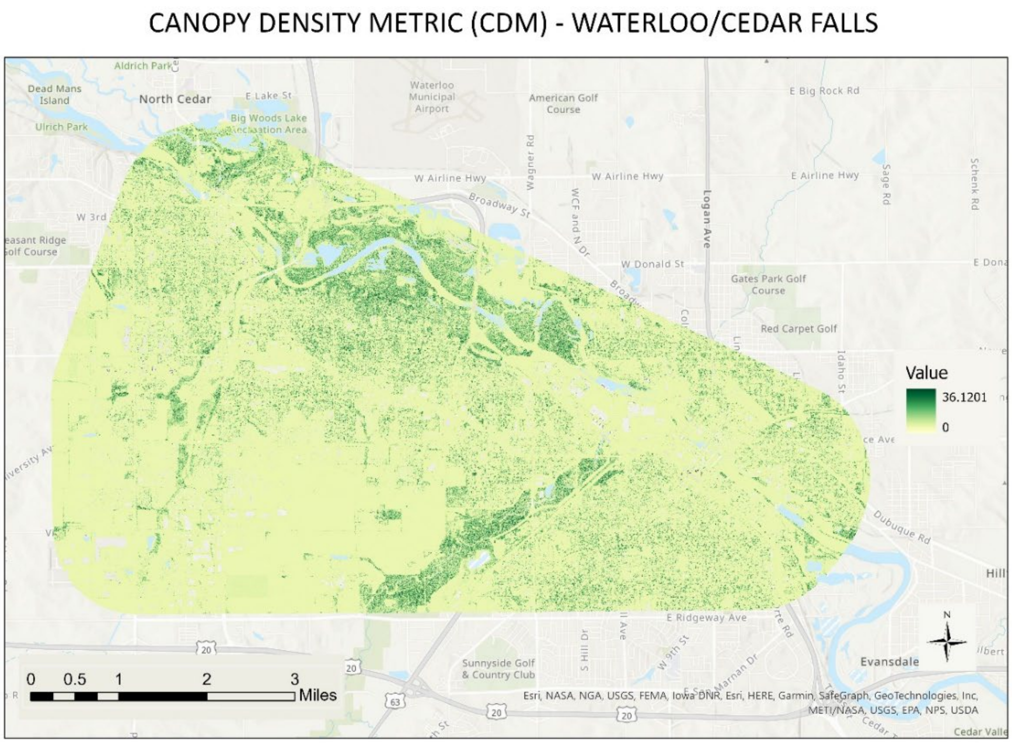
Figure 8
Illustration of Canopy Cover in Waterloo/Cedar Falls



Following a concept used in Voelkel and Shandas (2017), which considers not only the vegetation coverage of a given area but also its height, a new variable denominated as Canopy Density Metric was created and was derived from multiplying Canopy Cover by NDSM. Even though the calculation evokes the idea of volume, it is inaccurate to use the term since the result doesn't consider the individual and accurate shape of each object, therefore, the results for CDM are expressed only as a unitless measure displayed as Value, which can be seen in Figure 9.

Figure 9

Illustration of Canopy Density Metric in Waterloo/Cedar Falls



LiDAR points classified as buildings (rooftop points), and building footprints vector files provided by counties, cities, and public repositories (Microsoft 2018) were used to derive two variables that were used for the final model: Building Height (Figure 10) and Building Volume (Figure 11). The steps utilized to generate both variables were identical to what was used to generate the variable named Canoy Cover, identifying what points were classified as the goal feature (buildings points in this case), with the addition of building footprints vector files, which were used to reshape the surface area given that LiDAR points are not homogenously available for each built-up structure area

Figure 10

Illustration of Building Height (BH) in Waterloo/Cedar Falls

BUILDING HEIGHT (BH) - WATERLOO/CEDAR FALLS

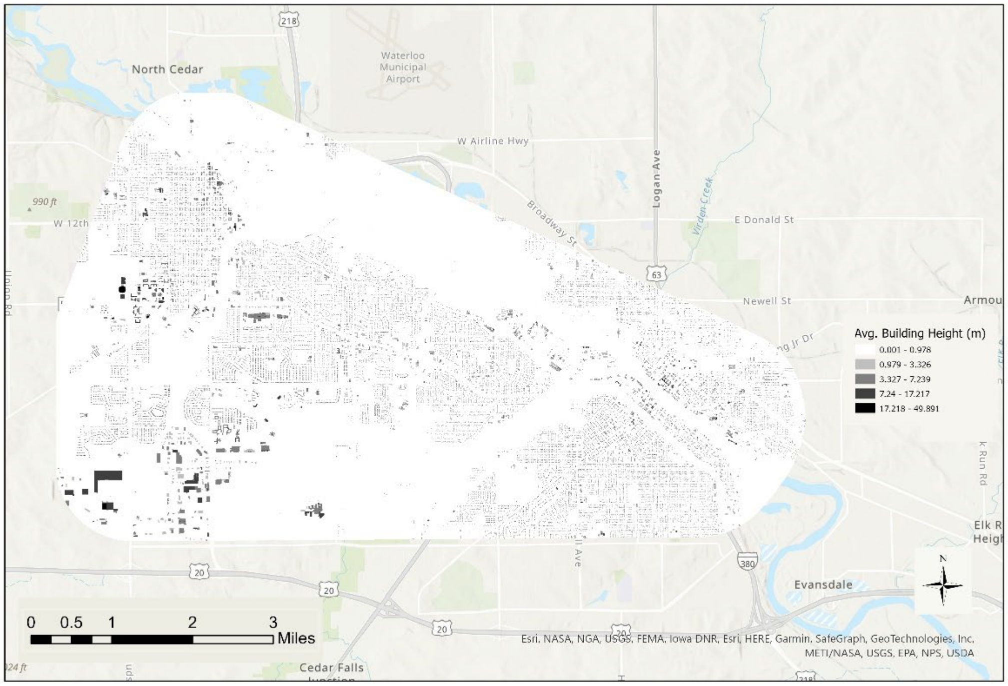
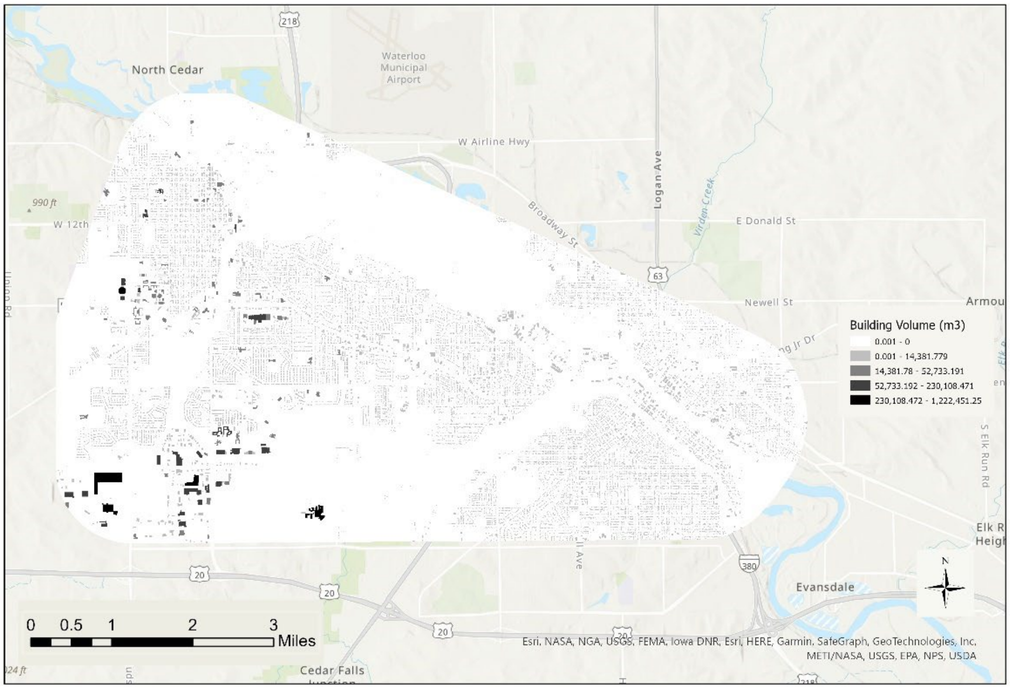


Figure 11

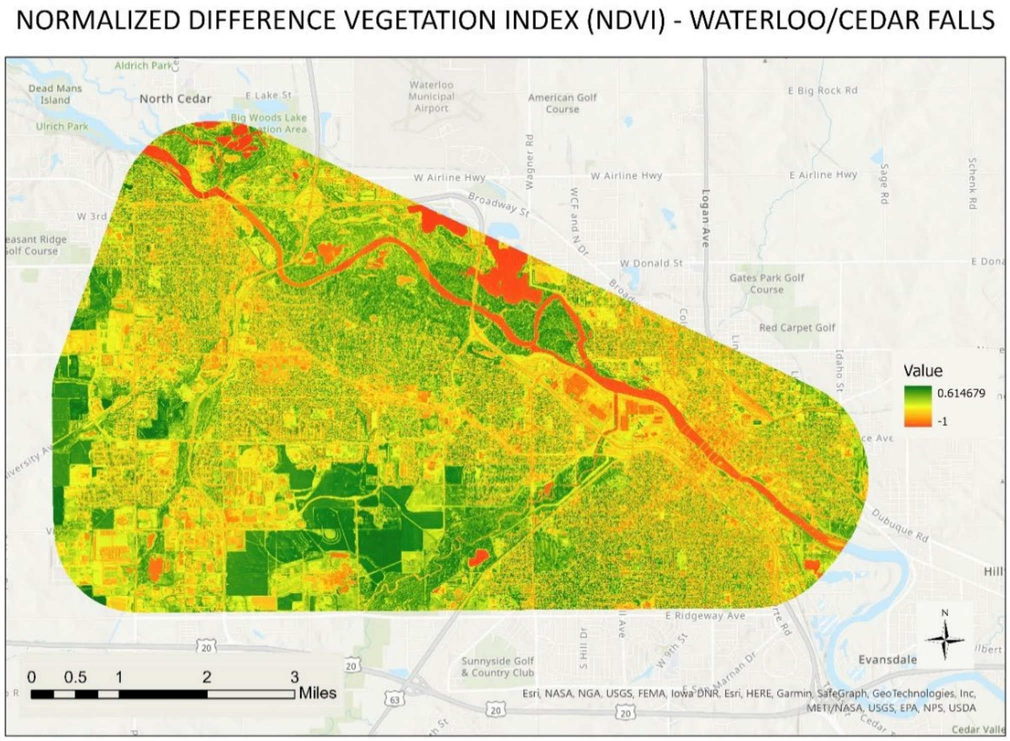
Illustration of Building Volume (BV) in Waterloo/Cedar Falls

BUILDING VOLUME (BV) - WATERLOO/CEDAR FALLS



To generate the fifth independent variable (NDVI), a further raster calculation of the multiple bands originating from NAIP imagery was needed. Since the downloaded imagery was presented as a four-band imagery (CIR + NIR), a simple raster calculation ((NIR – Red)/(NIR + Red)) was performed, where NIR is the Near Infrared channel/band, and Red is the Red channel/band. The low values in NDVI (e.g. < 0.1) reflect areas with little vegetation (e.g. water, barren land) and higher values (e.g. > 0.3) where there is a significant amount of vegetation (NASA 2000). An example of the calculated NDVI can be seen in Figure 12, which represents the area of Waterloo/Cedar Falls.

Figure 12
Illustration of NDVI in Waterloo/Cedar Falls

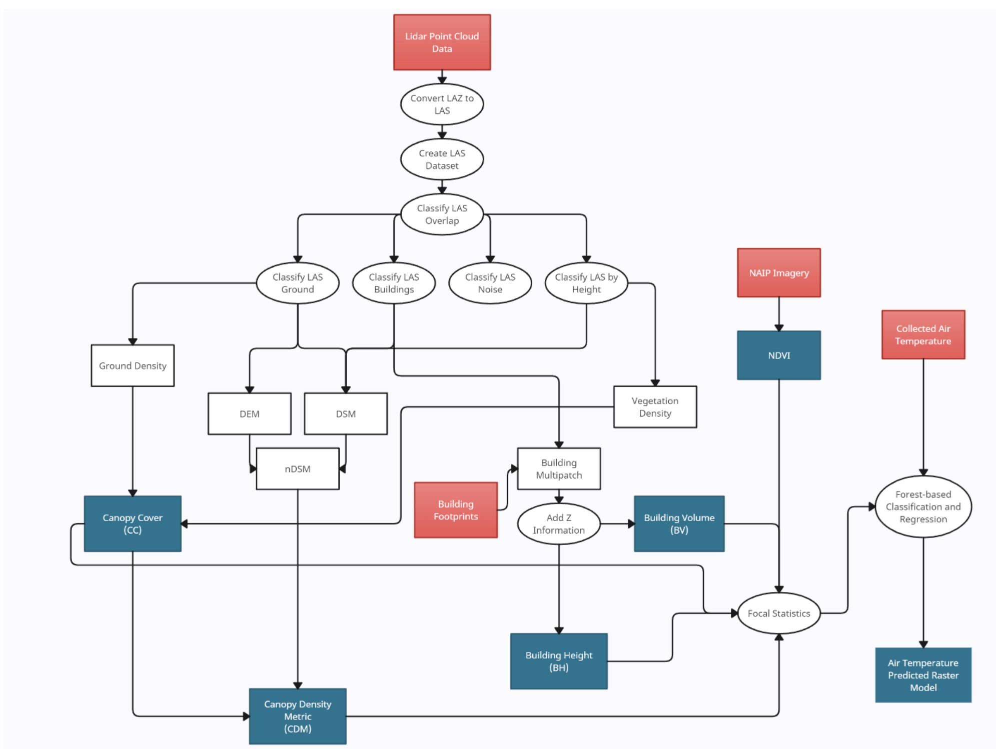


In order to evaluate the natural and morphological features surrounding measured temperature points, raster datasets were created for the five independent variables and five neighborhood distances (50, 100, 200, 400, and 800 meters) using the Focal Statistics tool. In addition, not one but two statistic types were used for each single variable and neighborhood distance: MEAN, which represents the average value of the cells in the defined neighborhood, and STD, which represents the standard deviation of the cells in the neighborhood. This method has been successfully implemented in other studies (e.g., Voelkel and Shandas 2017), providing alternatives for how values (based on different features) are distributed across given areas. A total of 50 different raster files were generated for each urban area (five variables * five neighborhood distances * two statistical measures).

A workflow of the process to derive the necessary raster datasets used as independent variables used in the Random Forest model can be seen in Figure 13. Ellipsoids represent functions and geoprocessing tools, rectangular red-filled objects represent the input datasets, rectangular white-filled objects represent the outcome files necessary to generate the final independent variables used in the Random Forest modeling, and rectangular blue-filled objects represent the generated independent variables and final outcome (predictive models).

Figure 13

Workflow to derive the final air temperature predicted raster models using Random Forest algorithm



To process the multiple datasets, three computers with two different hardware configurations were used. Computer one was powered by an AMD Ryzen 5 5600H (6 cores), 64GB of RAM, and a Nvidia RTX 3060 6GB. Computers two and three were powered by an Intel I7 12700 (12 cores), 32Gb RAM, and an Nvidia T400 4GB. All computers utilized NVME storage drives which significantly reduced the read and write speeds of different files necessary for the model. In addition, all computers had the same version of ArcGIS Pro (v3.03) at the time of processing.

The time taken to process all data varied across urban areas, either due to area size, or due to the differences in the quantity of features in each area (e.g., more rooftop points that needed to go through 3d multipatch tools). In addition, it is acknowledged that the total number of points derived from LiDAR data might have strongly influenced the necessary time to process each dataset since point spacing varied from 0.233 to 0.649 across multiple datasets. The total amount of time spent processing all data, not considering software errors, was 208 hours. Table 3 shows the total number of LiDAR points per urban area, the number of LiDAR tiles, the total storage of the processed data (not considering backup), and the time necessary to process the workflow.

Table 3
Amount of data and processing times for geospatial modeling

URBAN AREA	NUMBER OF LIDAR TILES	NUMBER OF LIDAR POINTS	TOTAL FINAL STORAGE SIZE	PROCESSING TIME
Burlington	73	617,284,346	44.5 GB	17 hours
Cedar Rapids	175	1,582,011,104	112.1 GB	37 hours
Council Bluffs	114	459,717,912	57.1 GB	19 hours
Des Moines	293	1,024,741,353	160.4 GB	52 hours
Fort Dodge	94	492,615,809	50.7 GB	16 hours
Marshalltown	78	258,124,400	35.4 GB	10 hours
Sioux City	93	719,962,046	48.9 GB	18 hours
WaterlooCF	119	1,108,188,995	77.6 GB	23 hours
Waverly	71	575,302,059	46.4 GB	16 hours
TOTAL	**1110**	**6,837,948,024**	**633.1 GB**	**208 hours**

The parameters set in RF for all models consisted of 1000 randomized trees built from all effective focal distance raster datasets to predict a new surface raster (air temperature values) in each urban area in the afternoon, evening, and night. A 70/30 hold-out method for model evaluation was used to assess the accuracy of all models. This technique separates the temperature measurements into two sets (training set and testing set) and was set as the percentage of data excluded for validation (=30%) before running each one of the models. Finally, only one number of runs was used as a validation option.

The RF model originally considered one dependent variable (predicted Air Temperature) and 50 different independent variables based on five variables (CC, CDM, BH, BV, NDVI) each calculated with two different statistical measurements (mean and standard deviation) at five different neighborhood distances (50 meters, 100 meters, 200 meters, 400 meters, and 800 meters) . However, Building Volume (BV) calculated at all neighbor distances as standard deviation (BV_STD) showed noticeable signs of banding (straight lines across the whole area of interest) with no calculated values, and thus was excluded from the modeling. Two other variables, Canopy Cover (CC), and Canopy Density Metric (CDM) presented the same banding issue using standard deviation at distances under 400 meters ((CC_STD<400m) and (CDM_STD<400m)). Their standard deviations computed at 50 meters, 100 meters, and 200 meters were also discarded from the model for all urban areas. Therefore, only 39 different independent variables were used for the final model in all urban areas (Table 4).

Table 4

List of independent variables utilized for Random Forest models

LIST OF INDEPENDENT VARIABLES (RASTER FILES) USED AS INPUT FOR RANDOM FOREST (RF)	
INDEPENDENT VARIABLE	**ALIAS**
Mean Building Height at 50m	BH_Mean50
Mean Building Height at 100m	BH_Mean100
Mean Building Height at 200m	BH_Mean200
Mean Building Height at 400m	BH_Mean400
Mean Building Height at 800m	BH_Mean800
Standard Deviation Building Height at 50m	BH_STD50
Standard Deviation Building Height at 100m	BH_STD100
Standard Deviation Building Height at 200m	BH_STD200
Standard Deviation Building Height at 400m	BH_STD400
Standard Deviation Building Height at 800m	BH_STD800
Mean Building Volume at 50m	BV_Mean50
Mean Building Volume at 100m	BV_Mean100
Mean Building Volume at 200m	BV_Mean200
Mean Building Volume at 400m	BV_Mean400
Mean Building Volume at 800m	BV_Mean800
Mean Canopy Cover at 50m	CC_Mean50
Mean Canopy Cover at 100m	CC_Mean100
Mean Canopy Cover at 200m	CC_Mean200
Mean Canopy Cover at 400m	CC_Mean400
Mean Canopy Cover at 800m	CC_Mean800
Standard Deviation Canopy Cover at 400m	CC_STD400
Standard Deviation Canopy Cover at 800m	CC_STD800
Mean Canopy Density Metric at 50m	CDM_Mean50
Mean Canopy Density Metric at 100m	CDM_Mean100
Mean Canopy Density Metric at 200m	CDM_Mean200
Mean Canopy Density Metric at 400m	CDM_Mean400
Mean Canopy Density Metric at 800m	CDM_Mean800
Standard Deviation Canopy Density Metric at 400m	CDM_STD400
Standard Deviation Canopy Density Metric at 800m	CDM_STD800
Mean NDVI at 50m	NDVI_Mean50
Mean NDVI at 100m	NDVI_Mean100
Mean NDVI at 200m	NDVI_Mean200
Mean NDVI at 400m	NDVI_Mean400
Mean NDVI at 800m	NDVI_Mean800
Standard Deviation NDVI at 50m	NDVI_STD50
Standard Deviation NDVI at 100m	NDVI_STD100
Standard Deviation NDVI at 200m	NDVI_STD200
Standard Deviation NDVI at 400m	NDVI_STD400
Standard Deviation NDVI at 800m	NDVI_STD800

Socio-demographic Data and Analysis

To examine whether high temperatures are located in the same areas as populations with specific demographics characteristics, income, and ethnicities/race data per census block group were compared to all the average modeled (predicted) temperature raster through the use of Pearson correlation coefficient and Linear Regression. The socioeconomic dataset was acquired through the American Community Survey from 2016 – 2021 (5-year estimate) administered and published by the United States Census Bureau and released on December 8th, 2022 (Census 2022). The aforementioned techniques allowed for the author to examine whether low-income and specific ethnic groups were statistically positively correlated to modeled higher temperatures in urban areas or not. The average modeled temperature was calculated by using Zonal Statistics as Table tool in ArcGIS Pro. This tool allowed for the calculation of statistical measures, such as the mean of modeled temperature in a given census unit which could then be compared statistically to the socioeconomic variables.

Chapter 4

Results

Chapter 4 describes the results obtained throughout the temperature data collection that occurred from May 2022 to September 2022; the spatial statistical predictive model respective results; and the socioeconomic analysis which seeks to explore the variation of temperature in different urban neighborhoods with varied socio-demographics.

Temperature data collection

Temperature data were collected in seven different urban areas at three consistent times of the day, referred to as afternoon (4 – 5 p.m.), evening (9 – 10 p.m.), and night (4 – 5 a.m.): Burlington, Cedar Falls, Cedar Rapids, Council Bluffs, Des Moines, Sioux City, and Waterloo. In three other cities, data were collected only in two periods, either due to logistical constraints (Waverly and Fort Dodge), or extreme weather conditions during the day of collection (Marshalltown).

Table 5 depicts the summary of dates, number of routes, time of collection, and number of measurements taken in each one of the study areas. The number of routes (sometimes referred to as a run) was somewhat dictated by the size of the study area and indicates how many individual vehicles were used on that given date/time.

Table 5
Summary of runs for air Temperature data collection

URBAN AREA	DATE	NUMBER OF ROUTES	TIME OF COLLECTION	TOTAL MEASURMENTS TAKEN
Waverly	14-Jun	1	Afternoon, Evening	7,160
Cedar Rapids	20-Jun, 21-Jun	3	Afternoon, Evening, Night	23,918
Marshalltown	05-July, 06-July	1	Afternoon, Night	4,114
Sioux City	18-July, 19-July	2	Afternoon, Evening, Night	9,647
Fort Dodge	19-July	2	Afternoon, Evening	5,151
Des Moines	02-Aug, 03-Aug	3	Afternoon, Evening, Night	29,817
Council Bluffs	13-Aug, 14-Aug	1	Afternoon, Evening, Night	11,805
Burlington	05-Aug, 06-Aug	1	Afternoon, Evening, Night	10,315
Cedar Falls/Waterloo	20-Sept, 21-Sept	1	Afternoon, Evening, Night	11,299

To minimize the impact that changes in weather (abrupt drops or spikes in temperature, winds, etc.), each run was planned to be performed within 60 minutes. Routes were defined utilizing Google Maps (satellite view, Street View) as a visual aid in identifying areas with a variety of land use/cover (e.g., residential, commercial, industrial, with varying vegetative and impervious cover). The temperature collection was scheduled on days that temperatures were expected to reach at least 90°F (32.2 °C) and sought to not only keep the expected run time of one hour but also to explore different natural and human-influenced land use in each urban area, going from vegetated parks and areas surrounded by bodies of water to residential, commercial, and industrial areas with varying levels of building and vegetation density. In specific circumstances, such as heavy traffic, road construction, or technical problems, some routes ended up slightly longer than one hour, which was the case for the Council Bluffs area (afternoon) with the longest route totaling 70 minutes and 55 seconds.

Considering the number of vehicles involved in all three times of the day and all urban areas where data was collected, the total number of individual collections (vehicle per route per time of collection per urban area) was 39. Out of the 39 individual collections, faulty/broken devices were seen on three occasions (7.7%). In one out of three situations, a backup device was able to replace the broken apparatus. In the other two cases, the devices didn't record/save the data that were being collected throughout the entire collection period, which was the case for the area of Cedar Rapids in one of the routes during the afternoon, and evening. In addition, a software update was required during the summer 2022, which led to an inadvertent change in data collection to every two seconds, as opposed to every one second. This change affected recordings in Fort Dodge, Marshalltown, and Sioux City. As further examinations suggested, the data was recorded correctly, even though less frequent measurements were acquired.

Table 6 shows the distribution of all 113,226 measurements taken in all urban areas and the number of points collected by all the devices at different times of the day.

Table 6
Total measurements in all urban areas

URBAN AREA	AFTERNOON	EVENING	NIGHT
Waverly	3,946	3,214	-
Cedar Rapids	7,275	7,341	9,302
Marshalltown	2,259	-	1,855
Sioux City	3,209	3,143	3,295
Fort Dodge	2,659	2,492	-
Des Moines	10,745	9,721	9,351
Council Bluffs	3,706	3,843	4,256
Burlington	3,444	3,423	3,448
Cedar Falls/Waterloo	4,120	3,479	3,700
Total	**41,363**	**36,656**	**35,207**

To avoid multiple measurements at the same location, data that were recorded when vehicles were stationary were excluded from the dataset. Table 7 shows the percentage of data that showed vehicle speeds = 0 therefore, being excluded from the dataset.

Table 7

Percentage of excluded data due to speed

URBAN AREA	% OF DATA RECORDED WHEN VEHICLES WERE STATIONARY
Burlington	-2.27%
Cedar Rapids	-3.64%
Council Bluffs	-4.11%
Des Moines	-3.59%
Fort Dodge	-4.39%
Marshalltown	-2.67%
Sioux City	-6.53%
WaterlooCF	-4.34%
Waverly	-2.74%

Contrary to some of the literature (Saaroni et al. 2000; Wong and Yu 2005; Voelkel and Shandas 2017), vehicle speeds ranging from non-stationary speeds to 55.11 mph showed no influence on air temperature when tested using linear regression of all measured temperatures using the software TIBCO Spotfire S+, results presented an R^2 = 0.0008034, and a p-value of 0.000 (Figure 14). Although the results were statistically significant by p-value, given the speeds at which the vehicles were driven (average of less than 20 miles per hour), each one-mile-per-hour increase in speed would reduce the temperature by only 0.015 Celsius degrees, which is considered a negligible variance in temperature when accounting the accuracy of the devices and driving speeds.

Figure 14

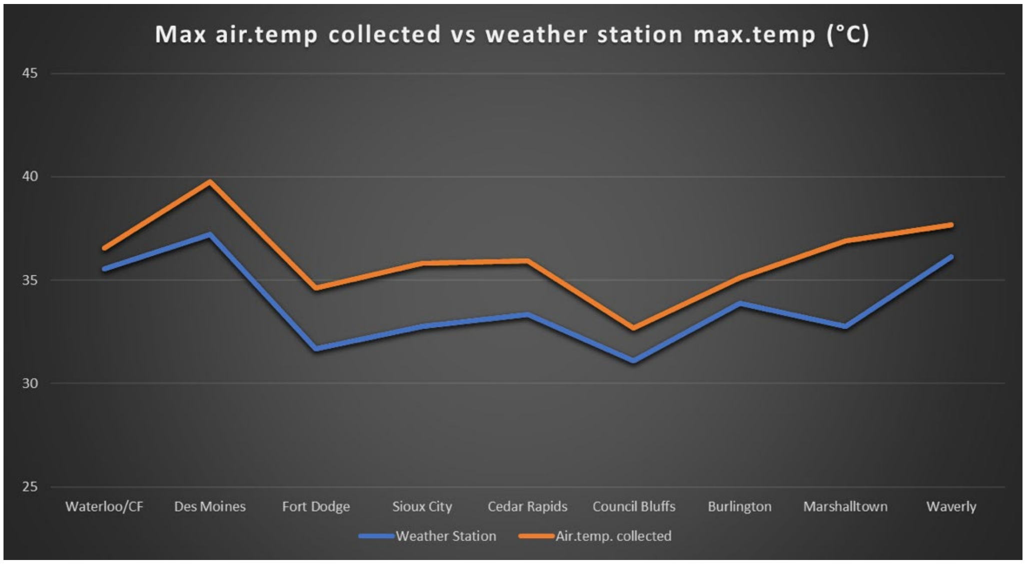
Linear regression for measured temperatures vs speed

```
      *** Linear Model ***

Call: lm(formula = airtemp ~ GPSspeed, data = Alltemp, na.action = na.exclude)
Residuals:
   Min     1Q  Median    3Q   Max
 -11.1 -3.989 -0.3791 4.121 9.914

Coefficients:
               Value Std. Error  t value  Pr(>|t|)
(Intercept)  29.9106    0.0332  900.9784    0.0000
   GPSspeed  -0.0151    0.0016   -9.4111    0.0000

Residual standard error: 4.768 on 108913 degrees of freedom
Multiple R-Squared: 0.0008126      Adjusted R-squared: 0.0008034
F-statistic: 88.57 on 1 and 108913 degrees of freedom, the p-value is 0
```

As a general reference, a comparison was made between measured data and those collected by locally established weather stations. When comparing the maximum recorded temperature for each urban area to the closest weather station's maximum measured air temperature for the day, the variation ranged between 1°C in Waterloo/CF (minimum difference) and 4.1°C in Marshalltown (maximum difference). In all instances, the lower temperature was recorded by weather stations (Figure 15). The local weather stations are often located in an open area without nearby buildings such as an airport, which could explain the variation of temperature between the stations and the mobile devices.

Figure 15

Maximum air temperature collected by mobile sensors vs. weather stations



The variation of air temperature in the period of one hour, regardless of the time of the day, never exceeded 5.66 °C, which was the case in Waterloo/CF area during the evening. Four out of eight urban areas (Burlington, Council Bluffs, Des Moines, and Waverly), experienced a higher thermal variation during the afternoon (compared to evening), four areas (Cedar Rapids, Fort Dodge, Waterloo, and Cedar Falls) experienced a higher thermal variation during the evening, and only one area (Sioux City) experienced the highest variation in temperature among all times of the day during the night period as seen in Table 8.

Table 8

Amplitude of air temperature measured in urban areas

URBAN AREA	AMPLITUDE OF AIR TEMPERATURE AFTERNOON	AMPLITUDE OF AIR TEMPERATURE EVENING (°C)	AMPLITUDE OF AIR TEMPERATURE NIGHT (°C)
Burlington	3.71	2.52	1.90
Cedar Rapids	3.47	3.52	3.21
Council Bluffs	4.80	3.20	2.93
Des Moines	4.15	3.16	2.35
Fort Dodge	4.26	5.51	-
Marshalltown	4.11	-	1.09
Sioux City	3.10	3.25	4.01
WaterlooCF	4.22	5.66	1.05
Waverly	2.93	2.25	-

The minimum and maximum temperatures, as well as the average recorded temperature of each urban area can be seen in tables 9, 10, and 11.

Table 9

Minimum measured temperature in urban areas

URBAN AREA	MINIMUM AIR TEMPERATURE AFTERNOON (°C)	MINIMUM AIR TEMPERATURE EVENING (°C)	MINIMUM AIR TEMPERATURE NIGHT (°C)
Burlington	31.41	28.47	25.92
Cedar Rapids	32.50	25.95	22.36
Council Bluffs	27.84	23.40	18.66
Des Moines	35.62	29.75	26.95
Fort Dodge	30.34	22.09	-
Marshalltown	32.79	-	21.33
Sioux City	32.75	27.15	22.04
WaterlooCF	32.35	23.15	22.14
Waverly	34.73	28.27	-

Table 10

Maximum measured temperature in urban areas

URBAN AREA	MAXIMUM AIR TEMPERATURE AFTERNOON (°C)	MAXIMUM AIR TEMPERATURE EVENING (°C)	MAXIMUM AIR TEMPERATURE NIGHT (°C)
Burlington	35.12	30.99	27.82
Cedar Rapids	35.97	29.47	25.57
Council Bluffs	32.64	26.60	21.59
Des Moines	39.77	32.91	29.30
Fort Dodge	34.60	27.60	-
Marshalltown	36.90	-	22.42
Sioux City	35.85	30.40	26.05
WaterlooCF	36.57	28.81	23.19
Waverly	37.66	30.52	-

Table 11
Average measured temperature in urban areas

URBAN AREA	AVERAGE AIR TEMPERATURE AFTERNOON (°C)	AVERAGE AIR TEMPERATURE EVENING (°C)	AVERAGE AIR TEMPERATURE NIGHT (°C)
Burlington	33.48	29.84	26.86
Cedar Rapids	34.10	28.07	24.12
Council Bluffs	29.86	25.05	20.40
Des Moines	37.33	31.44	28.47
Fort Dodge	32.39	25.16	-
Marshalltown	34.87	-	21.96
Sioux City	33.80	28.91	24.09
WaterlooCF	34.93	26.64	22.67
Waverly	35.93	29.81	-

In order to determine if temperature data was collected during unusually hot days, a brief analysis was conducted by comparing the temperature readings from mobile sensors on the observation day to the average high temperatures recorded over the past 10 years for both the specific month and day at the nearest local weather station. A total of seven out of 10 areas exceeded the 90th percentile, and five urban areas (Burlington, Cedar Falls, Des Moines, Waterloo, and Waverly) experienced the hottest day of the month.

Figures 16 to 39 display the routes and resulting measured temperatures displayed as the temperature in degrees Celsius (note, the upper and lower limit varies per map) in each urban area by the time of collection. The figures are presented in alphabetical order based on the urban area's name, and the time of collection follows a chronological order (afternoon to night).

Visually inspecting the temperatures collected in each urban area showed observations consistent with the literature. The lowest temperatures were usually collected close to or in highly vegetated areas, while the highest temperatures were often collected in areas such as downtown or with a large number of built-up structures (e.g., industrial parks).

This was observed in the case of Burlington (Figures 16 to 18), where the downtown area emerged as the hottest region. However, somewhat unexpectedly, the Mississippi River, located to the east of downtown, did not appear to exert a significant influence on temperature regulation. While it is well-known that large bodies of water aid in mitigating urban heat, and the density of built structures in the southern part of the city does not differ significantly from that in the downtown area, the absence of vegetation in the vicinity seemed to have the greatest impact when comparing the two regions. This disparity resulted in a substantial temperature difference, with the southernmost area exhibiting significantly cooler temperatures. The temperature patterns remained consistent for all times of collection.

Figure 16

Measured temperature in Burlington during the afternoon

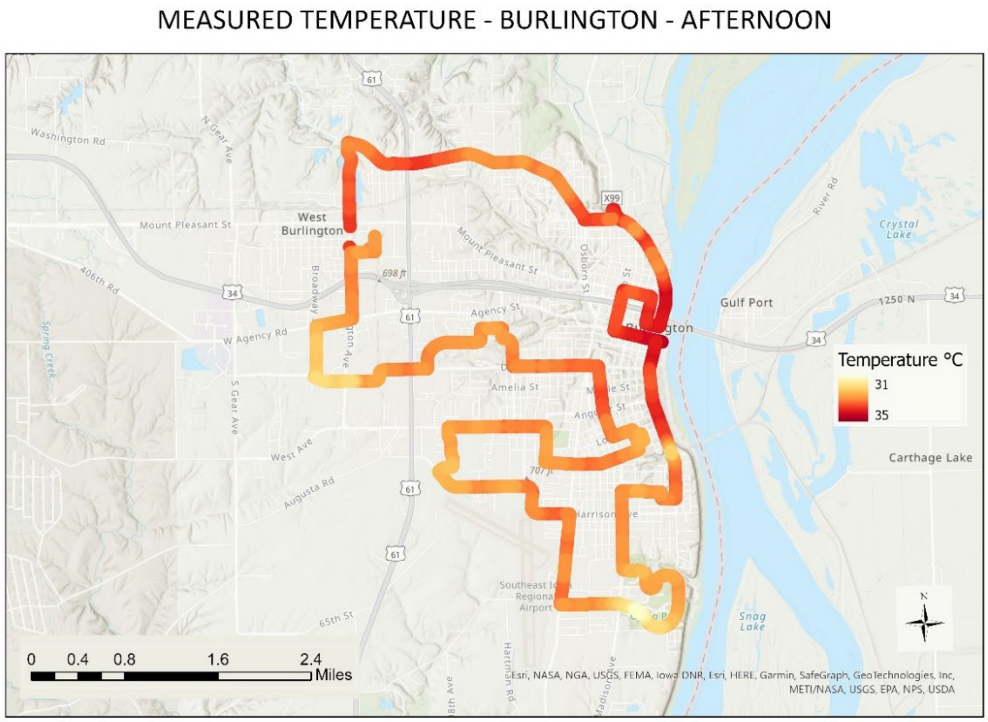


Figure 17

Measured temperature in Burlington during the evening

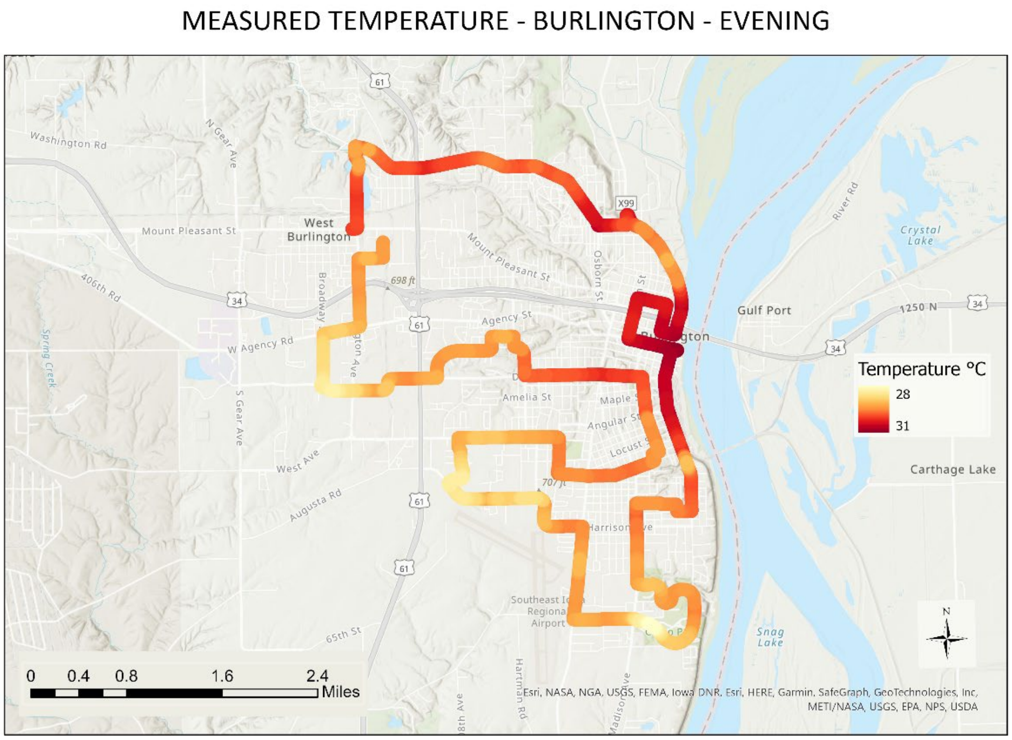
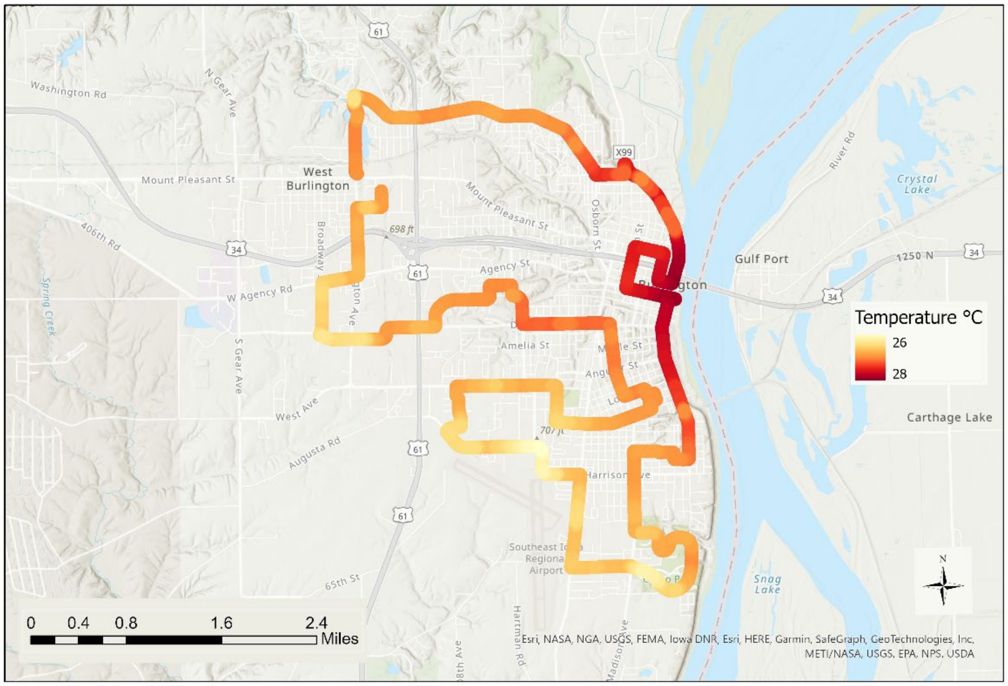


Figure 18
Measured temperature in Burlington during the night

MEASURED TEMPERATURE - BURLINGTON - NIGHT



Cedar Rapids (Figures 19 to 21) generally showed the same heat patterns at all times of collection, although the night period had one extra route (in the northeast area as seen in Figure 21). The areas close to downtown (especially in the night readings), and along Interstate 380 had the highest temperature measurements, regardless the time of collection.

Figure 19

Measured temperature in Cedar Rapids during the afternoon

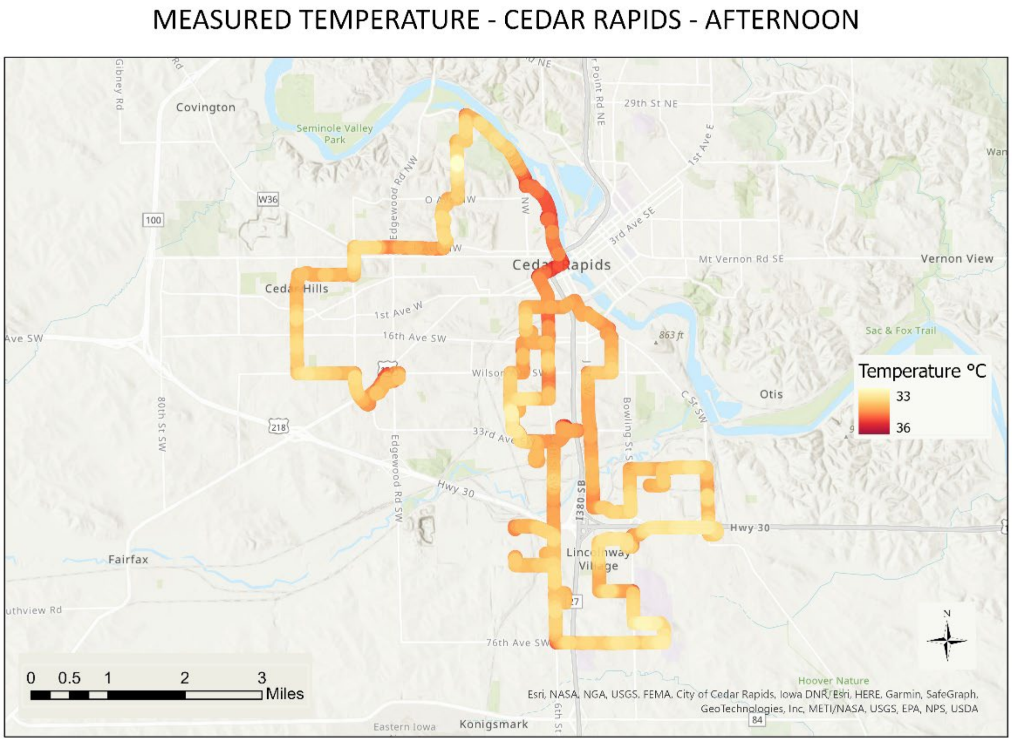


Figure 20

Measured temperature in Cedar Rapids during the evening

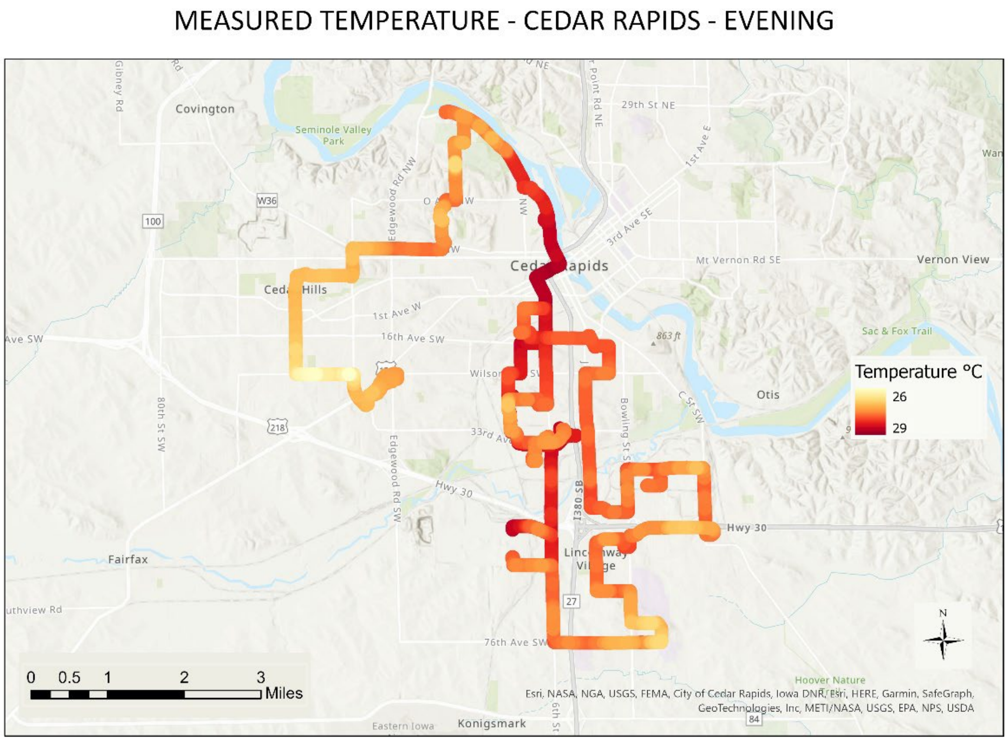
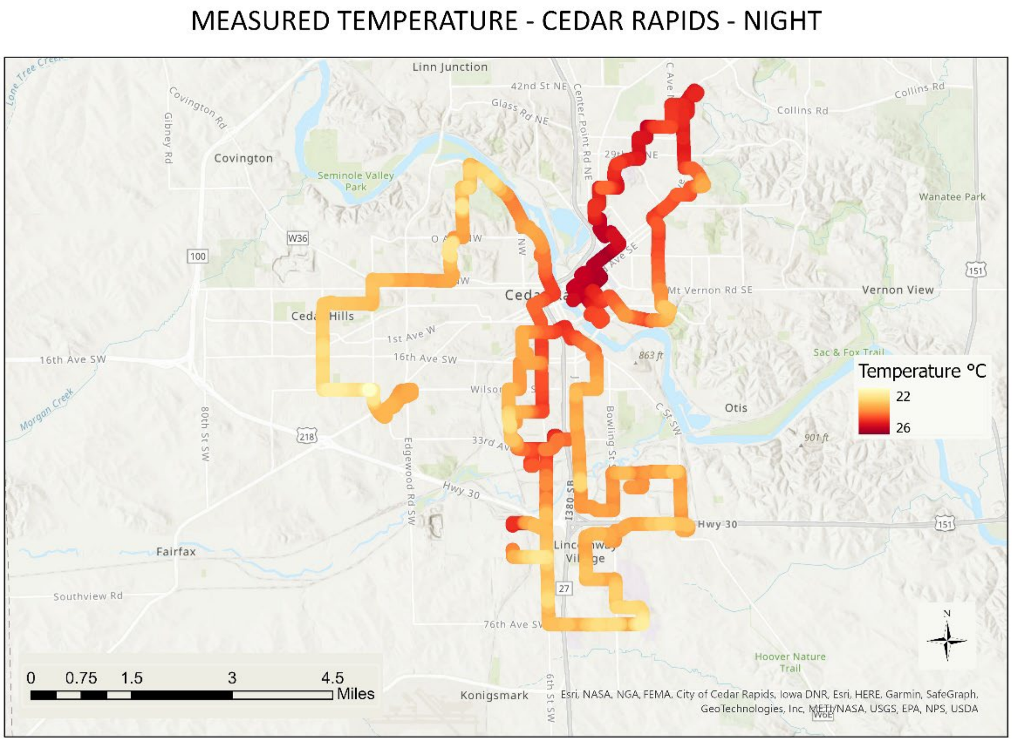


Figure 21
Measured temperature in Cedar Rapids during the night



In Council Bluffs (Figures 22 to 24) the coldest temperatures at all times of the day were seen in the northern part of the city. The results seem to be explained by higher altitudes (50+ meters compared to downtown) in addition to being highly vegetated areas. Interestingly, in the southern region of the city, regardless of the presence of a large body of water (Lake Manawa) and being the most vegetated area, temperatures were on par with the warmest areas of the town (downtown).

Figure 22

Measured temperature in Council Bluffs during the afternoon

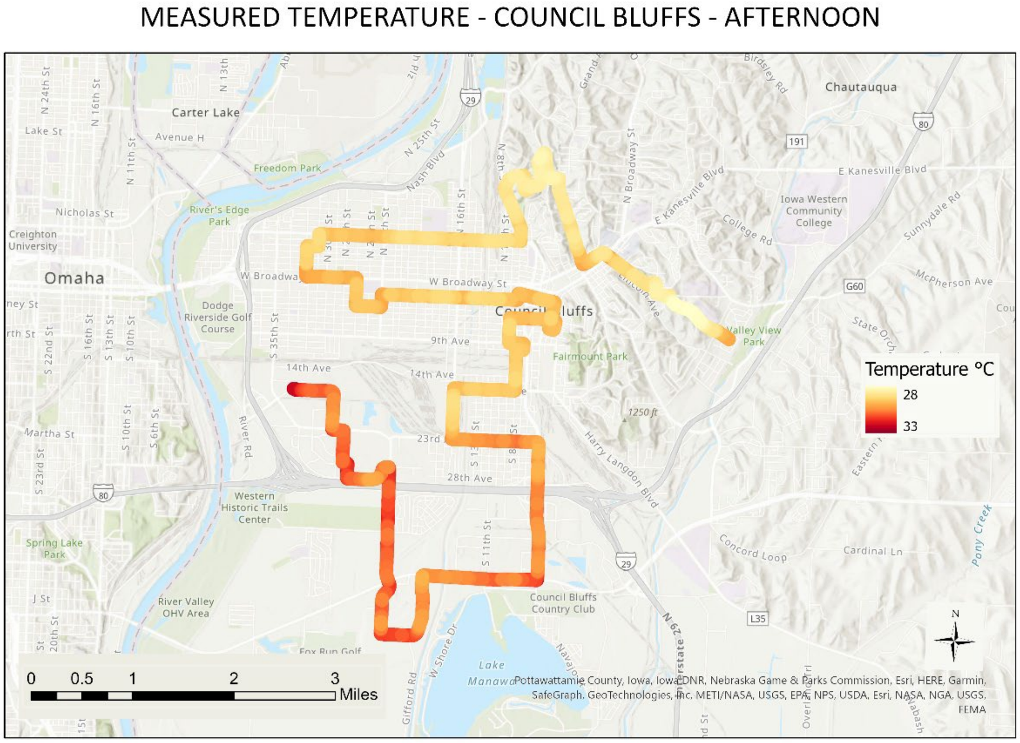


Figure 23

Measured temperature in Council Bluffs during the evening

MEASURED TEMPERATURE - COUNCIL BLUFFS - EVENING

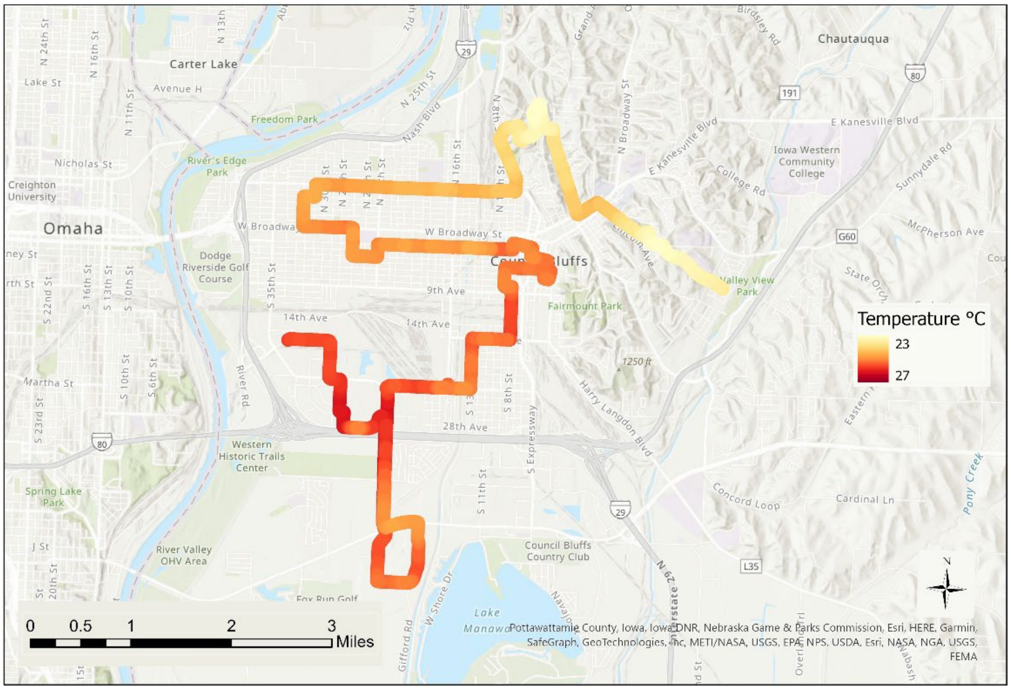
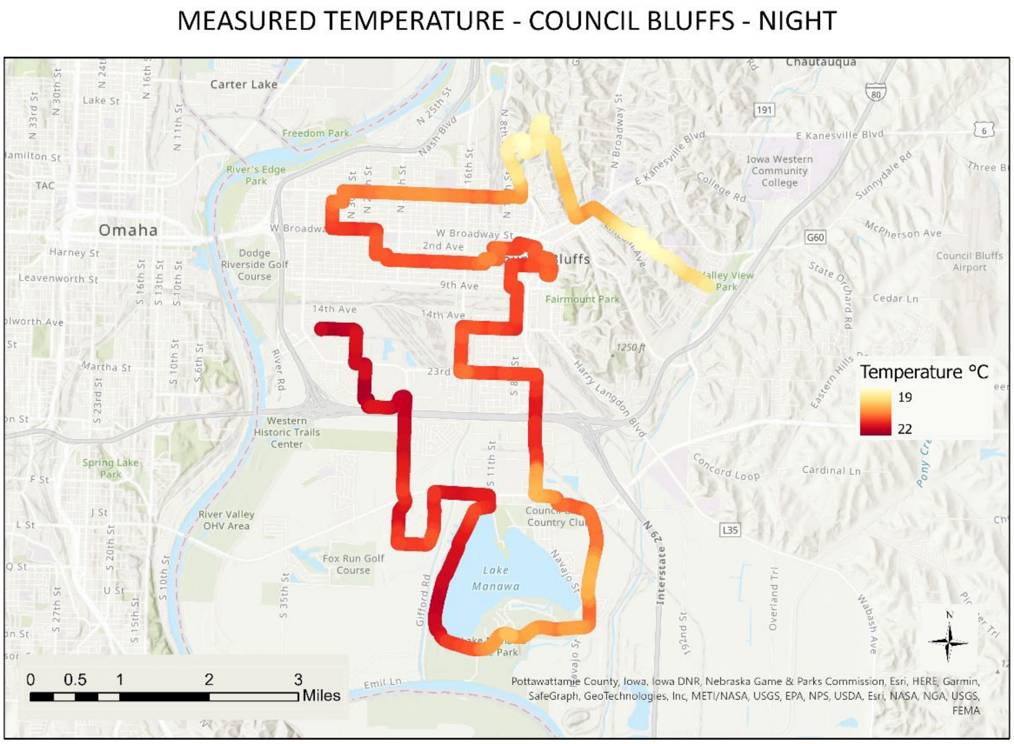


Figure 24

Measured temperature in Council Bluffs during the night



The capital Des Moines (Figures 25 to 27) seem to be a clear indication of the concept that large built-up structures, such as buildings, absorb heat during the day releasing it during the night through convection. While downtown, Capitol East, and areas close to Union Pacific Shortline Yard (train yard), were the warmest parts of the city throughout all times of collection, the heat in the same area seemed to be exacerbated (in area) by the number and volume of manmade structures during the night period, as can be seen in the Figure 26 and Figure 27. The lowest temperatures, however, were collected in areas such as Southwestern Hills close to Des Moines International Airport, and Des Moines Water Works Park.

Figure 25

Measured temperature in Des Moines during the afternoon

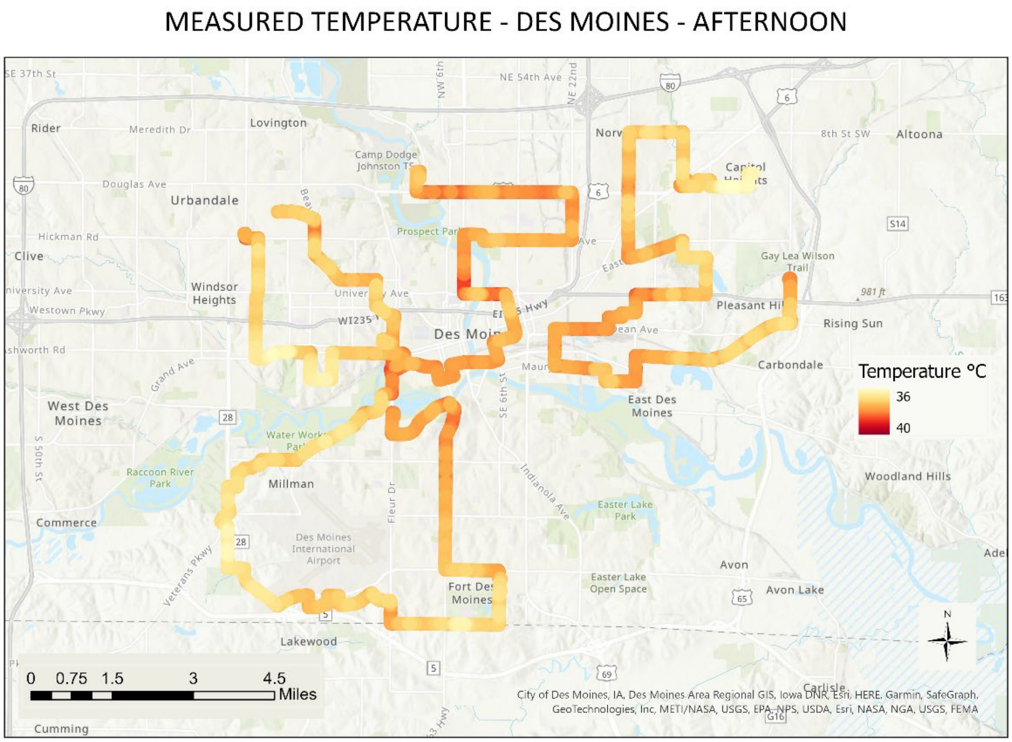


Figure 26

Measured temperature in Des Moines during the evening

MEASURED TEMPERATURE - DES MOINES - EVENING

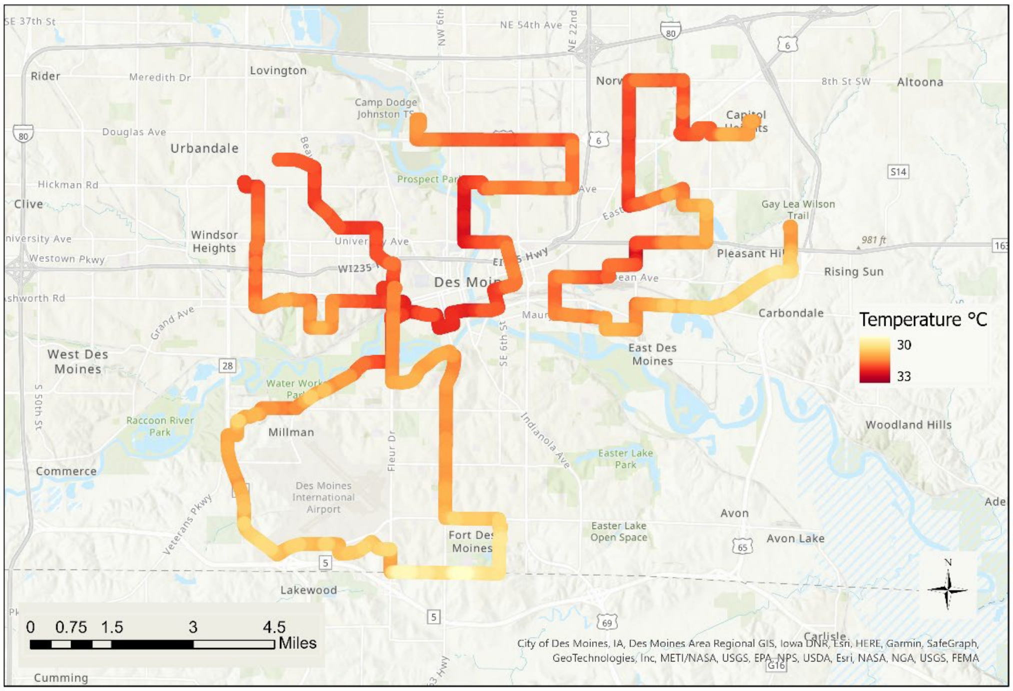
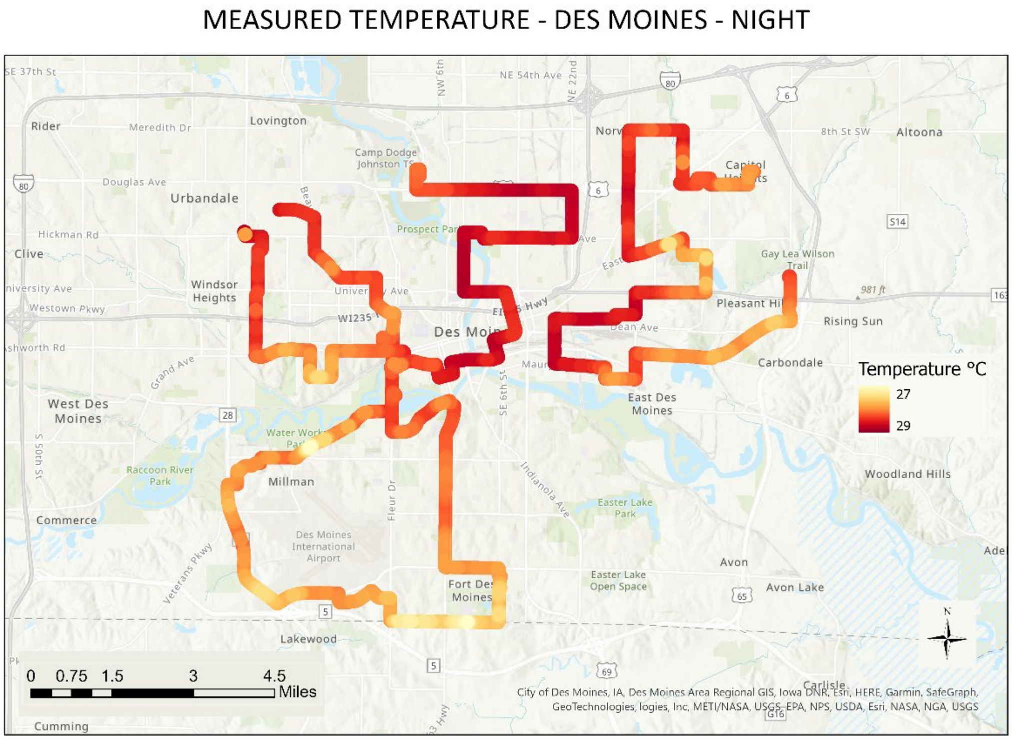


Figure 27

Measured temperature in Des Moines during the night



As with most cities, the airport is located in open areas and without large built-up structures in its surroundings, and thus it is not surprising that those surroundings presented low temperatures regardless of the time of collection. That was also the case for Fort Dodge during the afternoon and evening (Figure 28, and Figure 29, respectively). As expected, the warmest part of the city was in its center, which is occupied by many buildings and large parking lots.

Figure 28
Measured temperature in Fort Dodge during the afternoon

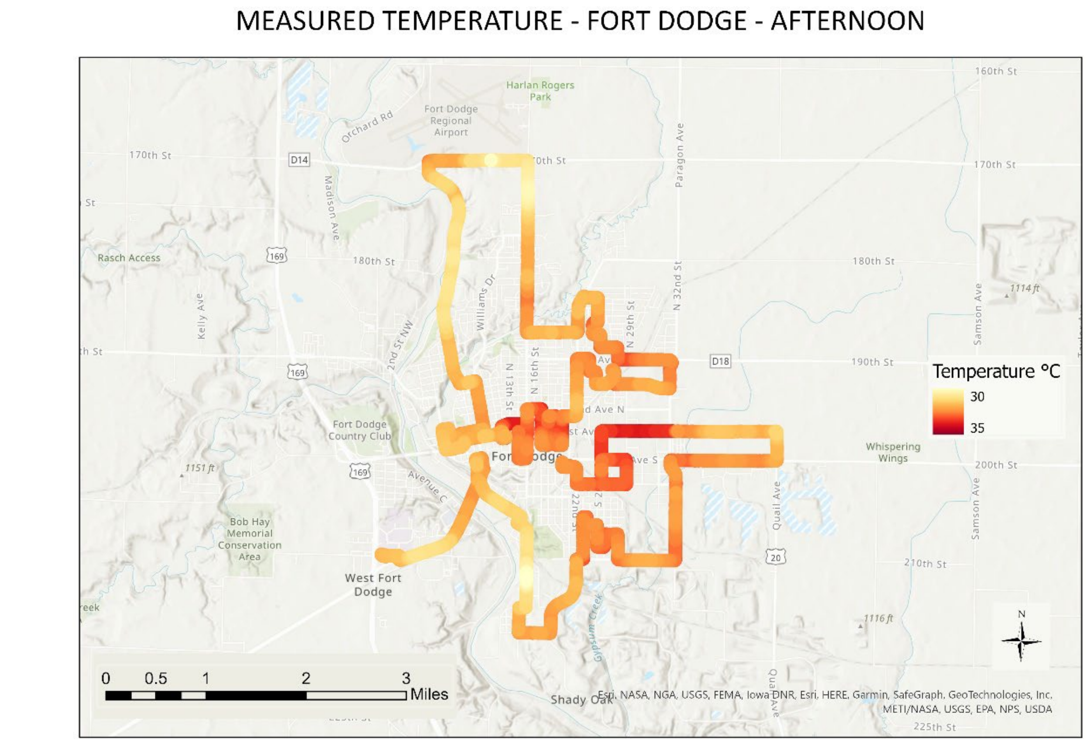
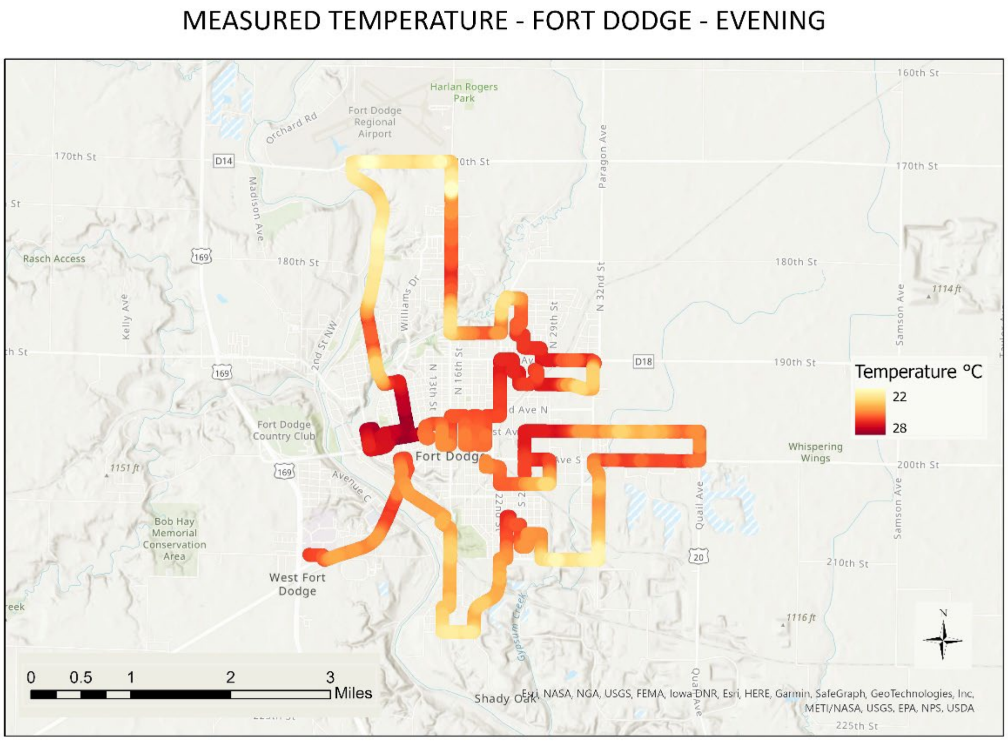


Figure 29

Measured temperature in Fort Dodge during the evening



Marshalltown (Figure 30, and Figure 31) presented consistent patterns throughout both collections (afternoon, and night), where downtown was where the highest temperatures were observed. Despite the fact that a strong rainfall hit the area during the evening, which prevented the second run to be executed, night collection showed the same areas as the highest and lowest temperatures, however, the differences between minimum and maximum temperatures collected were one of the smallest measured throughout all study areas, only 1.09°C.

Figure 30

Measured temperature in Marshalltown during the afternoon

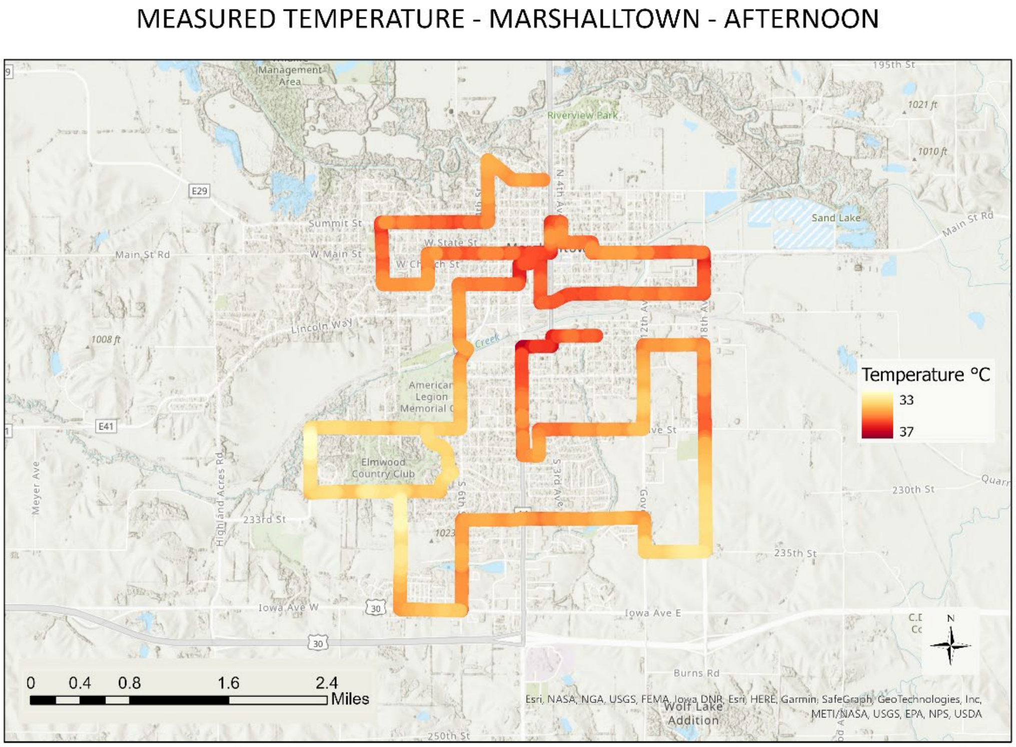
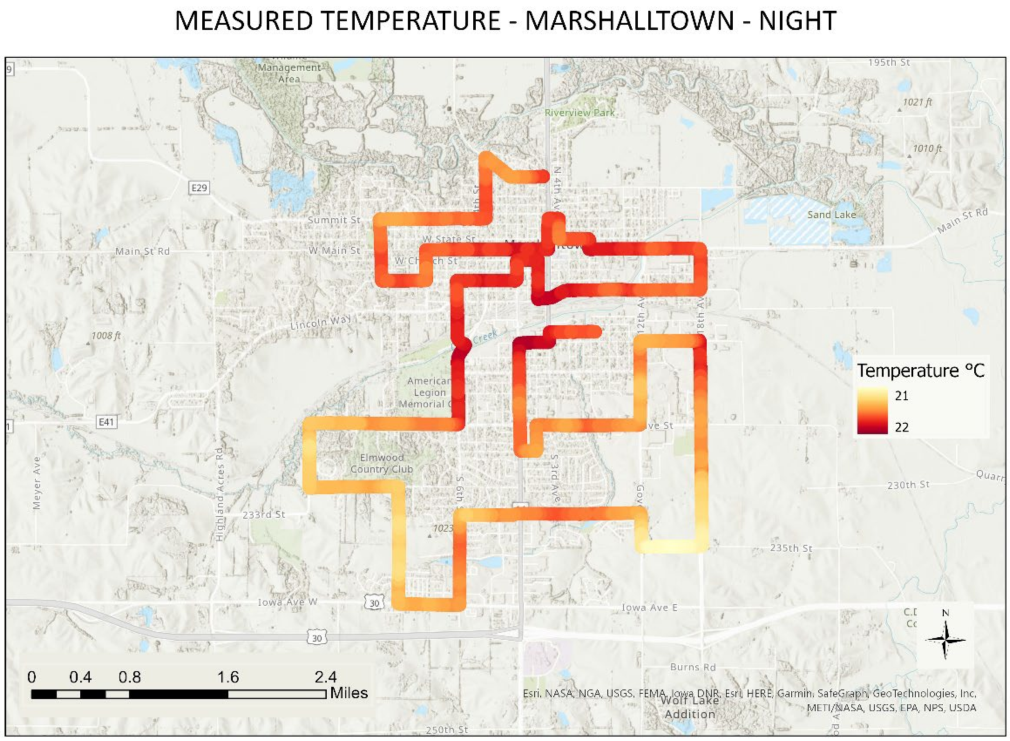


Figure 31
Measured temperature in Marshalltown during the night



Out of all 10 different urban areas, Sioux City showed somewhat the most unique results, where the area with the highest temperature in the afternoon was not the same compared to collections performed during the evening and night periods. In the afternoon (Figure 32), the highest temperature was collected close to a segment of the CN Railroad with Highway 20, whereas the highest temperature collected during the evening (Figure 33) was at the border with Nebraska (south) with Hamilton Boulevard, and the highest temperature collected during the nighttime (Figure 34) was in a residential area west of Isabella Street. The lowest temperatures were collected in the southeast part of the city, regardless the time of collection.

Figure 32

Measured temperature in Sioux City during the afternoon

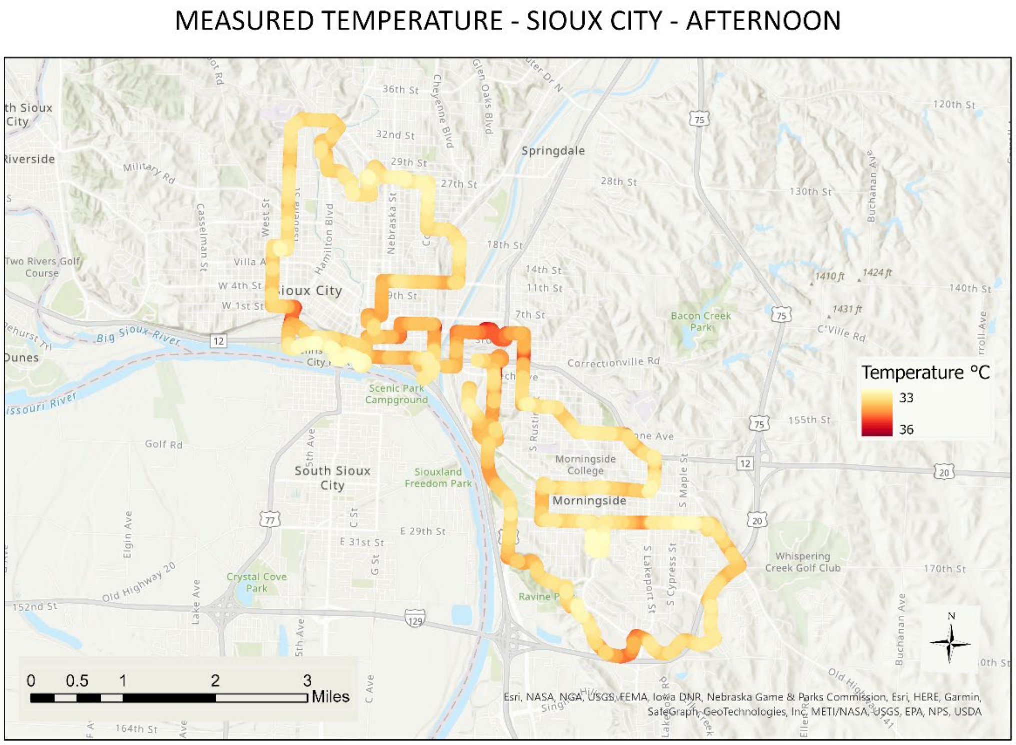


Figure 33

Measured temperature in Sioux City during the evening

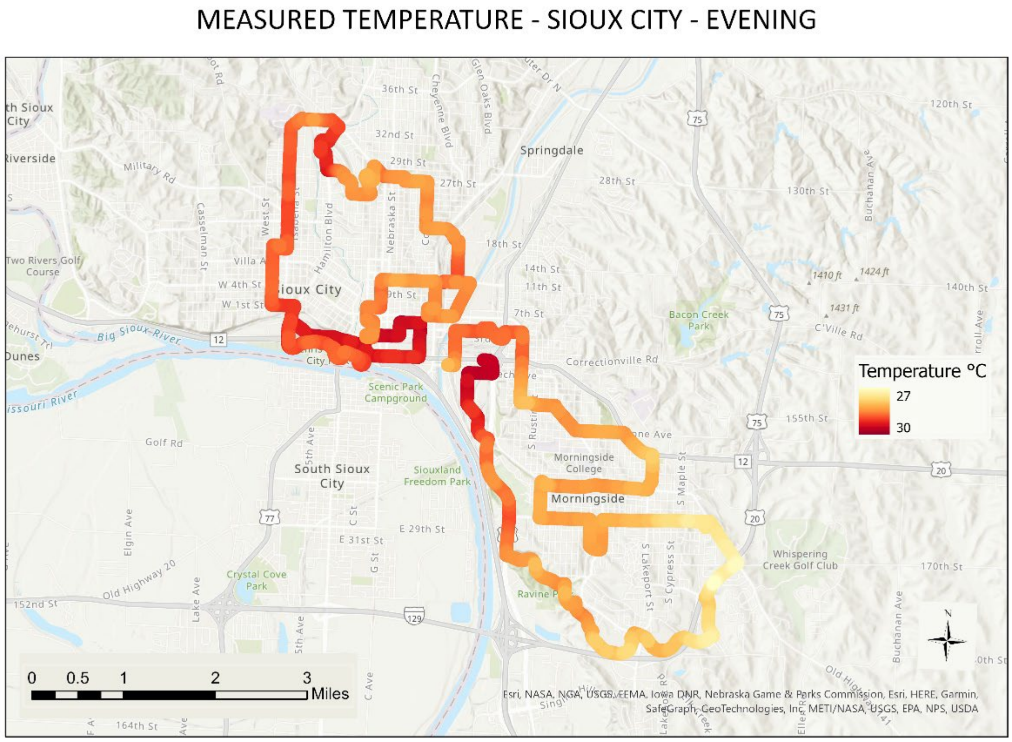
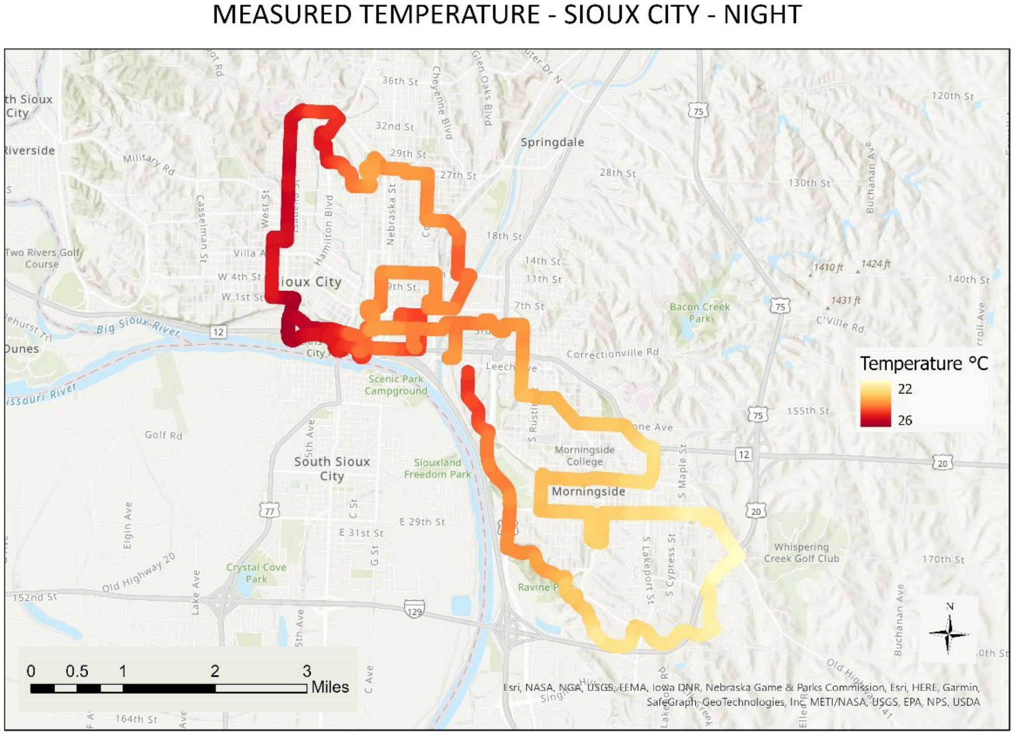


Figure 34

Measured temperature in Sioux City during the night



The study area named Waterloo-Cedar Falls (Figure 35 to Figure 37) showed consistent temperature patterns throughout all times of collection. The highest temperatures were measured in downtown Waterloo; along University Avenue, which connects the cities of Cedar Falls to Waterloo; and the area close to the University of Northern Iowa, which, although vegetated, presented a high volume for built-up structures (e.g., buildings, parking lots), which is thought to have contributed to the high temperatures measured by the mobile device. As expected, the lowest temperatures were collected in the most vegetated area, Hartman Reserve Nature Center, and in the southwest part of Cedar Falls by Greenhill Road, which presents low density for built-up structures.

Figure 35

Measured temperature in Waterloo/Cedar Falls during the afternoon

MEASURED TEMPERATURE - WATERLOO/CEDAR FALLS - AFTERNOON

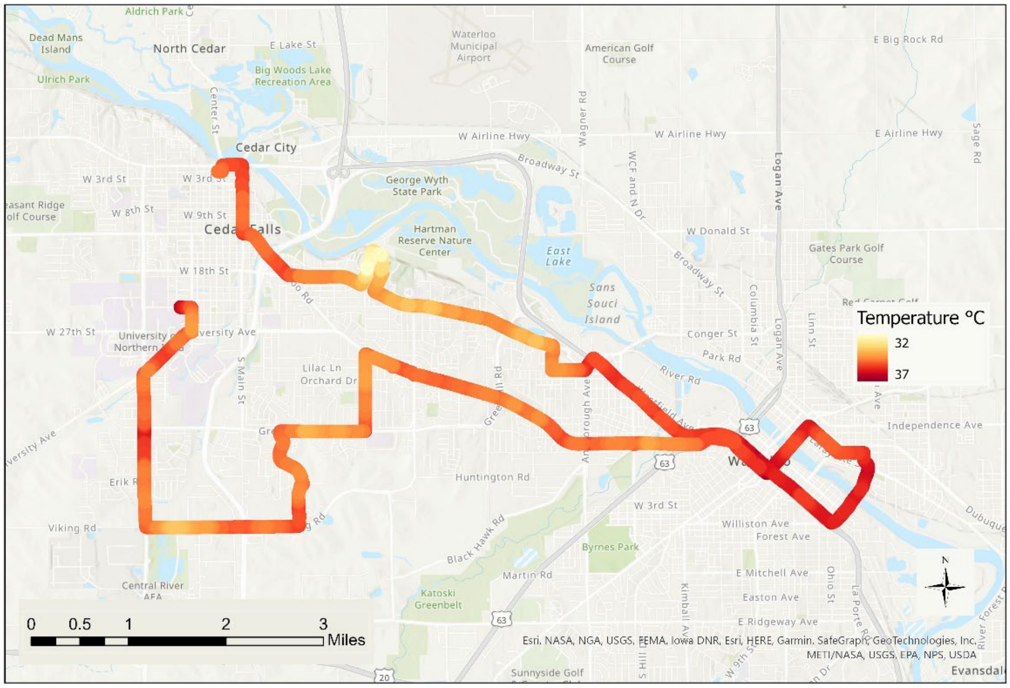


Figure 36

Measured temperature in Waterloo/Cedar Falls during the evening

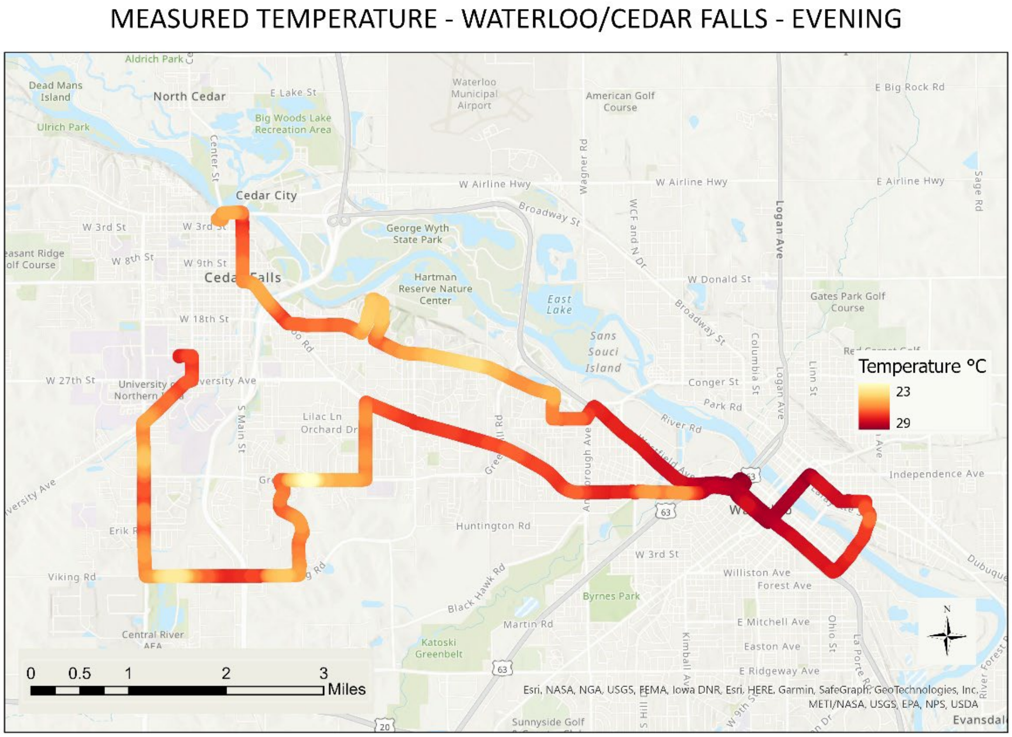
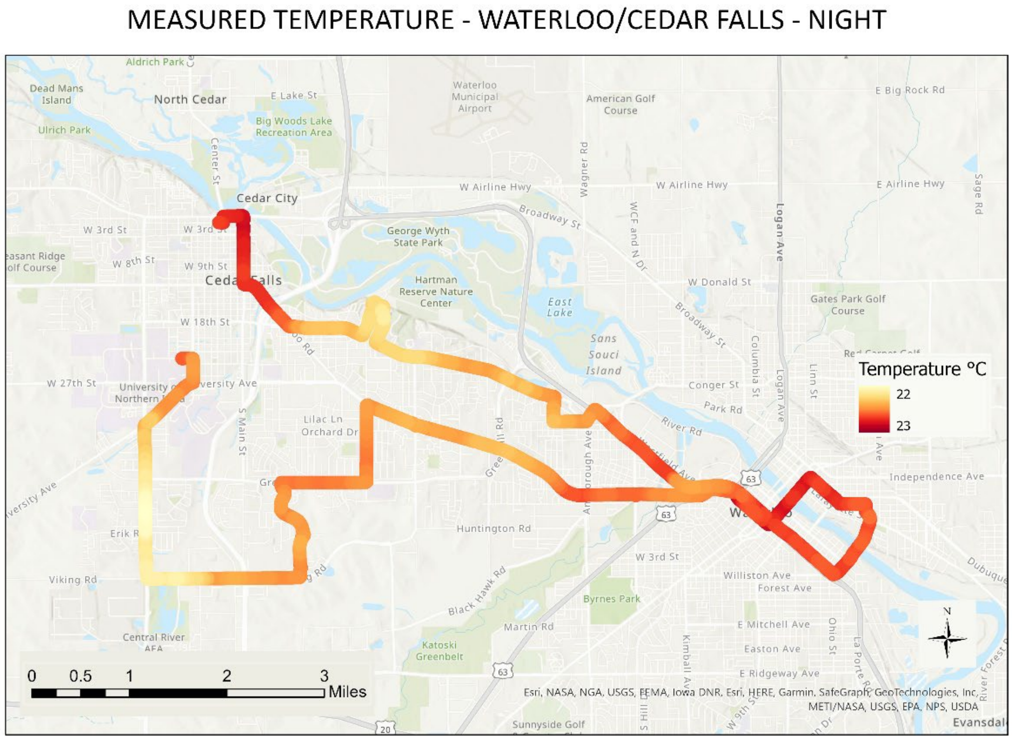


Figure 37

Measured temperature in Waterloo/Cedar Falls during the night



Finally, the City of Waverly (Figure 38, and Figure 39) seems to show a considerable influence of bodies of water regulating urban heat. Although the highest temperatures in the city were measured in areas with the highest amount of built-up structure, the lowest temperatures were measured in highly vegetated areas, especially close to Three Rivers Park by the Cedar River.

Figure 38

Measured temperature in Waverly during the afternoon

MEASURED TEMPERATURE - WAVERLY - AFTERNOON

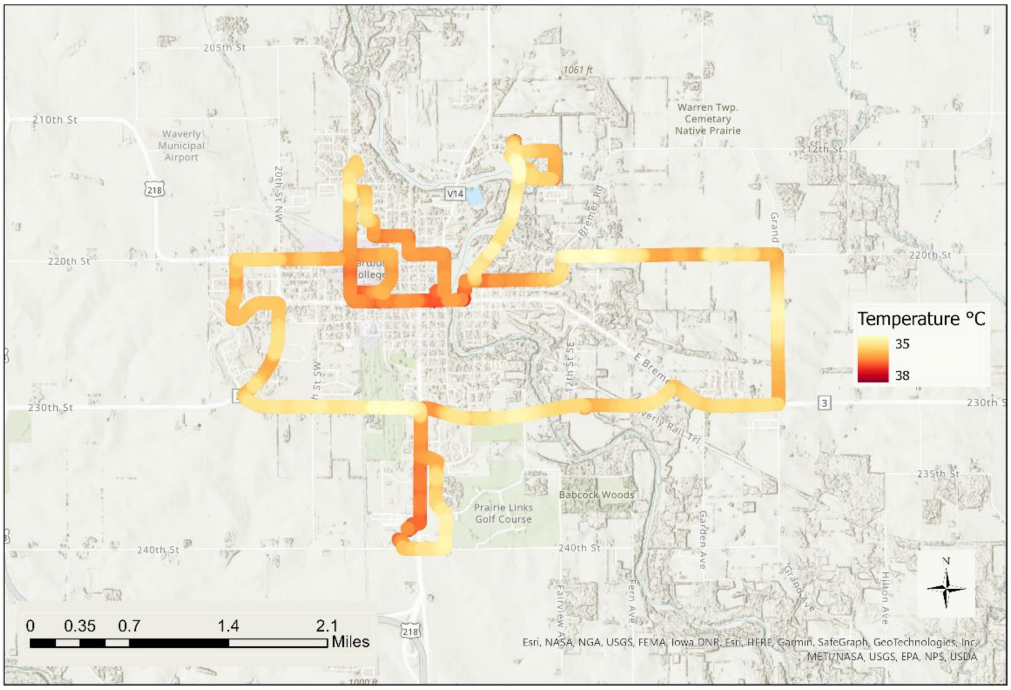
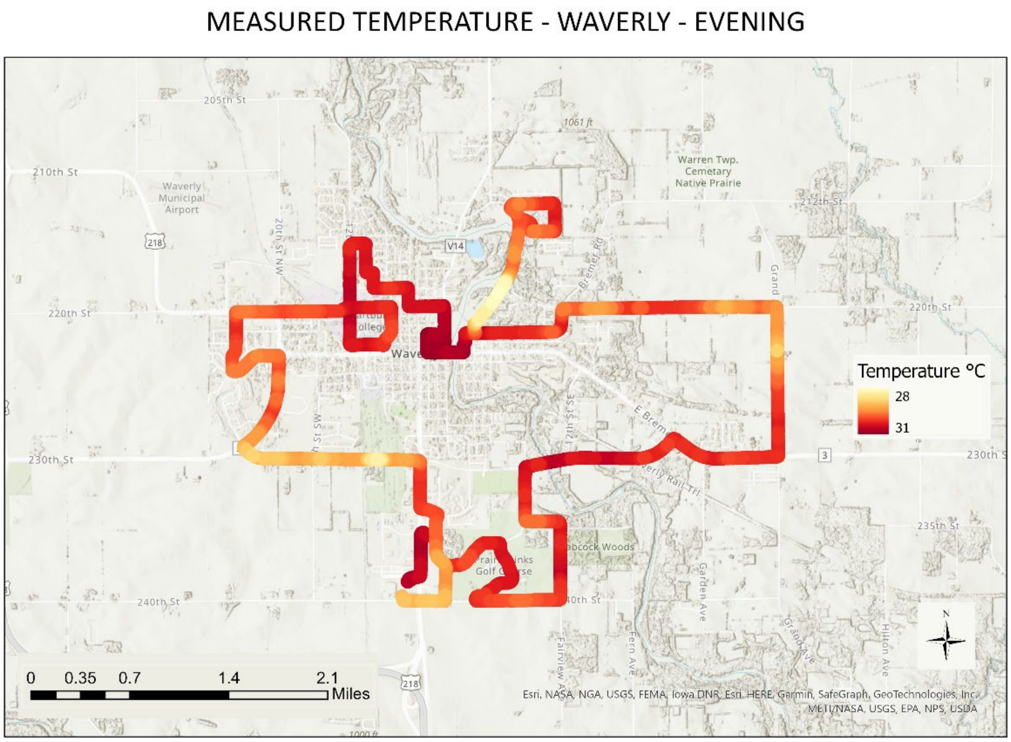


Figure 39
Measured temperature in Waverly during the evening



Random Forest Model Results

A total of 24 modeled (predicted) raster surface files, one for each urban area at each time of the day (afternoon, evening, and night) were produced, and are illustrated in alphabetical order and by time of the day.

The results showed an average R^2 of 0.947 for the afternoon models, R^2 of 0.973 for the evening models, and R^2 of 0.987 for the nighttime. Most study areas had models of over R^2 0.95 (20 out of 24 models), which is considered not only very strongly statistically significant but similar or higher R^2 values as compared to other studies that made use either of random forest (e.g., Voelkel and Shandas 2017; Shandas et al. 2019), or other statistic algorithms/techniques (e.g., Yokobori and Ohta 2009; Chen et al. 2020).

Table 12 presents the predictive model results (R^2) for every urban area by time of the day.

Table 12
Random Forest results for every urban area/time of the day

URBAN AREA	AFTERNOON R^2	EVENING R^2	NIGHT R^2
Burlington	0.978	0.995	0.994
CedarRapids	0.923	0.956	0.987
Council Bluffs	0.982	0.989	0.997
DesMoines	0.952	0.984	0.988
FortDodge	0.952	0.974	-
Marshalltown	0.956	-	0.963
Sioux City	0.879	0.974	0.987
WaterlooCF	0.975	0.994	0.99
Waverly	0.93	0.915	-

Considering all predictive surface raster models generated by Forest-based Classification and Regression algorithm results varied from an R^2 of 0.879 (p-value <0.00, standard error of 0.011), the lowest proportion of variance in air temperature explained by the independent variables, to an R^2 of 0.997 (p-value of 0.00, and standard error of 0.002), the highest proportion of variance in air temperature explained by the independent variables. These results were respectively Sioux City in the afternoon, and Council Bluffs in the night.

Analyzing p-values and standard errors for each result showed that all results are very statistically significant, with p-values of <0.000, and standard errors not higher than 0.011 as seen in table 13.

Table 13
Standard error of Random Forest for every urban area/time of the day

URBAN AREA	AFTERNOON STD. ERROR	EVENING STD. ERROR	NIGHT STD.ERROR
Burlington	0.005	0.002	0.002
CedarRapids	0.006	0.004	0.002
Council Bluffs	0.004	0.003	0.002
DesMoines	0.004	0.002	0.002
FortDodge	0.007	0.006	
Marshalltown	0.008		0.008
Sioux City	0.011	0.005	0.004
WaterlooCF	0.005	0.002	0.001
Waverly	0.007	0.009	

When accounting for the top 5 variables that explain the most variance in temperature by urban area and time of the day, NDVI, regardless of the statistical type used or distance, appears as the most consistent variable, showing as being the most important variable in 14 out of 24 models. The second most consistent variable in explaining the variation in temperature is BH (6 out of 24 models), followed by CDM (3 out of 24), and CC (1 out of 24 models). Regarding neighbor distances, when analyzing all 120 different most important variables (the 5 most important variables per each one of the 24 models), the 800 meters distance accounts for 52.5% of all variables, followed by 400 meters (20%), 200 meters (15.83%), 100 meters (8.33%), and 50 meters (3.33%). These results showed that the average occurrence of morphometric features (natural and built-up) in small neighbor distances has a smaller statistical influence on urban temperatures when compared to larger neighbor distances (e.g., the average building height in a 50 meters radius seems to have considerably less statistical significance/influence in generating the predictive models than the average for building height in 800 meters radius). Considering the statistic method of central tendency (mean vs. standard deviation), mean seems to have higher statistic influence on the predictive models than standard deviation has, appearing in 72.5% as one of the 5 most important variables per model as seen in Table 14.

Table 14

Table of most important variables by urban areas/time of the day

URBAN AREA		MOST IMPORTANT VARIABLES					PERCENTAGE OF EXPLANATION BY MOST IMPORTANT VARIABLE(S)	
CITY	TIME OF THE DAY	1ST VARIABLE	2ND VARIABLE	3RD VARIABLE	4TH VARIABLE	5TH VARIABLE	% EXPLAINED BY 1ST VARIABLE	% EXPLAINED BY 5 TOP VARIABLES
Council Bluffs	Afternoon	CC_Mean800	CC_STD800	CDM_STD400	CDM_STD800	NDVI_Mean800	19	55
Council Bluffs	Evening	CDM_STD800	CDM_Mean800	CDM_STD400	CC_Mean800	CDM_Mean400	24	62
Council Bluffs	Night	CDM_STD800	CDM_Mean800	CDM_STD400	NDVI_Mean200	NDVI_Mean800	26	59
DesMoines	Afternoon	NDVI_Mean400	NDVI_Mean100	NDVI_Mean800	NDVI_Mean50	NDVI_Mean200	12	47
DesMoines	Evening	BH_Mean800	NDVI_Mean400	NDVI_Mean200	CDM_STD800	NDVI_Mean800	11	40
DesMoines	Night	NDVI_Mean800	NDVI_Mean100	NDVI_Mean400	BH_Mean800	BV_Mean800	15	41
Marshalltown	Afternoon	NDVI_Mean400	NDVI_Mean800	NDVI_Mean200	BH_STD800	NDVI_Mean50	19	57
Marshalltown	Night	BH_Mean800	NDVI_Mean800	NDVI_Mean400	NDVI_Mean200	BH_STD800	19	64
CedarRapids	Afternoon	NDVI_Mean400	NDVI_Mean800	NDVI_Mean100	NDVI_Mean200	NDVI_STD800	12	42
CedarRapids	Evening	CDM_STD800	BH_STD800	NDVI_Mean800	CDM_Mean800	BV_Mean800	10	38
CedarRapids	Night	BH_Mean800	NDVI_Mean100	BH_Mean400	NDVI_STD800	NDVI_Mean200	15	46
Burlington	Afternoon	NDVI_Mean400	NDVI_Mean200	NDVI_Mean800	BH_STD800	NDVI_Mean100	23	58
Burlington	Evening	NDVI_Mean400	NDVI_Mean800	NDVI_STD800	CC_STD800	NDVI_Mean200	20	57
Burlington	Night	NDVI_Mean800	NDVI_Mean400	BH_STD800	NDVI_STD800	CC_STD800	27	65
FortDodge	Afternoon	NDVI_Mean800	BH_Mean400	BH_Mean200	BH_Mean800	BV_Mean800	16	40
FortDodge	Evening	BH_Mean400	BH_STD200	BH_Mean800	BH_Mean200	BH_STD400	18	57
Waverly	Afternoon	BH_STD400	BH_Mean200	BV_Mean200	BH_STD200	NDVI_Mean200	16	42
Waverly	Evening	BH_Mean400	CC_STD800	BH_STD400	NDVI_Mean800	BV_Mean800	10	38
WaterlooCF	Afternoon	NDVI_Mean50	NDVI_Mean800	CDM_Mean100	NDVI_Mean100	CC_Mean200	14	50
WaterlooCF	Evening	NDVI_Mean200	NDVI_Mean800	NDVI_Mean400	CDM_Mean100	NDVI_Mean50	20	63
WaterlooCF	Night	NDVI_Mean800	NDVI_Mean400	NDVI_Mean200	BV_Mean800	NDVI_Mean100	22	56
Sioux City	Afternoon	NDVI_STD400	NDVI_STD800	CC_Mean800	BV_Mean100	BH_Mean800	13	40
Sioux City	Evening	NDVI_Mean800	NDVI_Mean400	CC_Mean800	NDVI_STD800	CC_STD800	17	46
Sioux City	Night	NDVI_STD800	NDVI_Mean800	BH_STD800	CC_Mean800	CDM_STD800	17	53

The percentage that is explained by the five most important variables per model varied greatly, from 38% (Cedar Rapids/Evening and Waverly/Evening) to 65% (Burlington/Night), and the percentage explained by the most important variable varied from 10% (Cedar Rapids/Evening and Waverly/Evening) to 27% (Burlington/Night). Even though the exact same models (Cedar Rapids/Evening and Waverly/Evening) fall into the highest and lowest percentage explained by the five most important variables and the most important variables, this analysis falls outside of the scope of this study's objectives. However, the author believes that more testing would be needed since the models that have the higher and lower explanation by the five and most important variables are not necessarily the ones presenting the higher or lower R^2 coefficients.

Analyzing the predictive raster surface models generated for each study area, most of the results are seen as satisfactory (based on the coefficient R^2), and similar to other studies such as those conducted by Voelkel and Shandas (2017). Evaluating the models visually by study area and time of the day, all predictive raster surfaces (Figures 40 to 63) followed the patterns seen during the air temperature collection phase.

In Burlington (Figures 40 to 42), models showed the highest temperatures in the east part of the city, which is exactly what was seen when contrasted with the air temperature collected in the area.

Figure 40

Modeled raster surface for Burlington afternoon

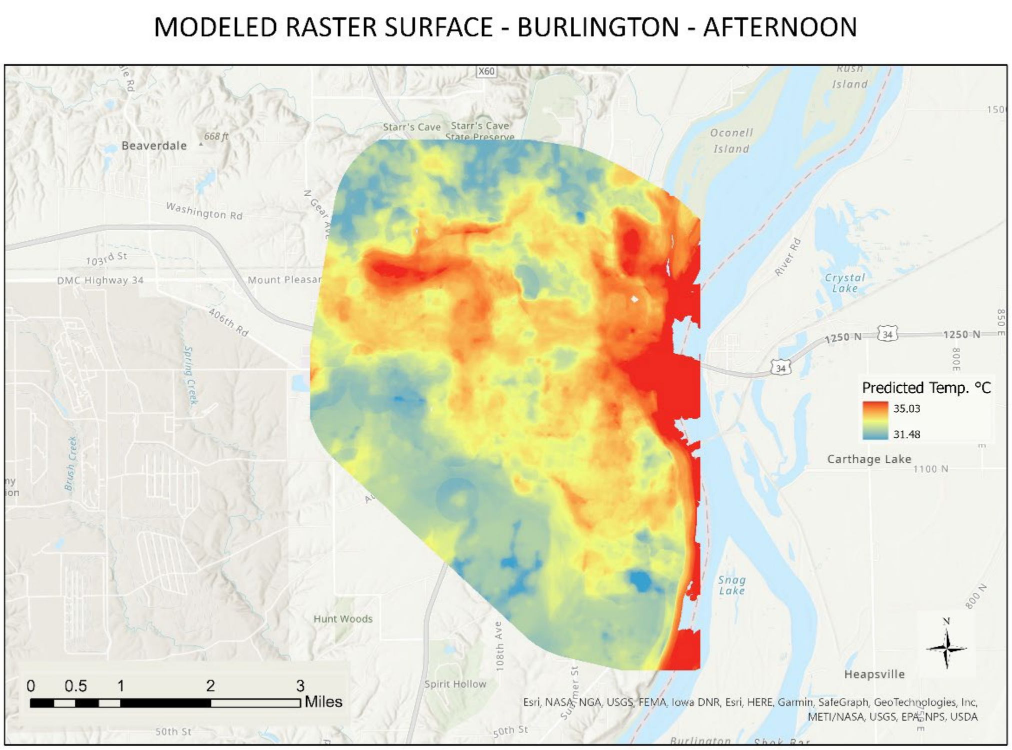


Figure 41

Modeled raster surface for Burlington evening

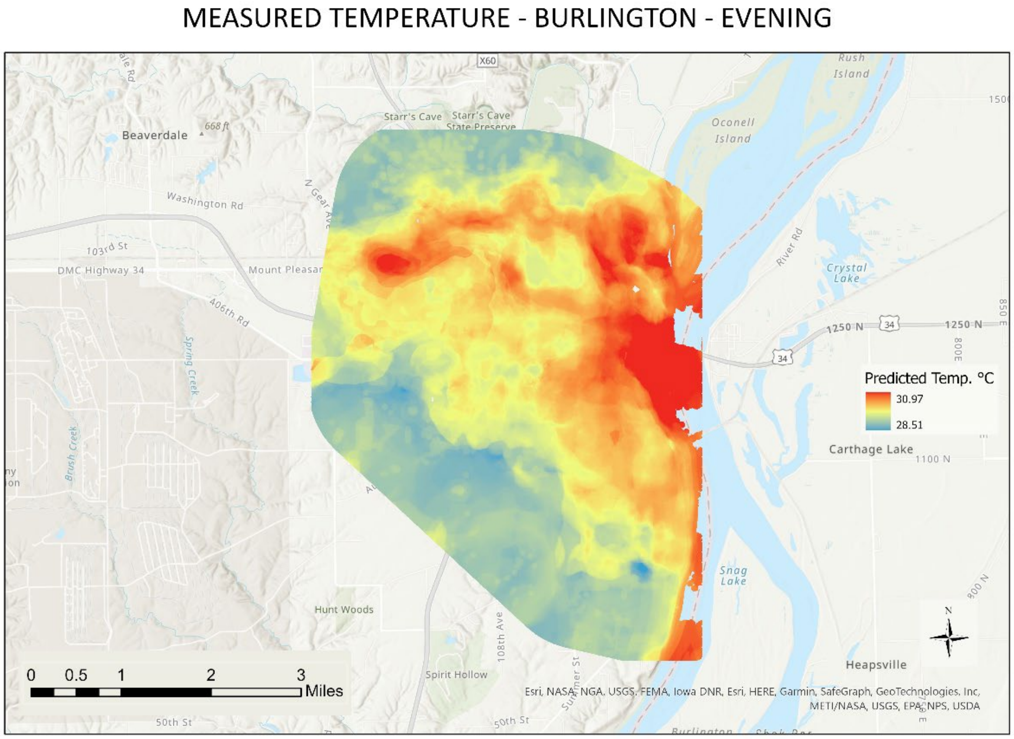
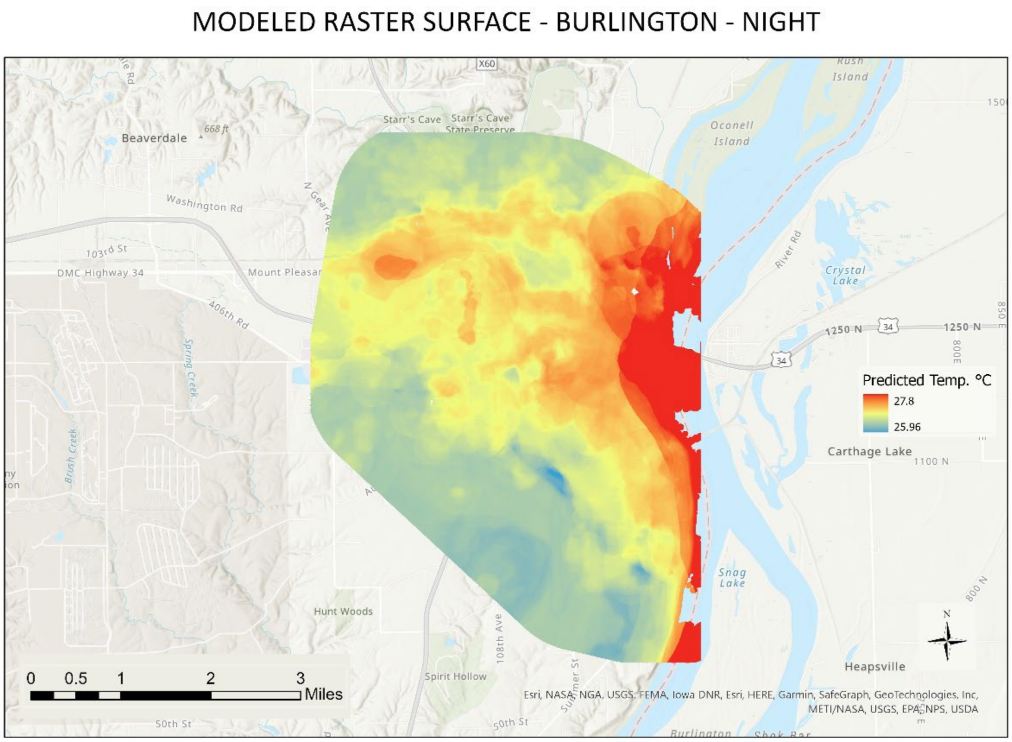


Figure 42
Modeled raster surface for Burlington night



The predictive raster surface model for Cedar Rapids showed no extraordinary findings, where similarly to Burlington and Council Bluffs, the warmest areas were seen in downtown or close to areas that are heavily occupied by built-up structures. Even though the afternoon (Figure 43) and evening run (Figure 44) in Cedar Rapids had no recordings for one out of 3 runs (33% of the projected air temperature collection area), the modeled temperature showed a smaller but acceptable decline in predictability of around 3.2% compared to the evening model (afternoon R^2 of 0.923 vs. evening R^2 of 0.956), and 6.4% compared to the night model (afternoon R^2 of 0.923 vs. night R^2 of 0.987). Figure 45 shows the night model for Cedar Rapids.

Figure 43

Modeled raster surface for Cedar Rapids afternoon

MODELED RASTER SURFACE - CEDAR RAPIDS - AFTERNOON

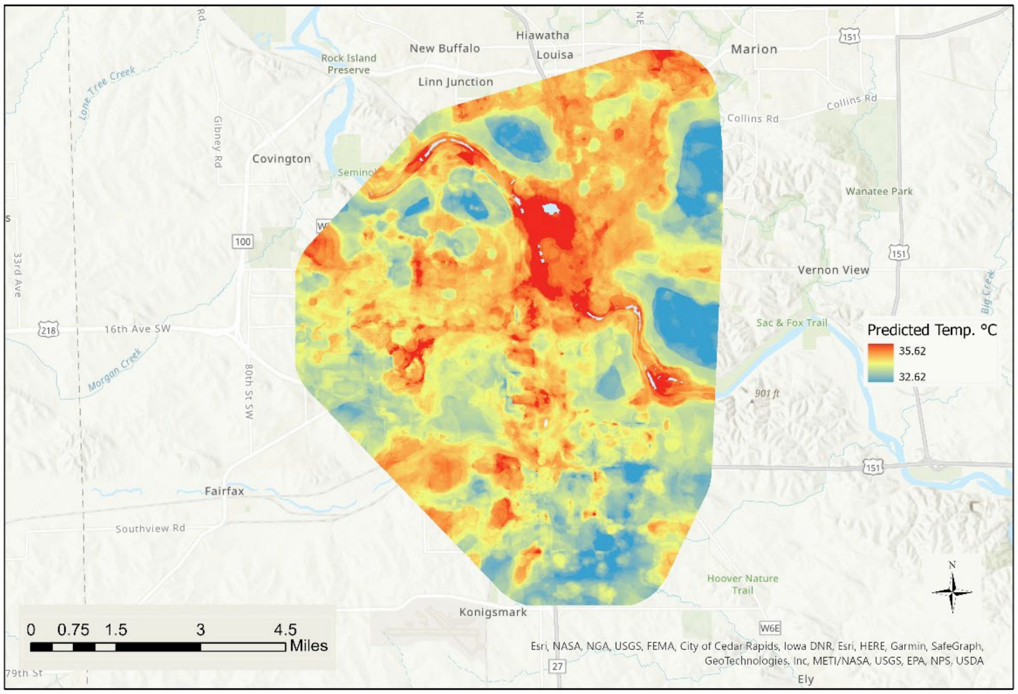


Figure 44

Modeled raster surface for Cedar Rapids evening

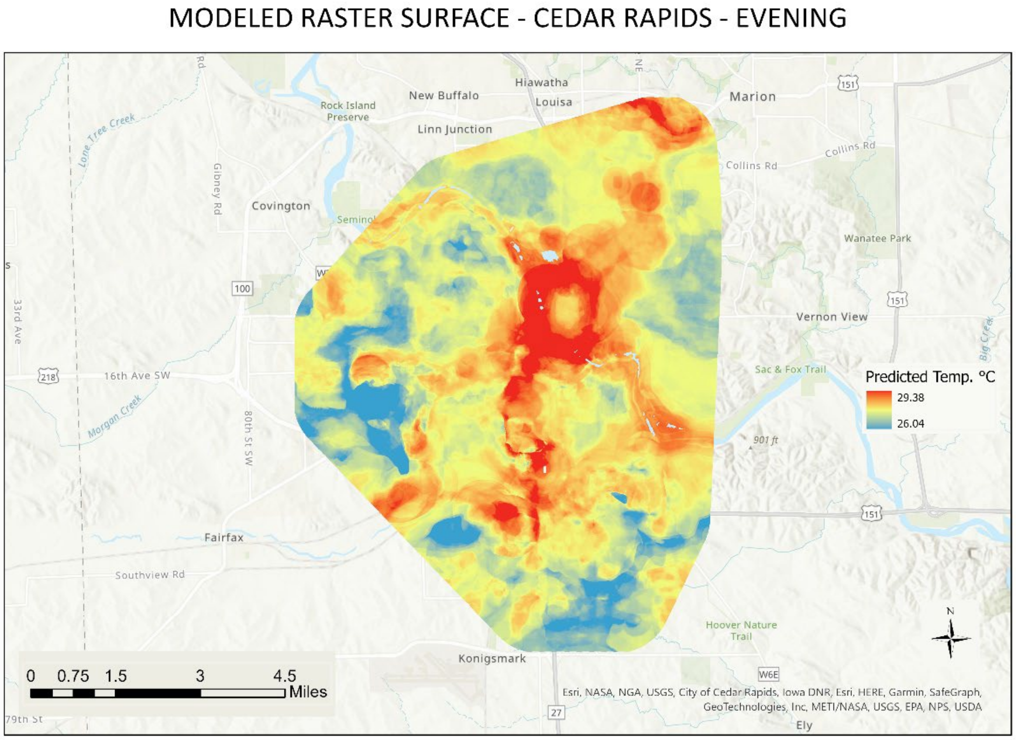
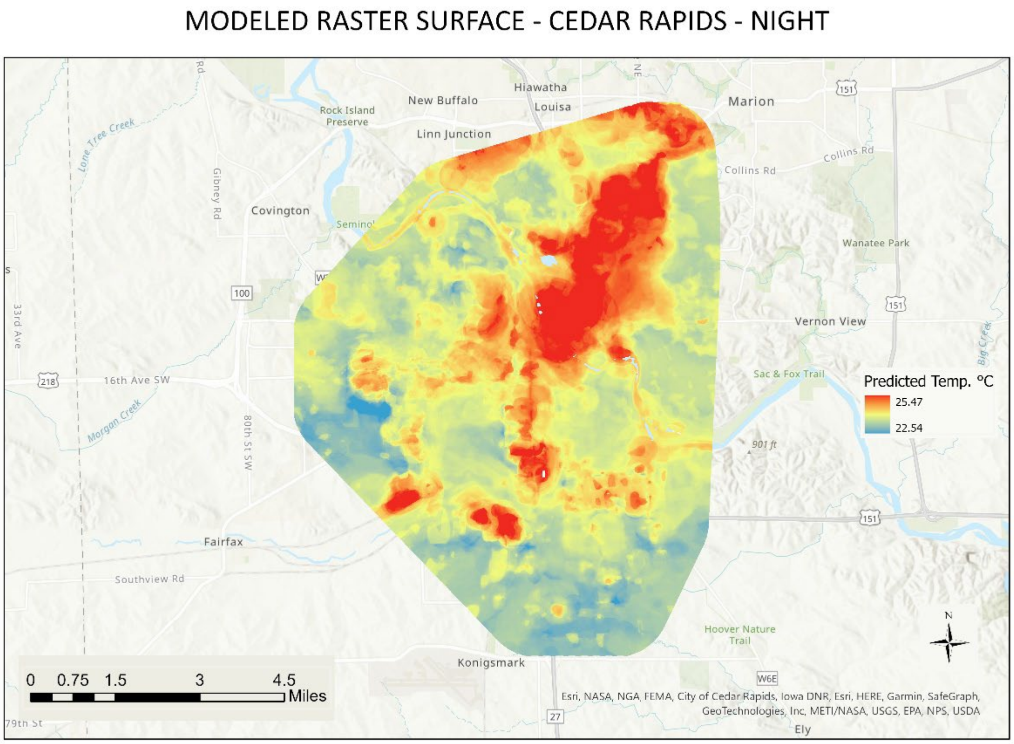


Figure 45
Modeled raster surface for Cedar Rapids night



Even though the study conducted by Hathway and Sharples (2012) examined the interaction of rivers in mitigating urban heat, their findings can be used to explain what was seen in the models of areas such as Burlington and Council Bluffs (Figures 46 to 48), where bodies of water did not show a great influence in cooling the temperatures off after certain distances (the mentioned studies showed no influences in areas farther than 40m to 100m from the river banks and in a city like Burlington where measurements took place hundreds of meters from the river in the relatively un-vegetated and high building volume downtown area it is not surprising the river showed little to no influence in the temperatures).

Figure 46
Modeled raster surface for Council Bluffs afternoon

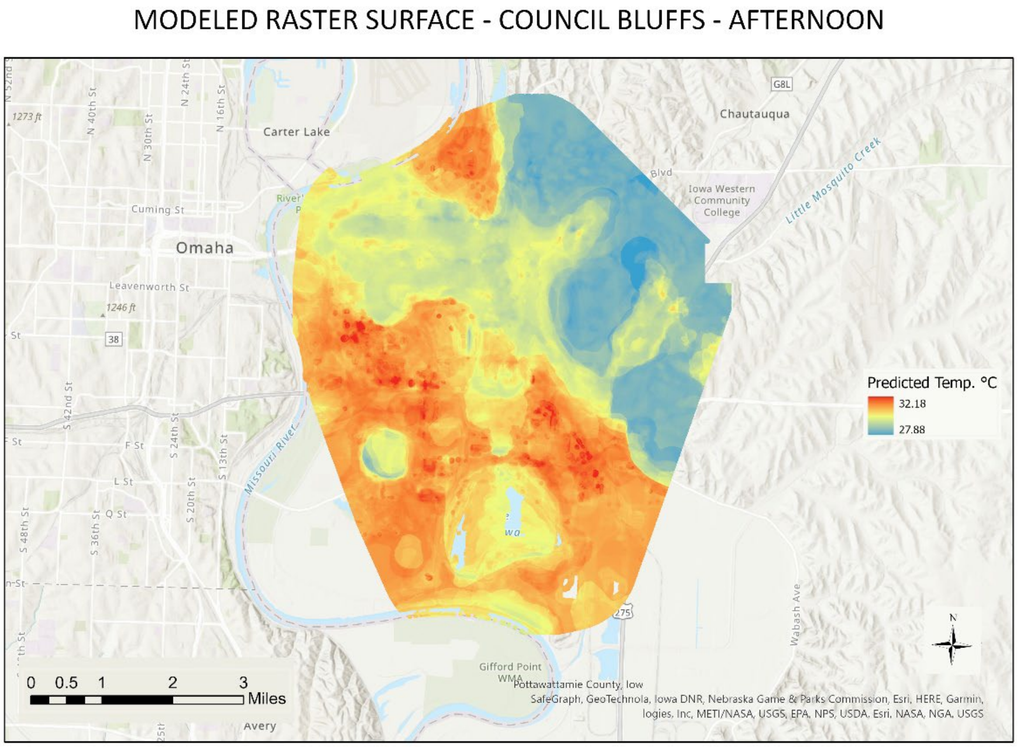


Figure 47
Modeled raster surface for Council Bluffs evening

MODELED RASTER SURFACE - COUNCIL BLUFFS - EVENING

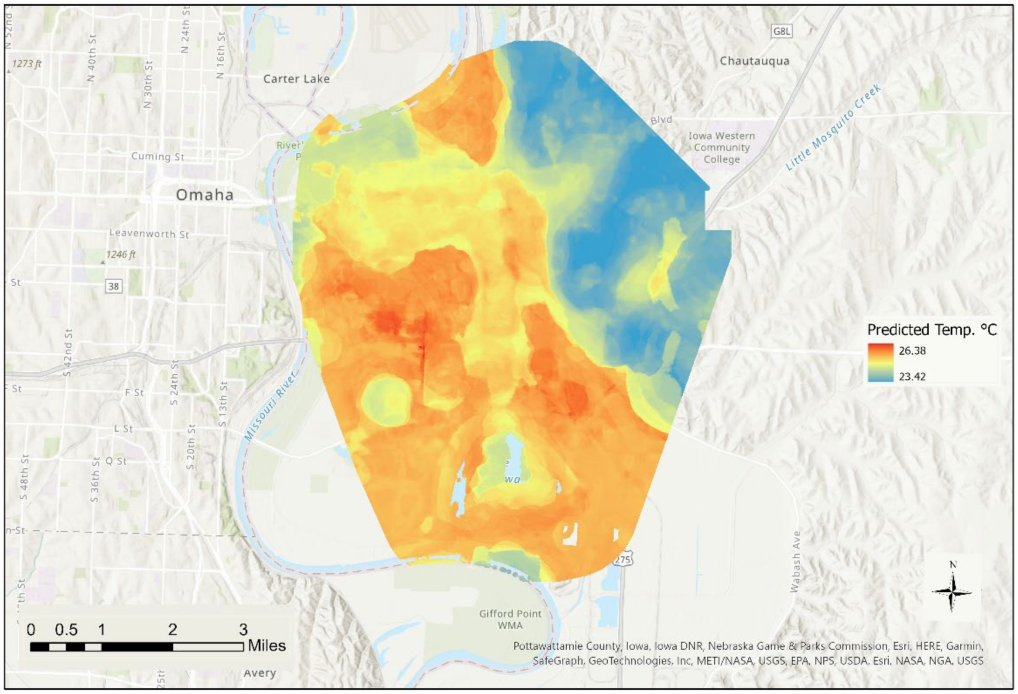
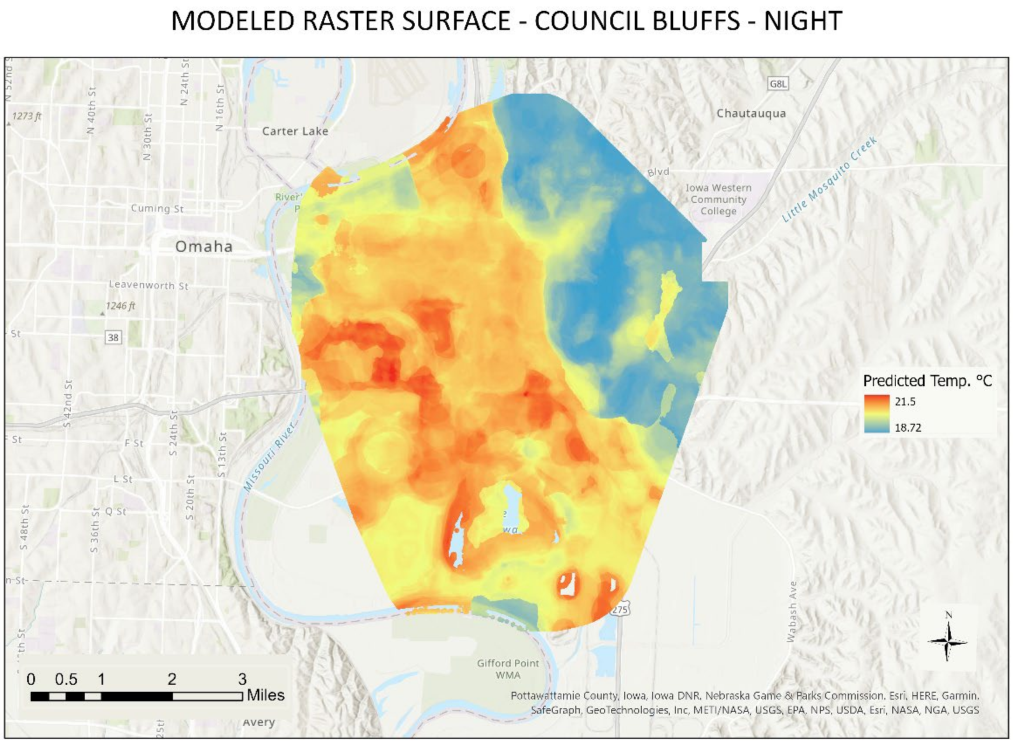


Figure 48

Modeled raster surface for Council Bluffs night



Due to the density of buildings and lack of highly vegetated areas, Des Moines (Figures 49 to 51) had resulted in three predictive models with areas of concentrated higher temperatures throughout most of the study area. However, in analyzing the models, it seems that the areas of extreme heat seen in the afternoon model extrapolate to a larger radius in the evening and night models, which is consistent with the radiative models of energetic basis suggested by the literature (e.g., Oke 1987) and consists in the urban canopy absorbing energy through solar radiation during the day, and losing energy through convection during the night, potentially warming up the air surrounding areas with high density of built-up structures.

Figure 49

Modeled raster surface for Des Moines afternoon

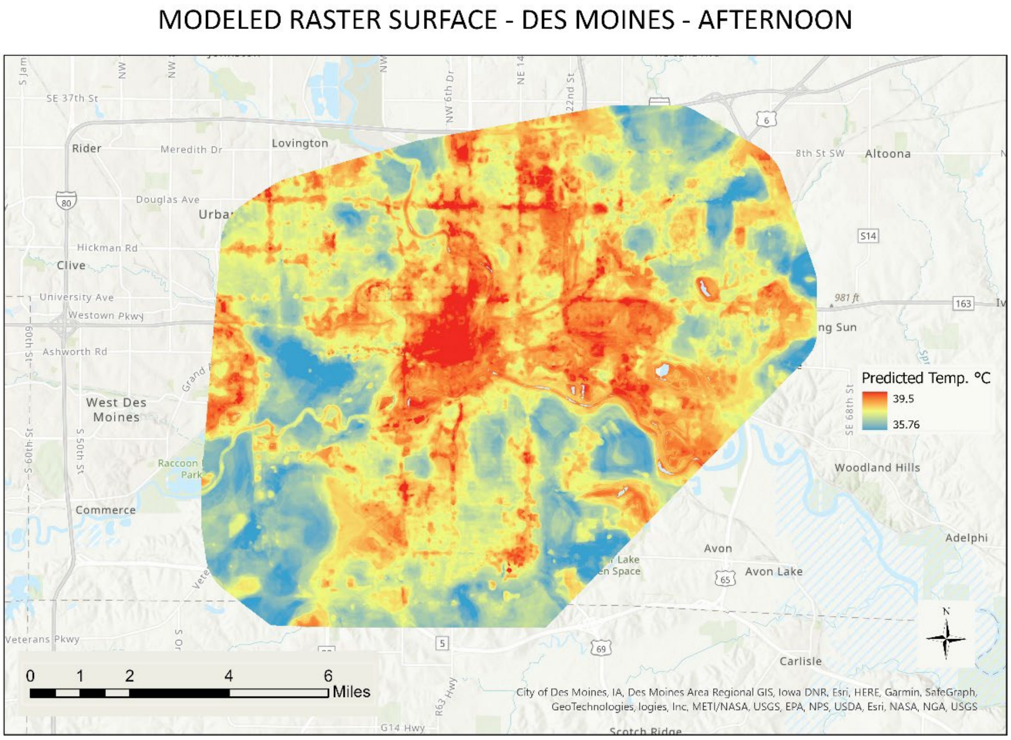


Figure 50

Modeled raster surface for Des Moines evening

MODELED RASTER SURFACE - DES MOINES - EVENING

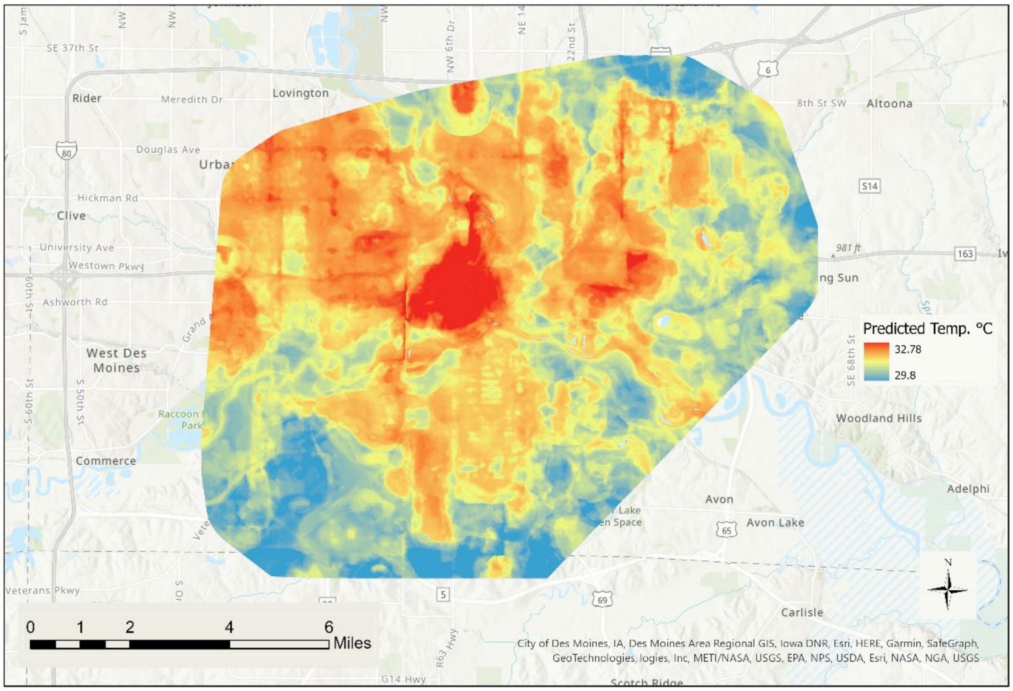
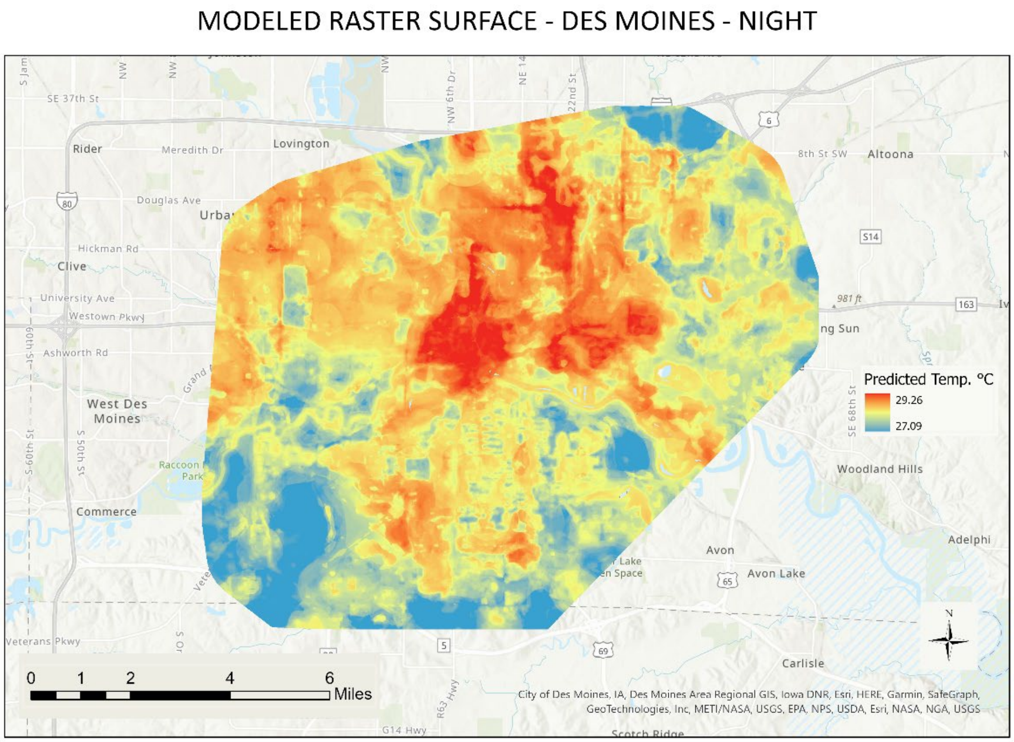


Figure 51
Modeled raster surface for Des Moines night



Although examining weather patterns in detail go beyond the scope of this study, besides the heat released by built-up structures, winds can be an active force in changes of air temperature during night periods (Oke 1983). It can be speculated that either winds or other atmospheric conditions could have influenced the shifts in air temperature patterns seen in the two models (afternoon, and evening) for Fort Dodge (Figure 52 and Figure 53), and the two models for Marshalltown (Figure 54 and Figure 55), which are consistent with the temperature measured during the collection stage of this study.

Figure 52

Modeled raster surface for Fort Dodge afternoon

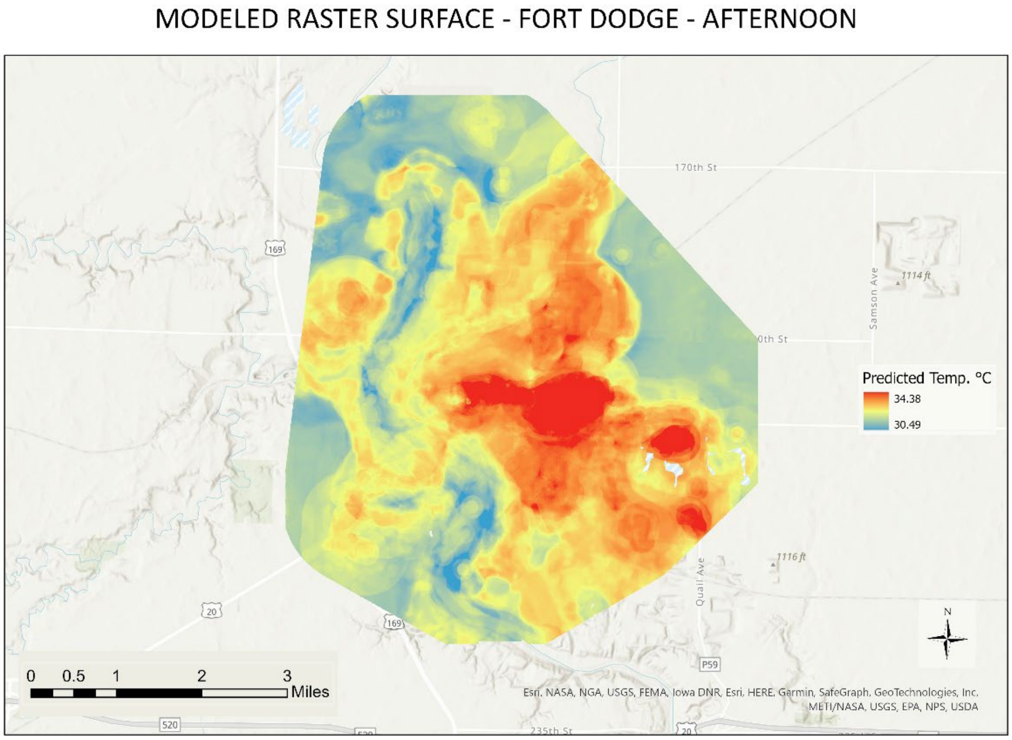


Figure 53

Modeled raster surface for Fort Dodge evening

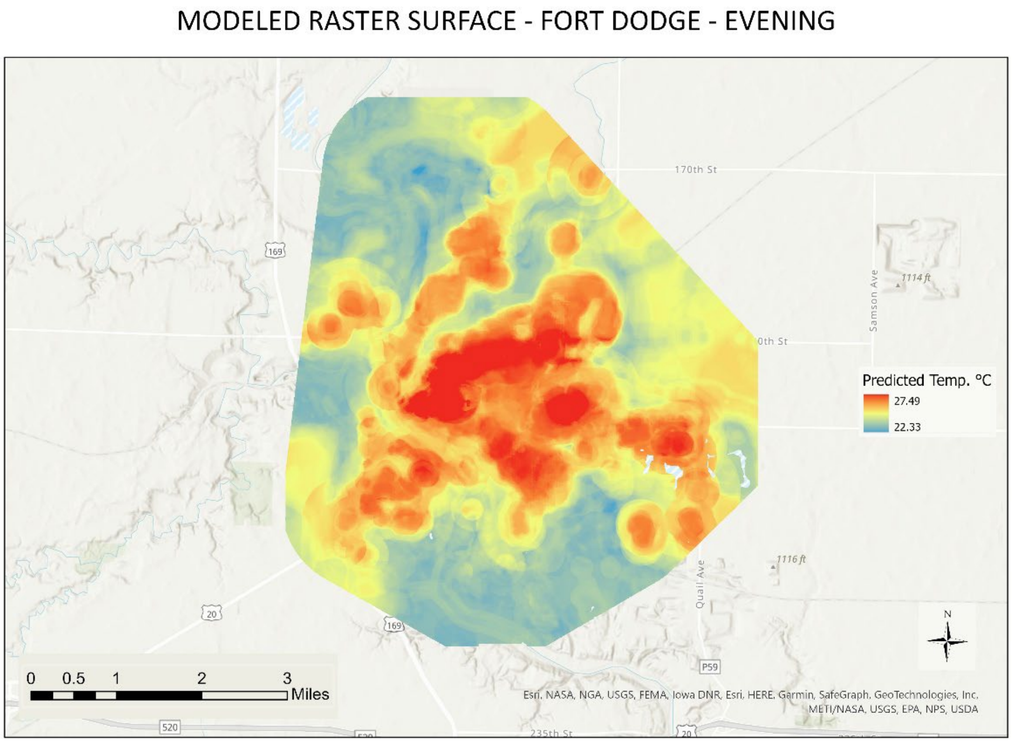


Figure 54

Modeled raster surface for Marshalltown afternoon

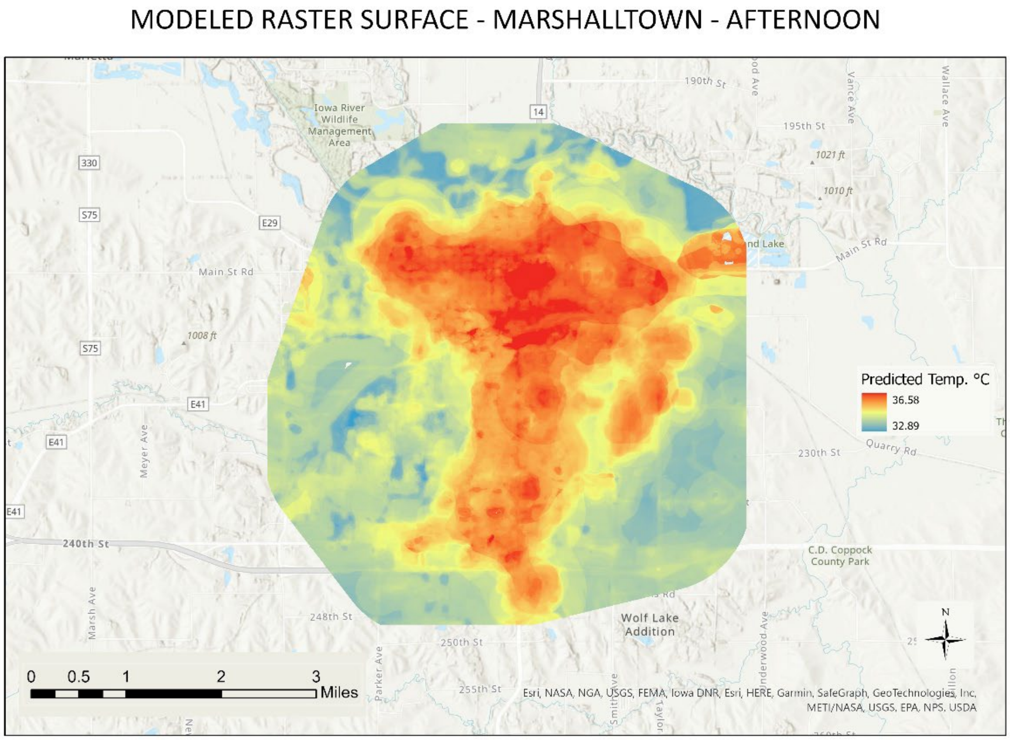
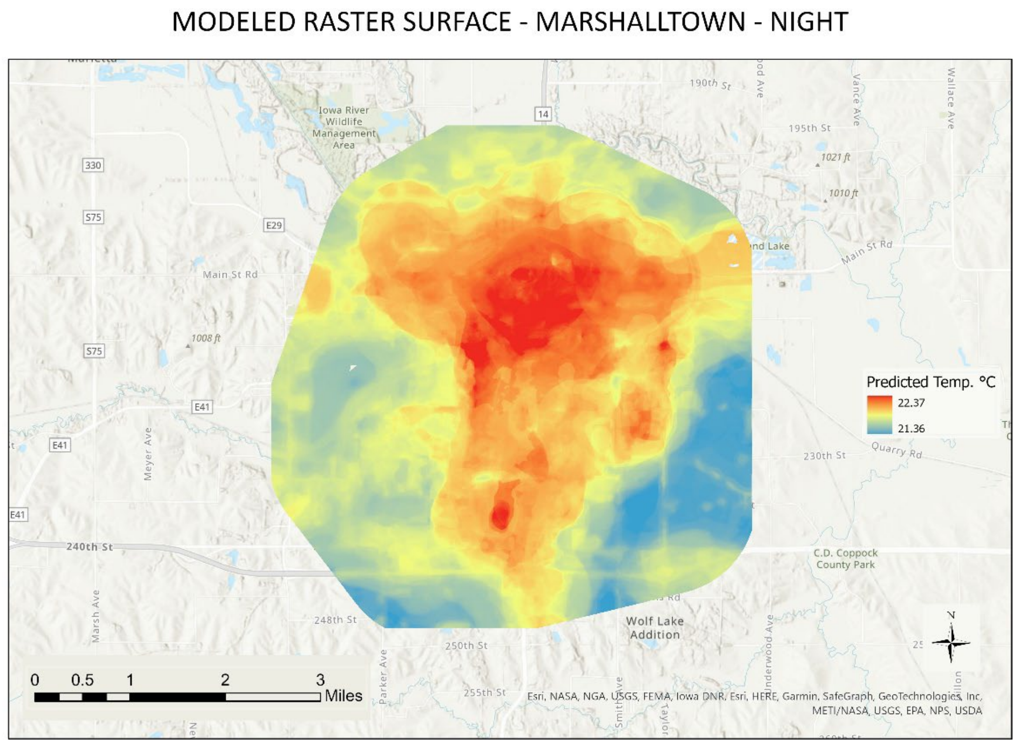


Figure 55
Modeled raster surface for Marshalltown night



Of all 24 predictive raster surface models, Sioux City afternoon (Figure 56) presented the lowest R^2 of 0.879. Differently than other models, small areas of low temperatures can be seen surrounded by warmer areas, which is unique to that study area and time of collection. The consistency of the routes performed in Sioux City (Figures 56 to 58), and the use of the exactly same devices for temperature data collection in later measurements eliminates the possibility of malfunctioning devices or that routes were performed in non-representative areas for the city. Results for the period of evening (R^2 0.974) and night (R^2 0.984) reinforce the idea that factors beyond the scope of this research might have influenced the data collection in the afternoon (e.g., weather).

Figure 56

Modeled raster surface for Sioux City afternoon

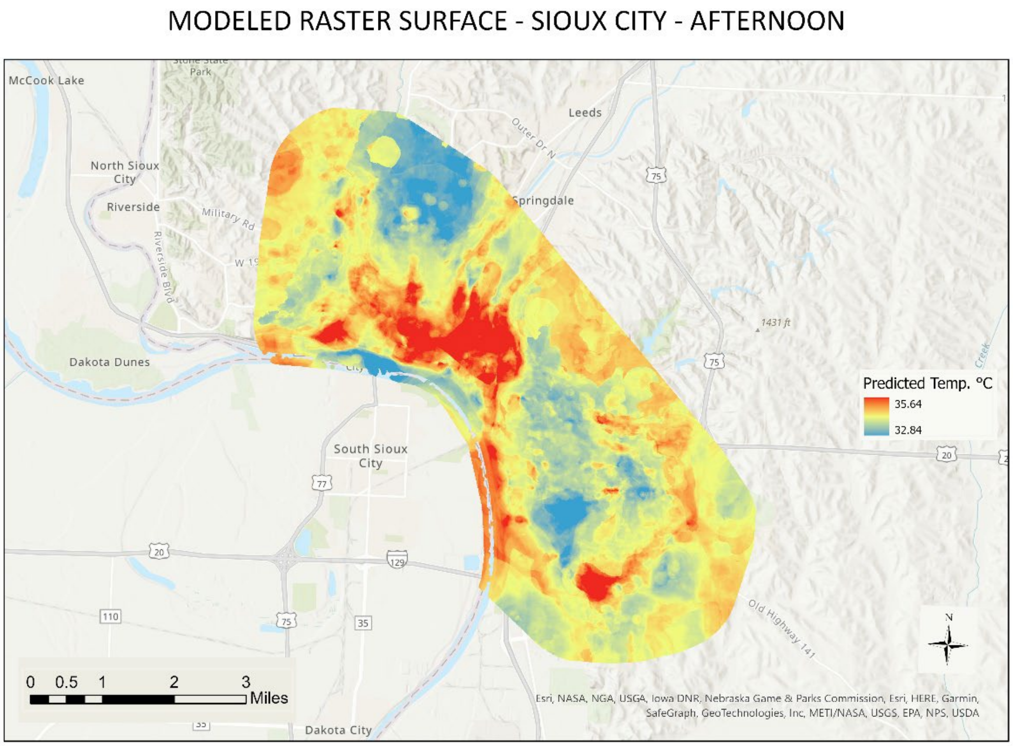


Figure 57
Modeled raster surface for Sioux City evening

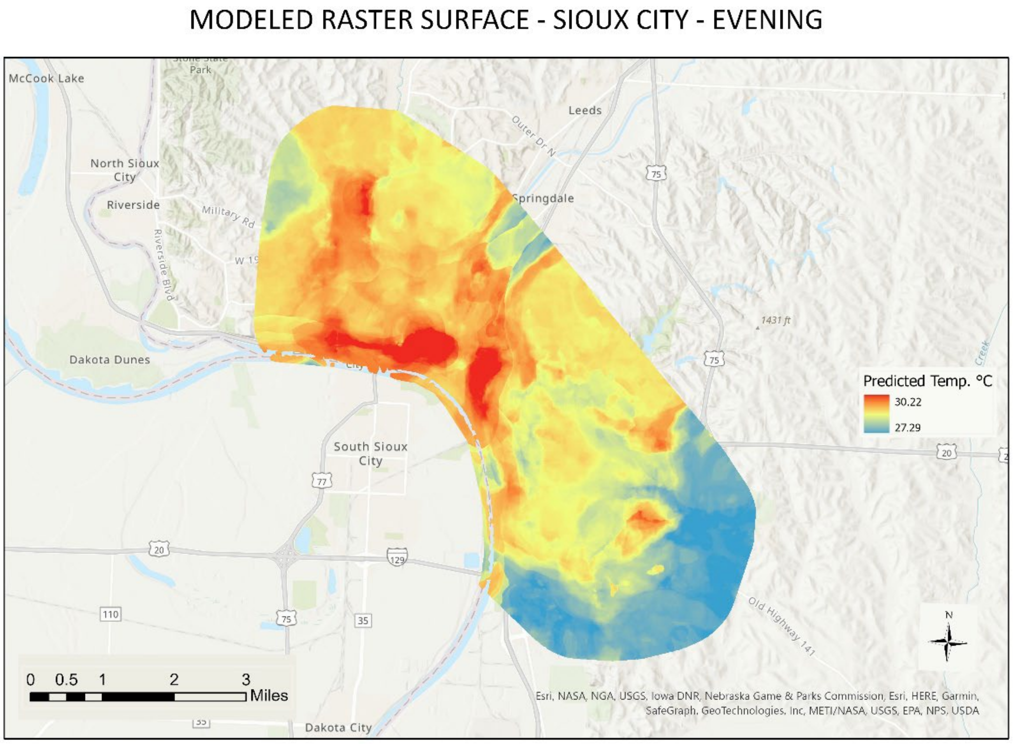
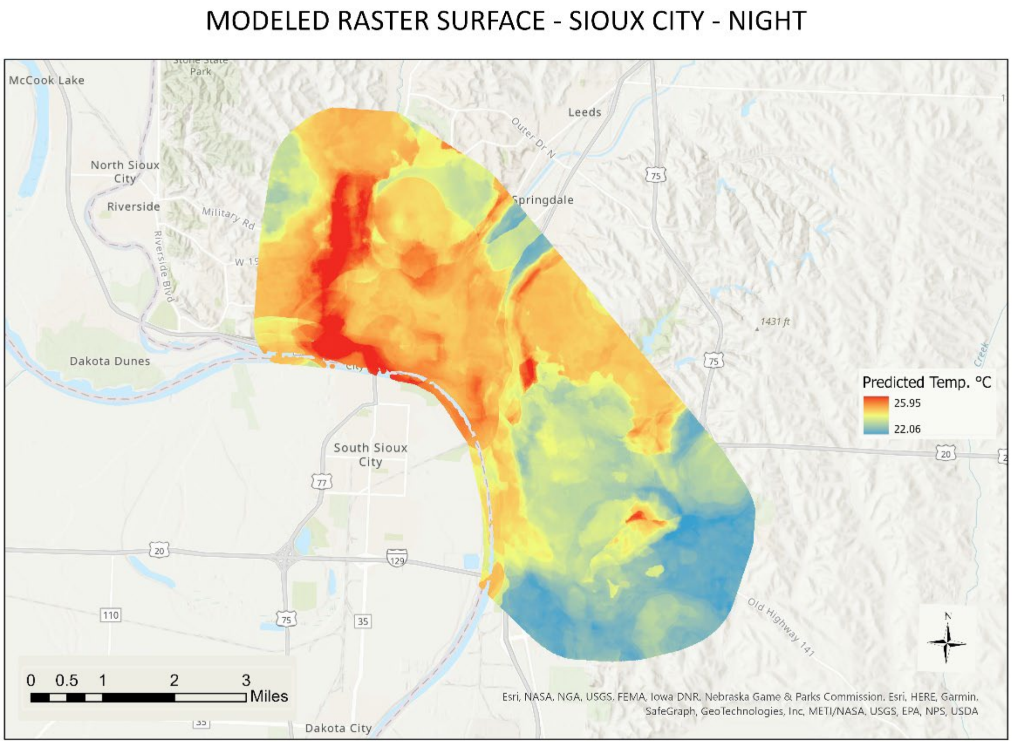


Figure 58
Modeled raster surface for Sioux City night



The only study area that comprised two different cities (Waterloo–Cedar Falls) showed modeled temperature rasters consistent with the temperature collected at all times of the day (Figures 59 to 61). The proportion of buildings in areas such as downtown Waterloo, downtown Cedar Falls, and the University of Northern Iowa campus influenced the high temperatures seen in those areas, especially during the nighttime, which is supported by the most important variables for that model (NDVI and BV).

Figure 59

Modeled raster surface for Waterloo/Cedar Falls afternoon

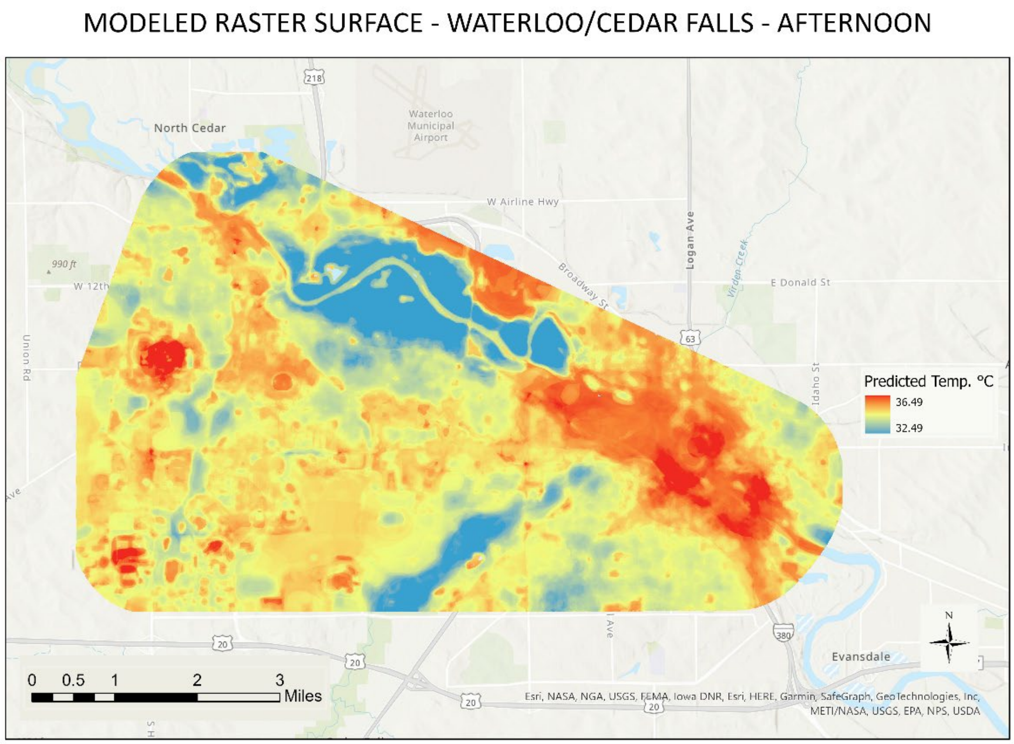


Figure 60

Modeled raster surface for Waterloo/Cedar Falls evening

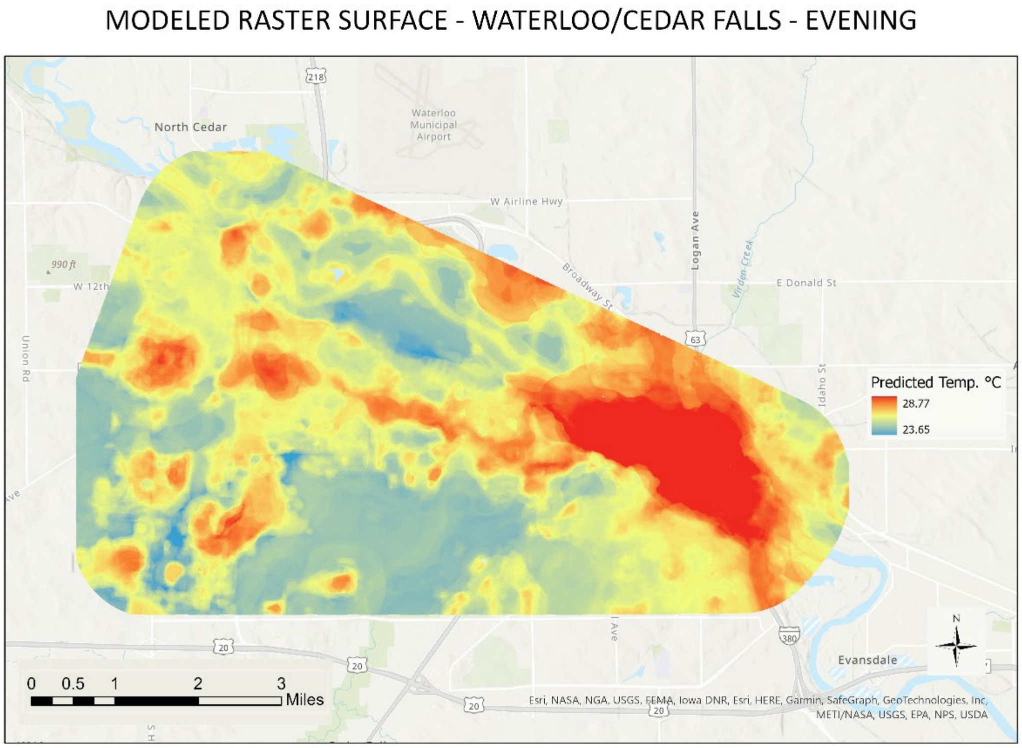
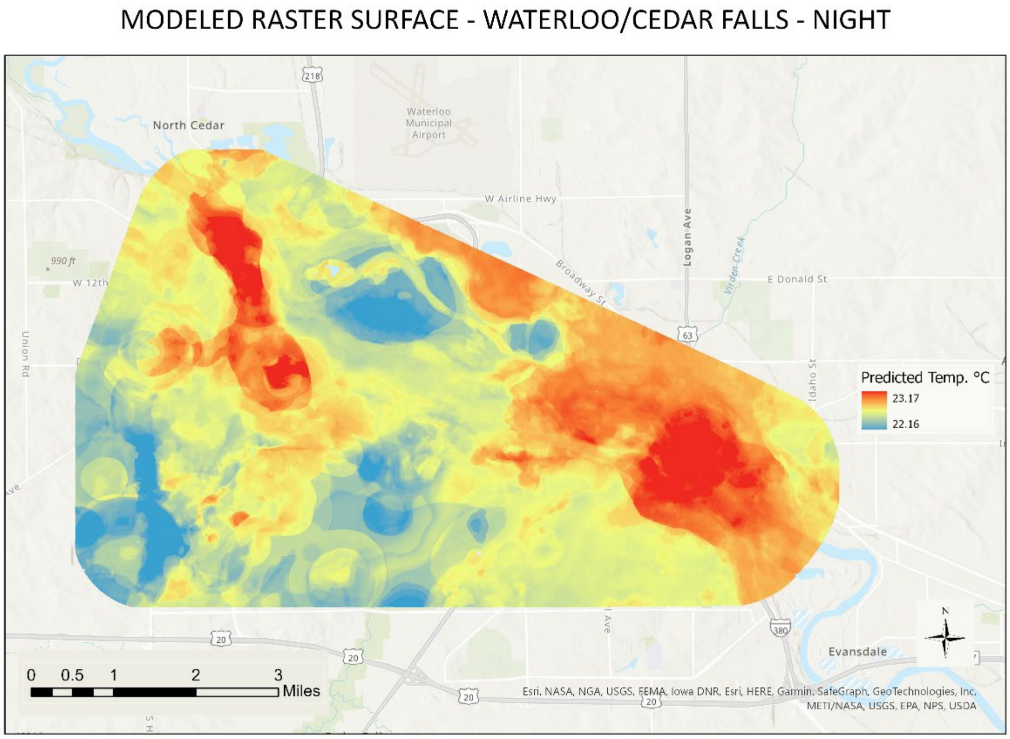


Figure 61

Modeled raster surface for Waterloo/Cedar Falls night



Finally, the models generated for Waverly (Figure 62 and Figure 63), show a distinctive pattern for the afternoon model, presenting small areas of high heat, although generally consistent with the temperature collected in the area. Given that building height calculated as standard deviation (BH_STD) was only seen as the most important variable in this model in specific (other models presented building height as the most important variable but calculated as Mean), it is believed that this pattern can be the result of the influence of that specific variable.

Figure 62

Modeled raster surface for Waverly afternoon

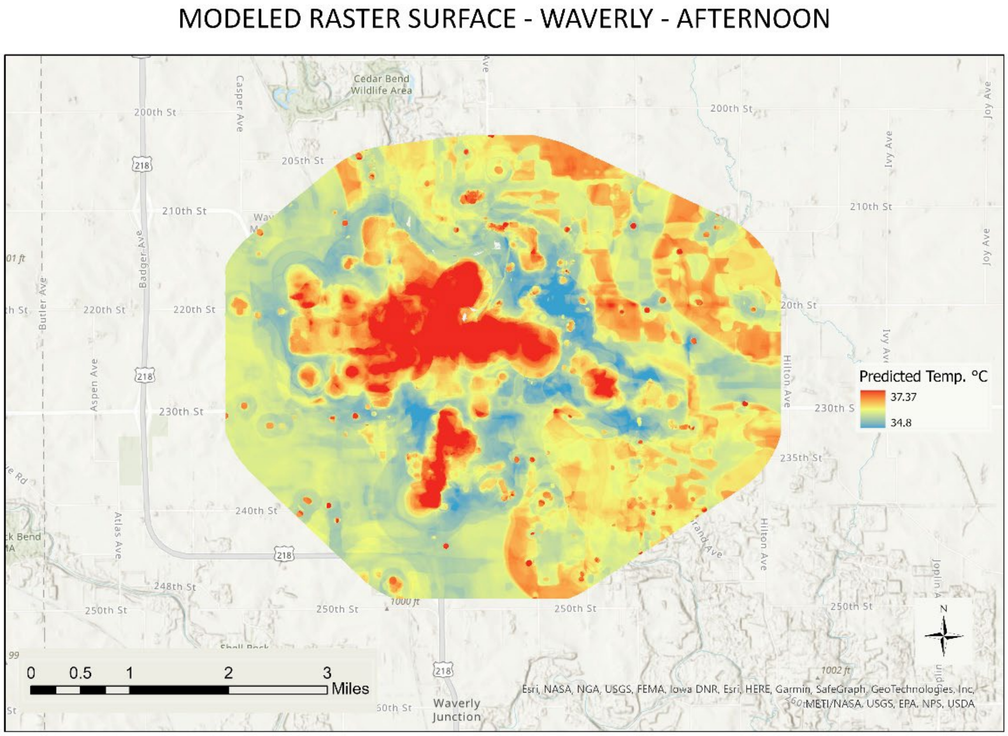
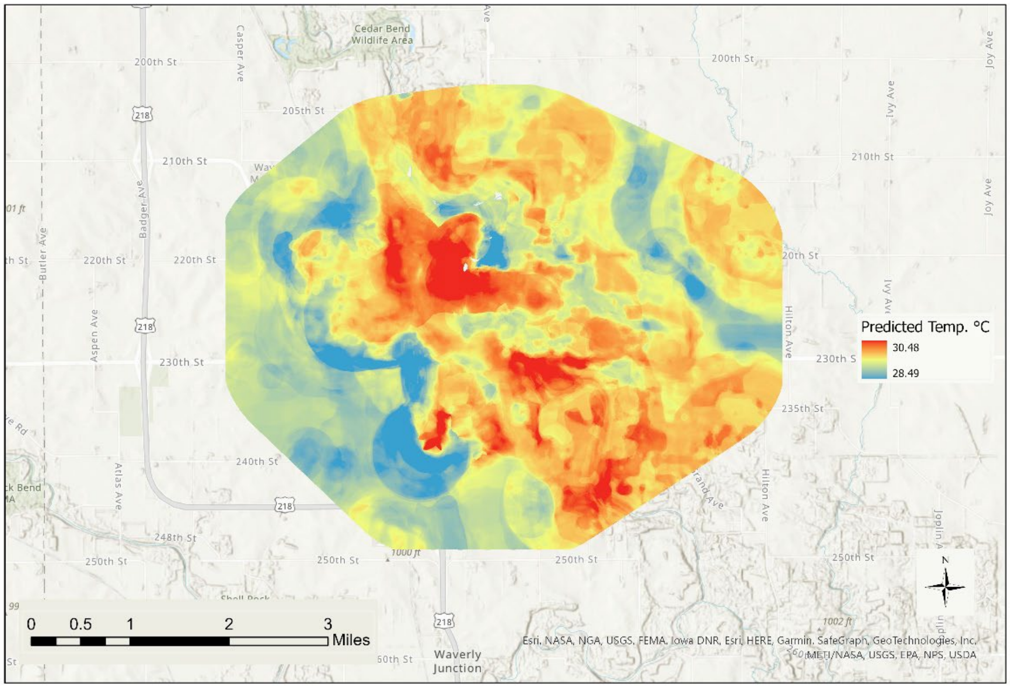


Figure 63
Modeled raster surface for Waverly night

MODELED RASTER SURFACE - WAVERLY - EVENING



In addition to the statistics output by the modeling, when comparing the predicted raster surface models with measured temperatures in each study area by the time of the day, values were presented with differences no higher than 0.5°C degrees for minimum temperatures (Waterloo/Cedar Falls - Evening) and no lower than 0.46°C degrees for maximum temperatures (Council Bluffs – Afternoon). Table 15 shows all cities and the time of collection with the difference of measured and predicted temperature for minimum and maximum temperatures.

Table 15

Measured vs predicted temperature by urban area/time of the day

URBAN AREA		MEASURED AND PREDICTED TEMPERATURES				VARIANCE IN MEASURED AND PREDICTED TEMPERATURES	
CITY	TIME OF THE DAY	MIN. MEASURED	MAX. MEASURED	MIN. PREDICTED	MAX. PREDICTED	VARIANCE IN MIN. TEMPERATURE	VARIANCE IN MAX. TEMPERATURE
Burlington	Afternoon	31.41	35.12	31.48	35.03	0.07	-0.09
Burlington	Evening	28.47	30.99	28.51	30.97	0.04	-0.02
Burlington	Night	25.92	27.82	25.96	27.8	0.04	-0.02
CedarRapids	Afternoon	32.5	35.97	32.62	35.62	0.12	-0.35
CedarRapids	Evening	25.95	29.47	26.04	29.38	0.09	-0.09
CedarRapids	Night	22.36	25.57	22.54	25.47	0.18	-0.1
Council Bluffs	Afternoon	27.84	32.64	27.88	32.18	0.04	-0.46
Council Bluffs	Evening	23.4	26.6	23.42	26.38	0.02	-0.22
Council Bluffs	Night	18.66	21.59	18.72	21.5	0.06	-0.09
DesMoines	Afternoon	35.62	39.77	35.76	39.5	0.14	-0.27
DesMoines	Evening	29.75	32.91	29.8	32.78	0.05	-0.13
DesMoines	Night	26.95	29.3	27.09	29.26	0.14	-0.04
FortDodge	Afternoon	30.34	34.6	30.49	34.38	0.15	-0.22
FortDodge	Evening	22.09	27.6	22.33	27.49	0.24	-0.11
Marshalltown	Afternoon	32.79	36.9	32.89	36.58	0.1	-0.32
Marshalltown	Night	21.33	22.42	21.36	22.37	0.03	-0.05
Sioux City	Afternoon	32.75	35.85	32.84	35.64	0.09	-0.21
Sioux City	Evening	27.15	30.4	27.29	30.22	0.14	-0.18
Sioux City	Night	22.04	26.05	22.06	25.95	0.02	-0.1
WaterlooCF	Afternoon	32.35	36.57	32.49	36.49	0.14	-0.08
WaterlooCF	Evening	23.15	28.81	23.65	28.77	0.5	-0.04
WaterlooCF	Night	22.14	23.19	22.16	23.17	0.02	-0.02
Waverly	Afternoon	34.73	37.66	34.8	37.37	0.07	-0.29
Waverly	Evening	28.27	30.52	28.49	30.48	0.22	-0.04

An example of measured temperatures overlapping predicted temperatures can be seen in the examples for Des Moines afternoon, Des Moines evening, and Des Moines night (Figures 64 to 66).

Figure 64

Overlap of measured temperature and predicted temperature in Des Moines (afternoon)

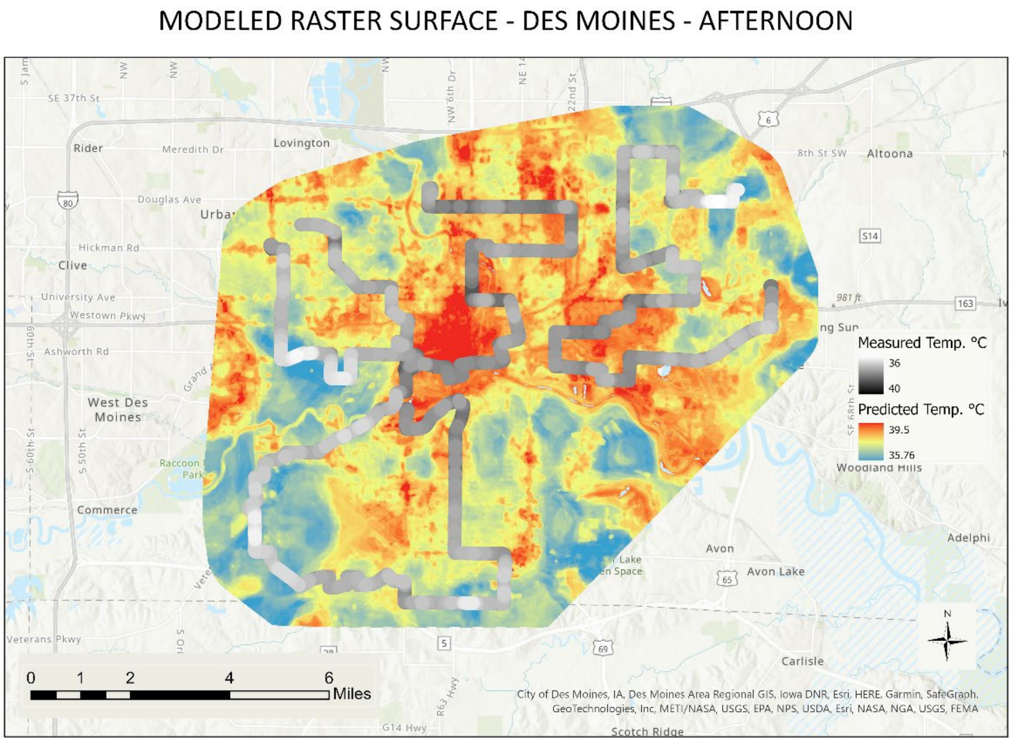


Figure 65

Overlap of measured temperature and predicted temperature in Des Moines (evening)

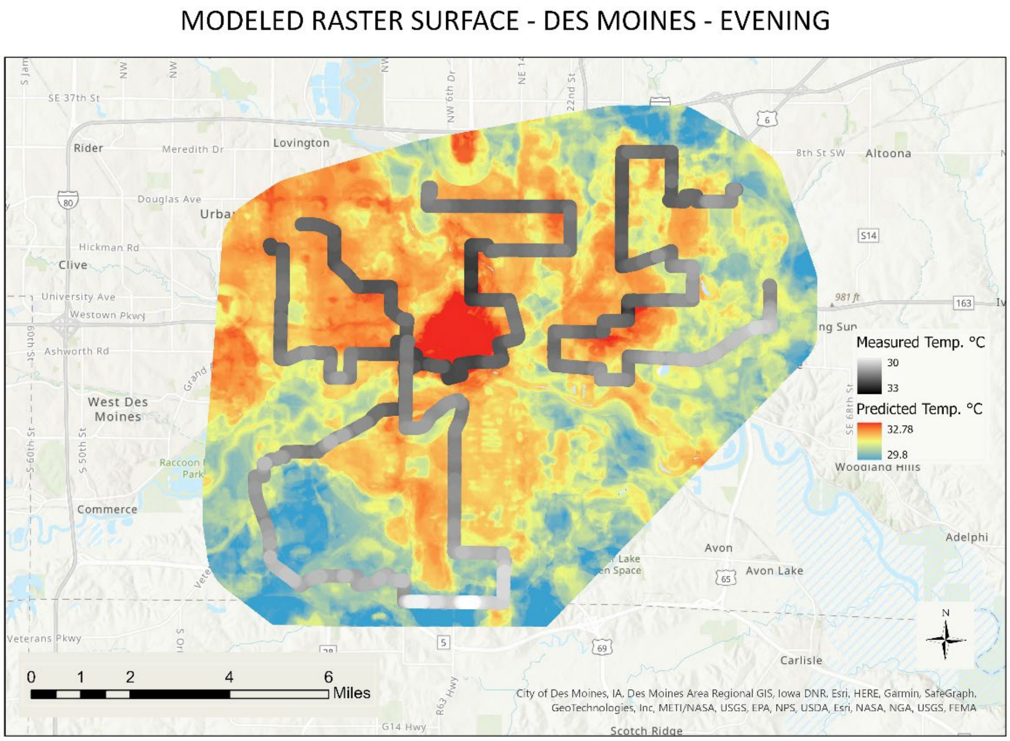
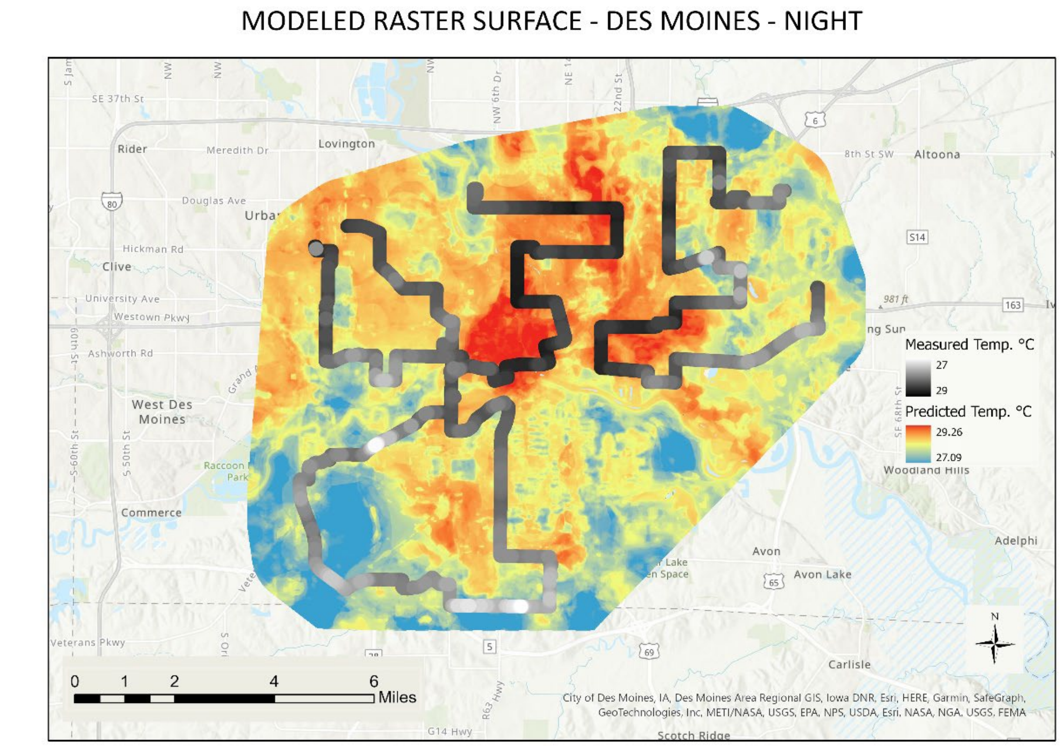


Figure 66

Overlap of measured temperature and predicted temperature in Des Moines (night)



Temperature and socio-demographics

In order to examine the variation in modeled air temperature across neighborhoods with different socio-demographic characteristics in each urban area, collected air temperature were summarized by block groups. Income data was downloaded directly from Census populational data website (Census 2022) and represented the Median Household Income per block group in 2021. This income data comes from the American Community Survey (ACS – 5 years estimate). To understand how underrepresented populational groups face urban heat, a dataset containing the total population per block group by race was also utilized.

A unique variable considering the percentage of non-white population per census block group was calculated since it has been seen through the literature (e.g., Hattis et al. 2012, Voelkel et al. 2018) that no particular minority group is exempt from experiencing higher temperatures than non-marginalized groups. The outcome provided by the tool Zonal Statistics as Table in ArcGIS Pro provided the mean modeled air temperature by block group in each one of the 10 Urban areas by all times of the day (when available) utilizing the predicted surface raster model. This allowed for a comparison of the median household income and percentage of non-white population which were already held per geographical area.

Results acquired from running the formula for Pearson correlation coefficient in Microsoft Excel Professional Plus 2021 for Income and modeled air temperature showed values ranging from -0.0125 (the lowest correlation) to -0.5948 (the highest correlation). Those two values represent the results for Waverly/Evening and Marshalltown/Afternoon respectively. As can be seen in Table 16, all values presented a negative correlation (all values highlighted in bold have a statistically significant p-value <0.05), which is consistent with the literature that suggests that areas with higher income experience less heat than areas with lower income. Out of 24 total results, one per each time of collection and urban area, 4 results fell into moderate negative correlation ($-0.5< r <-0.70$), with the remaining values falling into either low negative correlation ($-0.3<r<-0.50$) or very low negative correlation ($r<-.03$).

Table 16

Pearson correlation coefficient for income vs. air temperature

PEARSON CORRELATION COEFFICIENT FOR MEDIAN INCOME VS. AIR TEMPERATURE			
URBAN AREA	AFTERNOON	EVENING	NIGHT
Burlington	-0.3618	**-0.4175**	-0.3199
Cedar Rapids	**-0.3636**	**-0.2680**	**-0.3487**
Council Bluffs	**-0.3469**	**-0.4217**	**-0.4591**
Des Moines	**-0.4481**	**-0.2539**	**-0.3294**
Fort Dodge	-0.2586	-0.3547	
Marshalltown	**-0.5948**		**-0.5792**
Sioux City	**-0.3094**	**-0.3265**	-0.2403
Waterloo/CF	**-0.4751**	**-0.5521**	**-0.5378**
Waverly	-0.3731	-0.0125	

Results of a linear regression considering the median household income per block group as the dependent variable and mean modeled air temperature per block group as an independent variable showed mixed results. While most urban areas/times of the day showed statistically significant relation to income (p-value > 0.15), the adjusted R^2 presented values over 0.20 only in 5 areas/time of the day (Marshalltown and Waterloo/Cedar Falls in all collection times), showing a rather weak but significant correlation between income and measured air temperature as can be seen in Table 17. As Table 17 illustrates, some urban areas/times of the day provided a weak relationship between median household income and mean air temperature, with the strongest relationship in Marshalltown during the afternoon, with an R^2 of 0.328 and a p-value of 0.0011.

Table 17

Linear regression of median household income and mean air temperature by block group

LINEAR REGRESSION - MEDIAN AIR TEMP - MEDIAN INCOME (BY BLOCK GROUP)						
URBAN AREA	AFTERNOON R2	P-VALUE	EVENING R2	P-VALUE	NIGHT R2	P-VALUE
Burlington	0.1054	0.0301	0.1500	0.0113	0.0759	0.0572
Cedar Rapids	0.1227	0.0003	0.0613	0.0094	0.1120	0.0006
Council Bluffs	0.1057	0.0057	0.1641	0.0006	0.1976	0.0002
Des Moines	0.1969	0.0000	0.0599	0.0002	0.1042	0.0000
Fort Dodge	0.0263	0.2120	0.0878	0.0819		
Marshalltown	0.3280	0.0011			0.3089	0.0015
Sioux City	0.0835	0.0065	0.0945	0.0040	0.0450	0.0365
Waterloo/CF	0.2142	0.0000	0.2944	0.0000	0.2787	0.0000
Waverly	0.0316	0.2883	-0.1248	0.9727		

A table showing the highest adjusted R^2 (>0.19) found when examining the correlation between modeled air temperature and income by block group can be seen in Table 18, altogether with the corresponding slope value for each area/time of the day.

Table 18

Highest Random Forest adjusted R2 for income vs modeled air temperature

HIGHEST ADJUSTED R2 WITH STATISTICAL SIGNIFICANCE				
URBAN AREA	TIME OF THE DAY	R2	P-VALUE	SLOPE
Marshalltown	Afternoon	0.3280	0.0011	-22984.1114
Marshalltown	Night	0.3089	0.0015	-72702.2765
Waterloo/CF	Evening	0.2944	0	-42149.6499
Waterloo/CF	Night	0.2787	0	-28218.7518
Waterloo/CF	Afternoon	0.2142	0	-152507.3410

All cities with the highest adjusted R^2 presented a negative slope, which can be interpreted as the lower the income the higher temperature experienced by residents in the block groups of the urban areas. Even though some urban areas do present a negative value for R^2, those are not statistically significant to draw any conclusion.

When considering the proportion of non-white residents in each urban area per block group, the outputs of Pearson correlation coefficient and linear regression showed a variety of results, however, different than those seen in other studies (e.g., Shandas 2019). In this study, in most urban areas, ethnicity doesn't seem to be highly statistically significantly correlated to higher temperatures. Table 19 depicts the Pearson correlation coefficient results for all urban areas and time of the day with values ranging from -0.0078 to 0.4422 (all values highlighted in bold have a statistically significant p-value <0.05), which were the results obtained for Waverly/Evening and Sioux City/Evening respectively. It is worth mentioning that Waverly Evening was the only negative coefficient of all, and it is believed that due to the small sample size (only 10 block groups in total within the urban area), the results can't be considered as statistically significant in this situation.

Table 19
Person correlation coefficient for ethnicity vs. air temperature

PEARSON CORRELATION COEFFICIENT FOR ETHNICITY (NON-WHITES) VS. AIR TEMPERATURE			
URBAN AREA	AFTERNOON	EVENING	NIGHT
Burlington	0.1828	0.1821	0.1145
Cedar Rapids	**0.3052**	**0.2726**	**0.2864**
Council Bluffs	0.1191	0.1156	0.1302
Des Moines	**0.2837**	**0.2520**	**0.2644**
Fort Dodge	0.1858	0.2267	
Marshalltown	0.0836		0.0393
Sioux City	0.2474	**0.4422**	**0.4291**
Waterloo/CF	**0.3153**	**0.4081**	**0.4117**
Waverly	0.2232	-0.0078	

Out of 24 coefficients for all urban areas and time of collection, only 6 values presented a low positive correlation between ethnic minorities presence and air temperature (r> 0.3), while the remaining 18 values fell into very low correlation (r<0.3).

The linear regression results for each area/time of the day can be seen in Table 20, with its respective R^2 and statistical significance.

Table 20

Linear regression of median air temperature and proportion of minority groups by block group

LINEAR REGRESSION - MEDIAN AIR TEMP - MINORITY GROUP (BY BLOCK GROUP)						
URBAN AREA	AFTERNOON R2	P-VALUE	EVENING R2	P-VALUE	NIGHT R2	P-VALUE
Burlington	0.0073	0.2654	0.0070	0.2671	-0.0136	0.4875
Cedar Rapids	0.0836	0.0024	0.0646	0.0069	0.0724	0.0045
Council Bluffs	-0.0010	0.3370	-0.0018	0.3514	0.0018	0.2937
Des Moines	0.0775	0.0000	0.0556	0.0003	0.0630	0.0001
Fort Dodge	-0.0026	0.3439	0.0149	0.2461		
Marshalltown	-0.0285	0.6604			-0.0341	0.8366
Sioux City	0.0493	0.0260	0.1853	0.0000	0.1738	0.0001
Waterloo/CF	0.0867	0.0066	0.1548	0.0003	0.1578	0.0003
Waverly	-0.0690	0.5354	-0.1249	0.9829		

With the exception of a few areas which account for a small number of block groups, as for example Waverly (10), Fort Dodge (28), and Marshalltown (30), which were expected to present a low to non-statistical significance in the linear regression models, the urban area of Burlington and Council Bluffs, also presented a p-value that categorizes all their results as non-statistically significant (p-value > 0.15), even though the number of blocks in those areas was considerably higher than in the previously mentioned urban areas. The areas/time of the day that presented the highest adjusted R^2 and very strong statistical significance can be seen in Table 21, which also shows the slope values for each individual model.

Table 21

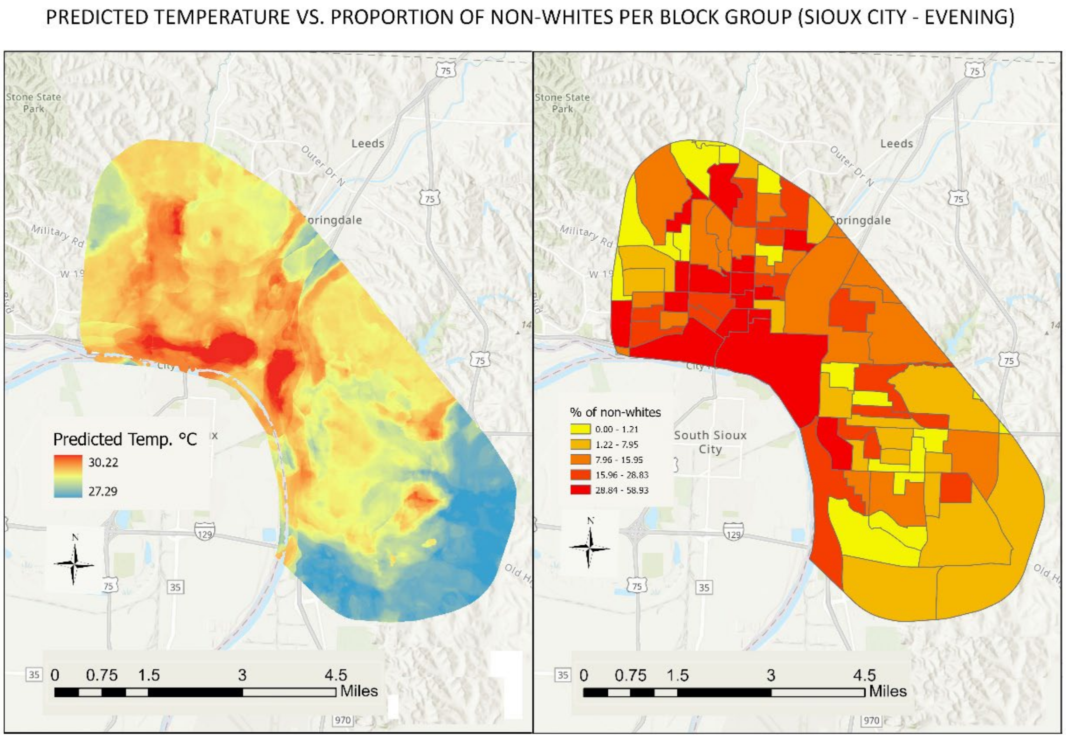
Highest adjusted R2 for proportion of minorities vs measured air temperature

HIGHEST ADJUSTED R2 WITH STATISTICAL SIGNIFICANCE				
URBAN AREA	TIME OF THE DAY	R2	P-VALUE	SLOPE
Sioux City	Evening	0.1853	0.0000	0.2652
Sioux City	Night	0.1738	0.0001	0.1235
Waterloo/CF	Night	0.1578	0.0003	0.8279
Waterloo/CF	Evening	0.1548	0.0003	0.1454
Waterloo/CF	Afternoon	0.0867	0.0066	0.1872

Out of all 24 models, Sioux City and Waterloo/Cedar Falls presented the most significant results. The highest adjusted R^2 of all models presented a value of 0.1853, a p-value of 0.0000, and a slope of 0.2652. This means for each degree increase in temperature (°C), the proportion of minority groups increases in 0.2652 (26.52%). Even though, the results presented in Table 21 are all statistically significant, the explanation capacity of the linear regression models suggests that there are other components that account for the relationship between variation of temperature and the presence of minority ethnical groups in each block group. Figure 67 presents a comparison between the modeled evening temperature in Sioux City (with the highest adjusted R^2 among all) and the distribution of ethnic minorities in the same region, allowing for a visual inspection of their relationship.

Figure 67

Predicted surface raster temperature and proportion of non-whites per block group in Sioux City (evening)



Chapter 5

Discussions and Conclusions

Discussion

This study resulted in a systematic examination of temperature patterns across a variety of urban areas in Iowa based on an extensive in-field temperature monitoring regime (collecting during the afternoon, evening and night, consistently for one hour in each data collection timeframe) and a highly detailed spatial statistical modeling effort, leading to unique results. The three selected times for data collection were based on existing literature aiming to comprehend urban heat patterns. Specifically, the study by Oke (1982) established that urban areas experience their highest air temperatures between 2 p.m. and 5 p.m., (corresponding to the afternoon data collection), and densely urban areas exhibit the most significant air temperature differences compared to non-built-up regions a few hours after sunset (corresponding to our evening data collection). Moreover, the nighttime period, known to worsen health issues related to heat exposure, has been emphasized in studies such as Basu (2009), Heaviside et al. (2017), and Voelkel et al. (2018), being the reason why air temperature data collection also occurred during the night (4 a.m. to 5 a.m.) in the current research.

The temperature data were measured across a variety of land uses in multiple Iowa cities at multiple times of the day resulting in the observation of varying temperature patterns. This sampling effort led to a unique and robust dataset for air temperature measurements (more than 110,000 temperature points were collected) in ten different areas of diverse urban features. Even though methods of air temperature collection measured by mobile devices have been researched in previous studies (e.g., Hart and Sailor, 2008; Yokobori and Ohta, 2009; Voelkel and Shandas, 2017), this study represents a novel effort to collect data across multiple urban areas in a single state. Temperature collection routes were designed to capture temperature data across a variety of urban land uses thus providing the necessary data required for the development of geospatial models with high predictive power throughout most, if not all urban areas.

Modeled results generated by random forest algorithms resulted in coefficients that varied from an R^2 of 0.879 to 0.997 (all statistically significant), with most coefficients resulting in R^2 higher than 0.95, which seemed consistent with other studies utilizing mobile sensors for temperature data collection, LiDAR data and/or aerial imagery to derive independent variables, and RF as machine learning algorithm as seen in Voelkel and Shandas (2017), Shandas et al. (2019), and Oukawa et al. (2022). These outcomes strongly indicate that the chosen independent variables and neighborhood distances were effective in generating modeled outcomes that are theoretically consistent with existing literature. In some cases, the predictive power of the results exceeds the ones found in previous studies, for example as in Voelkel et al. (2018), which results for the city of Portland presented an R^2 of 0.8199 for the model utilizing temperatures collected during the afternoon. Even though the collected air temperature data provided enough information for statistically significant predictive models to be developed, it is not certain how the model would perform if data were collected at different times of the day. Predicted surface raster temperatures in this study were found to be within a very similar range of collected air temperature, showing the expected consistency between collected air temperature and predicted air temperature. In addition, it is unclear how predictive models would have performed if air temperature data were collected throughout different routes in the chosen urban areas.

The results found by modeling and predicting the air temperature in different areas showed similarity with previous studies such as in Voelkel et al. (2018) and Shandas et al. (2019), where the highest predictive power was seen during the evening or night periods. The number of independent variables, even though they were divided into different neighborhood scales (50, 100, 200, 400, and 800 meters), showed that no statistically significant improvements in air temperature predictability would be seen if additional variables such as wind speed, albedo of buildings and roads, radiation, sky view factor, or housing density, were introduced.

Although a similar approach to air temperature collection (through the use of mobile sensors) and to derive independent variables were used in this study compared to the existent literature, the amount of temperatures collected in each study area and the number of independent variables selected to generate the air temperature models varied greatly from other studies.

Oukawa et al. (2022), temperature was collected in 12 different sites utilizing mobile sensors and models were developed through use of RF algorithm. However, in contrast to this study, more than 30 independent variables were utilized (not accounting for buffer sizes), considering atmospheric vertical indices (e.g., boundary layer height, total column water vapor), population and traffic (e.g., population density, road length), urban morphology (e.g., mean building height, building volume density, sky view factor), and weather data (e.g., relative humidity, wind speed, atmospheric pressure, incoming solar irradiation).

In Voelkel et al. (2018), the number of independent variables chosen to be part of the predictive model covered similar features as this study, accounting for natural features (e.g., canopy cover, canopy density), built-up structures (e.g., building height, building volume), and a calculated vegetation index (such as NDVI). In contrast, the number of air temperature measurements surpassed 50,000 points for a study area that occupies approximately 145mi^2 (US Census Bureau 2021), a considerable difference from the larger study area of this study (Des Moines), which consists in approximately 90mi^2 where an average of 10,000 air temperature points were collected per time of collection.

In Shandas et al. (2019), a study that used mobile temperature sensors to collect air temperature and similar number of independent variables to derive predictive models utilizing RF, collected an average (total number of collected points divided by each run) of 34,657, 24,682, and 26,435 air temperature measurements in areas of 59.93mi^2 (Richmond, VA), 80.95mi^2 (Baltimore, MD), and 61.13mi^2 (Washington, D.C.), respectively (US Census Bureau 2021). The hypothesis based on previous and the current study is that both the amount of air temperature collected per study area does play a role as well as the number of variables utilized as independent variables during the modeling phase, but extra efforts in collecting a larger sample seem to be unnecessary when comparing this study to the ones conducted by Voelkel et al. (2018) and Shandas et al. (2019). Specifically, considering just the number of air temperature measurements collected in this study, cities with fewer (-50%) measurements (e.g., Fort Dodge, Marshalltown, and Sioux City) showed minimal to no impact on the predictive models.

This could imply that shorter temperature collection times might yield comparable predictability for new models, leading to improved efficiency and the feasibility of executing data collection over shorter periods due to logistical and financial constraints.

Surprisingly, the consistency of specific independent variables as the most important, or explaining the most variation in the models (e.g., NDVI), in predicting temperature across different study areas showed that, in theory, even fewer independent variables could have been utilized. NDVI not only appears as the most important variable, it also appears as the most reoccurring important variable for the second, fourth, and fifth most important variable across all 24 models. In comparison to the study conducted by Voelkel and Shandas (2017), the most important variable for the three models provided (morning, afternoon, and evening), was building height, while in studies such as Oukawa et al. (2022) and Shandas et al. (2019) found the most important variables in explaining the modeled temperature were relative humidity and NDVI.

Regarding the five neighbor distances employed in this reserach, 800 meters is the one appeared the most frequently as the first most important variable (54.1%), second most important variable (54.1%), fourth most important variable (62.5%), and fifth most important variable (54.1%). Out of all neighbor distances, 50 meters had the least common appearance as the most important variable, regardless of the position (1^{st}, 2^{nd}, 3^{rd}, 4^{th}, or 5^{th} most important variable), with only 4 appearances, followed by 100 meters with 10 appearances, 200 meters with 19 appearances, 400 meters with 24 appearances, and 800 meters with 63 appearances. These results are somewhat similar to the ones found in Voelkel and Shandas (2017) in which neighborhood distances between 800 and

1000 were most commonly shown to be important overall (53.3% of all 5 most important variables of all models) and especially for morning and evening data collections.

For this study, to address the third objective ("To examine the temperature variation across urban neighborhoods with varying socio-demographic characteristics"), a different approach was used to carry out statistical analyses. In contrast to the second objective, where air temperature was treated as a dependent variable, the focus shifted to explore how air temperature varies regarding income and non-white population. For this purpose, income and non-white population were held constant, treating air temperature as the independent variable. For the results, interestingly and somewhat in contradiction to a number of studies (e.g., Shandas 2009, Hattis et al. 2012) that seek to understand how different demographics are affected by urban heat, in specific low-income neighborhoods, this study found that, even for statistically significant results, there was only low to moderate negative correlations between income and air temperature for the selected urban areas. Regarding the correlation between ethnical minorities and modeled air temperatures, cities such as Cedar Falls, Sioux City, and Waterloo presented evidence that racial minorities do tend to experience more heat than white populations in the urban areas. While the findings may not exhibit the same level of severity observed in other studies, the disparity in heat exposure among minority groups is a matter of concern that deserves heightened attention from local and state government.

Conclusions

The primary goal of this research was to monitor and model urban heat patterns in the state of Iowa by using a high temporal resolution mobile sensor and high spatial resolution geospatial data. Air temperature data were collected by mobile sensors in 10 urban areas and during multiple times of the day (afternoon, evening, and night), each for a period of approximately one hour, mainly during the summer of 2022. These data were leveraged by a machine learning algorithm (RF) to model urban temperature patterns in those 10 cities at different times while using highly detailed derived geospatial data on urban morphology (e.g., buildings) and greenness.

The use of the RF algorithm in predicting temperature across all urban areas resulted in significant R^2 coefficients, where 83% of the models exhibited values higher than 0.95, and 62.5% showed values over 0.97, all being statistically significant (p-value <0.01). This indicates that the approach of creating detailed geospatial temperature patterns by incorporating a diverse range of morphometric and natural features as independent variables (Canopy Cover, Canopy Density Metric, Building Height, Building Volume, and Normalized Difference Vegetation Index) in the urban environment proved to be effective. The highly statistically significant predictive power of the multiple developed models also offered valuable insights on the impacts that man-made and natural features (the foundation to derive most independent variables) have in contributing or preventing urban heat across the state of Iowa. These models also revealed that independent variables exert a stronger statistical influence when considering larger neighbor distances compared to small areas. Additionally, forest-based algorithms proved to be effective in this context, and the study highlighted how urban heat disproportionately affects different neighborhoods with varying sociodemographic characteristics.

As this research was conducted for a project (Iowa Economic Development Authority, Iowa Energy Center Grant Program, Agreement Number: 21-IEC-012) that has the purpose to ensure its replicability and produce beneficial socioeconomic outcomes for the communities in all urban areas where data was gathered, including every city in the state of Iowa, the measured data, modeled temperature data, and other resources from the study will be made publicly available to aid in their effective use moving forward. Urban planners, landscape architects, zoning specialists, and public health experts are just a few of the professionals that can potentially make use of the content provided in this thesis to develop ways of mitigating urban heat (e.g., by implementing highly vegetated areas) and establishing or improving environmentally equitable housing to prevent and respond to heat-related health issues. Moreover, continuous investment in understanding urban heat patterns will certainly allow for preventive measures to be taken at different scales, as well as mitigative actions that can develop novel techniques, significantly enhancing the overall quality of life for both individuals and the community as a whole. Ultimately, throughout collecting air temperature data in a variety of urban settings and by analyzing the differences in temperature across spatially distinct areas, all urban areas in this study experience urban heat regardless of their size and landscape features.

Limitations

Various limitations were encountered during the course of this research, and they can be categorized according to the three specific objectives of the study. The first objective, which pertains to the collection of air temperature data, presented limitations regarding the ability to characterize and utilize data on weather variation during the time of collection, for example, winds coming from different directions, humidity, cloudiness, and the angle at which the sun rays are directed and can be reflected by different morphometric features in the urban environment. While the author acknowledges that these limitations can exert a significant influence on urban temperatures, they fall outside the scope of this research. Unexpected circumstances such as malfunctioning devices, and heavy traffic during the time of collection were seen on a few occasions. Those situations have the potential to delay data collection, prevent the methods to be accurately applied in a consistent manner throughout different study areas, or postpone data acquisition, directly influencing the current planned routes or future collections. In situations where temperature variability is below 2 degrees Celsius throughout the whole route can also be a challenging situation. All mobile devices used during this research had an air temperature reading accuracy of 0.2°C, resulting in a potential variance of 0.4°C (-0.2°C to +0.2°C). For night temperature collection, which usually presented the lowest temperature variances during this study (e.g., Marshalltown/Night, Waterloo-Cedar Falls/Night), a margin of error of 0.4°C can represent a considerable amount of variance that might not be correctly explained by the different features encountered in the urban environment throughout the collection.

Even though the goal of this study was to collect air temperature data during hot days in different study areas, temperature patterns for urban areas during cold days are seldom explored by existent literature. While extensive research has been conducted on temperature patterns in urban environments, the focus has predominantly been on urban heat during extremely hot weather conditions. The lack of research in areas where temperatures frequently drop to sub-zero temperatures during winter warrants the need for additional studies.

Objective two, which focused on modeling and predicting temperature patterns across the whole urban area, has the potential to be affected by different factors, including the routes executed during the data collection phase not representing the true diversity of features found in the urban area; availability of high-resolution LiDAR, high-resolution satellite imagery, and accurate building footprints; and the possibility of unanticipated errors that might happen during the data processing phase, such as banding on a few raster files as it was experienced during this study. Furthermore, the performance of predictive models based on temperature collected at different timeframes or during longer or shorter routes remains uncertain, as the collection of air temperature utilizing different methods within the selected urban areas has not been conducted.

Objective three poses different challenges regarding data reliability and/or the lack of specific information that might be useful for accurate socioeconomic analyses. Reported income might not always best represent the income condition of a household, nor does it provide the information on how many people contribute to the published income for the unit, which can affect the statistics of a larger demographic area (e.g., census blocks, block groups, census tracts, etc.). Even though the chosen geographic area provided an acceptable framework for the comparison between air temperature and income/ethnicity, a more comprehensive study considering a smaller geographic unit (census blocks) could provide a better representation of how specific demographic groups experience heat in the urban environment, which would be supported by the high-resolution sensors and geospatial data utilized during this study. In addition, according to the literature (Shandas 2009, Voelkel et al. 2018, Alizadeh et al. 2022), it was expected that minority groups would be consistently shown as disproportionally affected by urban heat throughout most if not all study areas, which was not the case for most cities in this study. Innumerable factors could have contributed to the results, such as areas having a higher land value than their counterparts; gentrification; the existence of groups classified as bi- or multiracial; and also possibly due to the history of urban development in Iowa being different from some of the large metropolitan areas examined in other studies.

Future Directions

Even though the author believes this study accomplished its goal and objectives, further research would be beneficial to the understanding and modeling urban heat patterns in single or multi-study areas. Considering the high R^2 obtained from the temperature models in this study, the question of what is the minimum number of temperature points collected throughout the urban area that would give similar results is still a question to be answered. Besides, the number of variables and neighbor distances necessary to create accurate and reliable temperature models can potentially inspire new research using similar methods and geospatial data, extrapolating models to even larger areas outside of the urban environment (e.g.,>1,500 meters) , considering different variables that can portray the heterogeneity of the urban fabric (e.g., SVF, humidity, proportion of impervious surfaces, albedo of built-up structures), or reducing to a fewer number of variables that presented higher statistic influence in the air temperature models (e.g., NDVI at certain neighbor distances). Finally, other socioeconomic variables that go beyond income and ethnicity could be used to identify groups that can also be disproportionally affected by heat, such as elderly population or immigrants, given that limitations in finding housing can be posed as a bigger challenge to these demographics in the urban environment.

References

Ackerman, Edward A. 1941. “The Köppen Classification of Climates in North America.” *Geographical Review* 31, no. 1 (1941): 105. https://doi.org/10.2307/210420.

Adafruit. 2022. “Adafruit Sensirion SHT40 Temperature & Humidity Sensor.” adafruit industries blog RSS. Accessed November 2, 2022. https://www.adafruit.com/product/4885.

Alcoforado, Maria-João. 2010. Assessing and Modeling the Urban Climate in Lisbon. 10.1002/9781118557600.ch5.

Alizadeh, Mohammad Reza, John T. Abatzoglou, Jan F. Adamowski, Jeffrey P. Prestemon, Bhaskar Chittoori, Ata Akbari Asanjan, and Mojtaba Sadegh. 2022. “Increasing Heat-Stress Inequality in a Warming Climate.” *Earth's Future* 10, no. 2. https://doi.org/10.1029/2021ef002488.

Bailey, Harry P. 1964. “Toward a Unified Concept of the Temperate Climate.” *Geographical Review* 54, no. 4: 516. https://doi.org/10.2307/212980

Balázs, Bernadett, János Unger, Tamás Gál, Zoltán Sümeghy, János Geiger, and Sándor Szegedi. 2009. “Simulation of the Mean Urban Heat Island Using 2D Surface Parameters: Empirical Modelling, Verification and Extension.” *Meteorological Applications* 16, no. 3: 275–87. https://doi.org/10.1002/met.116.

Basu, Rupa. 2009. “High Ambient Temperature and Mortality: A Review of Epidemiologic Studies from 2001 to 2008” *Environmental Health* 8, no. 1. https://doi.org/10.1186/1476-069x-8-40.

Berrar, Daniel. 2019. “Cross-Validation.” https://doi.org/10.1016/b978-0-12-809633-8.20349-x.

Breiman, Leo. 2001. *Machine Learning* 45, no. 1:5–32. https://doi.org/10.1023/a:1010933404324.

Bryant, Edward. 1997. *Climate Process & Change*. New York: Cambridge University Press.

Chen, Deliang, and Hans Weiteng Chen. 2013. "Using the Köppen Classification to Quantify Climate Variation and Change: An Example for 1901–2010." *Environmental Development* 6: 69–79. https://doi.org/10.1016/j.envdev.2013.03.007.

Chen, Jike, Shuangen Jin, and Peijun Du. 2020. "Roles of Horizontal and Vertical Tree Canopy Structure in Mitigating Daytime and Nighttime Urban Heat Island Effects." *International Journal of Applied Earth Observation and Geoinformation* 89: 102060. https://doi.org/10.1016/j.jag.2020.102060.

Chun, Bumseok and Guldmann, Jean-Michel. 2012. Spatial Analysis of the Urban Heat Island Using a 3-D City Model. *Journal of Geographic Information System Association of Korea*. 20. 1-16. 10.12672/ksis.2012.20.4.001.

Chun, Bumseok, and Subhrajit Guhathakurta. 2016. "The Impacts of Three-Dimensional Surface Characteristics on Urban Heat Islands over the Diurnal Cycle." *The Professional Geographer* 69, no. 2: 191–202. https://doi.org/10.1080/00330124.2016.1208102.

Cohen, Li. 2022. "U.S. Cities Are 'Heat Islands' Boiling under Deadly Extreme Temperatures. It's Only Expected to Get Worse." *CBS News*. CBS Interactive, Accessed August 24, 2022. https://www.cbsnews.com/news/heat-islands-cities-extreme-temperatures-climate-change/.

Dahl, Kristina, Rachel Licker, John T Abatzoglou, and Juan Declet-Barreto. 2019. "Increased Frequency of and Population Exposure to Extreme Heat Index Days in the United States during the 21st Century." *Environmental Research Communications* 1, no. 7: 075002. https://doi.org/10.1088/2515-7620/ab27cf.

Degefu, Mekonnen Amberber, Mekuria Argaw, Gudina Legese Feyisa, and Sileshi Degefa. 2022. "Regional and Urban Heat Island Studies in Megacities: A Systematic Analysis of Research Methodology." *Indoor and Built Environment* 31, no. 7: 1775–86. https://doi.org/10.1177/1420326x211061491.

Environmental Protection Agency. 2022. "Heat Island Effect." Accessed May 10, 2022. https://www.epa.gov/heatislands.

ESRI. 2022. "Forest-Based Classification and Regression (Spatial Statistics)." Forest-based Classification and Regression (Spatial Statistics)-ArcGIS Pro | Documentation. Accessed November 1, 2022. https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/forestbasedclassificationregression.htm.

Fischer, E. M., S. Sippel, and R. Knutti. 2021. "Increasing Probability of Record-Shattering Climate Extremes." *Nature Climate Change* 11, no. 8: 689–95. https://doi.org/10.1038/s41558-021-01092-9.

Gallo, K. P., A. L. McNab, T. R. Karl, J. F . Brown, J. J. Hood, and J. D. Tarpley. 1993. "The use of NOAA AVHRR data for assessment of the urban heat island effect". *Journal of Applied Meteorology* 32 (5): 899-908.

Gartland, Lisa. 2011. *Heat Islands: Understanding and Mitigating Heat in Urban Areas*. London: Earthscan.

Gkatsopoulos, Panagiotis. 2017. “A Methodology for Calculating Cooling from Vegetation Evapotranspiration for Use in Urban Space Microclimate Simulations.” *Procedia Environmental Sciences* 38: 477–84. https://doi.org/10.1016/j.proenv.2017.03.139.

Hart, Melissa A., and David J. Sailor. 2008. “Quantifying the Influence of Land-Use and Surface Characteristics on Spatial Variability in the Urban Heat Island.” *Theoretical and Applied Climatology* 95, no. 3-4: 397–406. https://doi.org/10.1007/s00704-008-0017-5.

Hathway, E.A., and S. Sharples. 2012. “The Interaction of Rivers and Urban Form in Mitigating the Urban Heat Island Effect: A UK Case Study.” *Building and Environment* 58: 14–22. https://doi.org/10.1016/j.buildenv.2012.06.013.

Hattis, David, Yelena Ogneva-Himmelberger, and Samuel Ratick. 2012. “The Spatial Variability of Heat-Related Mortality in Massachusetts.” *Applied Geography* 33: 45–52. https://doi.org/10.1016/j.apgeog.2011.07.008.

Heaviside, Clare, Helen Macintyre, and Sotiris Vardoulakis. 2017. “The Urban Heat Island: Implications for Health in a Changing Environment.” *Current Environmental Health Reports* 4, no. 3: 296–305. https://doi.org/10.1007/s40572-017-0150-3.

Hoffman, Jeremy S., Vivek Shandas, and Nicholas Pendleton. 2020. “The Effects of Historical Housing Policies on Resident Exposure to Intra-Urban Heat: A Study of 108 US Urban Areas.” *Climate* 8, no. 1: 12. https://doi.org/10.3390/cli8010012.

Iowa. 2022. “Statewide High-Resolution Elevation (LiDAR) Plan for Iowa.” ArcGIS web application. Accessed September 6, 2022.

https://iowa.maps.arcgis.com/apps/webappviewer/index.html?id=11fa0e43a46845 56b90d6faa19691aaa.

Iowa State University. 2020. "New Analysis Measures Economic Impact of Iowa Industries.". Accessed October 11, 2022. https://www.news.iastate.edu/news/2020/05/05/iowaecon.

Kunkel, K. E., Changnon, S. A., Reinke, B. C., & Arritt, R. W. 1996. The July 1995 Heat Wave in the Midwest: A Climatic Perspective and Critical Weather Factors, *Bulletin of the American Meteorological Society*, 77(7), 1507-1518.

Kusaka, Hiroyuki, and Fujio Kimura. 2004. "Coupling a Single-Layer Urban Canopy Model with a Simple Atmospheric Model: Impact on Urban Heat Island Simulation for an Idealized Case." *Journal of the Meteorological Society of Japan. Ser*. II 82, no. 1: 67–80. https://doi.org/10.2151/jmsj.82.67.

LiDAR for Iowa. 2019. https://iowageodata2.s3.us-east-2.amazonaws.com/_docs/Iowa%20Statewide%20High%20Resolution%20Elevation%20Plan%20%28LiDAR%20for%20Iowa%29.pdf.

Magazine, Smithsonian. 2019. "The Pioneering Maps of Alexander Von Humboldt." Smithsonian.com. Smithsonian Institution, October 15, 2019. https://www.smithsonianmag.com/history/pioneering-maps-alexander-von-humboldt-180973342/.

Mahtta, Richa, Michail Fragkias, Burak Güneralp, Anjali Mahendra, Meredith Reba, Elizabeth A. Wentz, and Karen C. Seto. 2022. "Urban Land Expansion: The Role of Population and Economic Growth for 300+ Cities." *npj Urban Sustainability* 2, no. 1. https://doi.org/10.1038/s42949-022-00048-y.

Manley, Gordon. 1958. "On the Frequency of Snowfall in Metropolitan England." *Quarterly Journal of the Royal Meteorological Society* 84, no. 359: 70–72. https://doi.org/10.1002/qj.49708435910.

Microsoft, 2018. "Building Footprints". Accessed March 22nd, 2023. https://www.microsoft.com/en-us/maps/building-footprints

Mirzaei, Parham A. 2015. "Recent Challenges in Modeling of Urban Heat Island." *Sustainable Cities and Society* 19: 200–206. https://doi.org/10.1016/j.scs.2015.04.001.

Nakata-Osaki, C.M, Souza, L.C.L, Rodrigues, D.S. 2016. 'Uma Ferramenta para Análise do Potencial de Formação de Ilhas de Calor' 7º Congresso Luso Brasileiro para o Planejamento Urbano, Regional, Integrado e Sustentável.

National Aeronautics and Space Administration (NASA) 2000. "Measuring Vegetation (NDVI & EVI). Earth observatory. Accessed July 13th, 2023. https://earthobservatory.nasa.gov/features/MeasuringVegetation

National Centers for Environmental Information (NCEI) - National Oceanic and Atmospheric Administration (NOAA). 2022. Accessed September 24, 2022. https://www.ncei.noaa.gov/cag/statewide/time-series/13/tmin/all/6/2012-2022?base_prd=true&begbaseyear=2012&endbaseyear=2022.

NOAA. 2020. "2019 Was 2nd-Hottest Year on Record for Earth Say NOAA, NASA." National Oceanic and Atmospheric Administration. Accessed October 10, 2022. https://www.noaa.gov/news/2019-was-2nd-hottest-year-on-record-for-earth-say-noaa-nasa.

NOAA. 2022a. NOAA and communities to map heat inequities in 14 U.S. cities and counties. Accessed September 26, 2022. https://www.noaa.gov/news-release/noaa-and-communities-to-map-heat-inequities-in-14-us-cities-and-counties.

NOAA. 2022b. 2021 Iowa NAIP 4-Band 8 Bit Imagery. Accessed June 22, 2023. https://www.fisheries.noaa.gov/inport/item/68084

Oke, T.R. 1982. "The energetic basis of the urban heat island." *Quarterly Journal of the Royal Meteorological Society*, 108, pp. 1-24

Oke, T. R. 1987. Boundary Layer Climates. London England: Routledge.

Oke, T. 2009. Classifying urban climate field sites by "local climate zones": The case of Nagano, Japan.

O'Malley, Christopher, Poorang A.E. Piroozfarb, Eric R.P. Farr, and Jonathan Gates. 2014. "An Investigation into Minimizing Urban Heat Island (UHI) Effects: A UK Perspective." *Energy Procedia* 62: 72–80. https://doi.org/10.1016/j.egypro.2014.12.368.

Oukawa, Gabriel Yoshikazu, Patricia Krecl, and Admir Créso Targino. 2022. "Fine-Scale Modeling of the Urban Heat Island: A Comparison of Multiple Linear Regression and Random Forest Approaches." *Science of The Total Environment* 815: 152836. https://doi.org/10.1016/j.scitotenv.2021.152836.

Paravantis, John & Tzavali, A. & Mihalakakou, Giouli & Fotiadi, Aleksandr & Stigka, Eleni. 2015. Urban heat island intensity: A literature review. 24. 4537-4554.

Rajasekar, Umamaheshwaran, and Qihao Weng. "Urban Heat Island Monitoring and Analysis Using a Non-Parametric Model: A Case Study of Indianapolis." *ISPRS*

Journal of Photogrammetry and Remote Sensing 64, no. 1 (2009): 86–96. https://doi.org/10.1016/j.isprsjprs.2008.05.002.

Renou, E. 1855. 'Instructions météorologiques'. *Annuaire Société Météorologie de France* **3**(1):73-160. Quoted in Gartland, Lisa. (2011) *Heat Islands: Understanding and Mitigating Heat in Urban Areas*. London: Earthscan.

Renou, E. 1862. 'Différences de température entre Paris et Choisy-le-Roi'. *Annuaire Société Météorologie de France* **3**:83-97. Quoted in Gartland, Lisa. (2011) *Heat Islands: Understanding and Mitigating Heat in Urban Areas*. London: Earthscan.

Saaroni, Hadas, Eyal Ben-Dor, Arieh Bitan, and Oded Potchter. 2000. "Spatial Distribution and Microscale Characteristics of the Urban Heat Island in Tel-Aviv, Israel." *Landscape and Urban Planning* 48, no. 1-2 (2000): 1–18. https://doi.org/10.1016/s0169-2046(99)00075-4.

Sabrin, Samain, Maryam Karimi, and Rouzbeh Nazari. 2020. "Developing Vulnerability Index to Quantify Urban Heat Islands Effects Coupled with Air Pollution: A Case Study of Camden, Nj." *ISPRS International Journal of Geo-Information* 9, no. 6: 349. https://doi.org/10.3390/ijgi9060349.

Sakakibara, Yasushi, and Emi Matsui. 2005. "Relation between Heat Island Intensity and City Size Indices/Urban Canopy Characteristics in Settlements of Nagano Basin, Japan." *Geographical Review of Japan* 78, no. 12 : 812–24. https://doi.org/10.4157/grj.78.812.

Santamouris, M., C. Cartalis, A. Synnefa, and D. Kolokotsa. 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: 119–24. https://doi.org/10.1016/j.enbuild.2014.09.052.

Saydelles, A.P. 2005. ‘Estudo do campo térmico das ilhas de calor urbano em Santa Maria-RS’ Dissertação (mestrado) – Universidade Federal de Santa Maria, 2005.

Shandas, Vivek, Jackson Voelkel, Joseph Williams, and Jeremy Hoffman. 2019. “Integrating Satellite and Ground Measurements for Predicting Locations of Extreme Urban Heat.” *Climate* 7, no. 1: 5. https://doi.org/10.3390/cli7010005.

Shandas, Vivek. 2009. Climate Justice: Towards a Proactive Response to Social Inequities. *Connections* 10, no. 11: 4-5.

Sonobe, Rei, Yuki Yamaya, Hiroshi Tani, Xiufeng Wang, Nobuyuki Kobayashi, and Kan-ichiro Mochizuki. 2018. “Evaluating Metrics Derived from Landsat 8 Oli Imagery to Map Crop Cover.” *Geocarto International* 34, no. 8: 839–55. https://doi.org/10.1080/10106049.2018.1425739.

Stewart, Iain Douglas. 2000. “Influence of Meteorological Conditions on the Intensity and Form of the Urban Heat Island Effect in Regina.” *The Canadian Geographer/Le Géographe canadien* 44, no. 3: 271–85. https://doi.org/10.1111/j.1541-0064.2000.tb00709.x.

Stewart, I.D. 2011. ‘Redefining the Urban Heat Island’. The University of British Columbia.

Stone, Brian. 2007. “Urban and Rural Temperature Trends in Proximity to Large US Cities: 1951–2000.” *International Journal of Climatology* 27, no. 13 (2007): 1801–7. https://doi.org/10.1002/joc.1555.

Stone, Brian, and John M. Norman. 2006. "Land Use Planning and Surface Heat Island Formation: A Parcel-Based Radiation Flux Approach." *Atmospheric Environment* 40, no. 19: 3561–3573. https://doi.org/10.1016/j.atmosenv.2006.01.015.

Szymanowski, Mariusz, and Maciej Kryza. 2011. "Local Regression Models for Spatial Interpolation of Urban Heat Island—an Example from Wrocław, SW Poland." *Theoretical and Applied Climatology* 108, no. 1-2: 53–71. https://doi.org/10.1007/s00704-011-0517-6.

Unger, J. 2006. "Modelling of the Annual Mean Maximum Urban Heat Island Using 2D and 3D Surface Parameters." *Climate Research* 30: 215–26. https://doi.org/10.3354/cr030215.

United Nations. 2019. Department of Economic and Social Affairs, Population Division. World Urbanization Prospects: The 2018 Revision (ST/ESA/SER.A/420). New York: United Nations.

U.S. Bureau of Economic Analysis. 2022. "GDP by State." Accessed October 11, 2022. https://www.bea.gov/data/gdp/gdp-state.

US Census Bureau. 2021. "U.S. Census Bureau Urban Areas QuickFacts." Census.gov. Accessed October 10, 2022. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural/ua-quickfacts.html

U.S. Census Bureau Data (Census). 2022. Accessed April 20, 2023. https://www.census.gov/data.html

U.S. Geological Survey (USGS). 2021. ScienceBase-Catalog. Accessed June 4, 2021. https://www.sciencebase.gov/catalog/item/5f21cef582cef313ed940043.

U.S. Geological Survey (USGS). 2022. - Highest and Lowest Elevations. Accessed September 24, 2022. https://www.usgs.gov/educational-resources/highest-and-lowest-elevations.

Vicedo-Cabrera, A. M., Scovronick, N., Sera, F., Royé, D., Schneider, R., Tobias, A., Astrom, C., Guo, Y., Honda, Y., Hondula, D. M., Abrutzky, R., Tong, S., de Sousa Zanotti Stagliorio Coelho, M., Saldiva, P., Lavigne, E., Correa, P. M., Ortega, N. V., Kan, H., Osorio, S., Kyselý, J., … Gasparrini, A. 2021. The burden of heat-related mortality attributable to recent human-induced climate change. *Nature climate change*, *11*(6), 492–500. https://doi.org/10.1038/s41558-021-01058-x

Voelkel, Jackson, and Vivek Shandas. 2017. "Towards Systematic Prediction of Urban Heat Islands: Grounding Measurements, Assessing Modeling Techniques." *Climate* 5, no. 2 :41. https://doi.org/10.3390/cli5020041.

Voelkel, Jackson, Dana Hellman, Ryu Sakuma, and Vivek Shandas. 2018. "Assessing Vulnerability to Urban Heat: A Study of Disproportionate Heat Exposure and Access to Refuge by Socio-Demographic Status in Portland, Oregon." *International Journal of Environmental Research and Public Health* 15, no. 4: 640. https://doi.org/10.3390/ijerph15040640.

Wong, Kaufui V., Andrew Paddon, and Alfredo Jimenez. 2013. "Review of World Urban Heat Islands: Many Linked to Increased Mortality." *Journal of Energy Resources Technology* 135, no. 2. https://doi.org/10.1115/1.4023176.

Wong, Nyuk Hien, and Chen Yu. 2005. "Study of Green Areas and Urban Heat Island in a Tropical City." *Habitat International* 29, no. 3 (2005): 547–58. https://doi.org/10.1016/j.habitatint.2004.04.008.

Yokobori, T, and S Ohta. 2009. “Effect of Land Cover on Air Temperatures Involved in the Development of an Intra-Urban Heat Island.” *Climate Research* 39: 61–73. https://doi.org/10.3354/cr00800.

Yu, Qiuyan, Wenjie Ji, Ruiliang Pu, Shawn Landry, Michael Acheampong, Jarlath O’ Neil-Dunne, Zhibin Ren, and Shakhawat Hosen Tanim. 2020. “A Preliminary Exploration of the Cooling Effect of Tree Shade in Urban Landscapes.” *International Journal of Applied Earth Observation and Geoinformation* 92: 102161. https://doi.org/10.1016/j.jag.2020.102161.

Zhou, Decheng, Jingfeng Xiao, Stefania Bonafoni, Christian Berger, Kaveh Deilami, Yuyu Zhou, Steve Frolking, Rui Yao, Zhi Qiao, and José Sobrino. 2018. “Satellite Remote Sensing of Surface Urban Heat Islands: Progress, Challenges, and Perspectives.” *Remote Sensing* 11, no. 1: 48. https://doi.org/10.3390/rs11010048.rs11010048.

Appendix A

Income Linear Regression (TIBCO Spotfire S+ v. 8.2.0)

```
*** Linear Model ***

Call: lm(formula = BAFI ~ BAF, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median   3Q   Max
 -32206 -10688  -2926 6767 55877

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  732829.2204  300194.8746       2.4412       0.0200
        BAF  -20370.1718    9000.9955      -2.2631       0.0301

Residual standard error: 18240 on 34 degrees of freedom
Multiple R-Squared: 0.1309      Adjusted R-squared: 0.1054
F-statistic: 5.122 on 1 and 34 degrees of freedom, the p-value is 0.03013
172 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = BEI ~ BE, data = Income, na.action = na.exclude)
Residuals:
    Min    1Q Median   3Q   Max
 -33433 -8370  -2352 5436 52302

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  757361.6528  262772.8057       2.8822       0.0068
         BE  -23665.4874    8834.3839      -2.6788       0.0113

Residual standard error: 17780 on 34 degrees of freedom
Multiple R-Squared: 0.1743      Adjusted R-squared: 0.15
F-statistic: 7.176 on 1 and 34 degrees of freedom, the p-value is 0.0113
172 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = BNI ~ BN, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median   3Q   Max
 -34897 -10942  -3213 7154 55396

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  637497.6859  296682.5132       2.1488       0.0389
         BN  -21795.5402   11071.8262      -1.9686       0.0572

Residual standard error: 18530 on 34 degrees of freedom
Multiple R-Squared: 0.1023      Adjusted R-squared: 0.07591
F-statistic: 3.875 on 1 and 34 degrees of freedom, the p-value is 0.0572
172 observations deleted due to missing values


       *** Linear Model ***
```

```
Call: lm(formula = CBAI ~ CBA, data = Income, na.action = na.exclude)
Residuals:
    Min    1Q Median    3Q   Max
 -39479 -9158  390.9 11580 39563

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  392285.9124  116433.0931       3.3692       0.0013
        CBA  -11342.1140    3958.3842      -2.8653       0.0057

Residual standard error: 17430 on 60 degrees of freedom
Multiple R-Squared: 0.1204      Adjusted R-squared: 0.1057
F-statistic: 8.21 on 1 and 60 degrees of freedom, the p-value is 0.005735
146 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = CBEI ~ CBE, data = Income, na.action = na.exclude)
Residuals:
    Min    1Q Median    3Q   Max
 -38726 -9914  398.2 11512 36052

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  401321.1964   95135.0396       4.2184       0.0001
        CBE  -13930.2334    3867.2972      -3.6021       0.0006

Residual standard error: 16860 on 60 degrees of freedom
Multiple R-Squared: 0.1778      Adjusted R-squared: 0.1641
F-statistic: 12.97 on 1 and 60 degrees of freedom, the p-value is 0.0006419
146 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = CBNI ~ CBN, data = Income, na.action = na.exclude)
Residuals:
    Min    1Q Median    3Q   Max
 -39075 -7846 -139.9 11053 35450

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  342474.6541   70918.7652       4.8291       0.0000
        CBN  -14202.7910    3548.2231      -4.0028       0.0002

Residual standard error: 16510 on 60 degrees of freedom
Multiple R-Squared: 0.2108      Adjusted R-squared: 0.1976
F-statistic: 16.02 on 1 and 60 degrees of freedom, the p-value is 0.0001747
146 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = CRAI ~ CRA, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q    Max
 -54253 -18603  -3272 13192 111841

Coefficients:
                    Value    Std. Error       t value      Pr(>|t|)
(Intercept)  1738678.3763   449142.2536        3.8711        0.0002
```

```
        CRA   -49170.3314    13205.9493       -3.7233        0.0003

Residual standard error: 28760 on 91 degrees of freedom
Multiple R-Squared: 0.1322      Adjusted R-squared: 0.1227
F-statistic: 13.86 on 1 and 91 degrees of freedom, the p-value is 0.0003404
115 observations deleted due to missing values


      *** Linear Model ***

Call: lm(formula = CREI ~ CRE, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q    Max
 -54266 -19193  -4982 10158 115692

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  885837.2080  308819.3505       2.8685       0.0051
        CRE  -29311.3344   11045.9795      -2.6536       0.0094

Residual standard error: 29740 on 91 degrees of freedom
Multiple R-Squared: 0.07182      Adjusted R-squared: 0.06162
F-statistic: 7.041 on 1 and 91 degrees of freedom, the p-value is 0.009399
115 observations deleted due to missing values


      *** Linear Model ***

Call: lm(formula = CRNI ~ CRN, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q    Max
 -48159 -17493  -4834 10027 122208

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  747232.1978  191833.9690       3.8952       0.0002
        CRN  -28453.9316    8016.3444      -3.5495       0.0006

Residual standard error: 28930 on 91 degrees of freedom
Multiple R-Squared: 0.1216      Adjusted R-squared: 0.112
F-statistic: 12.6 on 1 and 91 degrees of freedom, the p-value is 0.0006131
115 observations deleted due to missing values


      *** Linear Model ***

Call: lm(formula = DMAI ~ DMA, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q    Max
 -46779 -16855  -2000 11828 151198

Coefficients:
                    Value    Std. Error       t value      Pr(>|t|)
(Intercept)  1696067.7428   226949.2037        7.4733        0.0000
        DMA   -43865.4744     6097.1514       -7.1944        0.0000

Residual standard error: 23480 on 206 degrees of freedom
Multiple R-Squared: 0.2008      Adjusted R-squared: 0.1969
F-statistic: 51.76 on 1 and 206 degrees of freedom, the p-value is 1.144e-011


      *** Linear Model ***
```

```
Call: lm(formula = DMEI ~ DME, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q    Max
 -49501 -14011  -1960 10272 170684

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  785711.9086  191757.4984       4.0974       0.0001
        DME  -22964.4368    6095.7895      -3.7673       0.0002

Residual standard error: 25410 on 206 degrees of freedom
Multiple R-Squared: 0.06445      Adjusted R-squared: 0.05991
F-statistic: 14.19 on 1 and 206 degrees of freedom, the p-value is 0.0002154


       *** Linear Model ***

Call: lm(formula = DMNI ~ DMN, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q    Max
 -50994 -15191  -1938 10583 168339

Coefficients:
                    Value    Std. Error       t value      Pr(>|t|)
(Intercept)  1148711.8075   216784.7546        5.2989        0.0000
        DMN   -38153.4402     7620.2793       -5.0068        0.0000

Residual standard error: 24800 on 206 degrees of freedom
Multiple R-Squared: 0.1085      Adjusted R-squared: 0.1042
F-statistic: 25.07 on 1 and 206 degrees of freedom, the p-value is 1.186e-006


       *** Linear Model ***

Call: lm(formula = FDAI ~ FDA, data = Income, na.action = na.exclude)
Residuals:
    Min    1Q Median   3Q   Max
 -35913 -7388  -3266 7720 44489

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  480739.0598  332485.8688       1.4459       0.1617
        FDA  -13291.2975   10352.1844      -1.2839       0.2120

Residual standard error: 19460 on 23 degrees of freedom
Multiple R-Squared: 0.06688      Adjusted R-squared: 0.02631
F-statistic: 1.648 on 1 and 23 degrees of freedom, the p-value is 0.212
183 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = FDEI ~ FDE, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q   Max
 -30693 -14158  -2684 10291 44614

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  253149.2548  109571.7566       2.3104       0.0302
        FDE   -8008.6553    4401.2218      -1.8196       0.0819

Residual standard error: 18840 on 23 degrees of freedom
```

```
Multiple R-Squared: 0.1258      Adjusted R-squared: 0.08784
F-statistic: 3.311 on 1 and 23 degrees of freedom, the p-value is 0.08186
183 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = MAI ~ MA, data = Income, na.action = na.exclude)
Residuals:
    Min    1Q Median   3Q   Max
 -29871 -8241  -2888 5970 42401

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  857839.7570  214747.4286       3.9946       0.0005
         MA  -22984.1114    6211.9962      -3.7000       0.0011

Residual standard error: 16910 on 25 degrees of freedom
Multiple R-Squared: 0.3538      Adjusted R-squared: 0.328
F-statistic: 13.69 on 1 and 25 degrees of freedom, the p-value is 0.001066
181 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = MNI ~ MN, data = Income, na.action = na.exclude)
Residuals:
    Min    1Q Median   3Q   Max
 -27555 -8813  -4332 6650 44546

Coefficients:
                    Value    Std. Error       t value      Pr(>|t|)
(Intercept)  1655218.1722   448038.9163        3.6944        0.0011
         MN   -72702.2765    20462.1697       -3.5530        0.0015

Residual standard error: 17140 on 25 degrees of freedom
Multiple R-Squared: 0.3355      Adjusted R-squared: 0.3089
F-statistic: 12.62 on 1 and 25 degrees of freedom, the p-value is 0.001545
181 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = SCAI ~ SCA, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q    Max
 -51101 -21126  -5191 16454 105283

Coefficients:
                    Value    Std. Error       t value      Pr(>|t|)
(Intercept)  1852808.6991   638681.3630        2.9010        0.0049
        SCA   -53220.8785    19011.7747       -2.7994        0.0065

Residual standard error: 28610 on 74 degrees of freedom
Multiple R-Squared: 0.09576      Adjusted R-squared: 0.08354
F-statistic: 7.836 on 1 and 74 degrees of freedom, the p-value is 0.006526
132 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = SCEI ~ SCE, data = Income, na.action = na.exclude)
Residuals:
```

```
    Min     1Q Median    3Q    Max
 -44330 -19038  -5135 14137 108611

Coefficients:
                    Value    Std. Error       t value      Pr(>|t|)
(Intercept)  1106766.4223   350608.5724        3.1567        0.0023
        SCE   -36332.0789    12226.2907       -2.9716        0.0040

Residual standard error: 28440 on 74 degrees of freedom
Multiple R-Squared: 0.1066      Adjusted R-squared: 0.09454
F-statistic: 8.831 on 1 and 74 degrees of freedom, the p-value is 0.003994
132 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = SCNI ~ SCN, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q    Max
 -48665 -19847  -5491 15037 110547

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  370453.3222  143518.8222       2.5812       0.0118
        SCN  -12785.3432    6004.2596      -2.1294       0.0365

Residual standard error: 29210 on 74 degrees of freedom
Multiple R-Squared: 0.05774      Adjusted R-squared: 0.045
F-statistic: 4.534 on 1 and 74 degrees of freedom, the p-value is 0.03655
132 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = WCAI ~ WCA, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q   Max
 -40138 -15405  -2686 11294 67760

Coefficients:
                    Value    Std. Error       t value      Pr(>|t|)
(Intercept)  1519608.0906   330529.8137        4.5975        0.0000
        WCA   -42149.6499     9537.4292       -4.4194        0.0000

Residual standard error: 24080 on 67 degrees of freedom
Multiple R-Squared: 0.2257      Adjusted R-squared: 0.2142
F-statistic: 19.53 on 1 and 67 degrees of freedom, the p-value is 0.00003709
139 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = WCEI ~ WCE, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q   Max
 -43753 -12716  -3475 11729 68842

Coefficients:
                   Value   Std. Error      t value     Pr(>|t|)
(Intercept)  799880.2160  136732.9853       5.8499       0.0000
        WCE  -28218.7518    5206.3141      -5.4201       0.0000

Residual standard error: 22810 on 67 degrees of freedom
```

```
Multiple R-Squared: 0.3048      Adjusted R-squared: 0.2944
F-statistic: 29.38 on 1 and 67 degrees of freedom, the p-value is 8.754e-007
139 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = WCNI ~ WCN, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q   Max
 -39101 -16907  -1762 13601 71944

Coefficients:
                    Value    Std. Error       t value      Pr(>|t|)
(Intercept)  3513161.0345   661499.1395        5.3109        0.0000
        WCN  -152507.3410    29205.4532       -5.2219        0.0000

Residual standard error: 23070 on 67 degrees of freedom
Multiple R-Squared: 0.2893      Adjusted R-squared: 0.2787
F-statistic: 27.27 on 1 and 67 degrees of freedom, the p-value is 1.882e-006
139 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = WAI ~ WA, data = Income, na.action = na.exclude)
Residuals:
    Min    1Q Median    3Q   Max
 -47248 -4543   4350 10605 16715

Coefficients:
                    Value    Std. Error       t value      Pr(>|t|)
(Intercept)  1329673.7907  1106529.1673        1.2017        0.2639
         WA   -35284.2350    31021.7907       -1.1374        0.2883

Residual standard error: 19660 on 8 degrees of freedom
Multiple R-Squared: 0.1392      Adjusted R-squared: 0.0316
F-statistic: 1.294 on 1 and 8 degrees of freedom, the p-value is 0.2883
198 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = WEI ~ WE, data = Income, na.action = na.exclude)
Residuals:
    Min     1Q Median    3Q   Max
 -42472 -10683   7921 14019 18489

Coefficients:
                    Value    Std. Error       t value      Pr(>|t|)
(Intercept)   119788.2341  1377595.9761        0.0870        0.9328
         WE    -1635.0884    46287.3528       -0.0353        0.9727

Residual standard error: 21190 on 8 degrees of freedom
Multiple R-Squared: 0.000156      Adjusted R-squared: -0.1248
F-statistic: 0.001248 on 1 and 8 degrees of freedom, the p-value is 0.9727
198 observations deleted due to missing values
```

Ethnicity Linear Regression (TIBCO Spotfire S+ v. 8.2.0)

```
*** Linear Model ***

Call: lm(formula = BAFR ~ BAF, data = Race, na.action = na.exclude)
Residuals:
     Min       1Q  Median      3Q   Max
 -0.1604 -0.08714 -0.0328 0.04062 0.606

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -2.3880  2.1983    -1.0863  0.2844
        BAF  0.0745  0.0659     1.1308  0.2654

Residual standard error: 0.1343 on 37 degrees of freedom
Multiple R-Squared: 0.0334      Adjusted R-squared: 0.007278
F-statistic: 1.279 on 1 and 37 degrees of freedom, the p-value is 0.2654
177 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = BER ~ BE, data = Race, na.action = na.exclude)
Residuals:
     Min      1Q   Median      3Q    Max
 -0.1556 -0.0845 -0.02255 0.04294 0.6024

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -2.1095  1.9590    -1.0769  0.2885
         BE  0.0742  0.0658     1.1268  0.2671

Residual standard error: 0.1343 on 37 degrees of freedom
Multiple R-Squared: 0.03317      Adjusted R-squared: 0.007044
F-statistic: 1.27 on 1 and 37 degrees of freedom, the p-value is 0.2671
177 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = BNR ~ BN, data = Race, na.action = na.exclude)
Residuals:
     Min       1Q   Median      3Q    Max
 -0.1361 -0.08592 -0.03874 0.03269 0.6156

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -1.4049  2.1428    -0.6556  0.5161
         BN  0.0561  0.0800     0.7012  0.4875

Residual standard error: 0.1357 on 37 degrees of freedom
Multiple R-Squared: 0.01312      Adjusted R-squared: -0.01356
F-statistic: 0.4917 on 1 and 37 degrees of freedom, the p-value is 0.4875
177 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = CBAR ~ CBA, data = Race, na.action = na.exclude)
Residuals:
```

```
      Min       1Q   Median      3Q    Max
 -0.07045 -0.04977 -0.02851 0.02689 0.3186

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -0.4248  0.5022    -0.8459  0.4007
        CBA  0.0165  0.0171     0.9672  0.3370

Residual standard error: 0.07613 on 65 degrees of freedom
Multiple R-Squared: 0.01419      Adjusted R-squared: -0.0009783
F-statistic: 0.9355 on 1 and 65 degrees of freedom, the p-value is 0.337
149 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = CBER ~ CBE, data = Race, na.action = na.exclude)
Residuals:
      Min       1Q   Median      3Q    Max
 -0.06691 -0.04965 -0.02822 0.03085 0.3222

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -0.3366  0.4236    -0.7947  0.4297
        CBE  0.0162  0.0172     0.9385  0.3514

Residual standard error: 0.07616 on 65 degrees of freedom
Multiple R-Squared: 0.01337      Adjusted R-squared: -0.001809
F-statistic: 0.8808 on 1 and 65 degrees of freedom, the p-value is 0.3514
149 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = CBNR ~ CBN, data = Race, na.action = na.exclude)
Residuals:
      Min       1Q   Median      3Q    Max
 -0.07015 -0.04688 -0.02737 0.03183 0.3214

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -0.2769  0.3192    -0.8675  0.3889
        CBN  0.0169  0.0160     1.0586  0.2937

Residual standard error: 0.07602 on 65 degrees of freedom
Multiple R-Squared: 0.01695      Adjusted R-squared: 0.001823
F-statistic: 1.121 on 1 and 65 degrees of freedom, the p-value is 0.2937
149 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = CRAR ~ CRA, data = Race, na.action = na.exclude)
Residuals:
     Min       1Q   Median      3Q    Max
 -0.1726 -0.09915 -0.01927 0.06024 0.4415

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -5.7367  1.8780    -3.0546  0.0029
        CRA  0.1725  0.0552     3.1239  0.0024

Residual standard error: 0.1234 on 95 degrees of freedom
```

```
Multiple R-Squared: 0.09316      Adjusted R-squared: 0.08361
F-statistic: 9.759 on 1 and 95 degrees of freedom, the p-value is 0.002366
119 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = CRER ~ CRE, data = Race, na.action = na.exclude)
Residuals:
     Min      1Q   Median      3Q    Max
 -0.1547 -0.0872 -0.03477 0.05755 0.4919

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -3.4065  1.2806    -2.6601  0.0092
        CRE  0.1265  0.0458     2.7617  0.0069

Residual standard error: 0.1247 on 95 degrees of freedom
Multiple R-Squared: 0.07432      Adjusted R-squared: 0.06457
F-statistic: 7.627 on 1 and 95 degrees of freedom, the p-value is 0.006903
119 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = CRNR ~ CRN, data = Race, na.action = na.exclude)
Residuals:
     Min       1Q   Median      3Q    Max
 -0.1771 -0.08705 -0.04074 0.06407 0.4608

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -2.2048  0.8015    -2.7508  0.0071
        CRN  0.0976  0.0335     2.9133  0.0045

Residual standard error: 0.1242 on 95 degrees of freedom
Multiple R-Squared: 0.08201      Adjusted R-squared: 0.07235
F-statistic: 8.487 on 1 and 95 degrees of freedom, the p-value is 0.004458
119 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = DMAR ~ DMA, data = Race, na.action = na.exclude)
Residuals:
     Min      1Q   Median      3Q    Max
 -0.2551 -0.1269 -0.03072 0.07399 0.6058

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -7.0340  1.6619    -4.2325  0.0000
        DMA  0.1945  0.0446     4.3568  0.0000

Residual standard error: 0.1758 on 213 degrees of freedom
Multiple R-Squared: 0.08183      Adjusted R-squared: 0.07751
F-statistic: 18.98 on 1 and 213 degrees of freedom, the p-value is 0.00002051
1 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = DMER ~ DME, data = Race, na.action = na.exclude)
Residuals:
```

```
     Min      1Q   Median      3Q    Max
 -0.2439 -0.1339 -0.03834 0.08423 0.5904

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -4.6948  1.3296    -3.5311  0.0005
        DME  0.1558  0.0423     3.6865  0.0003

Residual standard error: 0.1778 on 213 degrees of freedom
Multiple R-Squared: 0.05998      Adjusted R-squared: 0.05557
F-statistic: 13.59 on 1 and 213 degrees of freedom, the p-value is 0.0002885
1 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = DMNR ~ DMN, data = Race, na.action = na.exclude)
Residuals:
     Min      1Q  Median      3Q    Max
 -0.2576 -0.1283 -0.0319 0.07023 0.6083

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -5.7538  1.5194    -3.7868  0.0002
        DMN  0.2095  0.0534     3.9228  0.0001

Residual standard error: 0.1771 on 213 degrees of freedom
Multiple R-Squared: 0.06738      Adjusted R-squared: 0.063
F-statistic: 15.39 on 1 and 213 degrees of freedom, the p-value is 0.0001181
1 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = FDAR ~ FDA, data = Race, na.action = na.exclude)
Residuals:
     Min       1Q   Median     3Q    Max
 -0.1208 -0.08288 -0.01713 0.0362 0.2709

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -1.5684  1.7321    -0.9055  0.3735
        FDA  0.0520  0.0539     0.9642  0.3439

Residual standard error: 0.1046 on 26 degrees of freedom
Multiple R-Squared: 0.03452      Adjusted R-squared: -0.002614
F-statistic: 0.9296 on 1 and 26 degrees of freedom, the p-value is 0.3439
188 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = FDER ~ FDE, data = Race, na.action = na.exclude)
Residuals:
     Min       1Q   Median      3Q    Max
 -0.1204 -0.07922 -0.02096 0.04461 0.2322

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -0.5548  0.5535    -1.0024  0.3254
        FDE  0.0263  0.0222     1.1867  0.2461

Residual standard error: 0.1036 on 26 degrees of freedom
```

```
Multiple R-Squared: 0.05138      Adjusted R-squared: 0.01489
F-statistic: 1.408 on 1 and 26 degrees of freedom, the p-value is 0.2461
188 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = MAR ~ MA, data = Race, na.action = na.exclude)
Residuals:
     Min     1Q    Median      3Q    Max
 -0.1637 -0.115 -0.009255 0.07814 0.2415

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -0.5306  1.5321    -0.3463  0.7317
         MA  0.0197  0.0443     0.4441  0.6604

Residual standard error: 0.128 on 28 degrees of freedom
Multiple R-Squared: 0.006993      Adjusted R-squared: -0.02847
F-statistic: 0.1972 on 1 and 28 degrees of freedom, the p-value is 0.6604
186 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = MNR ~ MN, data = Race, na.action = na.exclude)
Residuals:
     Min      1Q   Median      3Q    Max
 -0.1574 -0.1151 -0.01072 0.06717 0.2435

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -0.5057  3.1483    -0.1606  0.8735
         MN  0.0299  0.1437     0.2082  0.8366

Residual standard error: 0.1283 on 28 degrees of freedom
Multiple R-Squared: 0.001545      Adjusted R-squared: -0.03411
F-statistic: 0.04334 on 1 and 28 degrees of freedom, the p-value is 0.8366
186 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = SCAR ~ SCA, data = Race, na.action = na.exclude)
Residuals:
     Min      1Q   Median     3Q    Max
 -0.2058 -0.1238 -0.03391 0.1067 0.3967

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -7.3079  3.2915    -2.2202  0.0293
        SCA  0.2223  0.0980     2.2692  0.0260

Residual standard error: 0.1532 on 79 degrees of freedom
Multiple R-Squared: 0.06119      Adjusted R-squared: 0.04931
F-statistic: 5.149 on 1 and 79 degrees of freedom, the p-value is 0.02598
135 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = SCER ~ SCE, data = Race, na.action = na.exclude)
Residuals:
```

```
     Min      1Q   Median      3Q    Max
 -0.2805 -0.0986 -0.01958 0.09189 0.4068

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -7.4429  1.7355    -4.2886  0.0001
        SCE  0.2652  0.0605     4.3818  0.0000

Residual standard error: 0.1419 on 79 degrees of freedom
Multiple R-Squared: 0.1955      Adjusted R-squared: 0.1853
F-statistic: 19.2 on 1 and 79 degrees of freedom, the p-value is 0.00003583
135 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = SCNR ~ SCN, data = Race, na.action = na.exclude)
Residuals:
     Min       1Q Median      3Q    Max
 -0.3195 -0.09225 -0.027 0.08503 0.4096

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -2.7894  0.6989    -3.9910  0.0001
        SCN  0.1235  0.0292     4.2229  0.0001

Residual standard error: 0.1429 on 79 degrees of freedom
Multiple R-Squared: 0.1842      Adjusted R-squared: 0.1738
F-statistic: 17.83 on 1 and 79 degrees of freedom, the p-value is 0.00006403
135 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = WCAR ~ WCA, data = Race, na.action = na.exclude)
Residuals:
     Min      1Q   Median      3Q    Max
 -0.2767 -0.1132 -0.07278 0.06988 0.7363

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -6.3233  2.3183    -2.7276  0.0080
        WCA  0.1872  0.0669     2.7996  0.0066

Residual standard error: 0.181 on 71 degrees of freedom
Multiple R-Squared: 0.09942      Adjusted R-squared: 0.08673
F-statistic: 7.838 on 1 and 71 degrees of freedom, the p-value is 0.006585
143 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = WCER ~ WCE, data = Race, na.action = na.exclude)
Residuals:
     Min      1Q  Median     3Q    Max
 -0.2554 -0.1104 -0.0557 0.0775 0.7193

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -3.6513  1.0138    -3.6015  0.0006
        WCE  0.1454  0.0386     3.7666  0.0003

Residual standard error: 0.1741 on 71 degrees of freedom
```

```
Multiple R-Squared: 0.1665      Adjusted R-squared: 0.1548
F-statistic: 14.19 on 1 and 71 degrees of freedom, the p-value is 0.0003384
143 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = WCNR ~ WCN, data = Race, na.action = na.exclude)
Residuals:
     Min      1Q   Median      3Q    Max
 -0.2257 -0.1178 -0.05425 0.07786 0.7047

Coefficients:
               Value Std. Error  t value Pr(>|t|)
(Intercept) -18.5844   4.9262    -3.7726   0.0003
        WCN   0.8279   0.2175     3.8064   0.0003

Residual standard error: 0.1738 on 71 degrees of freedom
Multiple R-Squared: 0.1695      Adjusted R-squared: 0.1578
F-statistic: 14.49 on 1 and 71 degrees of freedom, the p-value is 0.0002962
143 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = WAR ~ WA, data = Race, na.action = na.exclude)
Residuals:
      Min       1Q   Median      3Q     Max
 -0.04432 -0.03313 -0.01003 0.02691 0.06596

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept) -1.5384  2.4344    -0.6319  0.5451
         WA  0.0442  0.0682     0.6476  0.5354

Residual standard error: 0.04326 on 8 degrees of freedom
Multiple R-Squared: 0.04982      Adjusted R-squared: -0.06895
F-statistic: 0.4194 on 1 and 8 degrees of freedom, the p-value is 0.5354
206 observations deleted due to missing values


       *** Linear Model ***

Call: lm(formula = WER ~ WE, data = Race, na.action = na.exclude)
Residuals:
      Min       1Q   Median      3Q     Max
 -0.03828 -0.03495 -0.01124 0.02154 0.07477

Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept)  0.1018  2.8849     0.0353  0.9727
         WE -0.0021  0.0969    -0.0221  0.9829

Residual standard error: 0.04438 on 8 degrees of freedom
Multiple R-Squared: 0.00006078      Adjusted R-squared: -0.1249
F-statistic: 0.0004862 on 1 and 8 degrees of freedom, the p-value is 0.9829
206 observations deleted due to missing values
```

Python Code to Predict Air Temperature Raster Surface

```
'''The code below was created by Clemir Abbeg Coproski on May 2023 and was intended to use to generate a geospatial model of temperature
patterns in urban areas. Temperature data was collected using mobile sensors over Summer 2022 by the author and imagery, LiDAR
point cloud data, and building footprints were acquired from USGS, State of Iowa, Iowa Geodata, counties, cities, and public repositories.
The pre-requirements to run all the code below are: LAS dataset must be created and it must be in the same directory
of the geodatabase/environment; NDVI raster file must be in the geodatabase; temperature vector file must be in the geodatabase, building
footprints must be in the geodatabase, and buffer distance file must be in the geodatabase. Please change the name of the directories
and the study area accordingly.
This project was funded by the Iowa Economic Development Authority, Iowa Energy Center Grant Program (21-IEC-012).
All Rights Reserved.'''

import arcpy
from arcpy import env
from arcpy.ia import *
from arcpy.sa import *

env.workspace = r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb"

arcpy.ddd.ClassifyLasOverlap("StudyArea_LAS.lasd", "1 Meters", "DEFAULT", "PROCESS_EXTENT", "COMPUTE_STATS", "UPDATE_PYRAMID")

arcpy.ddd.ClassifyLasGround("StudyArea_LAS.lasd", "STANDARD", "RECLASSIFY_GROUND",
None, "COMPUTE_STATS", "DEFAULT", None, "PROCESS_EXTENT",
"UPDATE_PYRAMID")

arcpy.management.MakeLasDatasetLayer('StudyArea_LAS.lasd', 'StudyArea_Ground', '2',
'LAST; FIRST_OF_MANY; LAST_OF_MANY; SINGLE; 1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14; 15',
'INCLUDE_UNFLAGGED', 'INCLUDE_SYNTHETIC', 'INCLUDE_KEYPOINT', 'EXCLUDE_WITHHELD', 'EXCLUDE_OVERLAP')
arcpy.conversion.LasDatasetToRaster("StudyArea_LAS.lasd", r"c:\users\abbegcc\documents\arcgis\projects\StudyArea\StudyArea.gdb\StudyArea_DEM",
"ELEVATION", "TRIANGULATION NATURAL_NEIGHBOR NO_THINNING","FLOAT", "CELLSIZE", 1, 1)
StudyArea_DEMOk = ExtractByMask("StudyArea_DEM", r"D:\IowaEnergyCenter\spatial\Additionaldata.gdb\poly_StudyArea_buffer.shp", "INSIDE")
StudyArea_DEMOk.save("StudyArea_DEMOk")

arcpy.ddd.ClassifyLasNoise("StudyArea_LAS.lasd", "RELATIVE_HEIGHT", "CLASSIFY", "NO_WITHHELD",
"COMPUTE_STATS", "StudyArea_DEM", "-2 Meters", None, 10, "8
```

```
Meters", "8 Meters",
"DEFAULT", "PROCESS_EXTENT", None, "UPDATE_PYRAMID")
arcpy.analysis.Clip(r"D:\IowaEnergyCenter\spatial\BUILDING
FOOTPRINTS\StudyArea\StudyArea.shp",
r"D:\IowaEnergyCenter\spatial\Additionaldata.gdb\poly_StudyArea_buffer",
"StudyArea_BF_Clip")
arcpy.ddd.ClassifyLasBuilding("StudyArea_LAS.lasd", "2 Meters", "6 SquareMeters",
"COMPUTE_STATS", "DEFAULT",
None, "PROCESS_EXTENT", None, "RECLASSIFY_BUILDING",
"NOT_PHOTOGRAMMETRIC_DATA",
"AGGRESSIVE", "NO_CLASSIFY_ABOVE_ROOF", "1.5 Meters", 6,
"CLASSIFY_BELOW_ROOF", 99,
"UPDATE_PYRAMID")
arcpy.ddd.LasBuildingMultipatch("StudyArea_LAS.lasd", "StudyArea_BF_Clip",
"StudyArea_DEMOk",
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\St
udyArea.gdb\StudyArea_3DBuildings",
"BUILDING_CLASSIFIED_POINTS", "0.5 Meters", None)
arcpy.ddd.AddZInformation("StudyArea_3DBuildings",
"Z_MIN;Z_MAX;Z_MEAN;SURFACE_AREA;VOLUME;MIN_SLOPE;MAX_SLOPE;AVG_SLOPE", "0.001")
arcpy.ddd.ClassifyLasByHeight("StudyArea_LAS.lasd", "GROUND", "3 5;4 25;5 50",
"NONE", "COMPUTE_STATS", "DEFAULT", "PROCESS_EXTENT", None,
"UPDATE_PYRAMID")
arcpy.management.MakeLasDatasetLayer('StudyArea_LAS.lasd', 'StudyArea_Surface', "2; 3;
4; 5; 6; 9; 10; 11; 13; 14; 15; 16; 17",
'FIRST_OF_MANY; 1',
'INCLUDE_UNFLAGGED', 'INCLUDE_SYNTHETIC',
'INCLUDE_KEYPOINT', 'EXCLUDE_WITHHELD',
'EXCLUDE_OVERLAP')
arcpy.conversion.LasDatasetToRaster("StudyArea_Surface",
r"c:\Users\abbegcc\documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_dsm",
"ELEVATION", None, "FLOAT", "CELLSIZE", 1, 1)
StudyArea_DSMOk = ExtractByMask("StudyArea_dsm",
r"D:\IowaEnergyCenter\spatial\Additionaldata.gdb\poly_StudyArea_buffer.shp", "INSIDE")
StudyArea_DSMOk.save("StudyArea_DSMOk")
StudyArea_nDSM = arcpy.ia.Minus("StudyArea_DSMOk", "StudyArea_DEMOk");
StudyArea_nDSM.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_nDSM")
arcpy.management.MakeLasDatasetLayer("StudyArea_LAS.lasd", 'StudyArea_GroundClass', '2')
arcpy.management.LasPointStatsAsRaster("StudyArea_GroundClass", "StudyArea_BEDensity",
"POINT_COUNT", "CELLSIZE", "1")
```

```

StudyArea_BE_IsNull = arcpy.sa.IsNull("StudyArea_BEDensity");
StudyArea_BE_IsNull.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_BE_IsNull"
)

StudyArea_BE_OK = arcpy.sa.Con("StudyArea_BE_IsNull", 0, "StudyArea_BEDensity", '');
StudyArea_BE_OK.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_BE_OK")

arcpy.management.MakeLasDatasetLayer("StudyArea_LAS.lasd", 'StudyArea_VegClass', '3; 4;
5')

arcpy.management.LasPointStatsAsRaster("StudyArea_VegClass", "StudyArea_AGDensity",
"POINT_COUNT", "CELLSIZE", "1")

StudyArea_AG_IsNull = arcpy.sa.IsNull("StudyArea_AGDensity");
StudyArea_AG_IsNull.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_AG_IsNull"
)

StudyArea_AG_OK = arcpy.sa.Con("StudyArea_AG_IsNull", 0, "StudyArea_AGDensity", '');
StudyArea_AG_OK.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_AG_OK")

StudyArea_Density_Composite = arcpy.sa.Plus("StudyArea_AG_OK", "StudyArea_BE_OK");
StudyArea_Density_Composite.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_Density_Co
mposite")

StudyArea_Density_Float = arcpy.sa.Float("StudyArea_Density_Composite");
StudyArea_Density_Float.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_Density_Fl
oat")

StudyArea_CanopyCover = arcpy.sa.Divide("StudyArea_AG_OK", "StudyArea_Density_Float");
StudyArea_CanopyCover.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_CanopyCove
r")

StudyArea_CC = ExtractByMask("StudyArea_CanopyCover",
r"D:\IowaEnergyCenter\spatial\Additionaldata.gdb\poly_StudyArea_buffer.shp", "INSIDE")
StudyArea_CC.save("StudyArea_CC")

StudyArea_CDM = Times("StudyArea_nDSM", "StudyArea_CC")
StudyArea_CDM.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_CDM")

arcpy.conversion.MultipatchToRaster("StudyArea_3DBuildings",
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyAre
a\StudyArea.gdb\StudyArea_Min3d", 1, "MINIMUM_HEIGHT"
)
```

```
arcpy.conversion.MultipatchToRaster("StudyArea_3DBuildings",
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyAre
a\StudyArea.gdb\StudyArea_Max3d", 1, "MAXIMUM_HEIGHT"
)

arcpy.ddd.Minus("StudyArea_Max3d", "StudyArea_Min3d",
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_Absheight"
)

StudyArea_Absheight_IsNull = arcpy.ia.IsNull("StudyArea_Absheight");
StudyArea_Absheight_IsNull.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_Absheight_
IsNull")

StudyArea_Absheight_BHeight = arcpy.ia.Con("StudyArea_Absheight_IsNull", 0,
"StudyArea_Absheight", "Value = 1");
StudyArea_Absheight_BHeight.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_Absheight_
BHeight")

StudyArea_BH = ExtractByMask("StudyArea_Absheight_BHeight",
r"D:\IowaEnergyCenter\spatial\Additionaldata.gdb\poly_StudyArea_buffer.shp", "INSIDE")
StudyArea_BH.save("StudyArea_BH")

arcpy.ddd.MultiPatchFootprint("StudyArea_3DBuildings",
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\Stud
yArea.gdb\StudyArea_MultipatchVolume", "ORIG_OID")

arcpy.conversion.ExportTable("StudyArea_3DBuildings",
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\Study
Area.gdb\StudyArea_Volume_Table",
'', "NOT_USE_ALIAS", 'Z_Min "Z_Min" true true false 8
Double 0 0,First,#,StudyArea_3DBuildings,Z_Min,-1,-1;Z_Max
"Z_Max" true true false 8 Double 0
0,First,#,StudyArea_3DBuildings,Z_Max,-1,-1;Z_Mean "Z_Mean"
true true false 8 Double 0
0,First,#,StudyArea_3DBuildings,Z_Mean,-1,-1;SArea "SArea"
true true false 8 Double 0
0,First,#,StudyArea_3DBuildings,SArea,-1,-1;Volume "Volume"
true true false 8 Double 0
0,First,#,StudyArea_3DBuildings,Volume,-1,-1;Min_Slope
"Min_Slope" true true false 8 Double 0
0,First,#,StudyArea_3DBuildings,Min_Slope,-1,-1;Max_Slope
"Max_Slope" true true false 8 Double 0
0,First,#,StudyArea_3DBuildings,Max_Slope,-1,-1;Avg_Slope
"Avg_Slope" true true false 8 Double 0
0,First,#,StudyArea_3DBuildings,Avg_Slope,-1,-1;BldgID
"BldgID" true true false 2 Short 0
0,First,#,StudyArea_3DBuildings,BldgID,-1,-1;BldgUse
```

```
"BldgUse" true true false 2 Short 0
0,First,#,StudyArea_3DBuildings,BldgUse,-1,-1;BldgType
"BldgType" true true false 25 Text 0
0,First,#,StudyArea_3DBuildings,BldgType,0,25;BldgFloors
"BldgFloors" true true false 2 Short 0
0,First,#,StudyArea_3DBuildings,BldgFloors,-1,-1;BldgHeight
"BldgHeight" true true false 2 Short 0
0,First,#,StudyArea_3DBuildings,BldgHeight,-1,-1;BldgName
"BldgName" true true false 50 Text 0
0,First,#,StudyArea_3DBuildings,BldgName,0,50;YearBuilt
"YearBuilt" true true false 2 Short 0
0,First,#,StudyArea_3DBuildings,YearBuilt,-1,-1;PrimaryAddre
ss "PrimaryAddress" true true false 50 Text 0
0,First,#,StudyArea_3DBuildings,PrimaryAddress,0,50;FloodsAt
"FloodsAt" true true false 2 Short 0
0,First,#,StudyArea_3DBuildings,FloodsAt,-1,-1;Class
"Class" true true false 254 Text 0
0,First,#,StudyArea_3DBuildings,Class,0,254;Confidence
"Confidence" true true false 8 Double 0
0,First,#,StudyArea_3DBuildings,Confidence,-1,-1;ORIG_OID
"ORIG_OID" true true false 4 Long 0
0,First,#,StudyArea_3DBuildings,ORIG_OID,-1,-1;STATUS
"STATUS" true true false 4 Long 0
0,First,#,StudyArea_3DBuildings,STATUS,-1,-1;Shape_Leng
"Shape_Leng" true true false 8 Double 0
0,First,#,StudyArea_3DBuildings,Shape_Leng,-1,-1;ORIG_OID_1
"ORIG_OID_1" true true false 4 Long 0
0,First,#,StudyArea_3DBuildings,ORIG_OID_1,-1,-1', None)

arcpy.management.AddJoin("StudyArea_MultipatchVolume",
"ORIG_OID", "StudyArea_Volume_Table", "ORIG_OID", "KEEP_ALL",
"NO_INDEX_JOIN_FIELDS")

arcpy.conversion.ExportFeatures("StudyArea_MultipatchVolume",
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_BuildingVo
lume", '', "NOT_USE_ALIAS", 'release "release" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.release,-1,-1;capture_da
"capture_da" true true false 254 Text 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.capture_da,0,254;ORIG_OID
"ORIG_OID" true true false 4 Long 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.ORIG_OID,-1,-1;Z_Min
"Z_Min" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.Z_Min,-1,-1;Z_Max
"Z_Max" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.Z_Max,-1,-1;Z_Mean
"Z_Mean" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.Z_Mean,-1,-1;SArea
"SArea" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.SArea,-1,-1;Min_Slope
"Min_Slope" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.Min_Slope,-1,-1;Max_Slope
"Max_Slope" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.Max_Slope,-1,-1;Avg_Slope
```

"Avg_Slope" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.Avg_Slope,-1,-1;Z_Min_1
"Z_Min_1" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.Z_Min_1,-1,-1;Z_Max_1
"Z_Max_1" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.Z_Max_1,-1,-1;Shape_Lengt
h "Shape_Length" false true true 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.Shape_Length,-1,-1;Shape_
Area "Shape_Area" false true true 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_MultipatchVolume.Shape_Area,-1,-1;OBJECTID
"OBJECTID" false true false 4 Long 0
9,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.OBJECTID,-1,-1;Z_Min "Z_Min"
true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.Z_Min,-1,-1;Z_Max "Z_Max"
true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.Z_Max,-1,-1;Z_Mean "Z_Mean"
true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.Z_Mean,-1,-1;SArea "SArea"
true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.SArea,-1,-1;Volume "Volume"
true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.Volume,-1,-1;Min_Slope
"Min_Slope" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.Min_Slope,-1,-1;Max_Slope
"Max_Slope" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.Max_Slope,-1,-1;Avg_Slope
"Avg_Slope" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.Avg_Slope,-1,-1;BldgID
"BldgID" true true false 2 Short 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.BldgID,-1,-1;BldgUse
"BldgUse" true true false 2 Short 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.BldgUse,-1,-1;BldgType
"BldgType" true true false 25 Text 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.BldgType,0,25;BldgFloors
"BldgFloors" true true false 2 Short 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.BldgFloors,-1,-1;BldgHeight
"BldgHeight" true true false 2 Short 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.BldgHeight,-1,-1;BldgName
"BldgName" true true false 50 Text 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.BldgName,0,50;YearBuilt
"YearBuilt" true true false 2 Short 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.YearBuilt,-1,-1;PrimaryAddres
s "PrimaryAddress" true true false 50 Text 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.PrimaryAddress,0,50;FloodsAt
"FloodsAt" true true false 2 Short 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.FloodsAt,-1,-1;Class "Class"
true true false 254 Text 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.Class,0,254;Confidence
"Confidence" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.Confidence,-1,-1;ORIG_OID
"ORIG_OID" true true false 4 Long 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.ORIG_OID,-1,-1;STATUS
"STATUS" true true false 4 Long 0

```
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.STATUS,-1,-1;Shape_Leng
"Shape_Leng" true true false 8 Double 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.Shape_Leng,-1,-1;ORIG_OID_1
"ORIG_OID_1" true true false 4 Long 0
0,First,#,StudyArea_MultipatchVolume,StudyArea_Volume_Table.ORIG_OID_1,-1,-1', None)

arcpy.conversion.FeatureToRaster("StudyArea_BuildingVolume",
"Volume", "StudyArea_BVolume", 1)

StudyArea_IsNull_Vol = arcpy.ia.IsNull("StudyArea_BVolume");
StudyArea_IsNull_Vol.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_IsNull_Vol
")

StudyArea_VolumeOk = arcpy.ia.Con("StudyArea_IsNull_Vol", 0, "StudyArea_BVolume", "Value
= 1");
StudyArea_VolumeOk.save(
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_VolumeOk")

StudyArea_BV = ExtractByMask("StudyArea_VolumeOk",
r"D:\IowaEnergyCenter\spatial\Additionaldata.gdb\poly_StudyArea_buffer.shp", "INSIDE")
StudyArea_BV.save("StudyArea_BV")

neighborhood = arcpy.sa.NbrCircle("50", "MAP")
StudyArea_BV_Mean50 = arcpy.sa.FocalStatistics("StudyArea_BV", neighborhood, "MEAN",
"DATA")
StudyArea_BV_Mean50.save("StudyArea_BV_Mean50")

neighborhood = arcpy.sa.NbrCircle("100", "MAP")
StudyArea_BV_Mean100 = arcpy.sa.FocalStatistics("StudyArea_BV", neighborhood, "MEAN",
"DATA")
StudyArea_BV_Mean100.save("StudyArea_BV_Mean100")

neighborhood = arcpy.sa.NbrCircle("200", "MAP")
StudyArea_BV_Mean200 = arcpy.sa.FocalStatistics("StudyArea_BV", neighborhood, "MEAN",
"DATA")
StudyArea_BV_Mean200.save("StudyArea_BV_Mean200")

neighborhood = arcpy.sa.NbrCircle("400", "MAP")
StudyArea_BV_Mean400 = arcpy.sa.FocalStatistics("StudyArea_BV", neighborhood, "MEAN",
"DATA")
StudyArea_BV_Mean400.save("StudyArea_BV_Mean400")

neighborhood = arcpy.sa.NbrCircle("800", "MAP")
StudyArea_BV_Mean800 = arcpy.sa.FocalStatistics("StudyArea_BV", neighborhood, "MEAN",
"DATA")
StudyArea_BV_Mean800.save("StudyArea_BV_Mean800")

neighborhood = arcpy.sa.NbrCircle("50", "MAP")
StudyArea_BV_STD50 = arcpy.sa.FocalStatistics("StudyArea_BV", neighborhood, "STD", "DATA"
)
StudyArea_BV_STD50.save("StudyArea_BV_STD50")

```

```
neighborhood = arcpy.sa.NbrCircle("100", "MAP")
StudyArea_BV_STD100 = arcpy.sa.FocalStatistics("StudyArea_BV", neighborhood, "STD", "DATA")
StudyArea_BV_STD100.save("StudyArea_BV_STD100")

neighborhood = arcpy.sa.NbrCircle("200", "MAP")
StudyArea_BV_STD200 = arcpy.sa.FocalStatistics("StudyArea_BV", neighborhood, "STD", "DATA")
StudyArea_BV_STD200.save("StudyArea_BV_STD200")

neighborhood = arcpy.sa.NbrCircle("400", "MAP")
StudyArea_BV_STD400 = arcpy.sa.FocalStatistics("StudyArea_BV", neighborhood, "STD", "DATA")
StudyArea_BV_STD400.save("StudyArea_BV_STD400")

neighborhood = arcpy.sa.NbrCircle("800", "MAP")
StudyArea_BV_STD800 = arcpy.sa.FocalStatistics("StudyArea_BV", neighborhood, "STD", "DATA")
StudyArea_BV_STD800.save("StudyArea_BV_STD800")

neighborhood = arcpy.sa.NbrCircle("50", "MAP")
StudyArea_BH_Mean50 = arcpy.sa.FocalStatistics("StudyArea_BH", neighborhood, "MEAN", "DATA")
StudyArea_BH_Mean50.save("StudyArea_BH_Mean50")

neighborhood = arcpy.sa.NbrCircle("100", "MAP")
StudyArea_BH_Mean100 = arcpy.sa.FocalStatistics("StudyArea_BH", neighborhood, "MEAN", "DATA")
StudyArea_BH_Mean100.save("StudyArea_BH_Mean100")

neighborhood = arcpy.sa.NbrCircle("200", "MAP")
StudyArea_BH_Mean200 = arcpy.sa.FocalStatistics("StudyArea_BH", neighborhood, "MEAN", "DATA")
StudyArea_BH_Mean200.save("StudyArea_BH_Mean200")

neighborhood = arcpy.sa.NbrCircle("400", "MAP")
StudyArea_BH_Mean400 = arcpy.sa.FocalStatistics("StudyArea_BH", neighborhood, "MEAN", "DATA")
StudyArea_BH_Mean400.save("StudyArea_BH_Mean400")

neighborhood = arcpy.sa.NbrCircle("800", "MAP")
StudyArea_BH_Mean800 = arcpy.sa.FocalStatistics("StudyArea_BH", neighborhood, "MEAN", "DATA")
StudyArea_BH_Mean800.save("StudyArea_BH_Mean800")

neighborhood = arcpy.sa.NbrCircle("50", "MAP")
StudyArea_BH_STD50 = arcpy.sa.FocalStatistics("StudyArea_BH", neighborhood, "STD", "DATA")
StudyArea_BH_STD50.save("StudyArea_BH_STD50")

neighborhood = arcpy.sa.NbrCircle("100", "MAP")
StudyArea_BH_STD100 = arcpy.sa.FocalStatistics("StudyArea_BH", neighborhood, "STD", "DATA")
```

```
StudyArea_BH_STD100.save("StudyArea_BH_STD100")

neighborhood = arcpy.sa.NbrCircle("200", "MAP")
StudyArea_BH_STD200 = arcpy.sa.FocalStatistics("StudyArea_BH", neighborhood, "STD", "DATA")
StudyArea_BH_STD200.save("StudyArea_BH_STD200")

neighborhood = arcpy.sa.NbrCircle("400", "MAP")
StudyArea_BH_STD400 = arcpy.sa.FocalStatistics("StudyArea_BH", neighborhood, "STD", "DATA")
StudyArea_BH_STD400.save("StudyArea_BH_STD400")

neighborhood = arcpy.sa.NbrCircle("800", "MAP")
StudyArea_BH_STD800 = arcpy.sa.FocalStatistics("StudyArea_BH", neighborhood, "STD", "DATA")
StudyArea_BH_STD800.save("StudyArea_BH_STD800")

neighborhood = arcpy.sa.NbrCircle("50", "MAP")
StudyArea_CC_Mean50 = arcpy.sa.FocalStatistics("StudyArea_CC", neighborhood, "MEAN", "DATA")
StudyArea_CC_Mean50.save("StudyArea_CC_Mean50")

neighborhood = arcpy.sa.NbrCircle("100", "MAP")
StudyArea_CC_Mean100 = arcpy.sa.FocalStatistics("StudyArea_CC", neighborhood, "MEAN", "DATA")
StudyArea_CC_Mean100.save("StudyArea_CC_Mean100")

neighborhood = arcpy.sa.NbrCircle("200", "MAP")
StudyArea_CC_Mean200 = arcpy.sa.FocalStatistics("StudyArea_CC", neighborhood, "MEAN", "DATA")
StudyArea_CC_Mean200.save("StudyArea_CC_Mean200")

neighborhood = arcpy.sa.NbrCircle("400", "MAP")
StudyArea_CC_Mean400 = arcpy.sa.FocalStatistics("StudyArea_CC", neighborhood, "MEAN", "DATA")
StudyArea_CC_Mean400.save("StudyArea_CC_Mean400")

neighborhood = arcpy.sa.NbrCircle("800", "MAP")
StudyArea_CC_Mean800 = arcpy.sa.FocalStatistics("StudyArea_CC", neighborhood, "MEAN", "DATA")
StudyArea_CC_Mean800.save("StudyArea_CC_Mean800")

neighborhood = arcpy.sa.NbrCircle("50", "MAP")
StudyArea_CC_STD50 = arcpy.sa.FocalStatistics("StudyArea_CC", neighborhood, "STD", "DATA")
StudyArea_CC_STD50.save("StudyArea_CC_STD50")

neighborhood = arcpy.sa.NbrCircle("100", "MAP")
StudyArea_CC_STD100 = arcpy.sa.FocalStatistics("StudyArea_CC", neighborhood, "STD", "DATA")
StudyArea_CC_STD100.save("StudyArea_CC_STD100")

neighborhood = arcpy.sa.NbrCircle("200", "MAP")
```

```
StudyArea_CC_STD200 = arcpy.sa.FocalStatistics("StudyArea_CC", neighborhood, "STD", "DATA")
StudyArea_CC_STD200.save("StudyArea_CC_STD200")

neighborhood = arcpy.sa.NbrCircle("400", "MAP")
StudyArea_CC_STD400 = arcpy.sa.FocalStatistics("StudyArea_CC", neighborhood, "STD", "DATA")
StudyArea_CC_STD400.save("StudyArea_CC_STD400")

neighborhood = arcpy.sa.NbrCircle("800", "MAP")
StudyArea_CC_STD800 = arcpy.sa.FocalStatistics("StudyArea_CC", neighborhood, "STD", "DATA")
StudyArea_CC_STD800.save("StudyArea_CC_STD800")

neighborhood = arcpy.sa.NbrCircle("50", "MAP")
StudyArea_CDM_Mean50 = arcpy.sa.FocalStatistics("StudyArea_CDM", neighborhood, "MEAN", "DATA")
StudyArea_CDM_Mean50.save("StudyArea_CDM_Mean50")

neighborhood = arcpy.sa.NbrCircle("100", "MAP")
StudyArea_CDM_Mean100 = arcpy.sa.FocalStatistics("StudyArea_CDM", neighborhood, "MEAN", "DATA")
StudyArea_CDM_Mean100.save("StudyArea_CDM_Mean100")

neighborhood = arcpy.sa.NbrCircle("200", "MAP")
StudyArea_CDM_Mean200 = arcpy.sa.FocalStatistics("StudyArea_CDM", neighborhood, "MEAN", "DATA")
StudyArea_CDM_Mean200.save("StudyArea_CDM_Mean200")

neighborhood = arcpy.sa.NbrCircle("400", "MAP")
StudyArea_CDM_Mean400 = arcpy.sa.FocalStatistics("StudyArea_CDM", neighborhood, "MEAN", "DATA")
StudyArea_CDM_Mean400.save("StudyArea_CDM_Mean400")

neighborhood = arcpy.sa.NbrCircle("800", "MAP")
StudyArea_CDM_Mean800 = arcpy.sa.FocalStatistics("StudyArea_CDM", neighborhood, "MEAN", "DATA")
StudyArea_CDM_Mean800.save("StudyArea_CDM_Mean800")

neighborhood = arcpy.sa.NbrCircle("50", "MAP")
StudyArea_CDM_STD50 = arcpy.sa.FocalStatistics("StudyArea_CDM", neighborhood, "STD", "DATA")
StudyArea_CDM_STD50.save("StudyArea_CDM_STD50")

neighborhood = arcpy.sa.NbrCircle("100", "MAP")
StudyArea_CDM_STD100 = arcpy.sa.FocalStatistics("StudyArea_CDM", neighborhood, "STD", "DATA")
StudyArea_CDM_STD100.save("StudyArea_CDM_STD100")

neighborhood = arcpy.sa.NbrCircle("200", "MAP")
StudyArea_CDM_STD200 = arcpy.sa.FocalStatistics("StudyArea_CDM", neighborhood, "STD", "DATA")
StudyArea_CDM_STD200.save("StudyArea_CDM_STD200")
```

```
neighborhood = arcpy.sa.NbrCircle("400", "MAP")
StudyArea_CDM_STD400 = arcpy.sa.FocalStatistics("StudyArea_CDM", neighborhood, "STD", "DATA")
StudyArea_CDM_STD400.save("StudyArea_CDM_STD400")

neighborhood = arcpy.sa.NbrCircle("800", "MAP")
StudyArea_CDM_STD800 = arcpy.sa.FocalStatistics("StudyArea_CDM", neighborhood, "STD", "DATA")
StudyArea_CDM_STD800.save("StudyArea_CDM_STD800")

neighborhood = arcpy.sa.NbrCircle("50", "MAP")
StudyArea_NDVI_Mean50 = arcpy.sa.FocalStatistics("StudyArea_NDVI", neighborhood, "MEAN", "DATA")
StudyArea_NDVI_Mean50.save("StudyArea_NDVI_Mean50")

neighborhood = arcpy.sa.NbrCircle("100", "MAP")
StudyArea_NDVI_Mean100 = arcpy.sa.FocalStatistics("StudyArea_NDVI", neighborhood, "MEAN", "DATA")
StudyArea_NDVI_Mean100.save("StudyArea_NDVI_Mean100")

neighborhood = arcpy.sa.NbrCircle("200", "MAP")
StudyArea_NDVI_Mean200 = arcpy.sa.FocalStatistics("StudyArea_NDVI", neighborhood, "MEAN", "DATA")
StudyArea_NDVI_Mean200.save("StudyArea_NDVI_Mean200")

neighborhood = arcpy.sa.NbrCircle("400", "MAP")
StudyArea_NDVI_Mean400 = arcpy.sa.FocalStatistics("StudyArea_NDVI", neighborhood, "MEAN", "DATA")
StudyArea_NDVI_Mean400.save("StudyArea_NDVI_Mean400")

neighborhood = arcpy.sa.NbrCircle("800", "MAP")
StudyArea_NDVI_Mean800 = arcpy.sa.FocalStatistics("StudyArea_NDVI", neighborhood, "MEAN", "DATA")
StudyArea_NDVI_Mean800.save("StudyArea_NDVI_Mean800")

neighborhood = arcpy.sa.NbrCircle("50", "MAP")
StudyArea_NDVI_STD50 = arcpy.sa.FocalStatistics("StudyArea_NDVI", neighborhood, "STD", "DATA")
StudyArea_NDVI_STD50.save("StudyArea_NDVI_STD50")

neighborhood = arcpy.sa.NbrCircle("100", "MAP")
StudyArea_NDVI_STD100 = arcpy.sa.FocalStatistics("StudyArea_NDVI", neighborhood, "STD", "DATA")
StudyArea_NDVI_STD100.save("StudyArea_NDVI_STD100")

neighborhood = arcpy.sa.NbrCircle("200", "MAP")
StudyArea_NDVI_STD200 = arcpy.sa.FocalStatistics("StudyArea_NDVI", neighborhood, "STD", "DATA")
StudyArea_NDVI_STD200.save("StudyArea_NDVI_STD200")

neighborhood = arcpy.sa.NbrCircle("400", "MAP")
StudyArea_NDVI_STD400 = arcpy.sa.FocalStatistics("StudyArea_NDVI", neighborhood, "STD",
```

```
"DATA")
StudyArea_NDVI_STD400.save("StudyArea_NDVI_STD400")

neighborhood = arcpy.sa.NbrCircle("800", "MAP")
StudyArea_NDVI_STD800 = arcpy.sa.FocalStatistics("StudyArea_NDVI", neighborhood, "STD", "DATA")
StudyArea_NDVI_STD800.save("StudyArea_NDVI_STD800")

arcpy.stats.Forest("PREDICT_RASTER", "StudyArea_Afternoon", "air_temp", None, None, None,
"StudyArea_BV_Mean50; StudyArea_BV_Mean100; StudyArea_BV_Mean200;
StudyArea_BV_Mean400; StudyArea_BV_Mean800; StudyArea_BH_Mean50;
StudyArea_BH_Mean100; StudyArea_BH_Mean200; StudyArea_BH_Mean400;
StudyArea_BH_Mean800; StudyArea_CC_Mean50; StudyArea_CC_Mean100;
StudyArea_CC_Mean200; StudyArea_CC_Mean400; StudyArea_CC_Mean800;
StudyArea_CDM_Mean50; StudyArea_CDM_Mean100; StudyArea_CDM_Mean200;
StudyArea_CDM_Mean400; StudyArea_CDM_Mean800; StudyArea_NDVI_Mean50;
StudyArea_NDVI_Mean100; StudyArea_NDVI_Mean200; StudyArea_NDVI_Mean400;
StudyArea_NDVI_Mean800; StudyArea_BH_STD50; StudyArea_BH_STD100;
StudyArea_BH_STD200; StudyArea_BH_STD400; StudyArea_BH_STD800;
StudyArea_BV_STD50; StudyArea_BV_STD100; StudyArea_BV_STD200;
StudyArea_BV_STD400; StudyArea_BV_STD800; StudyArea_CC_STD50;
StudyArea_CC_STD100; StudyArea_CC_STD200; StudyArea_CC_STD400;
StudyArea_CC_STD800; StudyArea_CDM_STD50; StudyArea_CDM_STD100;
StudyArea_CDM_STD200; StudyArea_CDM_STD400; StudyArea_CDM_STD800;
StudyArea_NDVI_STD50; StudyArea_NDVI_STD100; StudyArea_NDVI_STD200;
StudyArea_NDVI_STD400; StudyArea_NDVI_STD800",
None, None, r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_RF_Afternoon", None, None,
"StudyArea_BV_Mean50 StudyArea_BV_Mean50; StudyArea_BV_Mean100 StudyArea_BV_Mean100;
StudyArea_BV_Mean200 StudyArea_BV_Mean200; StudyArea_BV_Mean400 StudyArea_BV_Mean400;
StudyArea_BV_Mean800 StudyArea_BV_Mean800; StudyArea_BH_Mean50 StudyArea_BH_Mean50;
StudyArea_BH_Mean100 StudyArea_BH_Mean100; StudyArea_BH_Mean200 StudyArea_BH_Mean200;
StudyArea_BH_Mean400 StudyArea_BH_Mean400; StudyArea_BH_Mean800 StudyArea_BH_Mean800;
StudyArea_CC_Mean50 StudyArea_CC_Mean50; StudyArea_CC_Mean100 StudyArea_CC_Mean100;
StudyArea_CC_Mean200 StudyArea_CC_Mean200; StudyArea_CC_Mean400 StudyArea_CC_Mean400;
StudyArea_CC_Mean800 StudyArea_CC_Mean800; StudyArea_CDM_Mean50 StudyArea_CDM_Mean50;
StudyArea_CDM_Mean100 StudyArea_CDM_Mean100; StudyArea_CDM_Mean200 StudyArea_CDM_Mean200;
StudyArea_CDM_Mean400 StudyArea_CDM_Mean400; StudyArea_CDM_Mean800 StudyArea_CDM_Mean800;
StudyArea_NDVI_Mean50 StudyArea_NDVI_Mean50; StudyArea_NDVI_Mean100 StudyArea_NDVI_Mean100;
StudyArea_NDVI_Mean200 StudyArea_NDVI_Mean200; StudyArea_NDVI_Mean400
```

```
StudyArea_NDVI_Mean400;
StudyArea_NDVI_Mean800 StudyArea_NDVI_Mean800; StudyArea_BH_STD50
StudyArea_BH_STD50;
StudyArea_BH_STD100 StudyArea_BH_STD100; StudyArea_BH_STD200
StudyArea_BH_STD200;
StudyArea_BH_STD400 StudyArea_BH_STD400; StudyArea_BH_STD800
StudyArea_BH_STD800;
StudyArea_BV_STD50 StudyArea_BV_STD50; StudyArea_BV_STD100
StudyArea_BV_STD100;
StudyArea_BV_STD200 StudyArea_BV_STD200; StudyArea_BV_STD400
StudyArea_BV_STD400;
StudyArea_BV_STD800 StudyArea_BV_STD800; StudyArea_CC_STD50
StudyArea_CC_STD50;
StudyArea_CC_STD100 StudyArea_CC_STD100; StudyArea_CC_STD200
StudyArea_CC_STD200;
StudyArea_CC_STD400 StudyArea_CC_STD400; StudyArea_CC_STD800
StudyArea_CC_STD800;
StudyArea_CDM_STD50 StudyArea_CDM_STD50; StudyArea_CDM_STD100
StudyArea_CDM_STD100;
StudyArea_CDM_STD200 StudyArea_CDM_STD200; StudyArea_CDM_STD400
StudyArea_CDM_STD400;
StudyArea_CDM_STD800 StudyArea_CDM_STD800; StudyArea_NDVI_STD50
StudyArea_NDVI_STD50;
StudyArea_NDVI_STD100 StudyArea_NDVI_STD100; StudyArea_NDVI_STD200
StudyArea_NDVI_STD200;
StudyArea_NDVI_STD400 StudyArea_NDVI_STD400; StudyArea_NDVI_STD800
StudyArea_NDVI_STD800",
None, None, "TRUE", 1000, None, None, 100, None, 30, None, None,
"FALSE", 1, "FALSE")

arcpy.stats.Forest("PREDICT_RASTER", "StudyArea_Evening", "air_temp", None, None, None,
"StudyArea_BV_Mean50; StudyArea_BV_Mean100; StudyArea_BV_Mean200;
StudyArea_BV_Mean400; StudyArea_BV_Mean800; StudyArea_BH_Mean50;
StudyArea_BH_Mean100; StudyArea_BH_Mean200; StudyArea_BH_Mean400;
StudyArea_BH_Mean800; StudyArea_CC_Mean50; StudyArea_CC_Mean100;
StudyArea_CC_Mean200; StudyArea_CC_Mean400; StudyArea_CC_Mean800;
StudyArea_CDM_Mean50; StudyArea_CDM_Mean100; StudyArea_CDM_Mean200;
StudyArea_CDM_Mean400; StudyArea_CDM_Mean800; StudyArea_NDVI_Mean50;
StudyArea_NDVI_Mean100; StudyArea_NDVI_Mean200; StudyArea_NDVI_Mean400
;
StudyArea_NDVI_Mean800; StudyArea_BH_STD50; StudyArea_BH_STD100;
StudyArea_BH_STD200; StudyArea_BH_STD400; StudyArea_BH_STD800;
StudyArea_BV_STD50; StudyArea_BV_STD100; StudyArea_BV_STD200;
StudyArea_BV_STD400; StudyArea_BV_STD800; StudyArea_CC_STD50;
StudyArea_CC_STD100; StudyArea_CC_STD200; StudyArea_CC_STD400;
StudyArea_CC_STD800; StudyArea_CDM_STD50; StudyArea_CDM_STD100;
StudyArea_CDM_STD200; StudyArea_CDM_STD400; StudyArea_CDM_STD800;
StudyArea_NDVI_STD50; StudyArea_NDVI_STD100; StudyArea_NDVI_STD200;
StudyArea_NDVI_STD400; StudyArea_NDVI_STD800",
None, None,
r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\S
tudyArea_RF_Evening", None, None,
"StudyArea_BV_Mean50 StudyArea_BV_Mean50; StudyArea_BV_Mean100
```

```
StudyArea_BV_Mean100;
StudyArea_BV_Mean200 StudyArea_BV_Mean200; StudyArea_BV_Mean400 StudyArea_BV_Mean400;
StudyArea_BV_Mean800 StudyArea_BV_Mean800; StudyArea_BH_Mean50 StudyArea_BH_Mean50;
StudyArea_BH_Mean100 StudyArea_BH_Mean100; StudyArea_BH_Mean200 StudyArea_BH_Mean200;
StudyArea_BH_Mean400 StudyArea_BH_Mean400; StudyArea_BH_Mean800 StudyArea_BH_Mean800;
StudyArea_CC_Mean50 StudyArea_CC_Mean50; StudyArea_CC_Mean100 StudyArea_CC_Mean100;
StudyArea_CC_Mean200 StudyArea_CC_Mean200; StudyArea_CC_Mean400 StudyArea_CC_Mean400;
StudyArea_CC_Mean800 StudyArea_CC_Mean800; StudyArea_CDM_Mean50 StudyArea_CDM_Mean50;
StudyArea_CDM_Mean100 StudyArea_CDM_Mean100; StudyArea_CDM_Mean200 StudyArea_CDM_Mean200;
StudyArea_CDM_Mean400 StudyArea_CDM_Mean400; StudyArea_CDM_Mean800 StudyArea_CDM_Mean800;
StudyArea_NDVI_Mean50 StudyArea_NDVI_Mean50; StudyArea_NDVI_Mean100 StudyArea_NDVI_Mean100;
StudyArea_NDVI_Mean200 StudyArea_NDVI_Mean200; StudyArea_NDVI_Mean400 StudyArea_NDVI_Mean400;
StudyArea_NDVI_Mean800 StudyArea_NDVI_Mean800; StudyArea_BH_STD50 StudyArea_BH_STD50;
StudyArea_BH_STD100 StudyArea_BH_STD100; StudyArea_BH_STD200 StudyArea_BH_STD200;
StudyArea_BH_STD400 StudyArea_BH_STD400; StudyArea_BH_STD800 StudyArea_BH_STD800;
StudyArea_BV_STD50 StudyArea_BV_STD50; StudyArea_BV_STD100 StudyArea_BV_STD100;
StudyArea_BV_STD200 StudyArea_BV_STD200; StudyArea_BV_STD400 StudyArea_BV_STD400;
StudyArea_BV_STD800 StudyArea_BV_STD800; StudyArea_CC_STD50 StudyArea_CC_STD50;
StudyArea_CC_STD100 StudyArea_CC_STD100; StudyArea_CC_STD200 StudyArea_CC_STD200;
StudyArea_CC_STD400 StudyArea_CC_STD400; StudyArea_CC_STD800 StudyArea_CC_STD800;
StudyArea_CDM_STD50 StudyArea_CDM_STD50; StudyArea_CDM_STD100 StudyArea_CDM_STD100;
StudyArea_CDM_STD200 StudyArea_CDM_STD200; StudyArea_CDM_STD400 StudyArea_CDM_STD400;
StudyArea_CDM_STD800 StudyArea_CDM_STD800; StudyArea_NDVI_STD50 StudyArea_NDVI_STD50;
StudyArea_NDVI_STD100 StudyArea_NDVI_STD100; StudyArea_NDVI_STD200 StudyArea_NDVI_STD200;
StudyArea_NDVI_STD400 StudyArea_NDVI_STD400; StudyArea_NDVI_STD800 StudyArea_NDVI_STD800",
None, None, "TRUE", 1000, None, None, 100, None, 30, None, None, "FALSE", 1, "FALSE")

arcpy.stats.Forest("PREDICT_RASTER", "StudyArea_Night", "air_temp",
```

```
None, None, None,
"StudyArea_BV_Mean50; StudyArea_BV_Mean100; StudyArea_BV_Mean200;
StudyArea_BV_Mean400; StudyArea_BV_Mean800; StudyArea_BH_Mean50;
StudyArea_BH_Mean100; StudyArea_BH_Mean200; StudyArea_BH_Mean400;
StudyArea_BH_Mean800; StudyArea_CC_Mean50; StudyArea_CC_Mean100;
StudyArea_CC_Mean200; StudyArea_CC_Mean400; StudyArea_CC_Mean800;
StudyArea_CDM_Mean50; StudyArea_CDM_Mean100; StudyArea_CDM_Mean200;
StudyArea_CDM_Mean400; StudyArea_CDM_Mean800; StudyArea_NDVI_Mean50;
StudyArea_NDVI_Mean100; StudyArea_NDVI_Mean200; StudyArea_NDVI_Mean400;
StudyArea_NDVI_Mean800; StudyArea_BH_STD50; StudyArea_BH_STD100;
StudyArea_BH_STD200; StudyArea_BH_STD400; StudyArea_BH_STD800;
StudyArea_BV_STD50; StudyArea_BV_STD100; StudyArea_BV_STD200;
StudyArea_BV_STD400; StudyArea_BV_STD800; StudyArea_CC_STD50;
StudyArea_CC_STD100; StudyArea_CC_STD200; StudyArea_CC_STD400;
StudyArea_CC_STD800; StudyArea_CDM_STD50; StudyArea_CDM_STD100;
StudyArea_CDM_STD200; StudyArea_CDM_STD400; StudyArea_CDM_STD800;
StudyArea_NDVI_STD50; StudyArea_NDVI_STD100; StudyArea_NDVI_STD200;
StudyArea_NDVI_STD400; StudyArea_NDVI_STD800",
None, None, r"C:\Users\abbegcc\Documents\ArcGIS\Projects\StudyArea\StudyArea.gdb\StudyArea_RF_Night", None, None,
"StudyArea_BV_Mean50 StudyArea_BV_Mean50; StudyArea_BV_Mean100 StudyArea_BV_Mean100;
StudyArea_BV_Mean200 StudyArea_BV_Mean200; StudyArea_BV_Mean400 StudyArea_BV_Mean400;
StudyArea_BV_Mean800 StudyArea_BV_Mean800; StudyArea_BH_Mean50 StudyArea_BH_Mean50;
StudyArea_BH_Mean100 StudyArea_BH_Mean100; StudyArea_BH_Mean200 StudyArea_BH_Mean200;
StudyArea_BH_Mean400 StudyArea_BH_Mean400; StudyArea_BH_Mean800 StudyArea_BH_Mean800;
StudyArea_CC_Mean50 StudyArea_CC_Mean50; StudyArea_CC_Mean100 StudyArea_CC_Mean100;
StudyArea_CC_Mean200 StudyArea_CC_Mean200; StudyArea_CC_Mean400 StudyArea_CC_Mean400;
StudyArea_CC_Mean800 StudyArea_CC_Mean800; StudyArea_CDM_Mean50 StudyArea_CDM_Mean50;
StudyArea_CDM_Mean100 StudyArea_CDM_Mean100; StudyArea_CDM_Mean200 StudyArea_CDM_Mean200;
StudyArea_CDM_Mean400 StudyArea_CDM_Mean400; StudyArea_CDM_Mean800 StudyArea_CDM_Mean800;
StudyArea_NDVI_Mean50 StudyArea_NDVI_Mean50; StudyArea_NDVI_Mean100 StudyArea_NDVI_Mean100;
StudyArea_NDVI_Mean200 StudyArea_NDVI_Mean200; StudyArea_NDVI_Mean400 StudyArea_NDVI_Mean400;
StudyArea_NDVI_Mean800 StudyArea_NDVI_Mean800; StudyArea_BH_STD50 StudyArea_BH_STD50;
StudyArea_BH_STD100 StudyArea_BH_STD100; StudyArea_BH_STD200 StudyArea_BH_STD200;
StudyArea_BH_STD400 StudyArea_BH_STD400; StudyArea_BH_STD800 StudyArea_BH_STD800;
StudyArea_BV_STD50 StudyArea_BV_STD50; StudyArea_BV_STD100
```

```
StudyArea_BV_STD100;
StudyArea_BV_STD200 StudyArea_BV_STD200; StudyArea_BV_STD400 StudyArea_BV_STD400;
StudyArea_BV_STD800 StudyArea_BV_STD800; StudyArea_CC_STD50 StudyArea_CC_STD50;
StudyArea_CC_STD100 StudyArea_CC_STD100; StudyArea_CC_STD200 StudyArea_CC_STD200;
StudyArea_CC_STD400 StudyArea_CC_STD400; StudyArea_CC_STD800 StudyArea_CC_STD800;
StudyArea_CDM_STD50 StudyArea_CDM_STD50; StudyArea_CDM_STD100 StudyArea_CDM_STD100;
StudyArea_CDM_STD200 StudyArea_CDM_STD200; StudyArea_CDM_STD400 StudyArea_CDM_STD400;
StudyArea_CDM_STD800 StudyArea_CDM_STD800; StudyArea_NDVI_STD50 StudyArea_NDVI_STD50;
StudyArea_NDVI_STD100 StudyArea_NDVI_STD100; StudyArea_NDVI_STD200 StudyArea_NDVI_STD200;
StudyArea_NDVI_STD400 StudyArea_NDVI_STD400; StudyArea_NDVI_STD800 StudyArea_NDVI_STD800",
None, None, "TRUE", 1000, None, None, 100, None, 30, None, None, "FALSE", 1, "FALSE")
```