8-19-2024

Exploring Heat Variations in Columbia, South Carolina

Michael Graeme Hohlfeld
University of South Carolina

EXPLORING HEAT VARIATIONS IN COLUMBIA, SOUTH CAROLINA

by

Michael Graeme Hohlfeld

Bachelor of Science
Salisbury University, 2022

Submitted in Partial Fulfillment of the Requirements

For the Degree of Master of Science in

Geography

College of Arts and Sciences

University of South Carolina

2024

Accepted by:

April Hiscox, Director of Thesis

Gregory Carbone, Reader

Yuhao Kang, Reader

Ann Vail, Dean of the Graduate School

ACKNOWLEDGEMENTS

I am grateful to Dr. April Hiscox for her continuous support and guidance through this degree and thesis. Thank you, Drs. Carbone and Kang, for being on this committee and offering your time and expertise. Additionally, I am lucky to be a part of the HVRI lab and CSTAR project, as I have been exposed to a broad range of research topics, methods, and knowledgeable people about both. I benefited greatly from being a part of this environment.

ABSTRACT

As the leading cause of weather-related fatalities in the United States, the concept of heat as a hazard has been studied for decades, especially in the urban environment. Urban heat island research is limited by the fact that most investigate land surface temperature (LST) rather than air temperature and other variables that affect human comfort including humidity. This thesis utilizes a public network of weather stations in Richland County, South Carolina to investigate differences in heat between the city of Columbia, and more rural parts of the county. The study period was June, July, and August of 2022. Results indicate that through the summer average, there is a range of afternoon Heat Index deviations of 8.1°F. The afternoon hot spots lie in downtown Columbia, as well as the more rural southwestern portion of the county. Heat Index, temperature, relative humidity, and dew point deviations were investigated for both afternoon and morning. To understand what land use variables best predict these deviations, twelve variables were explored including Normalized Difference Vegetation Index (NDVI) and percent impervious surface each at three scales. Results indicate that afternoon values are difficult to predict, possibly a result of the 10- meter station height with afternoon wind and vertical mixing. Land use variables performed better with morning heat variables. Finally, multiple linear regressions were created to predict morning and afternoon heat variables, with successful models on temperature deviations, with other variables lacking predictability. This project is a step towards creating local heat models based on air temperature and humidity rather than LST.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Heat is the leading cause of weather fatalities in the United States (National Oceanic and Atmospheric Administration (NOAA) 2023). Heat waves have been linked to substantially increased emergency department visits in North Carolina (Fuhrmann et al 2015) and Atlanta (Chen et al 2017), and increased cardiovascular mortality in US counties (Khatana et al 2022). Furthermore, the frequency, intensity, and duration of heat waves have all increased over the past several decades (Kirkpatrick & Lewis 2020, Habeeb et al 2015, Perkins et al 2012), and are expected to rise under climate change scenarios (Meehl & Tebaldi 2004), potentially leading to compounding risk due to consecutive high heat events (Baldwin et al 2019). Urban cores are known to be hotter than surrounding rural and suburban areas, known as the Urban Heat Island (UHI), which is exaggerated during heart waves (Li and Bou- Zeid 2013). With an ever- increasing urban population, extreme heat events, and higher exposure in urban cores, analysis into urban areas deserve attention. This thesis contributes to understanding urban heat by investigating the differences in heat across a county with an urban core, and by exploring what land use variables best predict these differences. Urban heat is often studied through remotely sensed land surface temperature, but by utilizing an existing county network of weather stations, this project is able to explore air temperature and humidity in situ.

1.1 HEAT INDICES

With a heightened awareness and attention towards heat risks, there is ongoing scientific discussion concerning what heat indices to use, when to use them, and how much heat varies within geographic regions. Despite the importance of heat as an atmospheric variable, the ability of the current measurements to predict heat accurately and reliably has been called into question. The National Weather Service (NWS) currently uses three indices for heat- the Heat Index, Wet Bulb Globe Temperature (WBGT), and an experimental tool known as HeatRisk, although the Heat Index remains the only tool for official warnings (NWS 2023).

The Heat Index was originally conceptualized by Robert G. Steadman in the late 1970s, intended to calculate an apparent temperature based on ranges of dry bulb temperature and relative humidity that are likely to be present at the earth's surface (Steadman 1979). The Heat Index is commonly known as the 'feels like' temperature in public forecasts. At a given temperature and humidity, the Heat Index value is the temperature at which one would experience the same comfort with baseline levels of water vapor pressure of 1.6 kPa, wind speed of 5.6 mph, barometric pressure of 101.3 kPa, and in the shade (Steadman 1979). The NWS adopted the Heat Index for their weather communications, and it is their official product for their heat watches, advisories, and warnings (NWS 2023). The NWS standard Heat Index equation is shown in Equation [1] and visualized in Figure 1.1.

$$\mathbf{HI = -42.379 + 2.04901523T + 10.14333127RH - .22475541T*RH - .00683783T^2 - .05481717RH^2 + .00122874T^2RH + .00085282TRH^2 - .00000199*T^2RH^2} \quad [1]$$

Despite its wide acceptance and use, the Heat Index has documented weaknesses as its calculation does not consider solar radiation, pressure, or wind. The Heat Index is measured in the shade, so adding exposure to the sun adds unaccounted stress. It also does not consider wind, which has the capacity to increase evaporative cooling. Steadman did however, author a subsequent paper considering these effects on apparent temperature (Steadman 1979). Recently, the Heat Index was noted to yield unphysical results at extreme temperature and humidity combinations, and an extended thermoregulation model was created to allow for accurate apparent temperatures in extreme environments (Lu and Romps 2022). There were severe underestimations found in the existing Heat Index in high humidity environments when compared to the extended model. For example, the existing Heat Index was found to underestimate the apparent temperature of an 88°F, 96% RH environment by 26°F.

Another index, the Wet Bulb Globe Temperature (WBGT) is a more holistic measurement of heat stress. WBGT was introduced in the 1950's in response to the occurrence of heat illnesses in US Army and Marine Corps training camps (Yaglou & Minard 1956). The WBGT comprises the dry bulb, wet bulb, and black globe temperatures. The dry bulb thermometer is a shaded dry thermometer, also known as air temperature. The wet bulb thermometer includes a wet cloth covering and is exposed to the air, wind, and sun. This serves as a measurement of evaporative cooling capacity of the air. The black globe thermometer lies inside of a black plastic sphere, that heats the inside when exposed to solar radiation. From these measurements the WBGT is calculated using Equation [2].

$$\mathbf{WBGT = .7 * T_{WB} + .2 * T_{BG} + .1 * T_{DB}} \quad [2]$$

By using three thermometers, the WBGT incorporates the effects of temperature, solar radiation, wind, and humidity. Since its initial adoption by the US military, the WBGT has been adopted in other domains including sports medicine. WBGT continues to be used by OSHA (Occupational Safety and Health Administration, OSHA Technical Manual Section III: Chapter 4- Heat Stress 2017, OSHA Heat Hazard Recognition 2023), is mandated for use by the South Carolina High School League South Carolina High School League 2021), and the Georgia High School Association (Georgia High School Association 2024). WBGT is still heavily researched in sports medicine today (e.g. Grunstein et al 2012, Tripp et al 2020, Grundstein et al 2023, Grundstein et al 2022). Although there is widespread adoption of the WBGT for risk monitoring, different risk thresholds exist depending on organization, activity, and acclimation. One example, developed by the University of Georgia divides the United States into three geographical regions and provides different risk thresholds for each (Grundstein et al 2015).

There are limitations to the use of WBGT index. The most severe of which is that proper WBGT measurement requires a wet bulb thermometer with a water reservoir, and a black globe thermometer. Most weather stations, including the nationwide NWS automated stations, do not have these instruments. There are methods for estimating these variables from existing sensors, some of which are fairly accurate (Mullin 2022), but using these methods in place of proper measurements has been questioned (Budd 2007). Wet bulb temperature has been modeled from dry bulb temperature and relative humidity at sea level pressure (Stull 2011), and black globe temperature can be estimated from readily available NWS station data (Dimiceli et al 2011). These WBGT estimation

methods could serve as proxies that do not require new instrumentation at every weather station, but using them is sub- optimal.

WBGT has also been shown to underestimate the physiological effects of restricted evaporation at low winds and high humidity (Budd 2007, Ramanathan & Belding 1973). While its inclusion of solar radiation and wind are crucial for representing heat stress, the typical temporal and spatial fluctuations of wind and solar radiation cause large spatiotemporal heterogeneity. At a given station, wind and solar fluctuations can change a WBGT value by several degrees in seconds or minutes. Coupling these fluctuations with narrow risk categories (e.g., US Military risk categories are 2-5 degrees Fahrenheit) results in rapidly changing perceived risk and recommendations (NOAA 2023). Making an accurate forecast for WBGT that is volatile in the microenvironment that humans experience heat stress, is likely to be a challenge. Another challenge of WBGT is the number of varying risk charts, limiting consistency with risk messaging. Introducing WBGT to the public may be a concern too, because the highest risk category is 90°F, which is less intimidating than Heat Index values of the same risk, or even air temperature of the same risk.

Recently, the western regions of the NWS have been exploring a different indicator known as HeatRisk. This forecasting tool considers temperature deviation from climatological normal, the duration of heat, day and nighttime temperatures, and if those temperatures are expected to have heat-related impacts. The prototype is a product of a collaboration with the Centers for Disease Control and Prevention (CDC) and based on peer reviewed research on heat hospitalizations (NWS 2023, Vaidyanathan et al 2019), and the effects of nighttime temperatures, prolonged exposure, and time of year that has

been supported by other findings (Chen et al 2017). This HeatRisk product is currently experimental and is being adjusted for use in the eastern US. This product uniquely addresses the role of overnight low temperatures. Warm overnight lows cause accumulating heat exposure, and makes it harder and more expensive to cool the home at night, which exaggerate the prolonged effects of heat stress. This is of greater concern to urban residents, because urban areas typically do not cool off at night like their rural counterparts do. A study done by CAPA Strategies for the City of Portland Bureau of Emergency Management tracked temperatures inside 49 public or affordable housing units in Portland over two and a half months. They found that every unit exceeded indoor temperatures over 80°F, 82% of housing units exceeded 85°F, and 18% exceeded 90°F (CAPA Strategies 2023).

A challenging task of heat risk messaging is deciding which method(s) to use, for each time and place. Using different metrics in the same meteorological environment can result in different risk categories and can produce conflicting information. With different methods of measuring and communicating heat risk, there are arguments for which measurement yields the best results in health outcomes. The best measurement of heat for predicting health outcomes varies too much by place, demographic group, and seasons, to select one best overall performing index (Goldie et al 2018, Barnett et al 2010, Goldie et al 2017, Fletcher et al 2012).

1.2 URBAN PHENOMENA

Heat as a hazard is further complicated by the spatial variability in heat conditions. The existence and effects of Urban Heat Island (UHI) have been studied extensively, (e.g., Oke 1980, Arnfield 2003, Roth 2007). The UHI is mostly the result of

an imbalanced surface energy budget caused by artificial surfaces that store heat, anthropogenic heat releases, and limited outgoing longwave radiation (Zhou et al 2019).

More detailed quantification of the UHI effect continues to be an active area of research, with studies using remote sensing to understand land surface temperature (LST) patterns (Imhoff et al 2010, NASA 2022). There are model-based tools to calculate UHI intensity based on urban geometry which is an important mechanism of the Urban Heat Island in cities with dense architecture geometry (Nakata-Osaki et al 2018). Another study noted synergistic effects of the UHI during heat waves, compounding urban risk (Li & Bou-Zeid 2013).

The most common method by which to examine the UHI is by LST, or Surface Urban Heat Island (SUHI). The number of SUHI studies has risen to over 75 per year in 2018 (Zhou et al 2019). Rapid urbanization, enhanced ability of thermal imaging, and increased attention to climate change have driven this rise. Although the insights of remote sensing can give on surface temperature are important, the fundamental limitation of these studies for practical use is that surface temperature is not directly comparable to air temperature (Zhou et al 2019). A review of urban heat thermal remote sensing progress by Voogt and Oke (2003) highlights a lack of observational studies to validate remotely sensed data. The most valuable information to obtain is high resolution temperature, humidity, wind, and solar radiation, data at the height and location of human experience. If public health officials, urban planners, and decision makers want to understand the heat stress their citizens experience, understanding SUHI is insufficient. Validating the SUHI with air temperature data is crucial.

In recent years, vehicle- based measurements for heat island mapping have been done in over 60 cities, including Columbia, SC (CAPA Strategies, Voelkel et al 2016). In 2023, CAPA Strategies, funded by NOAA and NIHHIS (National Integrated Heat Health Information System) completed a 'Heat Watch Report' for the City of Columbia. The campaign, coordinated by the University of South Carolina (USC), had volunteers drive on designated routes throughout the city with a car mounted temperature sensor for three time periods in early August 2022. Sensor data was sent to CAPA Strategies, who used remotely sensed imagery to interpolate between the routes of data. The afternoon results of this effort, shown in Figure 1.2, indicate dramatic temperature differences across the city. The afternoon temperatures ranged from 82.3°F to 100.0°F in downtown Columbia (CAPA Strategies 2022). This degree of afternoon temperature differential was greater in Columbia than for replicated campaigns in other southeast cities such as Atlanta, Charleston, Nashville, Knoxville, Jacksonville, as well as western cities including Austin, Los Angeles, and Albuquerque (NIHHIS Urban Heat Island Mapping Campaign Cities, 2019- 2022).

Each UHI measurement method has its own imperfections, for example, LST does not investigate the air temperature differences directly, where humans are experiencing the heat. Vehicle- based measurements get observations from the air that humans are experiencing, but sensors attached to cars may be exposed to additional sources of heat from the car or aspiration from the car's movement that do not necessarily reflect the conditions otherwise. The exposure to vehicle heat may be exaggerated in urban areas with more traffic, more frequent stops, and lower speeds. In addition, interpolating the temperature for differing land uses between observations exclusively

measured on roads could introduce bias. Additionally, there is anecdotal evidence of heterogeneous cloud cover in Columbia the day of the afternoon measurements. Noting these imperfections does not intend to proclaim that those methods are incorrect or useless, just that each methodology has its flaws.

The UHI is typically defined by temperature differences between urban and surrounding rural areas, but using other variables to consider heat stress, such as humidity, can yield different results. In addition to the UHI there are other urban phenomenon based on moisture differences known as urban moisture island (UMI), or urban dry island (UDI). In an analysis of over 40,000 weather stations in roughly 600 urban clusters in Europe during a 2019 heatwave, researchers found that the UDI effect moderated outdoor Heat Index when coupled with an UHI (Chakraborty et al 2022). They also demonstrated that surface temperature is a poor proxy for daytime urban heat stress variaitons, and that vegetation moderates heat stress poorer than it mitigates surface temperature (Chakraborty et al 2022), which challenges UHI mitigation methods. A study on mitigation efforts in New York City, found that vegetation cooled air temperature more effectively than increasing albedo, and the most effective mitigation strategy was curbside planting (Rosenzweig et al 2006). The effectiveness of mitigation studies, particularly regarding vegetation have conflicting results, depending on what measure of UHI is being used, whether it be temperature, heat stress, or LST. Urban vegetation does contribute other benefits like shade, carbon sequestration, and visual appeal, so heat mitigation is only a fraction of the role of urban vegetation. Additionally, Chakraborty et al found that air temperature and surface temperature had a weak daytime correlation. This analysis demonstrates that measuring UHI intensity with LST, and predicating the

effectiveness of mitigation strategies on LST alone is limited. In a meta- analysis of 34 UMI and UDI studies, cities in humid climates are found to have lower humidity differences between urban and rural locations than arid cities (Huang and Song 2023). In mid latitude cities, urban humidity was found to be between 20% lower and 50% higher than rural humidity, showing a bias towards an UMI (Huang and Song 2023). Through incorporating analysis of humidity, these studies give a better understanding of heat stress than temperature studies alone. Based on varying results on the presence of urban heat stress islands, there is no generalizability, emphasizing the importance of local understanding.

Other methods to understand the role of land use and heat stress have investigated differences in WBGT between land uses over small distances, giving insight into the intra-urban heterogeneity. Measuring WBGT across different surfaces at an athletic complex in Georgia revealed no significant differences (Grundstein & Cooper 2020), but WBGT differences between land surfaces over a similar proximity in North Carolina *were* found, and were exaggerated in sunny versus shaded conditions (Clark 2023). The importance of local assessment of WBGT for decision makers is further emphasized when comparing observed versus modeled data (Tripp et al 2020, Grundstein et al 2022, Pryor et al 2017). There is no consensus on how much heat stress varies based on land use and location within the urban environment and between urban and rural environments. It is also unclear at what scale differences arise, or what scale is appropriate to model UHI.

Currently, the University of South Carolina is working on a project funded by NOAA to understand the geographic variability of heat and heat risk perception for the

Columbia forecast region. This has included collecting data from several temporary weather stations with varying land surfaces over the heat seasons in 2022 and 2023. Preliminary results show that these land use variations yield different counts of days in each risk category for Heat Index. Figure 1.3 shows the differences between five of the project's stations meeting NWS Excessive Heat Advisory Heat Index thresholds, in contrast to the actual number of advisories and warnings that were issued by the forecast office. For days exceeding the 'Excessive Heat Warning' threshold of 110°F Heat Index, three of the stations all had the same percent of 45%. One of these three stations (labelled "Saluda" in Figure 1.3) was an agricultural site in rural Saluda County, SC. "Fort Jackson", a rural site located in a clearing of a large forested area also had 45% of days exceeding the warning threshold, as well as "Barnwell", an urban station located on a parking lot. The station with the lowest percentage was Fairfield, an airport location in rural Fairfield County, SC. The station with the highest percentage of days was 29203, which is an urban station located on concrete. The hottest and coolest stations here fit the typical UHI expectations, but having two rural vegetated sites yielding the same results as a parking lot urban site is curious.

Urban studies have aimed at understanding patterns of heat stress across space in a high- resolution manner. LST studies optimize spatial resolution at the cost of limiting insight into air temperature, humidity, and heat stress. These studies create a snapshot of high- resolution land surface temperature, but leave us clueless as to how much variance we feel in the air above, or temporal analysis. Other studies provide such temporal analysis of air qualities but have a limited number of stations, often just pairing one urban and one rural station. Other work, like the car mounted sensors provide great spatial

resolution and insight into air qualities, but have no temporal component. Therefore, the research gap in the urban heat literature is a temporal analysis of air qualities (both air temperature and humidity), using a high- density network of stations in an urban area. The dataset used in this thesis allows for analysis in this research gap.

1.3 RESEARCH QUESTIONS

The curious results from the ongoing University of South Carolina project, as well as the mixed results from UHI, UDI, UMI literature, and the dramatic UHI from the Heat Watch, led to the need for deeper investigation of city-wide variability analysis for Columbia. This thesis project addresses the topic of heat variability at the local scale by utilizing an existing weather station network in Richland County, South Carolina to answer the following research questions:

1) What are the differences in daily max Heat Index across stations around Columbia, South Carolina?

2) What variables may explain the differences in daily max Heat Index between stations around Columbia South Carolina?

3) Is a statistical model developed from past network observations able to improve Heat Index predictions at a local scale?

The work presented here focuses on the Heat Index and its components. Heat Index allows for a better investigation into heat stress than air temperature, but does not require the instrumentation required for WBGT measurements. Heat Index is also currently the product used for official warnings from the NWS, and is more familiar to the public as

either Heat Index or a ‘feels like’ temperature. Additionally, without considering solar radiation or wind, the heat stress data is less volatile, allowing for a better understanding of general trends.

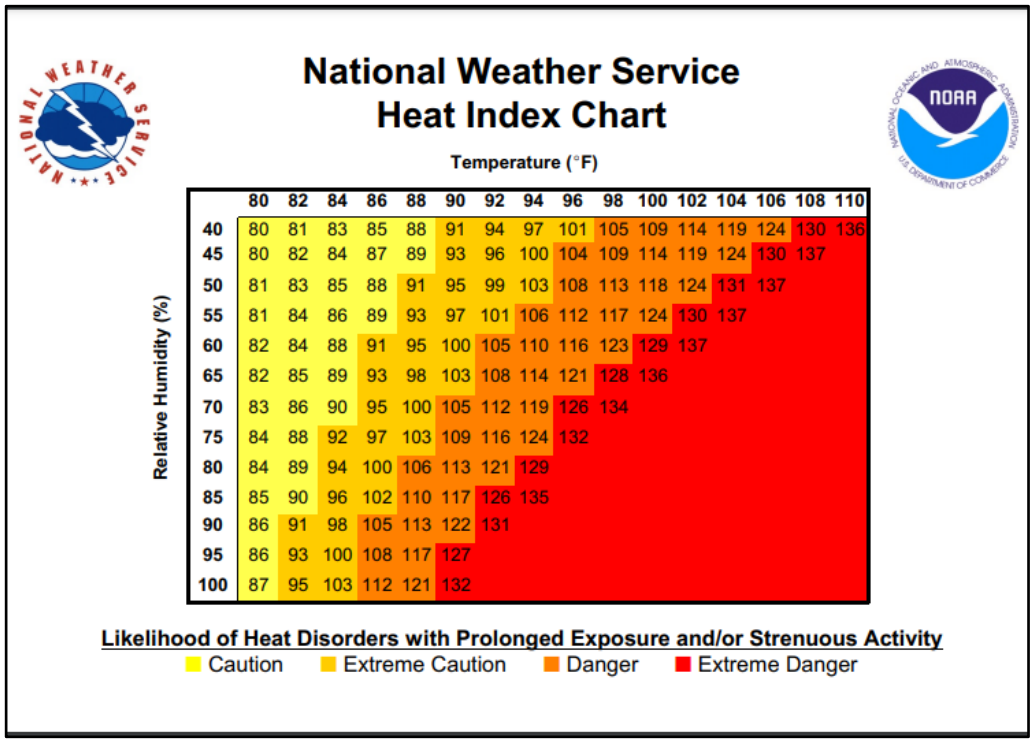

Figure 1.1 National Weather Service Heat Index Chart, from NWS.

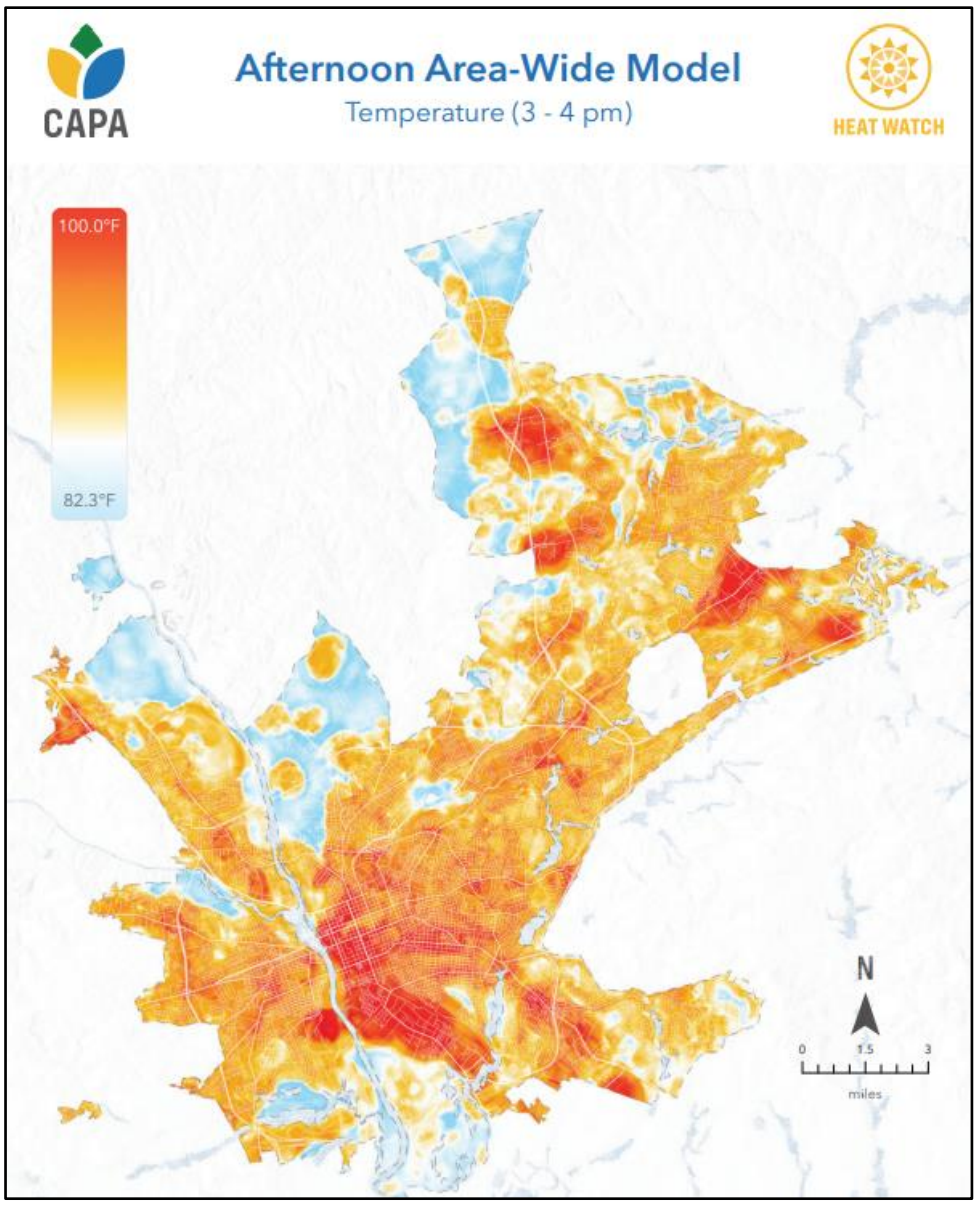

Figure 1.2 CAPA Heat Watch Model, from CAPA Strategies.

Count and Percent of days where 5 min peak met Heat Index advisory thresholds (6/13/23- 10/1/23), 67 days of data							Count of actual NWS Advisories (6/13/23-10/1/23), 111 days of data
NWS Heat Advisory	Flag Value	Saluda	Ft. Jackson	Fairfield County	Barnwell St. Lot	29203 Zip Code	Columbia Forecast Office
No Advisory	<104.9°F	31 (46%)	26 (39%)	42 (63%)	32 (48%)	19 (28%)	110 (99%)
Heat Advisory- (Heat Index Criteria)	105-109.9°F	6 (9%)	11 (16%)	8 (12%)	5 (7%)	7 (10%)	0 (0%)
Excessive Heat Warning- (Heat Index Criteria)	>110°F	30 (45%)	30 (45%)	17 (25%)	30 (45%)	41 (61%)	1 (.9%)

Figure 1.3 UofSC Data Heat Index Threshold Counts, NWS Data from Iowa Environmental Mesonet.

CHAPTER 2

METHODOLOGY

2.1 STUDY AREA

The study area for this thesis was Columbia, South Carolina. Within Richland County and located in the center of South Carolina, Columbia is home to the state capital as well as the University of South Carolina. The location of Richland County within the state and Columbia within the county is shown in Figure 2.1. The city trademarked the slogan ‘Famously Hot’ in 2010, which is backed by summer temperatures regularly in the mid 90’s coupled with oppressive humidity (Experience Columbia SC 2024, NCEI 2024). The record high temperature was a stifling 110°F, measured in June 2012 (NWS 2024). Home to around 130,000 residents, the city is the second largest in the state (US Census Bureau 2024). Demographically, Columbia is roughly half white and 40% black, with a poverty rate more than double the US average (US Census Bureau 2024). Coupling a financially vulnerable population with particularly oppressive heat is a recipe for high risk, especially if the vulnerable population lives within an UHI.

2.2 WEATHER DATA

To explore the differences in heat in Columbia and the relationship between heat and land use, both weather data and land use data were acquired. For the weather data, this research utilized an existing weather station network known as RC Winds (Richland County Weather Information Network Data System), which is a network of professional

grade weather monitors throughout Richland County, South Carolina, and surrounding areas (RC Winds 2023). Historical data is available by request through the RC Winds data manager at Richland County EMS and real time information is available through their public website.

This network has been used by emergency officials and local meteorologists since its origin in 2013 (Richland County SC 2023). The network contains over 60 stations, collecting data in five- minute intervals measuring air temperature, relative humidity, wind speed, solar radiation, barometric pressure and rainfall. Having this density of stations in one county gives officials a high-resolution insight into weather conditions. Without the RC Winds network, county officials would have to rely on the existing network of FAA (Federal Aviation Administration) and NWS automated weather stations at airports, and with only two stations in or near Richland County, the alternative density of information is much lower. Having this dense network is useful for emergencies like wildfires, chemical releases, hazmat incidents, winter weather, flooding, high winds, special events (RC Winds 2024), or in this case, research. The instrument heights range from four to 30 meters, with the mode being 10 meters.

For the work presented here, the 33 stations with a measurement height of 10 meters were used to minimize confounding variables. Figure 2.2 shows the locations of the 10- meter stations distributed across Richland County. Richland County, and therefore the RC Winds network, covers approximately 750 square miles in the heart of South Carolina, including the city of Columbia. The stations are spread throughout the county creating a somewhat uniform density. Most stations are located on the grounds of fire stations or emergency medical service stations, with a few others on other public

lands. Most of the instruments are on top of a dedicated tower in proximity to the one or two- story fire station buildings.

The majority of stations are surrounded by grass, near a building, often with trees higher than the instrument in proximity of the station. An example of a typical station is shown in Figure 2.3. Instrument guidelines for installation are to keep the instrument at least eight feet above peak roof height, for roof installations (Campbell Scientific 2024). Even though not on the roof, most stations are close to a roof. Based on station photos, it appears the majority of stations are well over eight feet above the roof peak, which the example station shown in Figure 2.3 demonstrates. Each station measures air temperature, relative humidity, barometric pressure, rainfall, solar radiation, wind speed, and wind direction at 5-minute time stamps. Each station sends data automatically to the user's computer.

2.3 WEATHER VARIABLES

The raw RC Winds data were used to compute a number of weather-related variables. Because the focus of this work is heat, all RC Winds data were subset to the months of June, July, and August of 2022, resulting in 92 days of data. Heat index was calculated for each available time interval using air temperature and relative humidity, based on the standard National Weather Service equation (Rothfusz 1990), as shown in Equation [1]. In addition, dew point was calculated for each time interval using August-Roche- Magnus approximation based on temperature and relative humidity to investigate UDI UMI (McNoldy 2024).

For initial investigation, the maximum calculated Heat Index for each day at each station was found, called Daily Max Heat Index (DMHI). Then, the daily average across 33 stations for each day was found, called DMHI Average (DMHIA). Further, the deviation for each station from that daily average was calculated, giving 92 days of deviations, named ΔDMHI. That variable was averaged, resulting in the Δ Daily Max Heat Index Average (ΔDMHIA). These variables can be viewed for further context in Table 2.1. This variable was a single value reflecting each stations' tendency to deviate from the network average, in degrees Fahrenheit. Positive values reflected a station tends to have relatively higher daily peaks, and lower values reflected a relatively cool station.

ΔDMHI was plotted in box plots to visualize the range of the differences that existed. The mean value in each of these box plots represents the variable ΔDMHIA. What was found is that the ΔDMHIA ranged from about -12°F to 17°F, producing a range of 29°F. The distribution showed six distinct outliers though, both based on their mean value and range. These outliers were Broad River, Elders Pond, Irmo Fire District, Landfill, Screaming Eagle, and Utilities. Further examination of these six stations indicated probable instrument errors due to excessively low or high relative humidity values, thus they were removed from further analysis. Additional discussion of this removal process can be found in Appendix A. The weather variables used for the final analysis were recalculated without the presence of the outlier station data found in this ΔDMHIA investigation.

For deeper investigation, and to involve more data, the hottest and coolest hours of the day were considered to understand diurnal patterns of temperature, humidity, and their relationships with spatial variables. It was found that the typical hottest hour of the

day through the summer was 3pm, and the coolest was 6am, based on 5- minute average temperature and Heat Index. The data were filtered into a 6am and a 3pm dataset. Deviations were calculated for relative humidity, temperature, Heat Index, dew point, as well as solar radiation and wind speed for each time stamp. For example, at station 'MLK Park', at 6:25 am June 3, the deviation for temperature was the MLK temperature at 6:25 minus the network average at 6:25 am June 3. From this, the median of these values was computed to represent the typical deviation for each station during the 6am hour. Therefore, each station had one median deviation value for temperature, relative humidity, Heat Index, dew point, wind speed, and solar radiation. These median deviation variables were used for the major analysis of this thesis. This allowed for a diurnal understanding of the UHI and UDI/ UMI based on air temperature, humidity, and heat stress. These deviations were assessed via summary statistics to get an understanding of the range and typical values, as well as visualizations including box plots and Inverse Distance weighting spatial interpolation to assess the deviations over space. Moran's I autocorrelation index was also performed to assess the spatial clustering, randomness, or dispersion of these deviations.

2.4 LAND COVER AND LOCATIONAL DATA

Based on the latitude and longitude given by RC Winds, the following variables were created for each station using ArcGIS Pro: elevation, latitude, longitude, multi-scale vegetation index, proximity to the urban center, multi scale impervious surface percentage, proximity to a water body, and soil available water supply, shown in Table 2.2. These served as the independent variables used to assess their prediction power of deviations in Heat Index, temperature, and other variables.

Multi scale vegetation was represented by NDVI (Normalized Difference Vegetation Index), calculated for the area around each station for three buffers, 5km, 1km, and 50m. The average value of the pixels within that circle represented the station's NDVI. The NDVI was downloaded from USGS EarthExplorer and was computed using Landsat 8 imagery from June 15, 2022, which was the best available imagery from the centermost date of the study period. While the NDVI value may change throughout the summer, only one value was calculated to represent each station's vegetation for the study period. The pixel size for this raster imagery is 30 meters. NDVI is a well- known index for assessing plant health and vegetation density (USGS 2023). Vegetation has the capacity to mitigate heat and add humidity. This index was particularly useful in testing its relationships to weather variable deviations.

Percent impervious surfaces was computed at the same buffers as NDVI, from the 2021 NLCD (National Land Cover Dataset). It is assumed that the land cover did not change enough between 2021 and 2022. The pixel size for this data was also 30 meters. Each land cover type in the dataset was given a value of 0 to 1 to represent it's percent perviousness, then the average value was calculated for each buffer area around each station. The buffers for this variable were the same as the NDVI, 5km, 1km, and 50m. A 50m buffer represented the immediate land use attributes near a station. The 1km buffer was chosen to represent intermediate land use attributes, relative to the study area. The 5km value was chosen to represent a broad land use attribute, without having too much overlap with other stations to avoid minimal inter- station variability. The extent of the 5km and 1km buffers are shown in Figure 2.4.

Proximity to the urban center was defined as the Euclidean distance to a center point of Columbia, the State House, in meters. The State House was chosen because it is a landmark that lies in the heart of the downtown Columbia area. Near it, the most densely urban land use exists. In addition, the State House lies in the heart of the hottest part of Columbia, as found by the CAPA Heat Watch (Figure 1.2).

Proximity to water body was defined as the Euclidean distance in meters to the nearest border of Lake Murray, which was considered the only water body close enough to the stations and large enough to have an effect on the station data. Lake Murray is a 48,000 acre reservoir located just west of Columbia. Upon the completion of the reservoir in the 1920's to make hydroelectric power, the 1.6-mile dam was the largest earthen dam in the world (South Carolina Department of Natural Resources 2024). The lake perimeter was chosen because similar to lake effect snow, or land and sea breezes, the phenomena are observed in higher correlation with the shoreline than the center of the water body. Although the severe lake effect snow and land and sea breezes are caused by much larger bodies of water than Lake Murray, a smaller scale effect is expected at Lake Murray, so the same perimeter logic was applied to compute the Proximity to Water Body variable.

Soil Available Water Supply (SAWS) was found through the USDA (United States Department of Agriculture) Soil Survey and was chosen to represent a soil type's capacity to hold moisture in the upper layer (USDA 2023). This variable served as a proxy for evaporation capacity of a soil. A soil with more evaporation will cool the soil, but add heat stress to the air above. The SAWS value for each station was the SAWS value for the soil type directly under the station point.

2.5 SPATIAL VARIABLES

To understand the distribution of the deviation of daily maximums several spatial techniques were used. Inverse Distance Weighting (IDW) was used as a spatial interpolation method to view the Median ΔDMHI variable on a continuum. From a set of points with a chosen variable to interpolate, IDW creates a raster of cell values by using a linearly weighted combination of a set of sample points (ESRI 2023). Using an appropriate color scheme, the IDW interpolation creates a heat map for the chosen variable. To further analyze the spatial distribution of this the ΔDMHIM, Moran's I Index was calculated. This tool gave insight into the spatial autocorrelation of a variable. Further, it revealed if the variable was clustered, random, or dispersed.

The relationship between the twelve land use and locational variables and daily max Heat Index deviations was measured through linear correlation. Using deviations was important because this research is curious to understand what explains the differences between the stations, rather than the values at the stations. Overall, the values at the stations like temperature, humidity, and Heat Index are results of meteorological conditions, air masses, the location of Columbia on the globe, and daily variations in meteorological conditions. The differences between stations, when averaged through the summer, are more likely to be the result of land use rather than weather conditions. For example, an afternoon thunderstorm that passed through the north western portion of the county may create large 3pm Heat Index differences, which could be incorrectly attributed to land use. Averaging these deviations for the summer gives more confidence that the differences that exist are a result of land use variables and not daily weather patterns.

UHI literature uses land use variables to model land surface temperature, often with multiple linear regression. After getting an understanding of the correlations that each land use variable had with the weather deviations, the next goal was to test the ability for a combination of these variables to predict the deviations in weather variables. Theoretically, a well- performing set of land use variables could be the inputs for a high spatial resolution heat stress model for Columbia. Then, interpolation between points would be based on an accurate regression, rather than linear interpolation. Ideally, a network of stations collecting heat data at human height would be ideal for model creation. However, it is the lack of these desired stations that led UHI research to rely on land surface temperature, and vehicle-based measurements rather than a network of permanent high- density stations. RC Winds provides a high-density network of stations, just at higher heights than human experience. With this network, a multiple linear regression was made initially for afternoon Heat Index deviations, then for other variable deviations to understand what could be predicted. Using Ordinary Least Squares (OLS) Multiple Linear Regression, and Exploratory Regression in ArcGIS Pro, as well as OLS from the scikit- learn machine learning library using python.

The performance evaluation metrics derives from these regressions will give insight into the possibility of using the land use variables in this study for creating a 10-meter heat stress model for Columbia. A VIF (Variance Inflation Factor) was computed to assess multicollinearity between variables. The VIF quantifies how much the variance of a predictor is inflated by the existence of correlation among the predictor variables in the model (PSU 2023). Models with a VIF over four require investigation, while models with a VIF over ten indicate serious multicollinearity (Kang, personal communications

2023). Other model performance metrics include Adjusted R^2, which lowers the r squared with higher model complexity, the closer this value to 1, the better the model. The AICc is the corrected Akike Information Criteria which is an estimation of prediction error, so minimizing this value is optimal. The Koenker (BP) Statistic p- value assesses the consistency of relationships over space, a value under .05 suggests that there are inconsistent relationships. The residuals of a regression should be randomly distributed. In the case that the residuals are clustered, measured by Moran's I, spatial error model is an appropriate method to account for this. If the coefficients vary across space, then a Geographically Weighted Regression should be used. In addition to ArcGIS Pro model evaluation, the model will be created and evaluated using Scikit- learn, a popular machine learning library in python (Pedregosa et al 2011). Using both a Decision Tree and OLS (Ordinary Least Squares) regression, the average MAE (Mean Absolute Error), and RMSE (Root Mean Square Error) will be found for twelve cross- validation folds using 80% of the points to train, and testing on the remaining 20%. This gave further insight to the applicability for these land use variables to improve local heat forecasting.

Table 2.1 Heat Index Variables

Variable Name	**Calculation**	**Number of Data Points**	**Explanation**
Heat Index (HI)	27 stations* 92 days* 288 observations/day	715,392	Calculated from 5-minute observation temperature and relative humidity
Daily Max Heat Index (DMHI)	27 stations* 92 days	2484	Maximum Heat Index found for each day and each station
Daily Max Heat Index Average (DMHIA)	92 days	92	Across network daily average of all stations' Daily Max Heat Index
Δ Daily Max Heat Index (ΔDMHI)	27 stations* 92 days	2484	Calculated from (Daily Max Heat Index) - (Daily Max Heat Index Average) to get each station's daily departure from mean
Δ Daily Max Heat Index Average (ΔDMHIA)	27 stations	27	The mean of 92 days of Δ Daily Max Heat Index for each station
Δ Daily Max Heat Index Median (ΔDMHIM)	27 stations	27	The median of 92 days of Δ Daily Max Heat Index for each station

Table 2.2 Land Use and Location Variables

Variable (Unit)	**Source**
Elevation (m)	USGS National Map
Latitude (Decimal Degrees)	RC Winds, personal communication
Longitude (Decimal Degrees)	RC Winds, personal communication
NDVI 5km (N/A)	USGS Earth Explorer
NDVI 1km (N/A)	USGS Earth Explorer
NDVI 50m (N/A)	USGS Earth Explorer
Proximity to Urban Environment (m)	Calculated Field
Percent Impervious 5km (%)	USGS NLCD
Percent Impervious 1km (%)	USGS NLCD
Percent Impervious 50m (%)	USGS NLCD
Proximity to Water Body (m)	Calculated Field
Soil Available Water Supply (Inches)	USDA Web Soil Survey

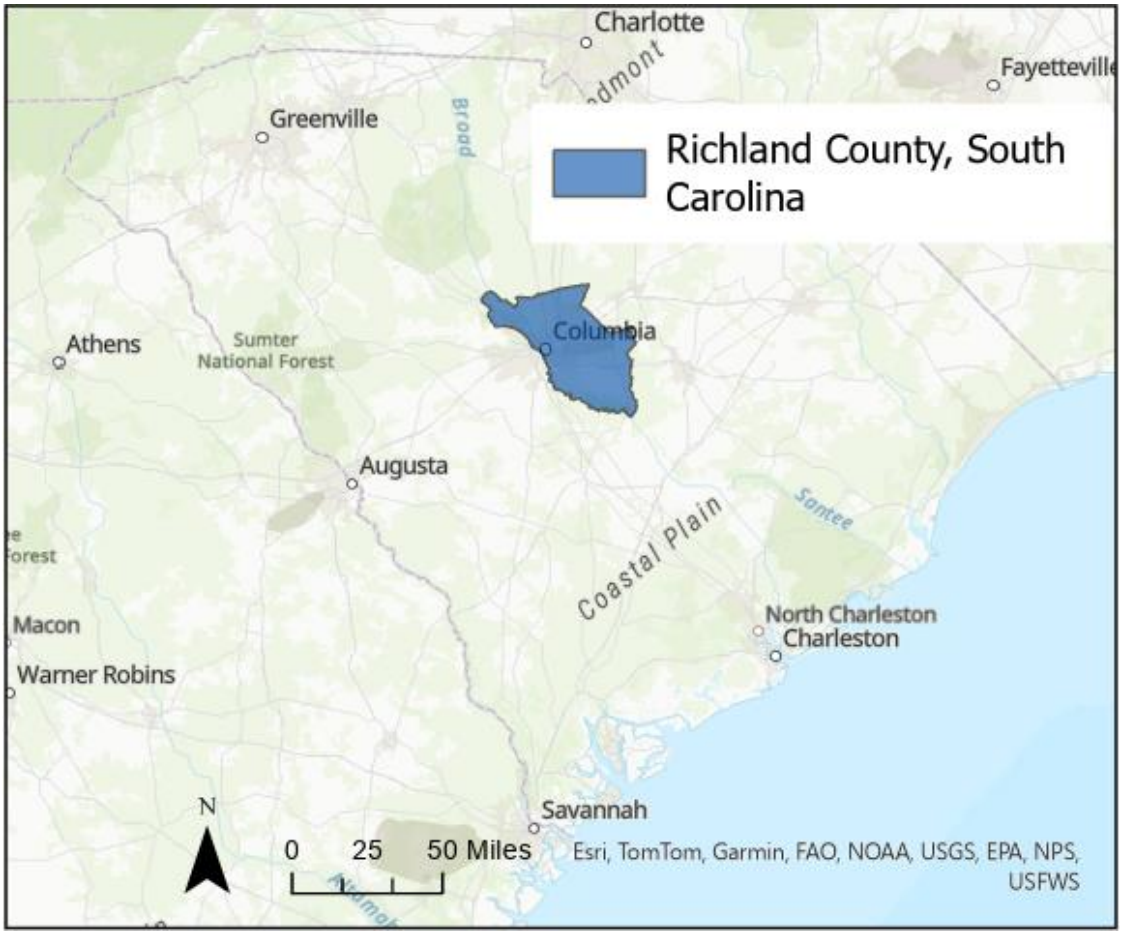

Figure 2.1 Richland County, South Carolina

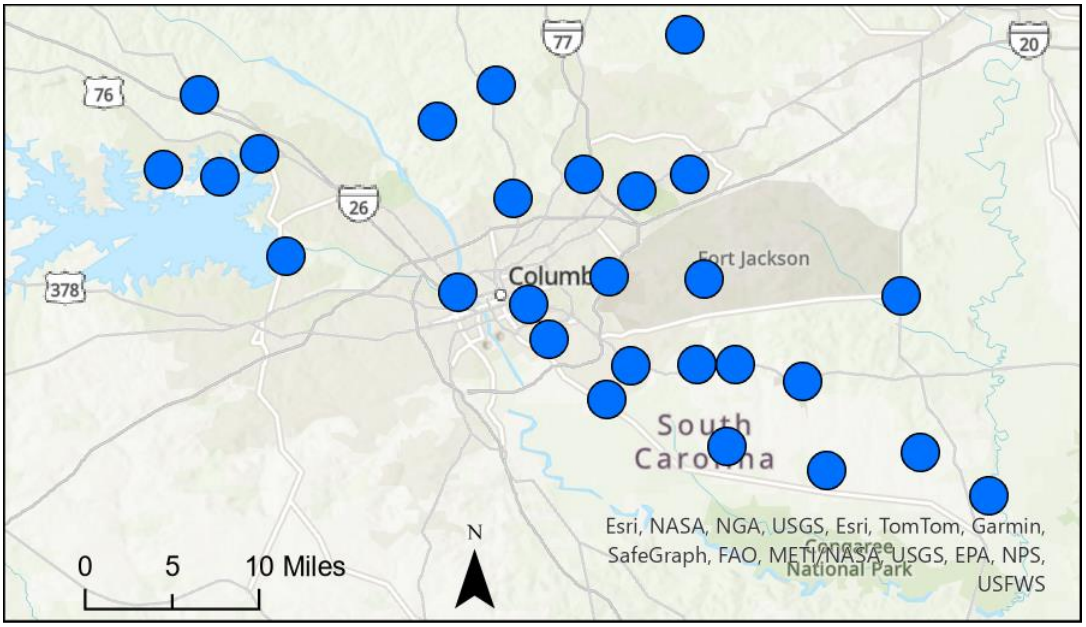

Figure 2.2 10-meter RC Winds Stations in Final Analysis.

Figure 2.3 Typical RC Winds Station, Left- Example RC Winds Station, Right- WeatherHawk weather station, from K. Aucoin, and Campbell Scientific.

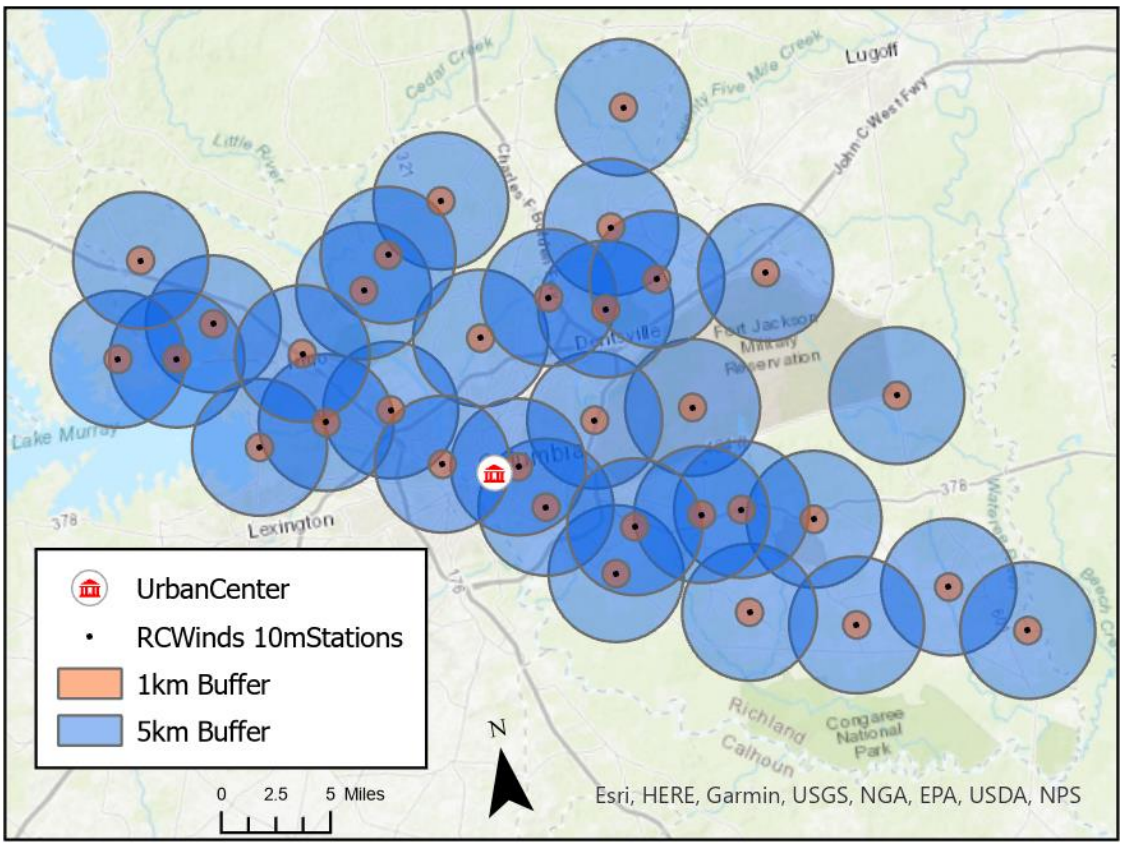

Figure 2.4 Buffer Visualization

CHAPTER 3

RESULTS

3.1 HEAT INDEX VARIABILITY

The first research question aimed to understand the deviations that existed across the network. Based on a distribution of the 3pm Heat Index deviations for all 5- minute time stamps, each station had time stamps where the value was higher than average, and each station had time stamps where the value was lower than average. Out of 27 stations, 14 had median values below 0, and 12 had median values above 0, leaving one station's median value at zero. Eleven stations' median value was outside a 2.5° deviation from zero. A visualization of the 5- minute Heat Index deviations during the 3pm hour is shown in Figure 3.1. For 3pm Heat Index, the median deviations ranged from -3.6°F to 4.5°F, a range of 8.1°F. This means that on an average day in the summer during the 3pm hour, one would expect a difference in Heat Index across the county of 8.1°F at 10 meters.

Using the 5-minute 3pm deviations, violin plots were made for all stations for temperature, relative humidity and heat index A sample of these violin plots are shown in Figure 3.2. These plots showed insight into the distributions of deviations for each station that supplemented the box plots. Several observations were made, which are discussed and hypothesized on here. First, the heat index and temperature deviation violin plots tended to have longer tails on the low (cooler) end, and shorter tails on the high (warmer)

end. Differential cloud cover and or rain patterns in the afternoon may allow for certain stations to cool off significantly compared to the network. Cold fronts and convective storms can create this rapid cooling for a set of stations. These reasons may explain the longer tails on the cooler side of the violins. While a cold front or localized rainfall may create sharp decreases in temperature at a station, it is much harder to quickly increase temperature and heat index at a given station, compared to the network. There are no frontal systems that can rapidly warm a station at 3pm, and although the rate of surface heating varies by land use, the effects of that dwindle with height, which may not be experienced at the 10m height. The violin plots for afternoon relative humidity deviations have longer tails on the high end, demonstrating that raising relative humidity compared to the network is more common than lowering it. This supports the observations of temperature plots, when temperature is quickly lowered by a front, cloud cover, rainfall or otherwise, the relative humidity will respond by increasing. Lake Murray station had the widest spread of afternoon deviations, especially for relative humidity and heat index. Being right on the eastern edge of Lake Murray, which has an east- west orientation, allows for different wind directions to significantly change the air qualities coming toward the station. Interestingly, this distribution is not seen for Coast Guard Island, which is in the northeast edge of Lake Murray. Fort Jackson had a uniquely bimodal relative humidity distribution, which is curious because the land use surrounding the station is consistent so there shouldn't be differences based on wind direction. This station is embedded within a large wooded area, so this bimodal humidity could reflect day-to day changes in vegetation and soil moisture, which the static independent variables would not capture. One final observation about the violin plots is that Owens

Field, a downtown airport station, had many outlier observation points for relative humidity on the high end. This was a curious finding, as the station is adjacent to a large area of low grass and asphalt, which would be expected to have higher temperatures and lower relative humidity.

Inverse Distance Weighting (IDW) was performed in ArcGIS Pro to visualize the interpolation of 3pm Heat Index deviations. The IDW creates a raster of data based on the bounding box of the input points, therefore the output is rectangular. As shown in Figure 3.3, hot spots are located in downtown Columbia, just north of Columbia, and in the southeast portion of the study area, in Eastover. Cool spots are located near Lake Murray, Fort Jackson, and Forest Acres, located just east of downtown Columbia.

In addition to spatial interpolation, Moran's I Spatial Autocorrelation assessment was completed for 3pm Heat Index deviations. The Moran's I index was .11, with a z-score of 1.19, and a p-value of .23, indicating that the pattern did not appear to be significantly different than random. With a positive z score however, the pattern skewed more likely to be clustered than to be dispersed.

A similar inspection into other variables was completed. Table 3.1 includes some descriptive statistics for the 6am median deviations, and Table 3.2 includes the same for 3pm median deviations. These tables were generated based on the having one median deviation for each station.

Moran's I Spatial Autocorrelation Index was computed for additional median deviation variables. The only significantly clustered variable was 3pm temperature deviations, as shown in the Moran's I results in Table 3.3. This variable was visualized

using IDW spatial interpolation, as shown by Figure 3.4. The major hot spot of the area was clearly downtown Columbia, but only with a range of temperature deviations of less than 6 degrees.

3.2 LAND USE AND VARIABILITY

The second research question is 'What variables may explain the differences in daily max Heat Index between stations around Columbia South Carolina?'. This question is aimed at understanding and quantifying which variables, and at which scale, are important for the possible creation of a high-resolution local heat model.

The twelve land use variables were tested for their strength and direction of correlation with 3pm Heat Index deviations. The only r-squared value above .1 was with Elevation (-.18). This implied a weak negative relationship between elevation and 3pm Heat Index deviation. All other land use variables displayed an r-squared of less than .06. The weak relationships that land use variables had with 3pm Heat Index was a surprising finding, given the literature on UHI suggesting the likelihood of afternoon heat to be more extreme in the urban areas. Additionally, the CAPA Heat Watch that took place in Columbia found afternoon differences of 18°F in the same study area. Granted, those observations were car- mounted and this study is analyzing 10-meter observations. Investigation into the components of Heat Index was necessary to explore a possible moderated heat stress due to lower urban humidity, similar to the one found in Chakraborty et al (2022). Afternoon temperature had little correlation to land use variables as well. The variables above an r-squared of .1 were Elevation, NDVI1km, NDVI50m, and %Imp5km. Elevation shows a negative weak correlation with 3pm temperature deviations. Two NDVI variables perform weakly but positive with 3pm

temperature, implying more vegetation means higher afternoon air temperatures. Although the correlation was weak, this was an unexpected result. Table 3.4 shows 3pm deviation r-squared values.

The afternoon deviations shown in Table 3.4 show mostly weak correlations. Wind was the only variable to have a single variable r-squared above .2. NDVI correlated negatively with wind deviations, implying that more vegetation is associated with lower afternoon wind. The mature height of the most common indigenous South Carolina trees is higher than 10 meters, so stations located near trees are likely below the treetops, decreasing wind (Coastal Expeditions 2021).

Even though the spatial interpolations shown in section 3.1 reveal hot spots in downtown Columbia, the correlation between land use variables and afternoon temperature and Heat Index are low, the highest of which is -.19 between temperature and Elevation. Additionally, the lack of correlation between afternoon dewpoint and relative humidity reveals no strong afternoon UMI or UDI. For the 3pm deviation correlations, the best performing variables in order were NDVI50m, NDVI1km, NDVI5km, Elevation, %Imperv5km, %Imperv1km, Proximity Water, Longitude, %Imperv50m, Proximity Urban, SAWS, and Latitude. The order of predictability among afternoon weather variables was wind, temperature, Heat Index, relative humidity, then dew point.

Investigating afternoon Heat Index was a starting point for this analysis, but going further into both the components of Heat Index, including other weather variables, as well as understanding diurnal patterns was necessary. The correlations found at 3pm were low, but more intriguing correlations were found for the 6am deviations. The two largest

NDVI buffers explained over 60% of the variance in 6am temperature deviations. Both had negative coefficients of determination, implying stations surrounded by more vegetation is associated with lower morning temperatures than network average. The 5km NDVI and 1km NDVI performed very similarly (-.63, -.62), but the 50-meter NDVI only explained half as much variance in 6am temperatures. This gives helpful insight as to what scale is important for understanding these differences, at least for 10-meter observations. The correlations for both NDVI5km and NDVI50m with 6am dew point deviations were relatively low at -.2 and -.24, respectively, but their direction implies a slight morning UMI effect. The mechanism that produces this is said to be due to higher rates of evapotranspiration among urban vegetation overnight due to higher temperatures (Huang and Song 2023).

Another notable finding was that the percent impervious variables performed weakly across the board, peaking at .28, and .25 for 1lm and 5km percent impervious surface, respectively. The direction of this relationship makes sense, as impervious surfaces are expected to retain and emit heat through the night, limiting nightly cooling. However, they explained less than half of their NDVI counterparts, demonstrating here that the presence of vegetation is more important for nighttime cooling than the absence of impervious surfaces. Upon initial consideration, NDVI and percent imperviousness may seem like opposite variables, but an area could have mostly impervious surfaces, with mature trees adding to the vegetation index. Additionally, an area could have 100% pervious surfaces but be covered by dead grass, dirt, or another low vegetation index cover. This is further demonstrated by their linear correlations, which show coefficients of determination between -.22 and -.42 between NDVI and percent impervious surface

variables compared at the same scale. Land use variables generally performed better with dew point deviations as well, and NDVI led the way again, but peaked at the 50- meter buffer this time.

For the 6am deviation correlations, the best performing variables in order were NDVI1km, NDVI5km, NDVI50m, %Imperv1km, Proximity Water, %Imperv5km, Longitude, Proximity Urban, %Imperv50m, SAWS, Elevation, then Latitude. The order of predictability among afternoon weather variables was temperature, wind, Heat Index, dew point, then relative humidity.

In summary, NDVI performed better than all other variables, including percent impervious surface. 6am weather deviations were much easier to predict from these land use variables than 3pm weather deviations, with coefficients of determination more than doubled in the morning than afternoon. Ranking the NDVI coefficients reveal that morning deviations are better predicted by smaller scale buffers of NDVI, rather than the larger scale buffers.

3.2 MODEL BUILDING

The third research question, ‘Is a statistical model developed from past network observations able to improve Heat Index predictions at a local scale?’ attempts to understand the possibility of using land use variables as the framework for a local heat model. Quantifying the explanatory power that a group of land use variables has on heat deviations, can help to understand the possibility of a local heat model. If the model created here explained much of the variance in the 3pm Heat Index, it could be used to create a high-resolution afternoon heat model for Columbia.

Using ArcGIS Pro's Exploratory Regression tool, one can select a dependent variable, candidate independent variables, and compare the model results of different combinations. Similar to previous research questions, the starting point for this analysis was the 3pm Heat Index deviations. The results of the 3pmHeatIndex exploratory regression indicate that the best performance comes from using Elevation alone, with an adjusted r squared of .15. Using two variables decreased the adjusted r squared to .13. Table 3.6 summarizes the results of the best model for each number of explanatory variables. Results indicate that a 3pm Heat Index model may not be significantly improved by the incorporation of land use variables at the 10- meter height.

Because the 6am temperature deviations had the best individual correlations with the land use variables, it was necessary to perform exploratory regression on it as well. Contrary to 3pm Heat Index, 6am temperature was able to be modeled accurately, with adjusted r squared values up to .71. The set of explanatory variables that reached this explanatory power were elevation, NDVI5km, NDVI50m, percent impervious 1km, and percent impervious 50m, all of which besides elevation were significant at the .05 level or better. The other metrics, like AICc, K(BP), and VIF were in acceptable ranges, meaning this model was successful. An assumption of a regression is that the residuals are randomly distributed. To test this, in ArcGIS Pro, the Generalized Linear Regression tool was run on 6am temperature with the five variables that produced the best result. Then, Moran's I spatial autocorrelation index was run on the residuals. The results of this test showed with a z-score of -.4, a p value of .65, and a Moran's Index of -.09, the pattern of residuals was randomly distributed. Figure 3.5 illustrates the five variable ArcGIS Pro regression model predicted versus observed 6am temperature deviation.

In addition to the 3pm Heat Index and 6am temperature, other morning and afternoon heat variables were put through exploratory regression as well. The model that maximized explanation power for 6am Heat Index, 6am dew point, 3pm temperature, and 3pm dew point resulted in .37, .32, .6, and .03 adjusted r squared values, respectively. Therefore, 3pm temperature was put into Generalized Linear Regression on ArcGIS Pro. The model that created the .6 adjusted r squared for 3pm temperature is shown in Table 3.6, and the residuals visualized in Figure 3.6. 6am and 3pm temperature were the only two variables that were able to output an adjusted r squared over .4, indicating temperature is better modeled than Heat Index.

These regression models were recreated using Python, to understand the performances further. Scikit- learn Ordinary Least Squares (OLS) regression and Decision Tree regression were used. Model performance was tested using twelve validation folds with an 80% training sample and 20% testing sample, and finding the average RMSE and MAE. The results of these outputs are shown in Table 3.7. In both the ArcGIS Generalized Linear Regression, and the Scikit- learn regressions, 3pm Heat Index was not able to be accurately and consistently modeled with the input variables given. The adjusted r squared was low (.15), showing weak explanatory power. Additionally, the RMSE and MAE were high (1.92 to 6.7), considering the range of values for this dependent variable was only 8.1°F. As for the 6am temperature regression model, the model accuracy is high, but the range of deviations for morning temperatures was only 4.1°F. Finally, for 3pm temperature, RMSE and MAE were both between the values for 3pm Heat Index and 6am temperature. The MAE for OLS regression for 3pm temperature was less than a degree Fahrenheit, and with a range of 5.5°F, this was

moderately successful. The OLS performed better for each deviation variable than the Decision Tree regression. This is likely because of the size of the input data. Training the Decision Tree with 80% of 27 stations is not optimal to maximize accuracy. The Decision Tree results may have been improved with more stations, or using 5-minute deviations, or daily medians rather than summer median deviations.

The results of this section indicate that Heat Index is not well modeled by the selected land use variables for this study, at 10 meters. However, morning and afternoon temperature were both found to be successfully modeled by the chosen land use variables. Dew point, relative humidity, nor Heat Index were successfully modeled, meaning that UMI or UDI patterns were not shown here.

Table 3.1 6am Weather Variable Deviation Summary Statistics

	6amRH_Dev %	**6amTemp_Dev °F**	**6amHI_Dev °F**	**6amDewP_Dev °F**
Mean	.18	-.08	.14	.02
Std dev.	4.77	1.04	2.30	1.73
Min	-15.00	-1.50	-3.70	-4.60
25%	.05	-.95	-2.10	.05
50%	1.60	0.00	-.30	.40
75%	2.95	.60	1.30	1.20
Max	5.70	2.60	5.00	1.70
Range	20.7	4.1	8.7	6.3

Table 3.2 3pm Weather Variable Deviation Summary Statistics

	3pmRH_Dev %	**3pmTemp_Dev °F**	**3pm HI_Dev °F**	**3pmDewP_Dev °F**
Mean	-.66	.13	-.15	-.15
Std dev.	5.30	1.25	2.44	2.80
Min	-13.20	-2.70	-3.60	-7.10
25%	-1.70	-.55	-2.50	-.65
50%	.20	.00	-.20	.70
75%	2.20	.95	1.3	1.35
Max	9.30	2.80	4.50	3.40
Range	22.5	5.5	8.1	10.5

Table 3.3 Moran's I Results

	Moran's I	**Z-score**	**p-value**	**Conclusion**
6amTemp_Dev	.14	1.38	.07	Random.
6amDewP_Dev	-.14	-.88	.38	Random.
6amHI_Dev	.11	1.21	.23	Random.
3pmTemp_Dev	.27	2.5	.01	Clustered, 95% confidence.
3pmDewP_Dev	-.13	-.73	.46	Random.
3pmHI_Dev	.11	1.19	.23	Random.

Table 3.4 3pm R-squared Correlation Matrix

	3RH_Dev	3Wnd_Dev	3Temp_Dev	3HI_Dev	3DewP_Dev
Latitude	0.00	0.02	-0.03	0.04	0.00
Longitude	0.01	-0.11	0.02	0.04	0.01
Elevation	0.00	0.01	-0.19	-0.18	-0.01
NDVI5km	-0.05	-0.42	0.06	0.00	-0.03
NDVI1km	-0.08	-0.39	0.16	0.02	-0.02
NDVI50m	-0.08	-0.57	0.12	0.01	-0.04
%Imp5km	-0.01	0.09	0.11	0.05	0.00
%Imp1km	-0.01	0.11	0.06	0.03	0.00
%Imp50m	0.00	0.06	0.05	0.03	0.01
SAWS	0.00	0.00	-0.07	-0.03	0.00
ProxWater	0.00	-0.13	0.02	0.03	0.00
ProxUrban	0.00	-0.02	-0.06	-0.03	0.00

Table 3.5 6am R-squared Correlation Matrix

	6RH_Dev	6Wnd_Dev	6Temp_Dev	6HI_Dev	6DewP_Dev
Latitude	0.00	0.04	0.01	0.00	0.00
Longitude	0.00	-0.14	-0.11	-0.05	-0.02
Elevation	0.00	0.08	0.02	0.01	0.01
NDVI5km	0.00	-0.28	-0.63	-0.20	-0.20
NDVI1km	-0.02	-0.42	-0.62	-0.27	-0.09
NDVI50m	-0.04	-0.23	-0.31	-0.04	-0.24
%Imp5km	0.00	0.00	0.25	0.09	0.04
%Imp1km	-0.01	0.00	0.28	0.12	0.03
%Imp50m	0.05	-0.02	0.07	0.00	0.11
SAWS	0.00	0.06	-0.05	-0.01	-0.02
ProxWater	0.00	-0.15	-0.15	-0.06	-0.04
ProxUrban	0.00	0.00	-0.17	-0.03	-0.06

Table 3.6 Exploratory Regression Results. Variable Significance (* = .10, ** = .05, *** = .01)

Variables	**Adj R^2**	**AICc**	**K(BP)**	**VIF**	**Model**
3pm Heat Index	.15	125.63	.05	1.00	-Elevation**
6am Temp	.71	58.42	.93	3.56	+Elevation -NDVI5km*** -NDVI50m** +PrcntImperv1km*** +PrcntImperv50m***
3pm Temp	.60	77.25	96	2.75	+Latitude -Elevation** +NDVI1km** +NDVI50m*** +PrcntImperv5km***

Table 3.7 Scikit- learn Regression Performance

Dependent Variable	**Method**	**RMSE average**	**MAE average**
6amTemperature	Decision Tree	.74	.94
6amTemperature	OLS	.48	.55
3pmHeatIndex	Decision Tree	6.7	2.32
3pmHeatIndex	OLS	2.16	1.92
3pmTemperature	Decision Tree	1.80	1.10
3pmTemperature	OLS	1.28	.81

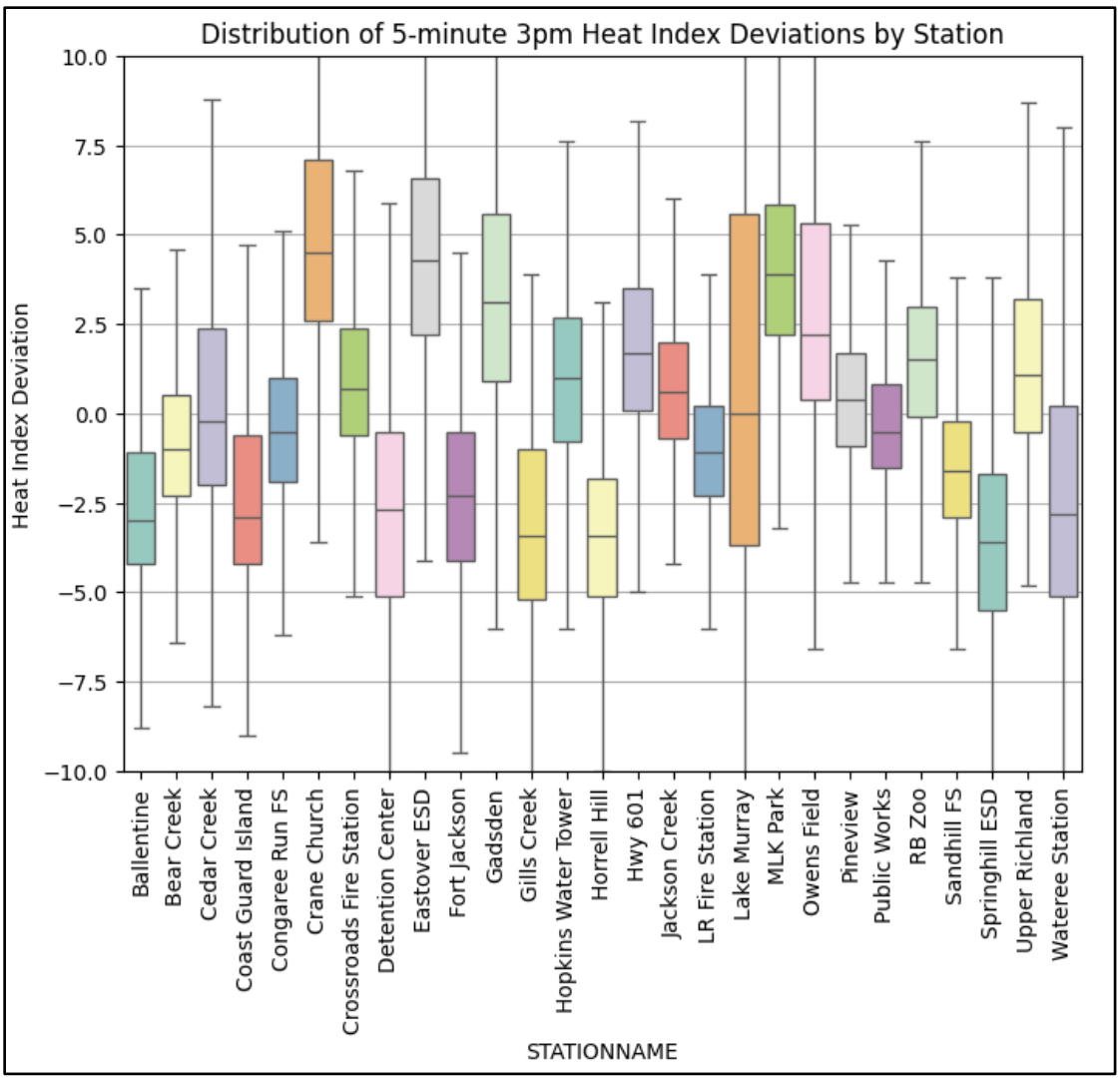

Figure 3.1 3pm Heat Index Deviation Box Plots

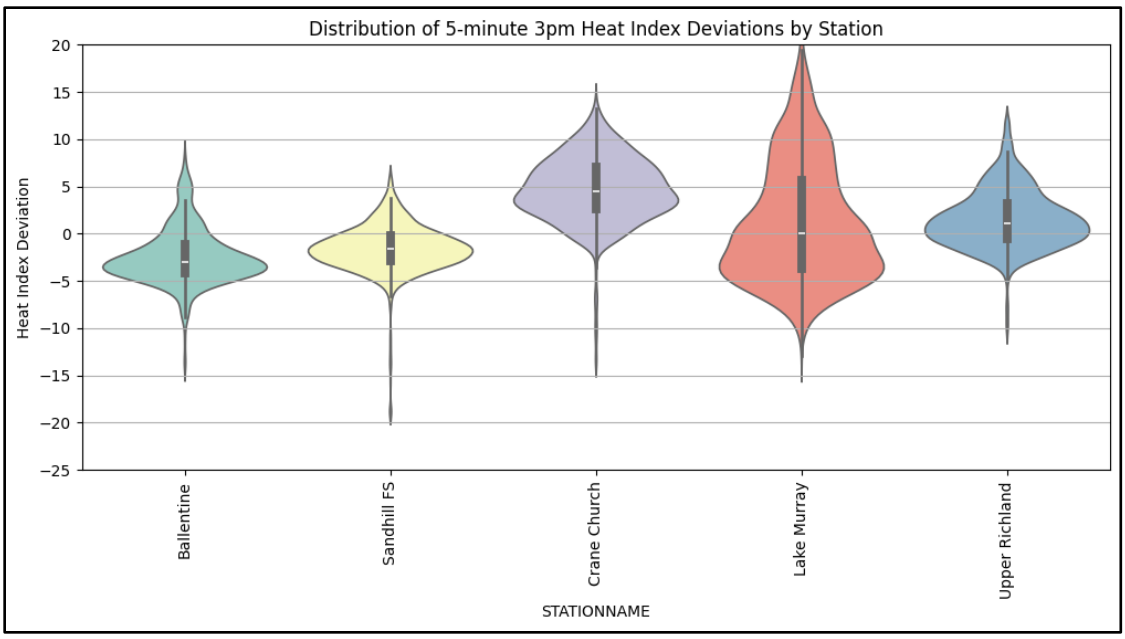

Figure 3.2 Sample of 3pm Heat Index Deviation Violin Plots

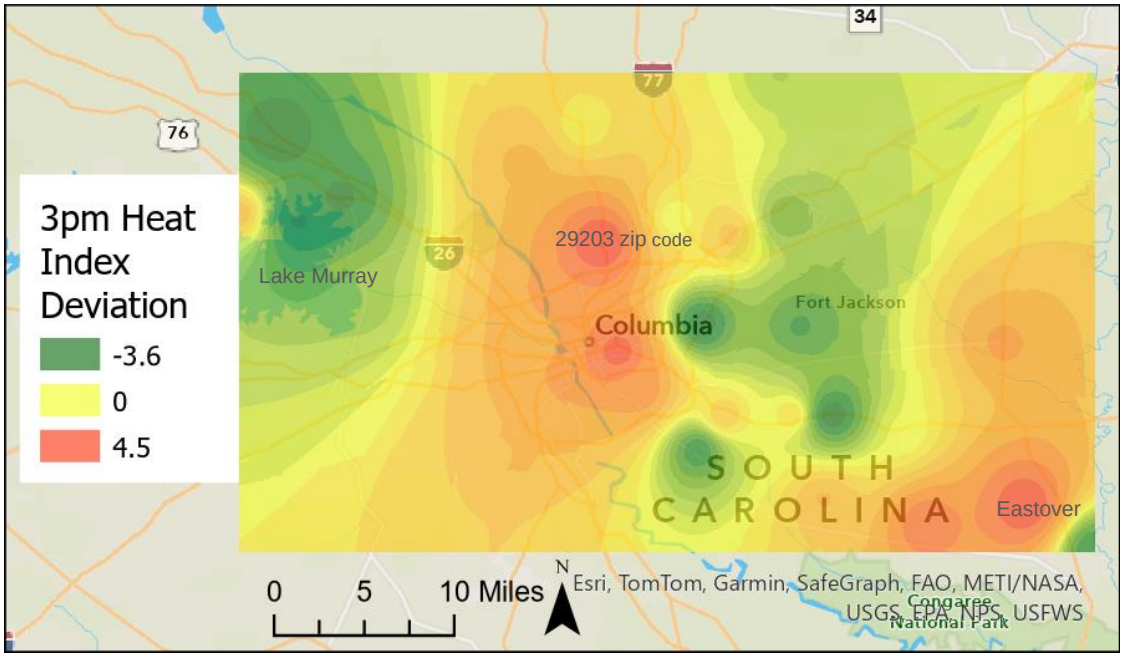

Figure 3.3 3pm Heat Index Interpolation

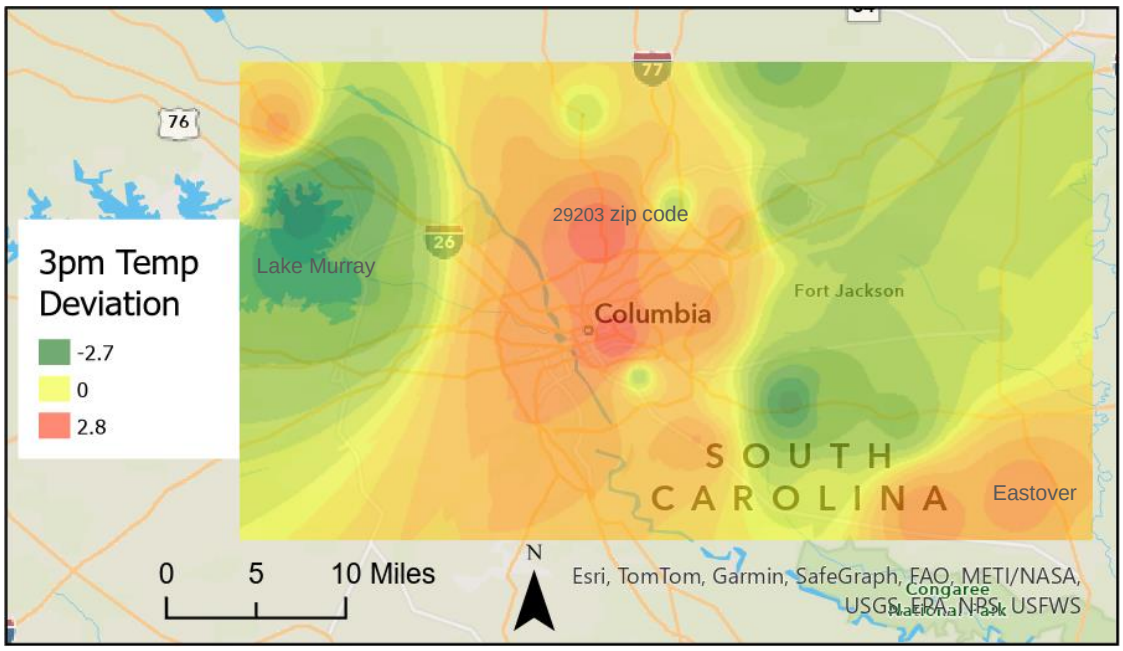

Figure 3.4 3pm Temperature Interpolation

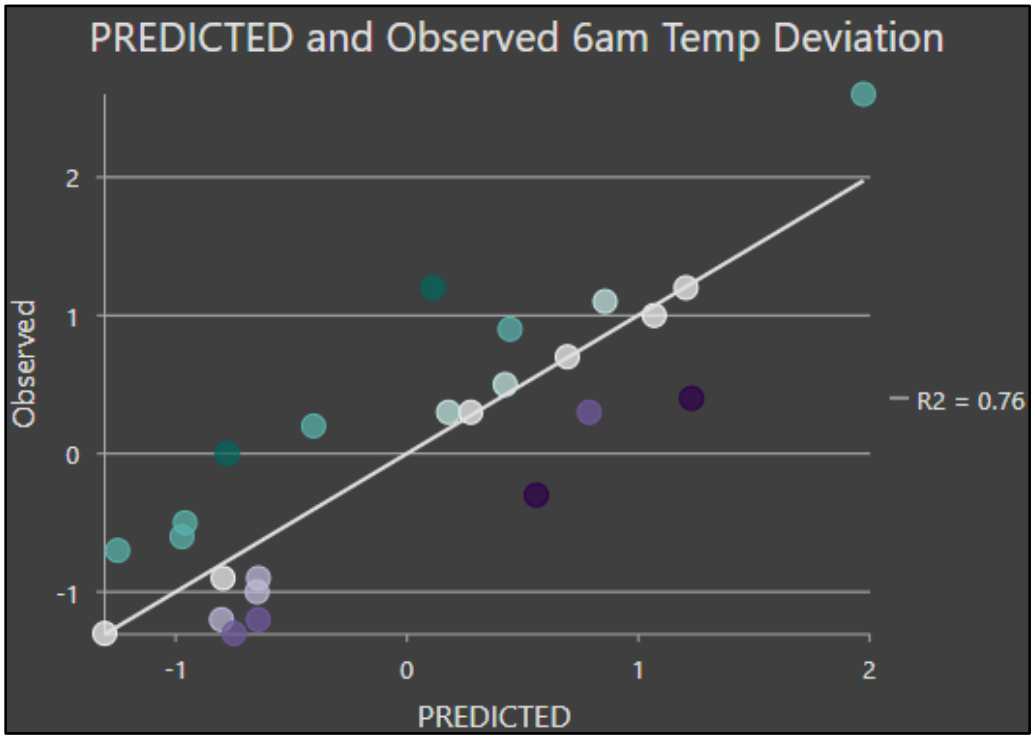

Figure 3.5 Predicted and Observed 6am Temperature Deviation

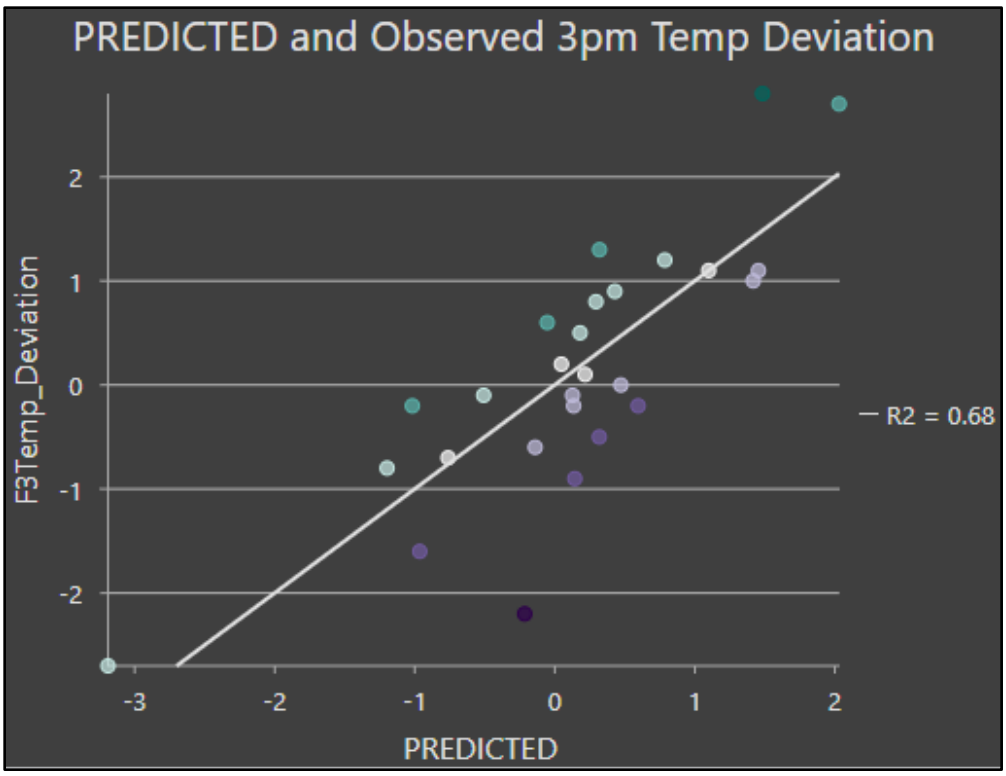

Figure 3.6 Predicted and Observed 3pm Temperature Deviation

CHAPTER 4
CONCLUSIONS

4.1 DISCUSSION

The most significant limitation to this study is the 10- meter height of weather stations, as this is not the height of human experience, and therefore cannot make conclusions on the patterns at other heights. The optimal urban heat stress study incorporates a high-density network of human height weather stations over varying land uses, with great spatial and temporal coverage. Unfortunately, such a network does not exist for most places, which is why land surface temperature is a popular alternative. Here, using the best available network, the understanding of local heat stress in Columbia, was improved because of the temporal or spatial coverage the RC Winds network provided. The number of stations, 27, was also not optimal for regression or statistical analysis.

Another limitation is the presumption of static independent variables. While proximity to water body will not change, other variables do change within the summer. For example, NDVI changes throughout the year, and if captured throughout the summer of 2022, may have yielded different results. The use of one image to create the NDVI raster was done because of the availability of high quality remotely sensed images of the study area, as well as a general assumption that the NDVI values would not change so significantly through the summer to be able to alter the general conclusions of this thesis. Additionally, the percent impervious surface was created from a national dataset from a

previous year, and some places within the study area may have changed land uses between the impervious surface data capture and the weather data capture.

Analyzing summer averages allowed for the understanding of general climate forecasting in Richland County. The choice to use static land use variables corresponded with the temporal scale of the summer deviations. Future research would benefit from the inclusion of dynamic variables. For example, capturing weekly NDVI could improve the accuracy of the model as vegetation health varies with rainfall. These dynamic variables could help understand the patterns found in the violin plots in section 3.1. Another example would be to include a variable for antecedent moisture, instead of the static SAWS. This could help understand dynamic differences in soil moisture and therefore humidity and latent heat. Controlling for dynamic weather variables like cloud cover, rainfall, and wind would also contribute to creating better daily weather forecasts. For example, a calm wind sunny high- pressure system day would likely create higher variance in air temperature across the county, whereas rainy windy days likely limit the variance. The limited explanation of land use on afternoon heat is hypothesized to be afternoon wind, shown in Figure 4.1. Understanding these finer scale dynamic variances would benefit forecasters to make accurate predictions and warnings.

4.2 CONCLUSIONS

Heat is the leading cause of weather-related fatalities in the United States (National Oceanic and Atmospheric Administration (NOAA) 2023). As global temperatures continue to rise, urban populations grow, and heat waves grow stronger and more frequent, excessive heat exposure will rise (Kirkpatrick & Lewis 2020, Habeeb et al 2015, Perkins et al 2012, Meehl & Tebaldi 2004, Baldwin et al 2019). Urban areas have

higher temperatures than their surrounding rural and suburban counterparts, creating the urban heat island (Oke 1980). Researchers have modeled the urban heat island, but due to a limited number of weather stations in urban areas, they commonly use remotely sensed land surface temperature. This allows for high resolution modeling, but the fundamental limitation of this method is that surface temperature is not directly comparable, and is a poor proxy for air temperature (Zhou et al 2019, Chakraborty et al 2022). Furthermore, remotely sensed land surface temperature does not give insight into heat stress variables like the Heat Index, or Wet Bulb globe temperature, or patterns of urban humidity.

This thesis utilized a unique network of weather stations in Richland County, South Carolina to understand the geographic patterns of heat stress, using the Heat Index. With 27 stations at 10-meters surrounding the city of Columbia, South Carolina, this network allowed for the investigation of the urban heat island through air temperature and humidity. The overarching goal for this thesis was to understand the plausibility of land use variables that are available at high spatial resolutions to model heat stress. The necessity for heat stress to be modeled at a high resolution assumes heterogeneity in the urban environment. If temperature, humidity, and therefore Heat Index were found to be homogenous across the study area, then modeling these variables at a higher resolution than a city would be trivial. Therefore, the first research question aimed to understand what differences in Heat Index did exist across the network.

The results showed that on an average summer day, during the hottest hour, 3pm, one would expect Heat Index to vary across the county by 8.1°F. Heat Index is the official heat warning tool of the National Weather Service, and this 8.1°F difference could mean the difference in a warning or no warning, or two different risk categories in

the county. These differences were found at 10-meters, so it is likely that differences are even greater for human- height measurements. That is exactly what the Heat Watch in Columbia in 2022 found, with afternoon differences over 18°F air temperature across the city, measured at human height (CAPA Strategies 2022). When mapping the 3pm Heat Index deviations, a spatial interpolation showed hot spots in downtown Columbia, as well as the more rural southeastern portion of the county. These warmer rural southeastern stations seemed to be more humid than the network average, which helps understand their higher Heat Index. Cool spots were located near Lake Murray, and near Fort Jackson.

To investigate deeper, Heat Index, temperature, relative humidity, and dew point deviations were all calculated for afternoon (3pm) and the morning (6am). Each station was given a median deviation for each variable, representing the summer average of how much that station tended to be higher or lower than the network average. From interpolating these variables on a map, generally an UHI effect is shown. Using Moran's I spatial autocorrelation index though, only 3pm temperature was found to be clustered. After the differences in temperature, Heat Index, dew point, and relative humidity were quantified and visualized in both the morning and afternoon, the next question aimed at understanding why these differences may exist.

From the latitude and longitude that came with each station's metadata, ten other land use and location variables were calculated for each station. These variables were all created with the possibility that they could explain the spatial pattern of the weather variable deviations. Individual variable correlations were created to understand which land use variables could explain spatial deviations the best. For morning and afternoon weather deviations, the three NDVI variables performed the best, followed by the percent

impervious surface variables. This implies that the density of vegetation is more important for heat mitigation than the absence of impervious surfaces, which is similar to the findings of Rosenzweig et al in their 2006 study on heat mitigation in New York City. Overall, the correlations between 3pm deviations and the land use variables were low, the highest combination being 3pm temperature and elevation with a coefficient of determination of -.19. The morning deviations correlated much better with the land use variables, as NDVI5km and NDVI1km each reached a coefficient of determination of over .6. Overall, morning weather deviations were much better correlated with land use variables than their afternoon counterparts.

Two typical diurnal weather patterns may explain the difference in performance between morning and afternoon models. Morning wind speeds are typically lower than afternoon winds, due to lower pressure gradient force. Typical summer mornings have little to no wind, which allows for more heterogeneous conditions to develop. Figure 4.1 shows the average wind speed by hour for this network, and clearly, overnight wind speeds are low, typically around .5 m/s overnight. 3pm winds are typically around 1.5 m/s, with large variance. Vegetated surfaces cool faster at night than impervious surfaces, and combining this with low winds, it makes sense why 6am temperatures would be better explained by land use variables than 3pm. In addition to horizontal wind, there is vertical mixing. Vertical mixing is the convection of air vertically, especially strong in the summer when the sun heats the surface and drives vertical temperature gradients. Although not measured by these stations, the vertical mixing on a typical summer day may limit the temperature and Heat Index differences that would have been expected to be predicted by land use. Especially at the 10- meter height, these two mechanisms are

likely the cause for the lack of explanation the land use variables have on the afternoon weather deviations.

Finally, the last research question addressed the possibility of using these land use variables to build a model for heat in Columbia. Exploratory regression in ArcGIS Pro was used to configure a set of best performing independent variables to maximize explanatory power for the dependent variable. The set of variables that performed the best for each dependent variable were assessed using python with scikit- learn OLS and Decision Tree regression. Twelve cross validation folds were run, using 80% of the data points to train the model, and 20% to test the performance. This was completed initially for 3pm Heat Index, then for all deviation variables. Afternoon Heat Index was not well modeled, with the best linear regression model reaching an adjusted r- squared value of .15, which was created by elevation alone. 6am temperature deviations however, were well modeled by land use variables, resulting in a .71 adjusted r- squared with reasonable RMSE and MAE values. This was not unexpected because the individual variable correlations were high with morning temperature deviations. Also unexpected was the fact that even though 3pm temperature deviations had low correlations with individual land use variables, the multiple linear regression model explained approximately 60% of the variance. These results indicate that temperature deviations are able to be modeled by the selected land use variables, whereas Heat Index, relative humidity, and dew point are not well modeled by the selected land use variables at 10 meters. This implies, for the 10-meter height, that there is an UHI, but only for temperature, not heat stress, and that there is no significant UMI or UDI.

Overall, this study aids the knowledge of local heat by highlighting differences in heat that exist across Richland County. By generating deviations through the summer, the regression results show that typical 6am temperature deviations are accurately modeled by land use variables, and 3pm temperature is fairly accurately modeled by land use variables. This thesis is a step towards understanding the capacity to model human heat stress from land use by utilizing observational data.

The RC Winds network allowed for the unique analysis in this study, and can be used for much more research on different natural hazards. Having a dense observational network, a micronet, eliminates the guesswork of forecasters, especially for important events, and during hazardous conditions. These micronets could be the data source for many more fine- scale hazard and climate studies, that are specific to the local area. These micronets allow local officials to use their real- time observations rather than meso-scale forecasts and observations. Some hazards, particularly thunderstorms and tornadoes are difficult to accurately predict and monitor with the low-density airport stations. With access to RC Winds, officials in Richland County would have real time access to all the variables that would help warn and prepare for these smaller scale hazards. An area for improvement, as it relates to human heat stress, is the incorporation of human height measurements on these micronets. This would allow for the optimal urban heat research, and potentially better forecasting.

The practical implications of this thesis will be providing insight into the heterogeneity of heat as a hazard in Richland County, South Carolina. It will complement the other ongoing work in the region. A combination of these works helps build evidence for future heat warnings and decision making in Columbia surrounding heat risk. These

decisions may include special warnings for known hot areas, warnings to public health officials about vulnerable groups. For example, when temperatures are higher in urban areas, this could have implications for unhoused populations, and those without access to reliable air conditioning. High overnight temperatures coupled with no air conditioning are a dangerous combination, especially if those without air conditioning are physically vulnerable. The principle of utilizing higher resolution local data for site- specific analysis can be applied to other hazards, too. Accessing the RC Winds weather network for rainfall, for example, could be a powerful tool for predicting flooding. Certain portions of the county throughout the year are likely to have gotten more or less rainfall, making them more or less likely to flood during incoming rain events.

This thesis demonstrates the capacity for local datasets to be used to understand the characteristics of hazards at a high resolution. Local data provides officials with evidence to support or reject their intuitions, and to limit the guesswork associated with hazard warnings and decision making. It matches the hazard data observation with the population the data is meant to benefit.

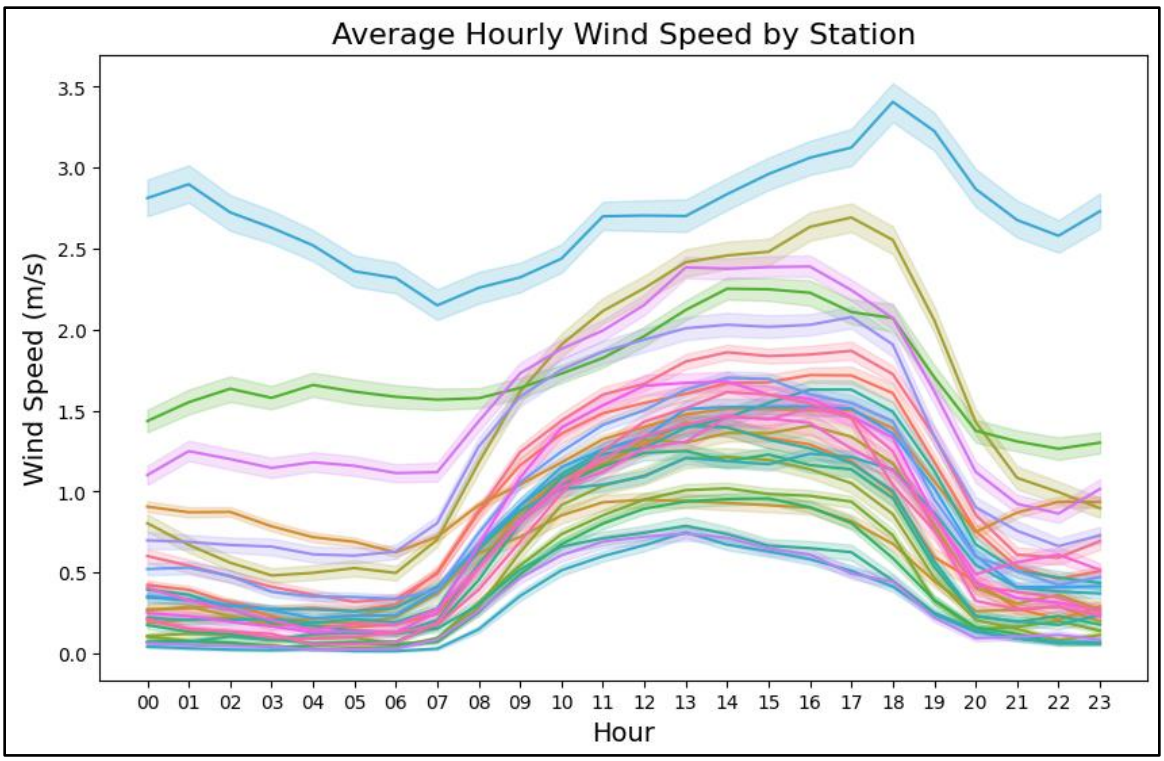

Figure 4.1 Average Hourly Wind Speed by Station

REFERENCES

"10.7 - Detecting Multicollinearity Using Variance Inflation Factors." *10.7 - Detecting Multicollinearity Using Variance Inflation Factors | STAT 462*, The Pennsylvania State University, online.stat.psu.edu/stat462/node/180/. Accessed 2 Nov. 2023.

Aucoin, K. (2024). *RC Winds Station Photos*.

Arnfield, A. John. "Two decades of urban climate research: A review of turbulence, exchanges of energy and water, and the urban heat island." *International Journal of Climatology*, vol. 23, no. 1, 2003, pp. 1–26, https://doi.org/10.1002/joc.859.

Baldwin, Jane Wilson, et al. "Temporally compound heat wave events and global warming: An emerging hazard." *Earth's Future,* vol. 7, no. 4, 2019, pp. 411–427, https://doi.org/10.1029/2018ef000989.

Barnett, Adrian, et al. "What measure of temperature is the best predictor of mortality?" *Epidemiology*, vol. 20, 2009, https://doi.org/10.1097/01.ede.0000362214.90491.53.

Budd, Grahame M. "Wet-bulb globe temperature (WBGT)—its history and its limitations." *Journal of Science and Medicine in Sport,* vol. 11, no. 1, 2008, pp. 20–32, https://doi.org/10.1016/j.jsams.2007.07.003.

Campbell Scientific. (n.d.). Weather Station for Home Weather, Education, Home Automation, Recreational, Environmental, and Industrial Applications. Edmonton, Alberta, Canada.

Campbell Scientific. (n.d.). WeatherHawk Signature Series Manual. Edmonton, Alberta, Canada.

CAPA Strategies. (2022, August). *NIHHIS Urban Heat Island Mapping Campaign Cities*. HEAT.gov - National Integrated Heat Health Information System. https://www.heat.gov/pages/nihhis-urban-heat-island-mapping-campaign-cities

CAPA Strategies. (2022). (rep.). *Home Forward Indoor Temperature Assessment*.

Chakraborty, T., et al. “Lower urban humidity moderates outdoor heat stress.” *AGU Advances*, vol. 3, no. 5, 2022, https://doi.org/10.1029/2022av000729.

Chen, Tianqi, et al. “Time-series analysis of heat waves and emergency department visits in Atlanta, 1993 to 2012.” *Environmental Health Perspectives*, vol. 125, no. 5, 2017, https://doi.org/10.1289/ehp44.

Clark, Jordan. “WET BULB GLOBE TEMPERATURE: IDENTIFYING RELATIONSHIPS WITH MORBIDITY AND DETERMINING THE IMPACT OF MICROCLIMATIC VARIABILITY ON ESTIMATES AND FORECASTS.” *University of North Carolina at Chapel Hill*, 2023.

“Climate Insights & Heat Zone Mapping.” *CAPA Strategies*, www.capastrategies.com/heat-watch. Accessed 2 Nov. 2023.

Coastal Expeditions. *Ten of the most common trees indigenous to South Carolina*. (2021, March 4). https://www.coastalexpeditions.com/blog/ten-of-the-most-common-trees-indigenous-to-south-carolina/

Coseo, Paul, and Larissa Larsen. “How factors of land use/land cover, building configuration, and adjacent heat sources and sinks explain urban heat islands in Chicago.” *Landscape and Urban Planning*, vol. 125, 2014, pp. 117–129, https://doi.org/10.1016/j.landurbplan.2014.02.019.

Dewitz, J., and U.S. Geological Survey, 2021, National Land Cover Database (NLCD) 2019 Products (ver. 2.0, June 2021): U.S. Geological Survey data release, https://doi.org/10.5066/P9KZCM54.

Dimiceli, Vincent E, et al. “Estimation of Black Globe Temperature for Calculation of the Wet Bulb Globe Temperature Index.” *Proceedings of the World Congress on Engineering and Computer Science*, vol. 2, Oct. 2011.

“Earthexplorer.” *EarthExplorer*, USGS - U.S. Geological Survey, earthexplorer.usgs.gov/. Accessed 2 Nov. 2023.

ESRI. (n.d.-a). *How IDW works*. How IDW works-ArcGIS Pro | Documentation. https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-idw-works.htm#:~:text=Inverse%20distance%20weighted%20(IDW)%20interpolation,of%20a%20locationally%20dependent%20variable.

Experience Columbia SC. (2010, December 10). *Columbia is famously hot and now registered*. Experience Columbia SC.

https://www.experiencecolumbiasc.com/articles/post/columbia-is-famously-hot-and-now-registered/

Fletcher, B. A., et al. “Association of summer temperatures with hospital admissions for renal diseases in New York State: A case-crossover study.” *American Journal of Epidemiology*, vol. 175, no. 9, 2012, pp. 907–916, https://doi.org/10.1093/aje/kwr417.

Fuhrmann, Christopher M., et al. “Impact of extreme heat events on emergency department visits in North Carolina (2007–2011).” *Journal of Community Health*, vol. 41, no. 1, 2015, pp. 146–156, https://doi.org/10.1007/s10900-015-0080-7.

“GHSA Practice Policy for Heat and Humidity.” *GHSA Practice Policy for Heat and Humidity | GHSA.Net*, www.ghsa.net/ghsa-practice-policy-heat-and-humidity. Accessed 2 Nov. 2023.

Goldie, James, Lisa Alexander, Sophie C. Lewis, and Steven Sherwood. “Comparative evaluation of human heat stress indices on selected hospital admissions in Sydney, Australia.” *Australian and New Zealand Journal of Public Health*, vol. 41, no. 4, 2017, pp. 381–387, https://doi.org/10.1111/1753-6405.12692.

Goldie, James, Lisa Alexander, Sophie C. Lewis, Steven C. Sherwood, et al. “Changes in relative fit of human heat stress indices to cardiovascular, respiratory, and renal hospitalizations across five Australian urban populations.” *International Journal of Biometeorology*, vol. 62, no. 3, 2017, pp. 423–432, https://doi.org/10.1007/s00484-017-1451-9.

Grundstein, Andrew J., et al. “A retrospective analysis of American football hyperthermia deaths in the United States.” *International Journal of Biometeorology*, vol. 56, no. 1, 2010, pp. 11–20, https://doi.org/10.1007/s00484-010-0391-4.

Grundstein, Andrew, and Earl Cooper. “Comparison of WBGTS over different surfaces within an athletic complex.” *Medicina*, vol. 56, no. 6, 2020, p. 313, https://doi.org/10.3390/medicina56060313.

Grundstein, Andrew, Earl Cooper, and Susan Yeargin. “Planning for heat: Time activity patterns, heat hazards, and exertional heat illnesses among American interscholastic football players.” *Theoretical and Applied Climatology*, 2023, https://doi.org/10.1007/s00704-023-04614-x.

Grundstein, Andrew, Earl R. Cooper, Olivia Cahill, et al. “Are local weather stations a feasible substitute for on-site measurements for Heat stress assessment in sports?”

International Journal of Sports Science & Coaching, 2022, p. 174795412211172, https://doi.org/10.1177/17479541221117240.

Grundstein, A., Williams, C., Phan, M., & Cooper, E. (2015). Regional heat safety thresholds for athletics in the contiguous United States. *Applied Geography, 56*, 55–60. https://doi.org/10.1016/j.apgeog.2014.10.014

Habeeb, Dana, et al. "Rising heat wave trends in large US cities." *Natural Hazards*, vol. 76, no. 3, 2015, pp. 1651–1665, https://doi.org/10.1007/s11069-014-1563-z.

"Heat - Heat Hazard Recognition." *Occupational Safety and Health Administration*, www.osha.gov/heat-exposure/hazards. Accessed 2 Nov. 2023.

"The Heat Index Equation." *Heat Index Equation*, National Weather Service Weather Prediction Center, www.wpc.ncep.noaa.gov/html/heatindex_equation.shtml. Accessed 2 Nov. 2023.

"How Spatial Autocorrelation (Global Moran's I) Works." *How Spatial Autocorrelation (Global Moran's I) Works-ArcGIS Pro | Documentation*, ESRI, pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm. Accessed 2 Nov. 2023.

Huang, Xinjie, and Jiyun Song. "Urban moisture and Dry Islands: Spatiotemporal variation patterns and mechanisms of urban air humidity changes across the globe." *Environmental Research Letters*, vol. 18, no. 10, 2023, p. 103003, https://doi.org/10.1088/1748-9326/acf7d7.

Imhoff, Marc L., et al. "Remote Sensing of the urban heat island effect across biomes in the continental USA." *Remote Sensing of Environment*, vol. 114, no. 3, 2010, pp. 504–513, https://doi.org/10.1016/j.rse.2009.10.008.

Jenerette, G. Darrel, et al. "Regional relationships between surface temperature, vegetation, and human settlement in a rapidly urbanizing ecosystem." *Landscape Ecology*, vol. 22, no. 3, 2006, pp. 353–365, https://doi.org/10.1007/s10980-006-9032-z.

Kang, Yuhao. "Lecture 8 Regression Analysis." GEOG 564 GIS-BASED MODELING. University of South Carolina, University of South Carolina.

Khatana, Sameed Ahmed, et al. "Association of extreme heat and cardiovascular mortality in the United States: A county-level longitudinal analysis from 2008 to 2017." *Circulation*, vol. 146, no. 3, 2022, pp. 249–261, https://doi.org/10.1161/circulationaha.122.060746.

"Landsat Normalized Difference Vegetation Index." *Landsat Normalized Difference Vegetation Index | U.S. Geological Survey*, United States Geological Survey, www.usgs.gov/landsat-missions/landsat-normalized-difference-vegetation-index. Accessed 2 Nov. 2023.

Li, Dan, and Elie Bou-Zeid. "Synergistic interactions between urban heat islands and heat waves: The impact in cities is larger than the sum of its parts." *Journal of Applied Meteorology and Climatology*, vol. 52, no. 9, 2013, pp. 2051–2064, https://doi.org/10.1175/jamc-d-13-02.1.

Lu, Yi-Chuan, and David M. Romps. "Extending the heat index." *Journal of Applied Meteorology and Climatology*, vol. 61, no. 10, 2022, pp. 1367–1383, https://doi.org/10.1175/jamc-d-22-0021.1.

McNoldy, B. (1995-2024). *Temperature, Dewpoint, and Relative Humidity Calculator*. Calculate Temperature, Dewpoint, or Relative Humidity. https://bmcnoldy.earth.miami.edu/Humidity.html

Meehl, Gerald A., and Claudia Tebaldi. "More intense, more frequent, and longer lasting heat waves in the 21st century." *Science*, vol. 305, no. 5686, 2004, pp. 994–997, https://doi.org/10.1126/science.1098704.

Mullin, Stafford. "An Analysis of Wet Bulb Globe Temperature Estimation Methods and Microclimate Heat Variability in Columbia, South Carolina." *University of South Carolina*, 2022.

NASA. (2022). *Arset - satellite remote sensing for measuring urban heat islands and constructing heat vulnerability indices*. NASA. https://appliedsciences.nasa.gov/get-involved/training/english/arset-satellite-remote-sensing-measuring-urban-heat-islands-and

Nakata-Osaki, Camila Mayumi, et al. "This – tool for heat island simulation: A GIS extension model to calculate urban heat island intensity based on urban geometry." *Computers, Environment and Urban Systems*, vol. 67, 2018, pp. 157–168, https://doi.org/10.1016/j.compenvurbsys.2017.09.007.

NIHHIS Urban Heat Island Mapping Campaign Cities. (2019-2022). Retrieved 2024.

NOAA/ NCEI. (n.d.). *U.S. Climate Normals*. NOAA NCEI U.S. Climate Normals Quick Access. https://www.ncei.noaa.gov/access/us-climate-normals/#dataset=normals-monthlyhttps://www.ncei.noaa.gov/access/us-climate-normals/#dataset=normals-

monthly&timeframe=15&location=SC&station=USW00013883&timeframe=15&location=SC&station=USW00013883

"NWS HeatRisk Prototype." *NWS Heatrisk*, www.wrh.noaa.gov/wrh/heatrisk/. Accessed 2 Nov. 2023.

Oke, T. R. "The energetic basis of the urban heat island." *Quarterly Journal of the Royal Meteorological Society*, vol. 108, no. 455, 1982, pp. 1–24, https://doi.org/10.1002/qj.49710845502.

"Osha Technical Manual (OTM) - Section III: Chapter 4." *Occupational Safety and Health Administration*, www.osha.gov/otm/section-3-health-hazards/chapter-4#heat_hazardassessment. Accessed 2 Nov. 2023.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(2011), 2825–2830.

Perkins, S. E., et al. "Increasing frequency, intensity and duration of observed global heatwaves and warm spells." *Geophysical Research Letters*, vol. 39, no. 20, 2012, https://doi.org/10.1029/2012gl053361.

Perkins-Kirkpatrick, S. E., and S. C. Lewis. "Increasing trends in regional heatwaves." *Nature Communications*, vol. 11, no. 1, 2020, https://doi.org/10.1038/s41467-020-16970-7.

Pryor, J. Luke, et al. "The heat strain of various athletic surfaces: A comparison between observed and modeled wet-bulb globe temperatures." *Journal of Athletic Training*, vol. 52, no. 11, 2017, pp. 1056–1064, https://doi.org/10.4085/1062-6050-52.11.15.

RAMANATHAN, N. L., and H. S. BELDING. "Physiological evaluation of the WBGT index for Occupational Heat Stress." *American Industrial Hygiene Association Journal*, vol. 34, no. 9, 1973, pp. 375–383, https://doi.org/10.1080/0002889738506866.

"RC Winds." *Richland County Weather Information Network Data System > Home*, rcwinds.com/#. Accessed 2 Nov. 2023.

"Richland County Showcases Unique Weather Monitoring Network – RC Winds." *Richland County*, 11 May 2015, www.richlandcountysc.gov/Home/News/ArtMID/479/ArticleID/680/Richland-

County-Showcases-Unique-Weather-Monitoring-Network-%E2%80%93-RC-WINDS.

Rosenzweig, Cynthia, et al. *MITIGATING NEW YORK CITY'S HEAT ISLAND WITH URBAN FORESTRY, LIVING ROOFS, AND LIGHT SURFACES*.

Roth, Matthias. "Review of Urban Climate Research in (sub)tropical regions." *International Journal of Climatology*, vol. 27, no. 14, 2007, pp. 1859–1873, https://doi.org/10.1002/joc.1591.

Rothfusz, Lans P. *The Heat Index "Equation" (or, More Than You Ever Wanted to Know About Heat Index)*, NWS Southern Region Headquarters, Fort Worth, TX, www.weather.gov/media/ffc/ta_htindx.PDF. Accessed 2 Nov. 2023.

"Soil Quality Indicators." *Soil Quality Physical Indicator Information Sheet Series*, USDA Natural Resources Conservation Service, June 2008.

South Carolina Department of Natural Resources. (n.d.). *South Carolina Lakes and Waterways*. SC Lakes and Waterways - Lake Murray. https://www.dnr.sc.gov/lakes/murray/description.html

South Carolina High School League, "Wet Bulb Globe Temperature Monitoring." *South Carolina High School League*, 20 June 2021, schsl.org/archives/2285.

Steadman, R. G. "The assessment of sultriness. part I: A temperature-humidity index based on human physiology and Clothing Science." *Journal of Applied Meteorology*, vol. 18, no. 7, 1979, pp. 861–873, https://doi.org/10.1175/1520-0450(1979)018<0861:taospi>2.0.co;2.

Stull, Roland. "Wet-bulb temperature from relative humidity and air temperature." *Journal of Applied Meteorology and Climatology*, vol. 50, no. 11, 2011, pp. 2267–2269, https://doi.org/10.1175/jamc-d-11-0143.1.

Tran, Duy X., et al. "Characterizing the relationship between land use land cover change and land surface temperature." *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 124, 2017, pp. 119–132, https://doi.org/10.1016/j.isprsjprs.2017.01.001.

Tripp, Brady, et al. "Comparison of wet bulb globe temperature measured on-site VS estimated and the impact on activity modification in high school football." *International Journal of Biometeorology*, vol. 64, no. 4, 2019, pp. 593–600, https://doi.org/10.1007/s00484-019-01847-2.

Tripp, Brady, et al. “Comparison of wet bulb globe temperature measured on-site VS estimated and the impact on activity modification in high school football.” *International Journal of Biometeorology*, vol. 64, no. 4, 2019, pp. 593–600, https://doi.org/10.1007/s00484-019-01847-2.

US Census Bureau. (n.d.). *U.S. Census Bureau quickfacts: Columbia City, South Carolina*. U.S. Census Bureau QuickFacts. https://www.census.gov/quickfacts/fact/table/columbiacitysouthcarolina/PST045222

US Department of Commerce, NOAA. “Heat Forecast Tools.” *National Weather Service*, NOAA’s National Weather Service, 8 Sept. 2023, www.weather.gov/safety/heat-index.

US Department of Commerce, NOAA. (n.d.). *NOAA Online Weather Data*. Climate. https://www.weather.gov/wrh/climate?wfo=cae

US Department of Commerce, NOAA. “Weather Related Fatality and Injury Statistics.” *National Weather Service*, NOAA’s National Weather Service, www.weather.gov/hazstat/. Accessed 2 Nov. 2023.

“USGS National Map.” *TNM Download V2*, apps.nationalmap.gov/downloader/. Accessed 2 Nov. 2023.

Vaidyanathan, Ambarish, et al. “Assessment of extreme heat and hospitalizations to inform early warning systems.” *Proceedings of the National Academy of Sciences*, vol. 116, no. 12, 2019, pp. 5420–5427, https://doi.org/10.1073/pnas.1806393116.

Voelkel, Jackson, et al. “Developing high-resolution descriptions of urban heat islands: A public health imperative.” *Preventing Chronic Disease*, vol. 13, 2016, https://doi.org/10.5888/pcd13.160099.

Voogt, J.A, and T.R Oke. “Thermal remote sensing of urban climates.” *Remote Sensing of Environment*, vol. 86, no. 3, 2003, pp. 370–384, https://doi.org/10.1016/s0034-4257(03)00079-8.

“Web Soil Survey.” *Web Soil Survey*, United States Department of Agriculture Natural Resources Conservation Service, websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx. Accessed 2 Nov. 2023.

YAGLOU, CONSTANTIN P., and DAVID MINARD. “Prevention of heat casualties at Marine Corps Training Centers.” *Archives of Industrial Health*, vol. 16, 1957, pp. 302–316, https://doi.org/10.21236/ad0099920.

Zhou, D., Xiao, J., Bonafoni, S., Berger, C., Deilami, K., Zhou, Y., Frolking, S., Yao, R., Qiao, Z., & Sobrino, J. (2018). *Remote Sensing*, *11*(1), 48. https://doi.org/10.3390/rs11010048

APPENDIX A

OUTLIER STATION REMOVAL PROSESS

After initial deviation plotting, further investigation of outlier stations was warranted. When plotting individual days throughout the entire summer, week segments, and longer stretches of time, a similar pattern became apparent in these stations. A particularly hot period from July 26th to July 30th 2022 was chosen to demonstrate the pattern. As shown in Figure A.1, the Heat Index for this period for the six outlier stations varied in daily peaks by about 70°F. The three low outliers are grouped closely for all four days and nights, and the three high outliers are well above the low grouping, but spread out. Next, air temperature was plotted for the same set of days and stations and as shown in Figure A.2, all six outlier stations had remarkably similar data pertaining to air temperature, with peak variations of single digit degrees Fahrenheit. So then, if temperature was grouped closely but Heat Index was not, then relative humidity is the variable creating the outliers. Plotting the same four days and six stations, revealed the relative humidity values were incredibly variant. Figure A.3 shows that two of the stations' relative humidity never broke 20%, and other stations' values often got 'stuck' at or near 100% relative humidity. Data patterns of this nature are indicative of instrument error. This was especially true for afternoon time stamps when the temperature measured over 90° Fahrenheit with the relative humidity staying over 90%, conditions that are physically unrealistic. Between the excessive differences in humidity between these stations, as well as these stations' variance from the ΔDMHIA, and the

excessive range shown in Figure A.1 of the ΔDMHI variable, these six stations were removed from the study.

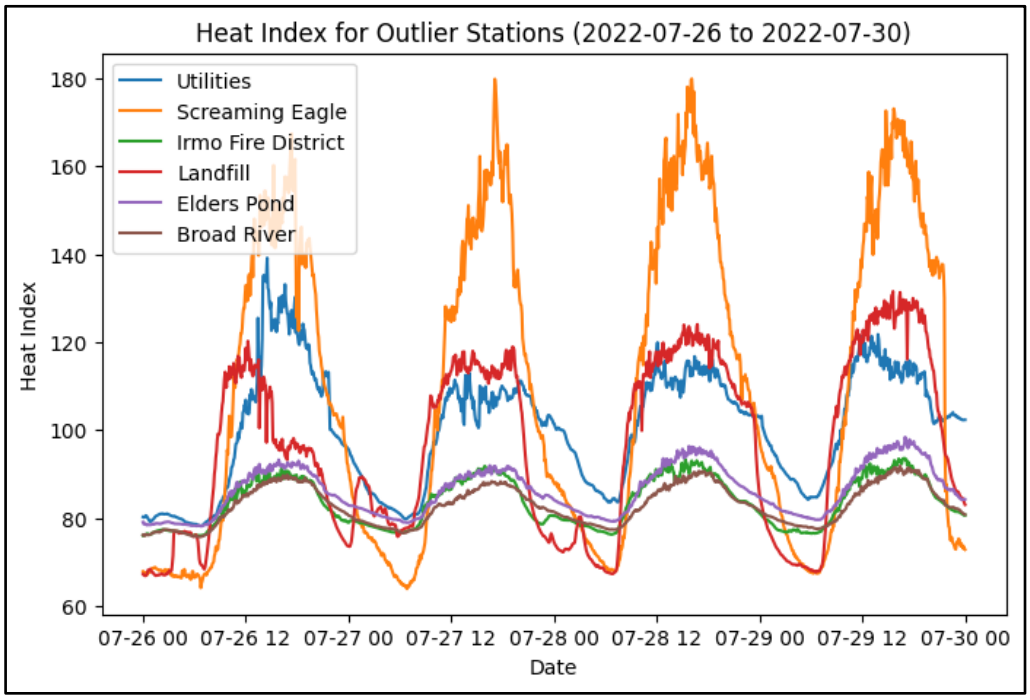

Figure A.1 Outlier Station Heat Index Time Series

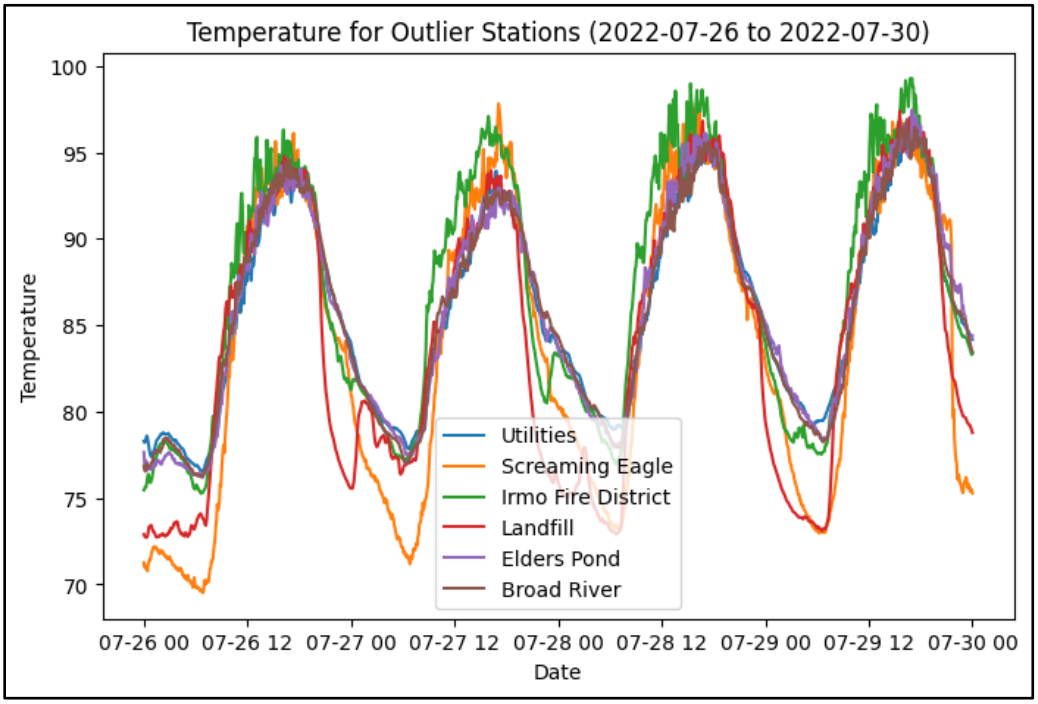

Figure A.2 Outlier Station Temperature Time Series

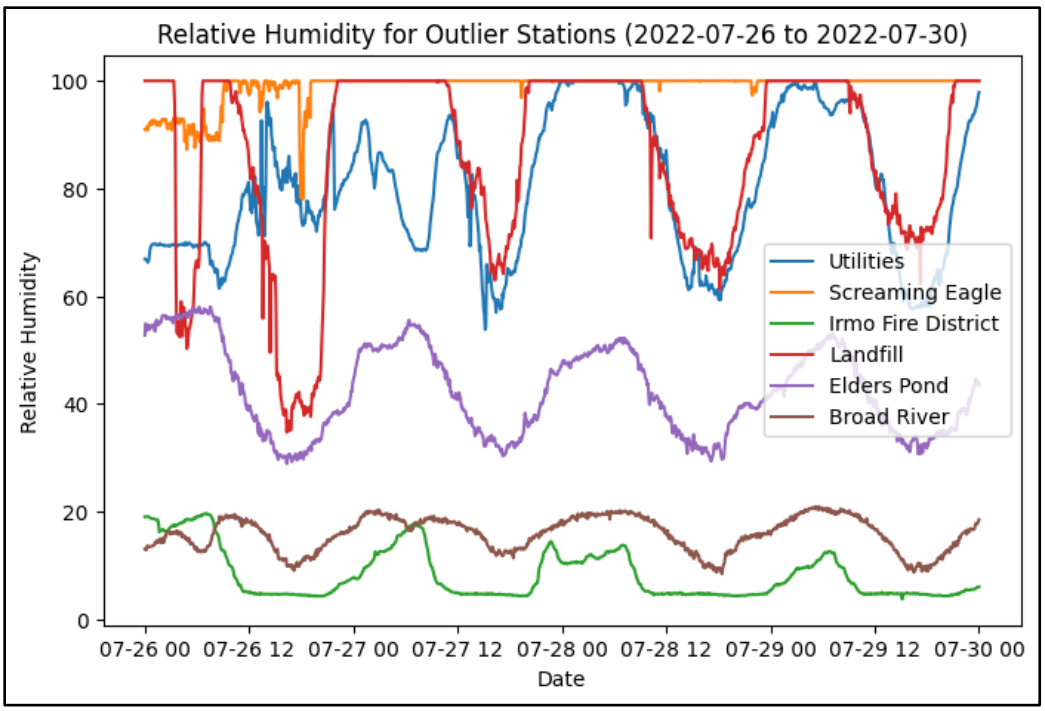

Figure A.3 Outlier Station Relative Humidity Time Series