A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Geography.

Chapel Hill
2023

Approved by:

Charles E. Konrad
Erika Wise
Conghe Song
Andrew Grundstein
Sandra Rayne
To my family and friends.
ABSTRACT


Heat is the leading cause of weather-related fatalities and is a grave concern for occupational health and athletic safety. Wet bulb globe temperature (WBGT) is a heat stress index that is becoming the gold standard for assessing heat stress as it accounts for the effects of air temperature, humidity, wind speed, and radiation on human body temperature. However, limited research has investigated the utility of WBGT in predicting morbidity at a population-level. Furthermore, WBGT varies widely across small distances and there has been limited validation of recent efforts to provide weather forecasts of WBGT.

In this dissertation, emergency department visits from North Carolina (2007-2016) are used to assess relationships between heat and morbidity. Microclimatic variations in WBGT across small distances (10s to 100s of meters) are quantified using WBGT data collected from high school campuses and suburban environments during the summers of 2019-2021. With this data, the accuracy of a WBGT forecast tool (developed by NOAA’s Southeast Regional Climate Center and Carolinas Integrated Sciences and Assessments) is evaluated and compared to a forecasting method operationalized in 2021 by the US National Weather Service.

This work revealed WBGT to be a more robust predictor of morbidity than other metrics, such as the Heat Index, and supports previous research on the higher heat-health burden in rural areas relative to urban areas in North Carolina. Large variations in WBGT were found across a
variety of surfaces and weather conditions. These variations on warm sunny days often result in different levels of danger across the same high school campus, such as between tennis courts and adjacent grassy fields. Lastly, the WBGT forecast was found to be within 1.5°F of observed WBGT on average and more accurate than other methods which had notable cold biases. Furthermore, incorporating high-resolution, sub-forecast grid level data to account for the influence of land cover on wind speed improved the forecast, particularly for more sheltered areas. These findings support the use of WBGT in heat-health warning systems, inform considerations when measuring the index, and demonstrate that WBGT can be accurately forecast to enable early warning of particularly hazardous periods.
ACKNOWLEDGEMENTS

I am extremely grateful to my advisor, Charles Konrad, for the support and guidance he has provided throughout my time as a graduate student. His mentorship and friendship have been invaluable, contributing greatly to my scholarship and professional development. I am extremely grateful to my other committee members. Dr. Erika Wise and Dr. Conghe Song provided inspiration which began when I was an undergrad taking their courses, continuing to foster my love for science and geography. Dr. Andrew Grundstein was integral to my work, sharing knowledge about WBGT as I began my research and allowing me to use his WBGT meter. Dr. Sandra Rayne also provided critical support and colleagueship that helped me persevere.

I would like to express my deepest appreciation to Carolinas Integrated Sciences and Assessments for funding. I am also thankful to the Department of Geography at Chapel Hill, as a student, both undergrad and graduate, and for travel funding. I would like to acknowledge the funding and experiences gained from the Weiss Urban Livability Fellowship.

Several external partners played an important role in my work. I am very appreciative of the North Carolina High School Athletic Association for opportunities to connect with high schools about WBGT and also the high schools that allowed me to do field work on their campuses: Cedar Ridge, Wake Forest, and Green Level High School. I would also like to acknowledge the US National Weather Service and Timothy Boyer for sharing information and insight into their efforts to forecast WBGT.

I am thankful for my fellow graduate students and friends, as our conversations, travels, and late nights and early mornings provided significant support, specifically Rachel Woodul, Manny
Hernandez, Sarah Schmidt, and Montana Eck. I would also like to extend my deepest gratitude to Dr. Ashley Ward for her mentorship which began when I was an undergraduate taking her classes and continues to this day. I am also extremely grateful for Dr. Rachel Willis who provided wisdom and encouragement, contributing significantly to my professional development. Lastly, I want to acknowledge other colleagues, William Schmitz and Jordan Mcleod.

I am forever indebted to the support of my family. My mom, Tracy, and father, Randy, gifted me with an incredibly supportive childhood which continues to this day, fostering my fascination with weather from a young age. My brother, Jacob, has and will always be a best friend, as well as my cousins that are truly siblings: Adam, Kristin, Josh, Garrett, Cameron, Lauren, and Allison. My aunts and uncles, Bryan, Darlene, Pam, and Jeff, and grandfather Loyd, have always been supportive.

Finally, Jacob Blount, and our precious dog, Miles, and cat, Loki, have always provided unwavering support and I would be nowhere without the love and support of the family we have cultivated over the last several years.

I am also grateful to NC DETECT for providing data. The North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT) is an advanced, statewide public health surveillance system. NC DETECT is funded with federal funds by North Carolina Division of Public Health (NC DPH), Public Health Emergency Preparedness Grant (PHEP), and managed through a collaboration between NC DPH and the University of North Carolina at Chapel Hill Department of Emergency Medicine's Carolina Center for Health Informatics (UNC CCHI). The NC DETECT Data Oversight Committee does not take responsibility for the scientific validity or accuracy of methodology, results, statistical analyses, or conclusions presented. The NC DETECT Data Oversight Committee (DOC) includes representatives from the NC DPH, UNC NC DETECT Team and NC Hospital Association.
# TABLE OF CONTENTS

LIST OF TABLES ......................................................................................................................... xi

LIST OF FIGURES ........................................................................................................................ xii

LIST OF ABBREVIATIONS .......................................................................................................... xiv

CHAPTER 1: INTRODUCTION ........................................................................................................ 1

Background ................................................................................................................................ 1

Research Questions and Objectives .......................................................................................... 4

Dissertation Outline .................................................................................................................... 6

CHAPTER 2: THE RELATIONSHIP BETWEEN HEAT STRESS INDICES AND MORBIDITY ACROSS NORTH CAROLINA .................................................................................. 8

Data ............................................................................................................................................ 12

Health and Population Data ....................................................................................................... 12

Climate Data ............................................................................................................................... 13

Methods ..................................................................................................................................... 15

Heat Index Calculation ............................................................................................................... 15

WBGT Estimation ..................................................................................................................... 16

Emergency Department Visit Time Series ................................................................................ 19

Statistical Analysis ..................................................................................................................... 19

Results ....................................................................................................................................... 22

Discussion and Conclusion ....................................................................................................... 32

CHAPTER 3: MICROCLIMATIC VARIATION OF WET BULB GLOBE TEMPERATURE .................. 37
CHAPTER 4: THE DEVELOPMENT AND ACCURACY ASSESSMENT OF WET BULB GLOBE TEMPERATURE FORECASTS

Background .................................................................................................................. 69

Data ............................................................................................................................. 76

Weather Forecast Data ............................................................................................... 76

Weather Station Measurements .................................................................................. 77

Observed WBGT Measurements ................................................................................ 77

Station WBGT ............................................................................................................. 78

Satellite Imagery and Land Cover ............................................................................. 80

Methods ...................................................................................................................... 80

Biases in the Station and Kestrel 5400 WBGT ......................................................... 54

WBGT Variations: Surface Types and Shade ............................................................. 59

Estimating Clear-Sky Radiation .................................................................................. 62

Discussion and Conclusion ....................................................................................... 63
LIST OF TABLES

Table 2.1. *WBGT Safety Guidelines from the North Carolina High School Athletic Association.* .................10
Table 2.2. *HRI Morbidity Thresholds.* ...........................................................................................................24
Table 2.3. *Mental Health Morbidity Thresholds.* ............................................................................................26
Table 2.4. *All Cause Morbidity Thresholds.* .....................................................................................................27
Table 2.5. *Cardiovascular Morbidity Thresholds.* ............................................................................................28
Table 2.6. *HRI Morbidity Thresholds by Age Group.* ......................................................................................30
Table 2.7. *Mental Health Morbidity Thresholds by Age Group.* .....................................................................30
Table 2.8. *All-Cause Morbidity Thresholds by Age Group.* ...........................................................................31
Table 2.9. *Cardiovascular Morbidity Thresholds by Age Group.* ....................................................................31
Table 3.1. *WBGT Safety Guidelines from the North Carolina High School Athletic Association.* ..........39
Table 3.2. *WBGT Thresholds: Wilcoxon Difference of Means.* .................................................................55
Table 3.3. *Natural Wet Bulb and Black Globe Temperature Difference in Means.* .....................................57
Table 3.4. *Differences in WBGT between Locations.* ....................................................................................61
Table 4.1. *WBGT Safety Guidelines from the North Carolina High School Athletic Association.* ........71
Table 4.2. *Surface Roughness Length Ranges for the Eight Land Cover Classes.* ......................................83
Table 4.3. *Forecast WBGT Flag Accuracy Compared to Observed WBGT.* ..................................................98
Table 4.4. *Forecast WBGT Flag Accuracy Compared to Station WBGT.* ......................................................98
Table 4.5. *Verification Scores for WBGT Flag Forecast by Estimation Method.* .......................................99
Table 4.6. *Gerrity Skill Scores for NBM WBGT Flag Forecasts.* .................................................................100
LIST OF FIGURES

Figure 2.1. Map of Weather Stations Utilized across North Carolina. .................................................................14

Figure 2.2. The Five County Regions of North Carolina. ..................................................................................20

Figure 2.3. Relationships between Heat Stress Metrics and HRI. .................................................................24

Figure 2.4. Relationships between Heat Stress Metrics and Mental Morbidity. ..........................................25

Figure 2.5. Relationships between Heat Stress Metrics and All-Cause Morbidity. ....................................27

Figure 2.6. Relationships between Heat Stress Metrics and Cardiovascular Morbidity. ..........................29

Figure 3.1. Relationship between WBGT and wind speed ..........................................................................41

Figure 3.2. Variation in wind speed as a function of height above ground. .............................................42

Figure 3.3. Field Work Instruments (WBGT meters and weather station). .............................................47

Figure 3.4. WBGT Biases by Device and by 2°F Increments of WBGT. ....................................................54

Figure 3.5. Kestrel 5400 WBGT Bias. ........................................................................................................56

Figure 3.6. WBGT Accuracy Metrics. ........................................................................................................58

Figure 3.7. Observed Skin Temperatures. .....................................................................................................59

Figure 3.8. Black Globe Temperature Biases Relative to Difference between Skin and Air Temperature. ..60

Figure 3.9. WBGT Sun versus WBGT Shade. ..............................................................................................61

Figure 3.10. Clear-sky Radiation Estimates. ..................................................................................................63

Figure 4.1. Relationship between Station WBGT and wind speed ............................................................73

Figure 4.2. Variation in wind speed as a function of height above ground ..............................................73

Figure 4.3. Field Work Instruments (WBGT meters and weather station). .............................................78

Figure 4.4. Map of the weather stations used (by network) for Station WBGT. ....................................79

Figure 4.5. WBGT Forecast Bias: NBM and NDFD. .....................................................................................90

Figure 4.6. WBGT Forecast Bias (Relative and Absolute). ........................................................................91

Figure 4.7. Daily Maximum WBGT Forecast Bias. ....................................................................................92
Figure 4.8. *WBGT Bias: Regional Variations.* ..........................................................93

Figure 4.9. *SC WBGT vs. SC WBGT Land Forecast Bias.* ......................................94

Figure 4.10. *NWS WBGT Bias: Regional Differences.* ...........................................95

Figure 4.11. *WBGT Forecast Bias by Hour of Day.* ..................................................96

Figure 4.12. *SC WBGT Bias Stratified by Meteorological Variables for Observed WBGT.* 97

Figure 4.13. *SC WBGT vs. SC WBGT Land Bias relative to Surface Roughness.* ..........101
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC</td>
<td>Akaike Information Criterion</td>
</tr>
<tr>
<td>ASOS</td>
<td>Automated Surface Observing System</td>
</tr>
<tr>
<td>ATC</td>
<td>Atmospheric Transmissivity Coefficient</td>
</tr>
<tr>
<td>AWOS</td>
<td>Automated Weather Observing System</td>
</tr>
<tr>
<td>CDC</td>
<td>Center for Disease Control</td>
</tr>
<tr>
<td>CISA</td>
<td>Carolinas Integrated Sciences and Assessments</td>
</tr>
<tr>
<td>ECONet</td>
<td>North Carolina Environment and Climate Observing Network</td>
</tr>
<tr>
<td>ED</td>
<td>Emergency Department</td>
</tr>
<tr>
<td>ERA5-Land</td>
<td>ECMWF Re-Analysis Fifth Generation–Land Model</td>
</tr>
<tr>
<td>GAM</td>
<td>Generalized Additive Model</td>
</tr>
<tr>
<td>HHWS</td>
<td>Heat-Health Warning System</td>
</tr>
<tr>
<td>HRI</td>
<td>Heat-Related Illness</td>
</tr>
<tr>
<td>NBM</td>
<td>National Blend of Models</td>
</tr>
<tr>
<td>NC DETECT</td>
<td>North Carolina Disease Event Tracking and Epidemiologic Collection Tool</td>
</tr>
<tr>
<td>NDFD</td>
<td>National Digital Forecast Database</td>
</tr>
<tr>
<td>NLCD</td>
<td>National Land cover Database</td>
</tr>
<tr>
<td>NWS WBGT</td>
<td>NWS WBGT Forecast</td>
</tr>
<tr>
<td>NWS</td>
<td>United States National Weather Service</td>
</tr>
<tr>
<td>SC WBGT</td>
<td>SERCC-CISA WBGT Forecast</td>
</tr>
<tr>
<td>SC WBGT LAND</td>
<td>SERCC-CISA WBGT Forecast using Surface Roughness</td>
</tr>
<tr>
<td>SERCC</td>
<td>Southeast Regional Climate Center</td>
</tr>
<tr>
<td>WBGT</td>
<td>Wet Bulb Globe Temperature</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

Background

Heat presents a significant danger to human health, with heat exposure being the leading cause of weather-related death in the United States (CDC, 2010; NOAA's National Weather Service, 2020) and within the context of climate change.

The effects of heat on human health have been thoroughly investigated. These investigations have advanced the understanding of the body's physiological response to increasing temperature and determined numerous underlying factors that increase a person's vulnerability to heat, including cardiovascular, respiratory, and renal disease (Gosling et al., 2009; Kovats & Hajat, 2008). The primary external environmental variables influencing the thermoregulation of the human body are air temperature, humidity, radiation, and wind speed. While all these variables act in concert to increase or decrease body temperature, air temperature and radiation, both from the sun and surfaces such as concrete, influence the body's skin temperature. Humidity is a significant determiner of one's sweat rate and thus evaporative cooling rate. Wind speed also impacts the evaporative cooling rate and, more generally, the airflow across the body and therefore convective heat loss. Heat stress indices account for the effect of these external environmental variables on body temperature by quantifying the danger posed by the meteorological environment. From the panoply of heat stress indices that have been developed, there is no universally accepted index to predict thermal stress (Brake & Bates, 2002; Jendritzky et al., 2012).

Common examples of environmental heat stress indices include the Heat Index, humidex, and universal thermal climate index. Wet bulb globe temperature (WBGT) is another heat stress
index. WBGT is becoming increasingly utilized in numerous sectors to safeguard health, such as in occupational settings and athletics. The benefits of using WBGT derive from it accounting for the entire suite of environmental variables impacting human body temperature: air temperature, humidity, wind speed, and radiation (Budd, 2008; Hondula et al., 2014). Comparatively, the Heat Index only accounts for the influence of air temperature and humidity on body temperature and is valid only for shaded conditions (since the air temperature sensor is shielded from the sun).

WBGT was developed in the 1950s by the United States military to address high rates of heat-related fatalities in training camps (Budd, 2008; Yaglou & Minard, 1957). WBGT is calculated from three components (1): dry bulb temperature (air temperature), natural wet bulb temperature, and black globe temperature (Budd, 2008). The black globe temperature is measured using a black globe thermometer, which is unshielded from radiation and serves as a measure of the radiant temperature incident on the human body given the effects of direct and diffuse short-wave radiation, long-wave radiation from the Earth’s surface, and wind speed (Kopec, 1977; Liljegren et al., 2008). The dry-bulb temperature is simply a standard measure of ambient air temperature, with the thermometer placed in a naturally ventilated radiation shield (Liljegren et al., 2008). Unlike the dry-bulb temperature, the natural wet-bulb temperature is measured using a thermometer unshielded from radiation, with a wetted wick wrapped around the bulb of the thermometer (Liljegren et al., 2008). This component thus accounts for the cooling effect of moisture evaporating from human skin, since the evaporation of the water from the wetted wick is dictated by the environmental conditions of wind speed, relative humidity, and solar radiation (Budd, 2008; Kopec, 1977; Liljegren et al., 2008; Yaglou & Minard, 1957). From these three components, WBGT is calculated according to the following formulation:

\[
WBGT = 0.7 \times NWB + 0.2 \times Tg + 0.1 \times Ta,
\]  (1)
where $NWB$ is the natural web bulb temperature, $T_g$ is the black globe temperature, and $db$ is the dry bulb temperature (air temperature).

Like other heat stress indices, WBGT has specific values corresponding to varying levels of danger, specifically four levels of risk, often referred to as WBGT flags. Each flag level has associated guidelines for activity modifications (e.g., work-rest ratios and hydration requirements), with the highest level (black flag) being the level at which all activity outdoor activity should be halted, which is commonly 90°F in occupational and athletic settings.

While WBGT is a robust indicator of environmental heat stress, several challenges exist with using, measuring, and interpreting the index.

- Heat-health warning systems (HHWS) safeguard public health during extreme heat, with varying levels of danger identified based on a selected heat stress index (Kovats & Kristie, 2006). WBGT cannot be effectively utilized as the basis for a Heat-Health Warning System since it does not satisfy two criteria of an effective HHWS (Kovats & Kristie, 2006): 1) limited research has been conducted on its ability to predict morbidity at a population-level (Vaneckova et al., 2011) and 2) WBGT forecasts are not common and limited research has assessed if it can be accurately forecast, which is important since it relies on numerous variables that each have their own forecast bias (Hondula et al., 2014).

- WBGT is not a standard meteorological variable measured at weather stations, and the accuracy and proper use of WBGT measurements are hindered by many cheaper, less accurate meters being on the market (Budd, 2008). Additionally, while methods have been developed that provide estimates of WBGT based on standard weather station measurements, these methods vary in accuracy. While the methodology developed in
Liljegren et al. (2008) has been found to be most accurate (Lemke & Kjellstrom, 2012; Patel et al., 2013), other methods with less robust verification are commonly used.

- WBGT is highly variable over small distances, requiring measurements to be conducted at the site of activity for robustness. The index varies substantially over small timescales, with literature recommending the use of an average WBGT over a variety of intervals, such as a 5-minute (Thorsson et al., 2007), 30-minute (Kopec, 1977), or 60-minute average (OSHA, 2017).

**Research Questions and Objectives**

This research addresses the following broad research questions:

1) What differences in heat-morbidity relationships exist when air temperature, the Heat Index, or WBGT are used to predict morbidity? How do relationships between the heat stress metrics and morbidity vary across demographics and across the state of North Carolina?

2) How does WBGT vary across small distances (10s to 100s of meters), over different surfaces (e.g., tennis courts), and in differing shade conditions (e.g., a shaded forest vs. shaded grass field)? What is the accuracy of estimating WBGT?

3) What is the accuracy of forecasting WBGT and how does this compare to other forecasts of the index? Can forecasts be improved by incorporating the varying influence of land cover on wind speed (i.e., variations in surface roughness)?

Through these questions, the research presented here contributes and expands upon existing literature in numerous ways:

1) Existing research in North Carolina on heat and health outcomes has revealed increases in morbidity (Sugg et al., 2016; Kovach et al., 2015) and mortality (Clark, 2019) with increasing heat. However, this research further contributes to this understanding by assessing the
impact of heat on other morbidities, including cardiovascular and respiratory morbidities, and by assessing the utility of using WBGT to predict morbidity. Importantly, many studies utilize air temperature (Quinn & Shaman, 2017; Sugg et al., 2016) and the Heat Index (Quinn et al., 2014) to predict morbidity. However, this research also assesses how the use of WBGT compares to the use of these more commonly used metrics in making these predictions, a notable gap in current literature (Hajat et al., 2010; Vaneckova et al., 2011). Lastly, this research addresses the need to validate the efficacy of the WBGT values commonly used as the thresholds for each flag level in relation to morbidity.

2) Compared to prior work, this research examines a more extensive sampling of days across microclimates in an array of athletic and suburban environments, particularly days with high thermal stress (WBGT > 88°F–90°F). The accuracy of WBGT estimated from a weather station and WBGTs measured by a commonly used WBGT meter (the Kestrel 5400 Heat Stress Tracker) are assessed. This assessment is based on ground truth WBGT observations made with a full-size WBGT meter manufactured by Kestrel to meet the specifications for a WBGT meter as outlined by the International Organization for Standardization, first used in Cooper et al. (2017). Additionally, comparing the WBGT over various surfaces (e.g., tennis courts vs. grass) furthers the understanding of these variations developed by existing literature (Grundstein & Cooper, 2020; Kopec, 1977; Pryor et al., 2017).

3) WBGT is not routinely forecast. The United States National Weather Service has a WBGT forecast that was operationalized in 2021 (NOAA, 2021) and had been an experimental product in the years prior (US NWS, 2019a). The development of another WBGT forecast tool is detailed here, which was operationalized in 2018. Limited verification has been conducted to assess the accuracy of WBGT forecasts and if it can be reliably forecast, since estimations require numerous forecast weather variables, each with their own bias
(Hondula et al., 2014). This study verifies these forecasts across multiple locations and microclimates for three summers (2019-2021). In addition to assessing the accuracy, this research assesses the utility of using surface roughness data that is on a much finer scale than the spatial scale of the forecast data to more accurately downscale wind speeds for calculating the WBGT forecast.

**Dissertation Outline**

Chapters 2–4 of this dissertation each focus on one of these three questions and outlines. 

Chapter 2 is titled “The Relationship between Heat Stress Indices and Morbidity across North Carolina”. This chapter assesses these relationships and compares the utility of using different heat stress indicators to predict morbidity at a population-level across North Carolina. Importantly, this chapter addresses the current lack of research on whether WBGT can be used as a heat stress indicator to reliably predict morbidity, and how its performance compares to more commonly used metrics, such as air temperature and the Heat Index. Differences in the heat-morbidity burden across age groups and the state’s three physiographic regions (the mountains, piedmont, and coastal plain) are also assessed. Lastly, the efficacy of existing thresholds for WBGT flag levels (e.g., 90°F for black flag) is investigated by comparing the relative changes in morbidity at these threshold values.

Chapter 3 is titled “Microclimatic Variation of Wet Bulb Globe Temperature”. The research presented within this chapter assesses the degree to which WBGT varies across small distances (10s to 100s of meters) and the differences in WBGT measured over a variety of surfaces and conditions, such as tennis courts, grass fields, and shaded forests. Additionally, the accuracy of estimates of WBGT is presented, and methods for estimating clear-sky radiation and modifying that radiation by percentage cloud cover are evaluated.
Chapter 4 is titled “The Development and Accuracy Assessment of Wet Bulb Globe Temperature Forecasts”. This chapter has two broad components. First, the research details the development of a tool to forecast WBGT out to five days, a product of a partnership between NOAA’s Southeast Regional Climate Center (SERCC) and Carolinas Integrated Sciences and Assessments (CISA). This chapter provides an assessment of the accuracy of these forecasts overall and in differing conditions. Additionally, the SERCC/CISA WBGT forecast tool (SERCC & CISA, 2023), which was operationalized in 2018, is compared to a WBGT forecast made by the United States National Weather Service using field work collected over the summers of 2019-2021. The second component of this chapter addresses efforts to improve the WBGT forecast by better accounting for the impact of surface roughness on wind speed using high-resolution satellite imagery (Sentinel 2A/2B and Landsat 8 ETM+) to derive surface roughness values.
CHAPTER 2: THE RELATIONSHIP BETWEEN HEAT STRESS INDICES AND MORBIDITY ACROSS NORTH CAROLINA

According to the US Center for Disease Control and US National Weather Service, exposure to extreme heat is the leading cause of weather-related death across the US (CDC, 2010; NOAA's National Weather Service, 2020). Given the projected warming due to climate change, the current threat posed by extreme heat will undoubtedly worsen as time progresses.

In order to safeguard public health during times of extreme heat, heat-health warning systems (HHWS) have been developed that link meteorological forecasts with public health messaging, warning the public of potentially dangerous environmental heat levels (Kovats & Kristie, 2006). Currently, the HHWS employed by the United States National Weather Service (NWS) utilizes the Heat Index, although this is supplemented with air temperature and other considerations based on region and NWS office (Hawkins et al., 2017). The Heat Index incorporates the effect of air temperature and humidity on human body temperature. However, this heat stress indicator has been critiqued for not incorporating all of the environmental variables that impact human body temperature, such as wind speed and solar radiation, which are accounted for by other heat stress indices such as wet bulb globe temperature (WBGT) (Budd, 2008; Hondula et al., 2014). While previous studies on the effect of heat on morbidity have primarily focused on defining these relationships by using air temperature (Quinn & Shaman, 2017; Sugg et al., 2016) and the Heat Index (Quinn et al., 2014), this research compares WBGT with these more commonly used indicators as a predictor of heat illness and other morbidities commonly associated with exposure to extreme heat.

**Wet Bulb Globe Temperature (WBGT)**

8
WBGT was developed in the 1950s by the United States military to decrease the number of heat-related casualties occurring at US military training camps (Budd, 2008; Yaglou & Minard, 1957). WBGT is calculated from the summation of three components (1): dry-bulb temperature, natural wet bulb temperature, and black globe temperature (Budd, 2008). The dry-bulb temperature is ambient air temperature, a commonly measured meteorological variable. Both the black globe temperature and natural wet bulb temperature account for the influences of multiple environmental variables on human body temperature: air temperature, radiation, humidity, and wind speed (Budd, 2008; Kopec, 1977; Liljegren et al., 2008). The black globe temperature is measured with a thermometer placed inside a black globe. This temperature is particularly sensitive to radiative forces, such as long wave radiation from the earth’s surface, ultimately indicating the radiant temperature impacting the human body (Kopec, 1977; Liljegren et al., 2008). The last component, the natural wet bulb temperature, is measured with a thermometer placed above or adjacent to a water reservoir. The thermometer is kept continuously moist by a wick being wrapped around the bulb of the thermometer, and the other end of the wick being submerged in the water reservoir. Importantly, this thermometer is not shielded from radiation, unlike the dry-bulb thermometer (Liljegren et al., 2008). Thus, this variable is akin to the interaction between the human body and environment with respect to sweat evaporation and the corresponding cooling effect. These three variables are summed to derive the WBGT using the following equation:

\[ WBGT = 0.7 \times NWB + 0.2 \times Tg + 0.1 \times Ta, \]

(1)

where \( NWB \) is the natural web bulb temperature, \( Tg \) is the black globe temperature, and \( Ta \) is the dry bulb temperature (air temperature).

To determine the danger posed by a given array of environmental conditions, the WBGT is measured and then cross-referenced with established ranges of values, each corresponding to
varying levels of danger. Each level of danger, or WBGT flag, has corresponding activity modifications and hydration recommendations to prevent overheating, such as seen in Table 2.1.

**Table 2.1. WBGT Safety Guidelines from the North Carolina High School Athletic Association.**

<table>
<thead>
<tr>
<th>WBGT (°F)</th>
<th>Athletic Activity Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 80</td>
<td>Unlimited activity with primary cautions for new or unconditioned athletes or extreme exertion; schedule mandatory rest/water breaks (5 min water/rest break every 30 min).</td>
</tr>
<tr>
<td>80–84.9</td>
<td>Normal practice for athletes; closely monitor new or unconditioned athletes and all athletes during extreme exertion. Schedule mandatory rest/water breaks (5 min water/rest break every 25 min).</td>
</tr>
<tr>
<td>85–87.9</td>
<td>New or unconditioned athletes should have reduced intensity practice and modifications in clothing. Well-conditioned athletes should have more frequent rest breaks and hydration as well as cautious monitoring for symptoms of heat illness. Schedule frequent mandatory rest/water breaks (5 min water/rest break every 20 min). Have cold or ice immersion pool on site for practice.</td>
</tr>
<tr>
<td>88–89.9</td>
<td>All athletes must be under constant observation and supervision. Remove pads and equipment. Schedule frequent mandatory rest/water breaks (5 min water/rest break every 15 min). Have cold or ice immersion pool on site for practice.</td>
</tr>
<tr>
<td>90+</td>
<td>SUSPEND PRACTICE/MUST INCLUDE MANDATORY BREAKS AS DIRECTED BY GAME DAY ADMINISTRATOR DURING CONTEST.</td>
</tr>
</tbody>
</table>

*Note. Adapted from Raleigh NWS WFO (2017).*

While WBGT is generally a better heat stress indicator relative to the Heat Index and standard air temperature, it is not typically used for heat warnings at a population-level for several reasons. WBGT is not a standard meteorological variable measured at weather stations, which makes real-time assessment of heat stress across broad spatial scales challenging, nor is it routinely forecasted by numerical weather prediction models. Further, limited research has been conducted linking WBGT to health outcomes, specifically when compared to other indices such as the Heat Index (Vaneckova et al., 2011), which this study addresses.

The first objective of this research is to investigate relationships between three heat stress indicators (WBGT, the Heat Index, and air temperature) and emergency department (ED) visits across the state of North Carolina for the 2007-2016 heat seasons (May-September). Heat-related emergency department visits as well as all-cause, cardiovascular-, respiratory-, and mental health-
related morbidity were analyzed. Data from weather stations is most commonly used in studies similar to this one. However, an oft-cited limitation with point weather station data is the possible mischaracterization of an individual’s environmental exposure, since environments in which people live and work differ greatly from the environment of an airport weather station (Kuras et al., 2017; Sugg et al., 2019). In addition to weather station data, this study addresses this possible concern by utilizing the ERA5-Land reanalysis data set, a robust, hourly gridded weather observation data set with a very fine spatial resolution (~9km) (Muñoz-Sabater, 2019) to model relationships with morbidity.

Prior work in North Carolina has revealed increases in morbidity (Fuhrmann et al., 2016; Kovach et al., 2015; Lippmann et al., 2013; Sugg et al., 2016) and mortality (Clark, 2019) with increasing heat, as well as the effects of heat on maternal health (Ward et al., 2019). However, the research presented here adds to this existing literature in three ways. First, it will also investigate the relationships between heat and cardiovascular, respiratory, mental health, and all-cause morbidities, as other research has reported strong associations between these morbidities and heat (Kovach et al., 2015; M. Li et al., 2015; Michelozzi et al., 2009; Páldy et al., 2005; Semenza et al., 1999; Song et al., 2017). Second, WBGT was not used to predict morbidity in these prior studies. Thus, there have been no comparisons of the relationships of WBGT and the Heat Index with morbidity. Third, the emergency department visit data used here spans ten years (2007-2016), while original research in Sugg et al. (2016) on morbidity in NC spanned only six years (2007-2012).

**Research Questions**

The following research questions were addressed:

1. Do the defined relationships between heat stress indices and ED visits differ when using weather station (point) data vs. the ERA5-Land data set (raster)? Are the relationships defined with one or the other stronger?
2. How do the three metrics (air temperature, the Heat Index, and WBGT) compare to one another with respect to predicting increased ED visits?
   
   a. Do certain morbidities have a stronger relationship with increasing temperature relative to the others under investigation? Does this differ across age groups?
   
   b. Do the relationships between the three heat stress metrics differ across the state, including urban vs. rural difference and climatic difference (mountains vs. piedmont vs. coast)?

3. Are current WBGT flag thresholds and NWS heat index guidelines adequate in categorizing increases in risk of HRI?

Data

Health and Population Data

The data containing ED visits for North Carolina were from the North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT), which is a statewide surveillance system created by the North Carolina Division of Public Health in 2004 (North Carolina Department of Public Health, 2017). The NC DETECT data set contains emergency department visits from 124 hospitals on a daily basis across the state of North Carolina; by 2008, NC DETECT captured more than 99% of total visits statewide (Hakenewerth et al., 2009; Lippmann et al., 2013; Rhea et al., 2012; Sugg et al., 2016). For each patient, the data contain their zip code and county of residence, age, sex, race, date, and time of emergency department visit, and up to eleven fields noting the patient’s diagnosis. From 2007-2015, these diagnosis fields are coded using the ninth revision of the International Classification of Disease (ICD), and beginning in 2016, the tenth revision of the ICD.

The relationships between the three heat stress indices and specific causes of morbidity were assessed by taking subsets of the NC DETECT data based on the ICD codes associated with each emergency department visit. The morbidities and associated ICD codes are as follows:
cardiovascular (ICD9, 390-459.9; ICD10, I00-I99), respiratory (ICD9, 460-519.9; ICD10, J00-J99), mental health (ICD9, 290-299.9; ICD10 F00-F99), and those related to heat exposure and heat-related illness (ICD-9, 2765, 992, E900; ICD-10, T67, X30). For an emergency department visit to be associated with one of these diseases or causes, at least one of the ICD codes was listed as one of the causes of morbidity.

County-level population data from the North Carolina State Demographer’s Office at the Office of State Budget and Management was used to model morbidity as a rate. The population data were segmented by age for each year from 2007-2016 and included mid-year population estimates for each county in North Carolina (North Carolina State Demographer, 2018). Additionally, the population-weighted centroid from the 2010 US Decennial Census was utilized to derive the distances between weather stations and counties (U.S. Census Bureau, 2010).

**Climate Data**

Two sources for the climate data were linked with emergency department visits. First, hourly weather observations were gathered from 115 weather stations across NC (Figure 2.1). Stations from three networks were utilized: 1) Automated Surface Observing System (ASOS) (19 stations), 2) Automated Weather Observing System (AWOS) (53 stations), and 3) North Carolina Environment and Climate Observing Network (ECONet) (43 stations). Station data from ASOS and AWOS networks were acquired through the Iowa Environmental Mesonet archive (IEM). The ECONet weather station data were retrieved from the Climate Retrieval Observations Network of the Southeast Database housed at the North Carolina State Climate Office (NC CRONOS).
Figure 2.1. Map of Weather Stations Utilized across North Carolina.

Note. The black dots correspond to locations of weather stations on the ASOS network. The green dots denote the locations of weather stations on the AWOS network. Lastly, the purple dots display the location of the weather stations on the ECONet network.

The second climate data source is the ERA5-Land data set, a gridded climate reanalysis data with a spatial resolution of 9 km and an hourly temporal resolution. The data were downloaded from the Copernicus Climate Data Store (Muñoz-Sabater, 2019).

The following variables from the hourly weather station data were utilized: barometric pressure, 2-meter ambient air temperature, 2-meter dew point temperature, relative humidity, and 10-meter wind speed. In addition, for stations that are a part of the ECONet network, solar radiation observations were used. For stations on the ASOS and AWOS networks, solar radiation is not reported, thus the reported cloud cover amount at multiple levels was used to estimate observed solar radiation.

For the ERA5-Land, the following variables were extracted: surface pressure, 2-meter ambient air temperature, 2-meter dew point temperature, 10-meter u-component of wind, 10-meter v-component of wind, and surface solar radiation downwards.
Methods

The Heat Index and WBGT were estimated, along with select input variables derived from the provided variables in each data set, detailed below.

Heat Index Calculation

For both the weather station and ERA5-Land data, the Heat Index was calculated using the equation from the US NWS (Rothfusz, 1990). For values of the Heat Index less than or equal to 80 degrees Fahrenheit, the results from (2) below were averaged with the air temperature:

\[
HI = 0.5 \times [T + 61.0 + ((T - 68) \times 1.2) + (RH \times 0.094)],
\]

where \(T\) is the air temperature in degrees Fahrenheit and \(RH\) is the relative humidity (%). For values of the Heat Index greater than 80°F, the following equation was used (Rothfusz, 1990):

\[
HI = -42.379 + 2.04901523 \times T + 10.14333127 \times RH - 0.022475541 \times T \times RH + 0.00683783 \times T^2 - 0.05481717 \times RH^2 + 0.00122874 \times T^2 \times RH + 0.00085282 \times T \times RH^2 - 0.00000199 \times T^2 \times RH^2.
\]

If relative humidity was below 13% and air temperature was 80°F–112°F, the calculated value from (4) was subtracted from (3) (Rothfusz, 1990):

\[
adj1 = \left[ \frac{13 - RH}{4} \right] \times \sqrt{17 - ABS(T - 95)}.
\]

If relative humidity is higher than 85% and air temperature was 80°F–112°F, the value from (3) was summed with (5) (Rothfusz, 1990):

\[
adj2 = \left[ \frac{RH - 85}{10} \right] \times \frac{87 - T}{5}.
\]

Since the ERA5-Land did not include relative humidity values, but instead dew point temperature. The relative humidity was calculated using the AERK Magnus formulation (Alduchov & Eskridge, 1996):

\[
RH = 100 \times \frac{e^{(17.625 \times Td)/(243.04 + Td)}}{e^{(17.65 \times T)/(243.04 + T)}},
\]

where \(Td\) is the dew point temperature.
where $T$ is the air temperature and $T_d$ is the dew point temperature.

**WBGT Estimation**

The Liljegren et al. (2008) methodological approach, which has been found to be most accurate at estimating WBGT (Lemke & Kjellstrom, 2012; Patel et al., 2013), was utilized to estimate WBGT using the R package “wbgt” (Lieblich & Spector, 2017).

Using the Liljegren et al. (2008) methodology, the equation for calculating the natural wet bulb temperature is given in (7):

$$
T_w = T_a - \frac{\Delta H}{C_p} \frac{M_{H2O}}{M_{Air}} \left( Pr \left( \frac{e_w}{Sc} \right)^{\alpha} (P - e_a) + \frac{\Delta F_{net}}{A h} \right),
$$

where $T_a$ is the ambient air temperature, $\Delta H$ is the heat of vaporization, $C_p$ is the specific heat at constant pressure, $M_{H2O}$ is the molecular weight of water vapor, $M_{Air}$ is the molecular weight of air, $Pr$ is the Prandtl number, $Sc$ is the Schmidt number, $\alpha$ is a constant (0.56), $e_w$ is the saturation vapor pressure of the wick, $e_a$ is the saturation vapor pressure of the air, $P$ is the barometric pressure, $\Delta F_{net}$ is the net radiant heat flux to the wick from the environment, $A$ is the surface area of the wick, and $h$ is the convective heat transfer coefficient (see Liljegren et al. (2008) for details) (Grundstein, 2018; Liljegren et al., 2008). The equation for black globe temperature is given in (8):

$$
T_g^4 = \frac{1}{2} (1 + \varepsilon_a) T_a^4 - \frac{h}{S \varepsilon_g} (T_g - T_a) 
+ \frac{S}{2 \varepsilon_g \alpha} \left[ \frac{1}{2 \cos(\theta)} - 1 \right] f_{dir} + \alpha_{sfc}.
$$

where $\varepsilon_a$ is the emissivity of air, $T_o$ is the ambient air temperature, $h$ is the convective heat transfer coefficient, $\varepsilon_g$ is the emissivity of the globe, $S$ is the total horizontal solar irradiance, $\alpha_g$ is albedo of the ground, $\theta$ is the solar zenith angle, $f_{dir}$ is the fraction of total horizontal irradiance ($S$) which is direct beam radiation, and $\alpha_{sfc}$ is the albedo of the surface (Grundstein, 2018; Liljegren et al., 2008). The following constants were used: $\varepsilon_g = 0.95$, $\alpha_g = 0.05$, and $\alpha_{sfc} = 0.45$ (Liljegren et al., 2008).
Since weather stations on the ASOS and AWOS networks did not record solar radiation, it was estimated based on the reported cloud cover. Cloud cover is reported at multiple levels in the atmosphere. The cloud cover amounts and corresponding percentage cloud cover are as follows: clear (0%–5%), few (5%–10%), scattered (25%–50%), broken (50%–87%), and overcast (87%–100%) (National Oceanic and Atmospheric Administration et al., 1998). Based on comparisons of estimated solar radiation from cloud cover against in situ solar radiation observations at ECONet stations, as well as comparing the ranges of estimated WBGT at ECONet and ASOS/AWOS stations, the reported cloud amount for each level was converted to percentage cloud cover using the maximum value from the ranges above: clear (5%), few (10%), scattered (50%), broken (87%), and overcast (100%). To derive a single percentage cloud cover value for each observation, the cloud level with the maximum amount of reported cloud cover was used, e.g., if an observation reported “few” at cloud level 1 and “scattered” at cloud level 2, scattered (50%) was the cloud amount used to calculate solar radiation for that observation. After determining the percentage cloud cover for each observation, the clear-sky direct radiation value corresponding to the location and time of observation was modified by the percentage cloud cover to derive an estimate of solar radiation using the following equation:

\[
S_{rad} = R_0 \times (1 - 0.75n^{.4}),
\]

where \(n\) is the cloud cover fraction (0.0–1.0) (Kasten & Czeplak, 1980) and \(R_0\) is the clear-sky direct radiation (w/m²) estimated using (10):

\[
R_0 = 990 \times \sin(\varnothing - 30),
\]

where \(\varnothing\) is solar elevation angle (Kasten & Czeplak, 1980).

For the ERA5-Land, wind speed was calculated from the u and v components using the following equation:
\[ \vec{V} = \sqrt{u^2 + v^2}, \]  
where \( u \) is wind speed in the horizontal and \( v \) is the wind speed in the vertical (Guillory, 2020).

After calculating the wind speed for ERA5-Land, both ERA5-Land wind speed and wind speed for the weather station data were logarithmically downscaled from 10 meters to 2 meters using the following function:

\[ U_z = U_r \left( \frac{Z}{Z_r} \right)^p, \]  
where \( U_r \) is the mean wind speed at height \( Z \) above ground, \( U_r \) is the wind speed at the reference height \( Z_r \), and \( p \) is the power-law exponent (US EPA, 2000). The power-law exponent was determined from Pasquill-Gifford Stability classes with the Solar Radiation Delta-T (SRDT) method. The exponents are provided in Appendix 2.1. The “Urban” exponents were utilized here. SRDT indicates atmospheric stability, and thus the degree to which wind speeds higher in the atmosphere mix down to lower altitudes, based on observed solar radiation (daylight hours) and low-level vertical temperature difference at night (US EPA, 2000). Since the sensitivity of all anemometers installed on the ASOS/AWOS and the ECONet weather station networks was 1 meter per second, wind speeds below this value were increased to 1 meter per second, following Liljegren et al. (2008) and Clark (2019).

Lastly, barometric pressure is needed for estimating WBGT. While the weather station data provided this variable directly, this variable had to be estimated for the ERA5-Land from surface pressure using the following equation:

\[ slp = p \times \left( 1 - \frac{0.0065 \times Z}{T + 0.0065 \times Z} \right)^{-5.257}, \]  
where \( p \) is the surface pressure, \( Z \) is altitude in meters, and \( T \) is the air temperature (NOAA, 2017). The altitude of each ERA5-Land pixel was obtained from the orography data provided alongside the ERA5-Land.
Emergency Department Visit Time Series

A daily time series of ED visits for each county in NC was created by summing the number of visits for each day for all-cause morbidity and for each morbidity type (cardiovascular-, respiratory-, mental health-, and heat-related) based on the listed county of residence in each ED visit record. The number of visits for each morbidity type was also calculated based on age group (ages 0–12, 13–19, 20–34, 35–54, 55–64, and 65+). The data were amalgamated at the county level because the available yearly population data used to model morbidity as a rate was at the county-level. This time series of daily ED visit counts was then date-matched and linked with daily statistics of air temperature, the Heat Index, and WBGT (max, min, and mean) derived from each ERA5-Land pixel and each weather station.

The daily statistics for each county were obtained from the closest weather station to the population-weighted centroid of a given county, using Euclidean distance. The population-weighted centroid was based on the 2010 US Decennial Census (U.S. Census Bureau, 2010). Missing data in the daily time series of heat stress indices were filled with data from the second closest weather station. For the ERA5-Land data, the statistics for a given day for each county were calculated based on averaging the values of all pixels falling within a given county's borders.

To assess the cumulative effect of increasing temperature, lags of the daily statistics were calculated, specifically a one-, two-, and three-day lag, as well as three-day moving averages of these statistics.

Statistical Analysis

Previous research in NC on the effects of heat on morbidity (Kovach et al., 2015; Sugg et al., 2016) and mortality (Clark, 2019) found significant differences in the relationships between increasing temperature and morbidity burden across the three physiographic regions of the state:
the mountains, piedmont, and coastal plain (Figure 2.2). Further, these studies found the health effects of increasing temperature differed with respect to the rurality of a county.

Consequently, this study segments the state into five distinct regions based on both physiography and rurality: (1) mountains, (2) rural piedmont, (3) urban piedmont, (4) urban coastal plain, and (5) rural coastal plain. The mountain region was not split into urban and rural components due to the relatively low population of the region, which would lead to issues with sample size, and due to the majority of the region being rural. Additionally, to avoid issues with sample size, the analysis across the different age groups was not segmented by region.

Figure 2.2. The Five County Regions of North Carolina.

Note. County regionalization first used in Clark (2019).

Generalized additive models (GAM) were used to define relationships between the heat stress metrics and morbidity. GAMs are similar to generalized linear models (GLM) where the linear predictor is a summation of smooth functions of the covariates (Wood, 2017). GAMs are a common statistical approach to modeling relationships between heat and morbidity (Bell et al., 2008; Hajat et al., 2007; Kovats et al., 2004; B. Li et al., 2012; Petkova et al., 2017; Sugg et al., 2016).
Multiple GAMs were created to model the relationships between each heat stress metric individually across the multiple morbidity types under investigation. The heat stress metric in each model was modeled using natural cubic splines with 1 degree of freedom per 5 degrees Celsius (Kovats et al., 2004; B. Li et al., 2012). To control and account for the underlying seasonality of the ED visits, terms corresponding to the day of the week and year were included in each model. These terms were modeled using random effects splines with 1 degree of freedom per year (10 total) and 6 degrees of freedom for day of the week. In addition, the effect of holidays on morbidity was accounted for through a factor variable corresponding to all US federal holidays. To assess and compare the relationships for each region separately, every GAM contained a factor variable corresponding to the county group membership of a given county. All GAMs were modeled using R with the R package “mgcv” (Wood, 2011).

To determine which source of data (weather station or ERA5-Land) and heat stress index were more robust, the Akaike Information Criterion (AIC) of models for each morbidity and heat stress index combination were compared. The conclusions derived from comparing AIC were validated by conducting a chi-squared test on the difference in minimized smoothing parameter selection score (fREML score) from each model, factoring in any difference between models in the number of degrees of freedom. These comparisons were conducted using the R package “itsadug” (van Rij et al., 2022).

**Identifying Threshold Temperatures**

The relative increase in morbidity associated with each WBGT flag threshold and at 2.5°F increments of Heat Index and air temperature were calculated to understand the utility of current WBGT flag thresholds and the thresholds of the Heat Index often used in the current NWS HWWS. To determine the critical thresholds where morbidity notably increased, there were two steps:
1. Temperature identified as threshold if morbidity was 1) greater than at lower temperatures and 2) decreased at higher temperatures.

2. If morbidity continued to increase at higher temperatures, but the 95% confidence intervals overlapped with the threshold identified in step 1, both the temperature in step 1 and the temperature in step 2 were highlighted as thresholds.

Additionally, the values of each heat stress index corresponding to peak morbidity were determined when assessing morbidity changes for different age groups.

**Results**

First, the overall assessment of which data set performed best will be discussed (ERA5-Land versus weather station data), followed by discussions of the relationships between heat indices and morbidity across regions, and finally, the relationships between the heat stress indices and morbidity across age groups. The results here will be presented by describing the character of the temperature-morbidity response curve, e.g., a “j-shaped” curve, with exponentially increasing morbidity at the highest temperatures (Gosling et al., 2014) for each heat stress index and by assessing the relative changes in morbidity at 2.5°F increments of the Heat Index and air temperature, 1°F increments of WBGT, and at each WBGT flag level. These relative changes will be presented as a decimal value, with, for example, a change of 1.15 corresponding to a 15% increase in morbidity.

**ERA5 vs. Weather Station Data**

The modeled relationships defined when using the ERA5-Land data and the weather station data were similar with respect to the overall character and magnitude of the relationships between the heat indices and morbidity. However, the models utilizing the weather station data to predict morbidity were more robust when comparing model AIC values. Thus, the results here will only include the relationships defined with the weather station data.
Comparisons of the models for each morbidity and heat stress index combination revealed daily maximum WBGT (non-lagged, one-day maximum) to be a stronger predictor than air temperature and the Heat Index, both when assessing regional differences and differences across age groups. Daily maximum Heat Index then followed, explaining more deviance in the morbidity data than daily maximum air temperature. The results presented here will focus on these one-day maximum, non-lagged variables, emphasizing the relationships defined with the Heat Index and WBGT.

**Relationships with Morbidity across Regions**

The relationships between morbidity and heat stress indices across all regions display strong, statistically significant relationships with a characteristic exponential increase in morbidity at higher temperatures, which is most apparent with HRI and mental health morbidity. HRI rapidly increased at WBGT values of 88°F–91°F, with relative increases of 1.13–1.19 (i.e., 13 to 19%) when comparing morbidity at 1°F increments (Appendix 2.2). Comparing the morbidity increases at WBGT flag levels, morbidity increases ranged from 1.34–1.75, and the largest increase in HRI occurred at a value of 88°F (WBGT red flag) in all regions except for the urban piedmont and rural coastal plain, where the largest increase occurred at 95°F (Table 2.2). The largest magnitude increases at these thresholds occurred in the rural piedmont and urban piedmont (Table 2.2). Across all WBGT values, the rural coastal plain had statistically significant higher rates of morbidity compared to other regions (Figure 2.3).
Table 2.2. *HRI Morbidity Thresholds*.

<table>
<thead>
<tr>
<th>Heat Index</th>
<th><em>°F</em></th>
<th>Mountains</th>
<th>Rural Piedmont</th>
<th>Urban Piedmont</th>
<th>Urban Coastal Plain</th>
<th>Rural Coastal Plain</th>
</tr>
</thead>
<tbody>
<tr>
<td>WBGT</td>
<td>88</td>
<td>1.45 (1.42–1.53)</td>
<td>1.34 (1.34–1.34)</td>
<td>1.63 (1.64–1.63)</td>
<td>1.40 (1.40–1.40)</td>
<td>1.51 (1.51–1.52)</td>
</tr>
<tr>
<td></td>
<td>95</td>
<td>2.23 (1.54–2.35)</td>
<td>1.53 (1.42–1.61)</td>
<td>1.79 (1.50–2.01)</td>
<td>1.69 (1.28–1.97)</td>
<td>1.75 (1.54–1.92)</td>
</tr>
<tr>
<td>Heat Index</td>
<td>103</td>
<td>1.24 (1.24–1.24)</td>
<td>1.24 (1.24–1.24)</td>
<td>1.35 (1.34–1.35)</td>
<td>1.27 (1.27–1.28)</td>
<td>1.22 (1.22–1.23)</td>
</tr>
<tr>
<td></td>
<td>108</td>
<td>1.29 (1.18–1.36)</td>
<td>1.20 (1.15–1.24)</td>
<td>1.31 (1.27–1.34)</td>
<td>1.16 (1.11–1.21)</td>
<td>1.20 (1.15–1.23)</td>
</tr>
</tbody>
</table>

Note. Regional thresholds for HRI. Relative increases (e.g., 1.45 = 45% increase) in morbidity for WBGT (compared at each WBGT flag level, e.g., morbidity at 88°F assessed relative to morbidity at 85°F) and the Heat Index (compared at 2.5°F increments). 95% confidence intervals included in parentheses. Bolded values represent identified thresholds.

Figure 2.3. *Relationships between Heat Stress Metrics and HRI.*

![Graph showing relationships between daily maximum Heat Index, WBGT, and air temperature](image)

Note. Relationships between HRI (per 100,000 persons) and the three heat indices (daily maximum Heat Index, WBGT, and air temperature) for each county region. Shaded regions represent 95% confidence interval.

HRI increased more linearly as modeled by Heat Index, compared to the more exponential patterns seen with WBGT and air temperature (Figure 2.3). Notable increases in HRI occurred at two Heat Index thresholds: 103°F and 108°F in all regions, with increases at these thresholds ranging from 1.2–1.35 (Table 2.2). The urban piedmont had the highest relative increase at these thresholds (Table 2.2).
Figure 2.4. Relationships between Heat Stress Metrics and Mental Morbidity.

Note. Relationships between mental health morbidity (per 100,000 persons) and the three heat indices (daily maximum Heat Index, WBGT, and air temperature) for each county region. Shaded regions represent 95% confidence interval.

Relationships in all regions were statistically significant and positive (increasing morbidity as temperature increased) between temperature and mental health morbidity, except WBGT in the rural piedmont and air temperature in the urban piedmont (Figure 2.4). Mental health morbidity in the urban piedmont and coastal plain increased only at the highest values of the Heat Index, with the largest increase in morbidity in the urban piedmont (1.25 [1.21–1.29]) and urban coastal plain (1.13 [1.10–1.16]) occurring at a Heat Index of 113°F (Table 2.3). In the other regions, the thresholds occurred at notably cooler values: 98°F in the mountains and 103°F in both the rural piedmont and rural coastal plain (Table 2.3).
Table 2.3. Mental Health Morbidity Thresholds.

<table>
<thead>
<tr>
<th>°F</th>
<th>Mountains</th>
<th>Rural Piedmont</th>
<th>Urban Piedmont</th>
<th>Urban Coastal Plain</th>
<th>Rural Coastal Plain</th>
</tr>
</thead>
<tbody>
<tr>
<td>88</td>
<td>1.23 (1.24–1.23)</td>
<td>1.00 (1.00–1.00)</td>
<td>1.01 (1.01–1.01)</td>
<td>1.13 (1.13–1.13)</td>
<td>1.10 (1.10–1.10)</td>
</tr>
<tr>
<td>90</td>
<td>1.23 (1.22–1.24)</td>
<td>1.00 (1.00–1.00)</td>
<td>1.01 (1.01–1.02)</td>
<td>1.08 (1.07–1.08)</td>
<td>1.08 (1.07–1.08)</td>
</tr>
<tr>
<td>95</td>
<td>1.59 (1.42–1.73)</td>
<td>1.01 (0.99–1.03)</td>
<td>1.03 (1.01–1.05)</td>
<td>1.10 (1.06–1.14)</td>
<td>1.14 (1.11–1.16)</td>
</tr>
<tr>
<td>98</td>
<td>1.05 (1.04–1.05)</td>
<td>1.07 (1.07–1.07)</td>
<td>1.00 (0.99–1.00)</td>
<td>1.08 (1.08–1.09)</td>
<td>1.07 (1.07–1.07)</td>
</tr>
<tr>
<td>103</td>
<td>1.01 (1.00–1.01)</td>
<td>1.11 (1.10–1.11)</td>
<td>0.92 (0.92–0.92)</td>
<td>1.10 (1.10–1.11)</td>
<td>1.10 (1.10–1.11)</td>
</tr>
<tr>
<td>108</td>
<td>0.94 (0.93–0.94)</td>
<td>1.04 (1.03–1.04)</td>
<td>1.01 (1.01–1.02)</td>
<td>1.12 (1.11–1.12)</td>
<td>1.08 (1.07–1.08)</td>
</tr>
<tr>
<td>113</td>
<td>0.86 (0.82–0.90)</td>
<td>0.92 (0.89–0.95)</td>
<td>1.25 (1.21–1.29)</td>
<td>1.13 (1.10–1.16)</td>
<td>1.01 (0.98–1.04)</td>
</tr>
</tbody>
</table>

Note. Regional thresholds for mental health morbidity based on WBGT and the Heat Index. Relative increases in morbidity for WBGT (compared at each WBGT flag level, e.g., morbidity at 88°F assessed relative to morbidity at 85°F) and the Heat Index (compared at 2.5°F increments). 95% confidence intervals are included in parentheses. Bolded values depict the identified thresholds.

The character of the relationship between mental health morbidity and increasing temperature notably differs from the patterns with all other morbidities, with WBGT displaying a more prominent “j-shaped” curve compared to air temperature and the Heat Index, except for the relationship between Heat Index and mental morbidity in the urban coastal plain (Figure 2.4). While mental health morbidity increased exponentially across all WBGT values higher than 85°F in the mountains, the largest increase occurred at a WBGT value of 95°F (relative to 92°F) (1.59 [1.42–1.73]) (Table 2.3), which was also a threshold found for the rural coastal plain (Table 2.3). Otherwise, except for the rural piedmont, where no statistically significant increase was defined, the thresholds in all other regions occurred at 88°F (red flag), with the lowest increase in the urban piedmont (Table 2.3). As with HRI, the thresholds in the two urban county regions were at higher Heat Index values than in rural regions (Table 2.3).

Similarly, thresholds for all-cause morbidity in the two urban regions were only at the highest values of the Heat Index for all-cause morbidity (113°F) (Table 2.4). These thresholds for both the rural piedmont and rural coastal plain were 103°F and 108°F, respectively (Table 2.4).
Table 2.4. All Cause Morbidity Thresholds.

<table>
<thead>
<tr>
<th>All-Cause Thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>°F</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>WBGT 85</td>
</tr>
<tr>
<td>95</td>
</tr>
<tr>
<td>Heat Index 103</td>
</tr>
<tr>
<td>108</td>
</tr>
<tr>
<td>113</td>
</tr>
</tbody>
</table>

Note. Regional thresholds for all-cause morbidity based on WBGT and the Heat Index. Relative increases in morbidity for WBGT (compared at each WBGT flag level, e.g., morbidity at 88°F assessed relative to morbidity at 85°F) and the Heat Index (compared at 2.5°F increments). 95% confidence intervals are included in parentheses. Bolded values depict the identified thresholds.

The relationship between all-cause morbidity and WBGT in the urban piedmont was significantly more 'j-shaped' than any other morbidity-index relationship in this region, with increasing morbidity beginning at WBGT values in the upper 80°Fs (Figure 2.5). WBGT thresholds were 85°F in the mountains, urban coastal plain, and rural coastal plain. The threshold for the urban piedmont was at a WBGT of 95°F, at which the largest regional morbidity increase occurred, 1.12 (1.11–1.13) (Table 2.4).

Figure 2.5. Relationships between Heat Stress Metrics and All-Cause Morbidity.

Note. Relationships between all-cause morbidity (per 100,000 persons) and the three heat indices (daily maximum Heat Index, WBGT, and air temperature) for each county region. Shaded regions represent 95% confidence interval.
Finally, cardiovascular morbidity increased with rising temperature, as modeled by all heat stress indices in all regions, except for the rural and urban piedmont, as modeled by WBGT (Figure 2.6). The lowest threshold occurred in the mountains and rural coastal plain (85°F), followed by the urban coastal plain (88°F) (Table 2.5).

**Table 2.5. Cardiovascular Morbidity Thresholds.**

<table>
<thead>
<tr>
<th>°F</th>
<th>Mountains</th>
<th>Rural Piedmont</th>
<th>Urban Piedmont</th>
<th>Urban Coastal Plain</th>
<th>Rural Coastal Plain</th>
</tr>
</thead>
<tbody>
<tr>
<td>WBGT</td>
<td>85</td>
<td>1.05 (1.05–1.05)</td>
<td>0.99 (0.99–0.99)</td>
<td>0.99 (0.99–0.99)</td>
<td>1.05 (1.05–1.05)</td>
</tr>
<tr>
<td></td>
<td>88</td>
<td>1.03 (1.03–1.04)</td>
<td>0.98 (0.98–0.98)</td>
<td>0.95 (0.95–0.95)</td>
<td>1.07 (1.07–1.07)</td>
</tr>
<tr>
<td>Heat Index</td>
<td>100</td>
<td>1.00 (1.00–1.00)</td>
<td>1.02 (1.02–1.02)</td>
<td>0.97 (0.97–0.97)</td>
<td>1.01 (1.01–1.01)</td>
</tr>
<tr>
<td></td>
<td>103</td>
<td>1.00 (1.00–1.01)</td>
<td>1.02 (1.02–1.03)</td>
<td>0.96 (0.95–0.96)</td>
<td>1.01 (1.01–1.01)</td>
</tr>
<tr>
<td></td>
<td>113</td>
<td>0.96 (0.95–0.98)</td>
<td>1.00 (0.99–1.02)</td>
<td>1.15 (1.13–1.17)</td>
<td>1.01 (1.00–1.02)</td>
</tr>
</tbody>
</table>

*Note. Regional thresholds for cardiovascular morbidity based on WBGT and the Heat Index. Relative increases in morbidity for WBGT (compared at each WBGT flag level, e.g., morbidity at 88°F assessed relative to morbidity at 85°F) and the Heat Index (compared at 2.5°F increments). Bolded values represent identified thresholds. 95% confidence intervals are included in parentheses.*

Similarly, the coolest Heat Index threshold for cardiovascular morbidity occurred in the mountains at 100°F. For other regions, the thresholds were 103°F in the rural piedmont and urban coastal plain and 113°F in the rural coastal plain and urban piedmont (Table 2.5). The character of the relationship defined by WBGT in the urban coastal plain was notably more ‘j-shaped’ when compared to all other relationships for cardiovascular morbidity (Figure 2.6).
Figure 2.6. Relationships between Heat Stress Metrics and Cardiovascular Morbidity.

Note. Relationships between cardiovascular morbidity (per 100,000 persons) and the three heat indices (daily maximum Heat Index, WBGT, and air temperature) for each county region. Shaded regions represent 95% confidence interval.

Relationships with Morbidity across Age Groups

Relationships between morbidity and heat stress indices across age groups revealed strong, positive relationships. As was the case when assessing relationships by region, HRI and mental health morbidity had the most characteristic increase as a function of temperature compared to other morbidities. Besides air temperature, which performed the poorest when comparing model AIC values, all age groups had similar HRI thresholds based on Heat Index (103°F) and WBGT (88°F–90°F) (Table 2.6). Further, the value at which increasing morbidity peaked occurred at 94°F for WBGT. Notably, peak morbidity increases defined by the Heat Index for ages 0–12 and 55–64 occurred at 108°F and 113°F, respectively (Table 2.6). The magnitude increases in morbidity at the WBGT thresholds were larger across all age groups than the thresholds identified for the Heat Index and air temperature, with the highest morbidity burden occurring in ages 20–34, 35–54, and 55–64 (Table 2.6).
Table 2.6. HRI Morbidity Thresholds by Age Group.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Threshold (°F)</th>
<th>Peak (°F)</th>
<th>Threshold (°F)</th>
<th>Peak (°F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-12</td>
<td>90 1.29 (1.24–1.31)</td>
<td>94 1.81 (1.26–2.36)</td>
<td>103 1.30 (1.19–1.35)</td>
<td>108 1.35 (1.00–1.46)</td>
</tr>
<tr>
<td>13-19</td>
<td>88 1.22 (1.20–1.26)</td>
<td>94 1.40 (0.50–1.73)</td>
<td>103 1.28 (1.21–1.32)</td>
<td></td>
</tr>
<tr>
<td>20-34</td>
<td>88 1.43 (1.41–1.45)</td>
<td>94 1.41 (0.92–1.66)</td>
<td>103 1.28 (1.26–1.29)</td>
<td></td>
</tr>
<tr>
<td>35-54</td>
<td>88 1.61 (1.55–1.66)</td>
<td>94 1.43 (1.12–1.59)</td>
<td>103 1.25 (1.23–1.26)</td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>88 1.53 (1.49–1.55)</td>
<td>94 1.65 (0.06–2.16)</td>
<td>103 1.24 (1.20–1.27)</td>
<td>113 1.37 (1.06–1.50)</td>
</tr>
<tr>
<td>65+</td>
<td>90 1.28 (1.28–1.28)</td>
<td>94 1.62 (0.74–1.92)</td>
<td>103 1.25 (1.21–1.28)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Thresholds for HRI for each age group based on the Heat Index and WBGT. Threshold and peak morbidity temperature were equal if peak morbidity is not specified. 95% confidence interval included in parentheses. Relative increases in morbidity for WBGT (compared at each WBGT flag level, e.g., morbidity at 88°F assessed relative to morbidity at 85°F) and the Heat Index (compared at 2.5°F increments).

Table 2.7. Mental Health Morbidity Thresholds by Age Group.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Threshold (°F)</th>
<th>Peak (°F)</th>
<th>Threshold (°F)</th>
<th>Peak (°F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-12</td>
<td>90 1.10 (1.06–1.12)</td>
<td>94 1.32 (0.49–1.69)</td>
<td>93 1.01 (1.01–1.01)</td>
<td></td>
</tr>
<tr>
<td>13-19</td>
<td>88 1.09 (1.08–1.11)</td>
<td>94 1.11 (0.88–1.26)</td>
<td>103 1.09 (1.06–1.11)</td>
<td></td>
</tr>
<tr>
<td>20-34</td>
<td>88 1.11 (1.10–1.11)</td>
<td>94 1.12 (1.01–1.20)</td>
<td>98 1.07 (1.07–1.08)</td>
<td>103 1.07 (1.06–1.08)</td>
</tr>
<tr>
<td>35-54</td>
<td>88 1.10 (1.10–1.11)</td>
<td>94 1.11 (1.02–1.18)</td>
<td>98 1.07 (1.07–1.07)</td>
<td>103 1.06 (1.05–1.07)</td>
</tr>
<tr>
<td>55-64</td>
<td>88 1.10 (1.09–1.11)</td>
<td>94 1.11 (1.06–1.17)</td>
<td>98 1.07 (1.06–1.07)</td>
<td>103 1.07 (1.06–1.09)</td>
</tr>
<tr>
<td>65+</td>
<td>88 1.05 (1.04–1.05)</td>
<td>103 1.04 (1.02–1.05)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Thresholds for mental health morbidity for each age group based on the Heat Index and WBGT. Threshold and peak morbidity temperature were equal if peak morbidity is not specified. 95% confidence interval included in parentheses. Relative increases in morbidity for WBGT (compared at each WBGT flag level, e.g., morbidity at 88°F assessed relative to morbidity at 85°F) and the Heat Index (compared at 2.5°F increments).

Similar to HRI, the thresholds for mental health morbidity were relatively similar across age groups (Table 2.7). Those under 64 had the largest morbidity increase, with the threshold occurring at 88°F WBGT, except for ages 0–12 (90°F) (Table 2.7).

The relationships between all-cause morbidity and heat indices differed from the patterns seen with HRI and mental health morbidity by having lower magnitude increases in morbidity at all thresholds (Table 2.8). Further, the Heat Index values at which these thresholds occurred were higher than those for HRI (Table 2.6) and mental health (Table 2.7). Ages 20–34, 35–54, and 55–64 had the largest increases in the WBGT thresholds for all-cause morbidity (Table 2.8).
Table 2.8. All-Cause Morbidity Thresholds by Age Group.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Threshold (°F)</th>
<th>Peak (°F)</th>
<th>Threshold (°F)</th>
<th>Peak (°F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-12</td>
<td>85 1.00 (1.00-1.00)</td>
<td>110 1.00 (1.00-1.01)</td>
<td>108 1.02 (1.01-1.02)</td>
<td>108 1.01 (1.01-1.02)</td>
</tr>
<tr>
<td>13-19</td>
<td>88 1.01 (1.01-1.02)</td>
<td>108 1.02 (1.01-1.02)</td>
<td>108 1.01 (1.01-1.02)</td>
<td></td>
</tr>
<tr>
<td>20-34</td>
<td>88 1.03 (1.03-1.03)</td>
<td>108 1.02 (1.01-1.02)</td>
<td>108 1.01 (1.01-1.02)</td>
<td></td>
</tr>
<tr>
<td>35-54</td>
<td>88 1.04 (1.04-1.04)</td>
<td>108 1.02 (1.01-1.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>85 1.04 (1.03-1.04)</td>
<td>108 1.02 (1.01-1.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65+</td>
<td>85 1.02 (1.02-1.02)</td>
<td>108 1.01 (1.01-1.02)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Thresholds for all-cause morbidity for each age group based on the Heat Index and WBGT. Threshold and peak morbidity temperature were equal if peak morbidity is not specified. 95% confidence interval included in parentheses. Relative increases in morbidity for WBGT (compared at each WBGT flag level, e.g., morbidity at 88°F assessed relative to morbidity at 85°F) and the Heat Index (compared at 2.5°F increments).

Similar to all-cause morbidity, the threshold for those aged 35 and older for cardiovascular morbidity was at higher values of the Heat Index (110°F) compared to younger age groups, although the morbidity increases at these thresholds were low, ranging from 1.01-1.02 (Table 2.9).

Table 2.9. Cardiovascular Morbidity Thresholds by Age Group.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Threshold (°F)</th>
<th>Peak (°F)</th>
<th>Threshold (°F)</th>
<th>Peak (°F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-12</td>
<td>85 1.00 (1.00-1.01)</td>
<td>96 1.17 (1.76-9.39)</td>
<td>93 1.00 (1.00-1.00)</td>
<td></td>
</tr>
<tr>
<td>13-19</td>
<td>88 1.00 (0.97-1.02)</td>
<td>100 1.02 (1.01-1.02)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-34</td>
<td>85 1.04 (1.03-1.04)</td>
<td>110 1.01 (1.01-1.02)</td>
<td>113 1.03 (1.00-1.06)</td>
<td></td>
</tr>
<tr>
<td>35-54</td>
<td>85 1.02 (1.02-1.02)</td>
<td>110 1.02 (1.01-1.03)</td>
<td>113 1.04 (1.01-1.08)</td>
<td></td>
</tr>
<tr>
<td>55-64</td>
<td>85 1.02 (1.02-1.02)</td>
<td>110 1.01 (1.00-1.01)</td>
<td>113 1.02 (1.00-1.04)</td>
<td></td>
</tr>
<tr>
<td>65+</td>
<td>85 1.02 (1.02-1.02)</td>
<td>110 1.01 (1.00-1.01)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Thresholds for cardiovascular morbidity for each age group based on the Heat Index and WBGT. Threshold and peak morbidity temperature were equal if peak morbidity is not specified. 95% confidence interval included in parentheses. Relative increases in morbidity for WBGT (compared at each WBGT flag level, e.g., morbidity at 88°F assessed relative to morbidity at 85°F) and the Heat Index (compared at 2.5°F increments).

Contrasting this, WBGT flag thresholds occurred at lower temperatures for cardiovascular morbidity compared to other morbidities for ages 20–54 (85°F). Although the WBGT threshold for ages 0–12 was 85°F, morbidity for this age group peaked at a WBGT of 96°F. However, the 95% confidence interval ranges for these values were very wide (Table 2.9).
Relationships between the heat stress indices and respiratory morbidity were not significant and thus are not included here.

**Discussion and Conclusion**

This research addressed four specific questions. First, do the defined relationships between heat stress and ED visits differ when using weather station (point) data vs. ERA5-Land (raster) data, and are the relationships defined with one or the other stronger? The research presented here found the relationships were more statistically robust when utilizing weather station data, based on comparing model AIC values and chi-squared tests on the difference in minimized smoothing parameter selection score (fREML score). This result, particularly as it relates to WBGT, is unsurprising due to the challenge of adequately accounting for differences in microclimate when interpolating observations across a landscape. Specifically, wind speed biases arising during interpolation are not only likely due to the wide variations in surface roughness across land cover types (Verkaik et al., 1998), but particularly critical since WBGT is extremely sensitive to wind (Budd, 2008; Havenith & Fiala, 2016). Biases with wind speed would be most amplified during periods of low wind since this is when WBGT would be highest and most variable due to small wind speed fluctuations. Failing to capture these extremes (i.e., account for this variability) occurring when heat stress risk would also be the greatest would significantly impact the ability of a statistical model to explain variance in morbidity. Additionally, the proximity of weather stations to population centers may explain the statistical superiority of using weather station data versus the ERA5-Land data set.

Second, the three heat stress indices (air temperature, the Heat Index, and WBGT) showed variations in the character and strength of their relationship with morbidity. Daily maximum WBGT was the strongest predictor of all morbidity types, followed by daily maximum Heat Index and daily maximum air temperature. The more linear relationships defined between the Heat Index and HRI relative to the relationships defined with the other heat stress indices are hypothesized to result
from the broader numerical range of more dangerous values of the Heat Index (e.g., 100°F–113°F) relative to WBGT (e.g., 88°F–92°F). Therefore, WBGT is better at characterizing increases in morbidity at higher temperatures.

Furthermore, two morbidities were found to have more characteristic relationships with increasing temperatures, i.e., a “j-shaped” curve: HRI and mental health morbidity. This pattern persisted when assessing the relationships between heat and age-specific morbidity. WBGT thresholds for HRI were similar across all age groups. This is a particularly interesting finding since the threshold for older age groups would be expected to be at lower values given that population's higher vulnerability (Fuhrmann et al., 2016; Kovats & Hajat, 2008; Páldy et al., 2005; Petkova et al., 2017; Semenza et al., 1999). For the other morbidities investigated, older populations did have lower WBGT thresholds and values at which peak morbidity occurred. For cardiovascular morbidity, the finding of higher Heat Index thresholds for older relative to younger populations is hypothesized to be related to the smaller sample size in younger ages, thus possibly confounding the results. Furthermore, this also may result from the less defined and smaller increases in morbidity with temperature for younger populations, particularly in comparison to HRI (Table 2.2) and mental health (Table 2.3).

Although there were different threshold values across ages and morbidities, the magnitudes of morbidity increases were relatively similar. However, this was not the case for WBGT thresholds for ages 20–34 and 35–54, which had larger morbidity increases than other age groups. It is hypothesized that WBGT may better correlate with occupation-related exposure, as this age group has a higher proportion of outdoor workers. Investigating this pattern by assessing differences across sex may further explain this finding.

Third, distinct regional differences in the relationships between heat and morbidity were identified. These differences varied based on the heat stress index used and were ultimately a
function of underlying population differences between regions. Overall, and particularly with HRI, this research confirmed the findings of Kovach et al. (2015) and Sugg et al. (2016), with rural areas having a higher morbidity burden than their urban counterparts. This is specifically seen in the rural coastal plain, with WBGT being the most statistically robust differentiator of the higher morbidity burden in this region. Evidence of the complexity of the identified relationships, however, the urban piedmont had the highest burden for all-cause morbidity at extremely warm values (e.g., WBGT > 91°F). Overall, rural regions generally experienced larger increases at lower temperatures relative to their urban counterparts. It is hypothesized that the rural-urban differences are tied to the proportion of the population engaged in outdoor labor. Rural areas have a higher number of outdoor workers, such as in agriculture (Hanna et al., 2011; Sheridan & Dolney, 2003) and higher rates of energy poverty, which is when households cannot afford the cost of energy to meet basic needs, e.g., air conditioning (Nussbaumer et al., 2012). Although the heat-morbidity burden is highest in rural areas in all other cases, this finding aligns with other research about urban heat-mortality burden in NC being highest (Clark, 2019).

A few relationships between morbidity and heat stress measures were weak and did not exhibit a characteristic form. There are a few possible explanations. The relatively higher topographic variations in these regions (mountains and rural piedmont) make it more challenging to capture variations in WBGT across the landscape adequately, and thus accurately characterize exposure. Furthermore, there are fewer weather stations that directly measure solar radiation, specifically ECONet stations in the rural piedmont (Figure 2.1), which provides the most robust basis for estimating WBGT compared to derived solar radiation from cloud cover.

The final research question assessed the utility of current WBGT flag thresholds and heat index thresholds commonly used by the US NWS. While the thresholds for the issuance of heat products differ across the seven forecast offices covering the state, a majority of the state is
covered by offices following the NWS national procedural guidance, where the Heat Index thresholds for heat advisories and warnings are 105°F and 110°F, respectively (Hawkins et al., 2017; NWS Instruction 10-515, 2018). The results presented here emphasize the need for region-specific thresholds based on the Heat Index and reassessing the existing thresholds overall, since the largest increases in all morbidities under investigation occurred at cooler values. For example, the Heat Index thresholds for HRI across all age groups were 103°F and the regional thresholds identified were 103°F and 108°F (relative to 100°F and 105°F, respectively). Therefore, lowering the values of the current thresholds used to issue heat products by several degrees would more soundly advise against activity when conditions are becoming most dangerous and limit morbidity.

The findings here determined that existing WBGT flag thresholds for the two highest risk categories, red and black flag, overwhelmingly corresponded to the largest increases in HRI and mental health morbidity for all age groups, and all-cause morbidity for those between the ages of 13 and 54. While a WBGT red flag (88°F) corresponded to the largest HRI increases in all county regions, assessing changes in morbidity at 1°F increments revealed regional differences in the thresholds and associated magnitudes of morbidity increases (Appendix 2.2). This emphasizes the importance of both age and region-specific WBGT guidelines.

Modifications have been made to existing WBGT thresholds in some applications, for example, with the Georgia High School Athletic Association increasing the threshold for the most dangerous category from 90°F to 92°F (Georgia High School Athletic Association, 2020). While the results here show peak morbidity occurring at values greater than 90°F and 92°F in some instances, primarily with HRI (Table 2.6) and mental health morbidity (Table 2.7), confidence at these values, which are on the tail of the distribution, was low, with wide 95% confidence intervals, relative to the confidence intervals at the identified thresholds of largest morbidity increase. There were no instances of 92°F relative to 90°F instead of 90°F relative to 88°F representing a statistically
significant (non-overlapping confidence intervals) larger increase in morbidity. Thus, this research provides support for continued recognition of WBGT black flag conditions beginning at 90°F.

It is important to note that this study had several limitations. First, the use of WBGT as a predictor of morbidity is complicated by the lack of in situ WBGT observations, thus WBGT was estimated from weather station and ERA5-Land data. Further, there were differences in how the WBGT was estimated based on the weather station network used, as some stations directly recorded solar radiation while others reported only cloud cover. As is the case in similar studies, the weather data utilized may not accurately capture the conditions to which a given individual experiencing morbidity was exposed and assumes that environmental heat was the source of their exposure. Lastly, particularly with morbidities such as all-cause and cardiovascular morbidity, there are numerous confounding variables that vary by individual which could not be accounted for in this study.

The use of WBGT to assess heat stress risk continues to expand in adoption and in different applications, such as in occupational health (ISO, 2017; OSHA, 2017) and athletic safety (Grundstein et al., 2015; NCHSAA, 2016). The results of this study reveal that WBGT effectively predicts morbidity at a population-level and is a stronger predictor than the Heat Index and air temperature. This finding satisfies one of the criteria to consider using WBGT in a Heat Health Warning System (HHWS): the identification of the conditions most dangerous at a population-level (Kovats & Kristie, 2006). An additional criterion for an effective HHWS is the ability to reliably forecast the heat stress index, presenting a focus for further research on WBGT.
CHAPTER 3: MICROCLIMATIC VARIATION OF WET BULB GLOBE TEMPERATURE

The paramount importance of investigating environmental heat stress is evident with the fact that exposure to heat is the leading cause of weather-related death in the United States (CDC, 2010; NOAA’s National Weather Service, 2020). To determine how dangerous a given set of environmental conditions are to human health, heat stress indices are utilized, with the most common example in the United States being the Heat Index (Hawkins et al., 2017). While the Heat Index is commonly used in the United States, wet bulb globe temperature (WBGT) is a more robust heat stress index since it accounts for the effect of air temperature, humidity, solar radiation, and wind speed on human body temperature (Budd, 2008; Hondula et al., 2014). Comparatively, the Heat Index does not account for the effect of solar radiation nor wind speed on body temperature, which is a commonly cited limitation (Budd, 2008; Hondula et al., 2014).

Background

Wet bulb globe temperature (WBGT)

WBGT was developed in the 1950s by the United States military in response to numerous heat-related casualties at training camps (Budd, 2008; Yaglou & Minard, 1957). To calculate the WBGT, three components are utilized: dry-bulb temperature, natural wet bulb temperature, and black globe temperature (Budd, 2008).

Dry-bulb temperature is simply the ambient air temperature. The natural wet bulb temperature is similar to the dry bulb temperature, except it is 1) not measured within a radiation shield and 2) has a wet wick wrapped around the bulb of the thermometer (Liljegren et al., 2008). Thus, the natural wet bulb temperature is able to account for the influence of solar radiation since
it is unshielded from the radiative environment and also mimics the cooling effect of sweat evaporating off of the skin. The natural wet bulb temperature accounts for the following environmental variables influence on human body temperature: solar radiation, wind speed, humidity, and air temperature (Budd, 2008; Liljegren et al., 2008). Similarly, the black globe temperature is also sensitive to these environmental variables. However, the black globe temperature is a particularly better indicator of the radiative forcings active on the body, i.e., the radiant temperature (both shortwave and longwave radiation) (Budd, 2008; Kopec, 1977; Liljegren et al., 2008). These three components are then summed with the following weightings to derive WBGT:

\[
WBGT = 0.7 \times NWB + 0.2 \times Tg + 0.1 \times Ta,
\]

where \( NWB \) is the natural web bulb temperature, \( Tg \) is the black globe temperature, and \( Ta \) is the dry bulb temperature (air temperature).

Similar to other heat stress indices, each WBGT value has an associated level of danger, often referred to as “WBGT Flag Level”. The values defining the thresholds for each flag level vary. For example, the threshold for a black flag is set as 92°F by some organizations but is most commonly 90°F. The most common thresholds are presented in the following table:
Table 3.1. *WBGT Safety Guidelines from the North Carolina High School Athletic Association.*

<table>
<thead>
<tr>
<th>WBGT (<em>F)</em></th>
<th>Athletic Activity Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 80</td>
<td>Unlimited activity with primary cautions for new or unconditioned athletes or extreme exertion; schedule mandatory rest/water breaks (5 min water/rest break every 30 min).</td>
</tr>
<tr>
<td>80-84.9</td>
<td>Normal practice for athletes; closely monitor new or unconditioned athletes and all athletes during extreme exertion. Schedule mandatory rest/water breaks (5 min water/rest break every 25 min).</td>
</tr>
<tr>
<td>85-87.9</td>
<td>New or unconditioned athletes should have reduced intensity practice and modifications in clothing. Well-conditioned athletes should have more frequent rest breaks and hydration as well as cautious monitoring for symptoms of heat illness. Schedule frequent mandatory rest/water breaks (5 min water/rest break every 20 min). Have cold or ice immersion pool on site for practice.</td>
</tr>
<tr>
<td>88-89.9</td>
<td>All athletes must be under constant observation and supervision. Remove pads and equipment. Schedule frequent mandatory rest/water breaks (5 min water/rest break every 15 min). Have cold or ice immersion pool on site for practice.</td>
</tr>
<tr>
<td>90+</td>
<td>SUSPEND PRACTICE/MUST INCLUDE MANDATORY BREAKS AS DIRECTED BY GAME DAY ADMINISTRATOR DURING CONTEST.</td>
</tr>
</tbody>
</table>

Note. Adapted from Raleigh NWS WFO (2017).

Although developed back in the 1950s, WBGT is utilized in numerous applications and its use continues to rise. For example, while not required, WBGT is the recommended method for measuring heat stress by the United States Occupational Safety and Health Administration (OSHA, 2017). Additionally, an increasing number of US states require WBGT measurements to determine if athletic practices can continue safely (Grundstein et al., 2015; NCHSAA, 2016), including North Carolina, South Carolina, Georgia, Florida, Minnesota, Massachusetts, New Jersey, Vermont, and Arkansas.

Despite the growing utilization of WBGT, several challenges with the index remain: First, WBGT is typically measured using a device that either directly measures or approximates each individual component of the index. Devices measuring each component directly are the most accurate but are more expensive. Furthermore, some cheaper devices use the psychrometric wet bulb temperature instead of the natural wet bulb temperature, which is problematic since the psychrometric wet bulb temperature can be significantly cooler, particularly at low wind speeds (Kopec, 1977). However, various empirical and statistical models have been developed to estimate...
WBGT based on variables that weather stations widely record. The developed methodologies vary in accuracy and computational complexity, with the physically based model in Liljegren et al. (2008) found to be the most accurate (Lemke & Kjellstrom, 2012; Patel et al., 2013). While estimates of WBGT can provide useful information in evaluating the heat stress of an environment, direct measurement of the conditions onsite at the time of any outdoor activity remains critical since WBGT is very sensitive to slight changes in environmental conditions, such as wind speed and solar radiation (Budd, 2008).

**Microclimatic influences on Heat Stress**

The high sensitivity of WBGT to slight changes in environmental conditions does not represent an issue with the index since human body temperature is equally as sensitive. A large source of small-scale, local variations in the environmental variables determining thermal stress (humidity, wind speed, and radiation) is land cover.

Local wind speed is strongly influenced by the density, location, and height of trees and buildings in the immediate vicinity (US EPA, 2000). Above buildings and trees (30–50 feet (9–15 meters) above the ground), there are typically only small differences in the wind speed across a region (e.g., 2 miles or more) (Oke, 1987). However, objects extending from the Earth's surface, primarily trees and buildings, result in increasing friction as height above the surface decreases, slowing wind speeds down relative to the wind higher in the atmosphere. The difference in the amount of friction, referred to as surface roughness, between locations corresponds to differences in the degree of wind deceleration at the surface. This, in turn, can lead to marked increases in WBGT and thermal stress on hot days. The high sensitivity of WBGT to changes in wind speed can be seen in Figure 3.1, which displays an example where there can be differences of up to two flag levels when wind speeds decrease from 3 to 1.5 mph on sunny days.
Figure 3.1. *Relationship between WBGT and wind speed.*

Note. WBGT was estimated using Liljegren et al. (2008) method. Relationship depicted for three categories of cloud cover, with an air temperature of 86°F (30°C) and dew point of 70°F (21.1°C).

To estimate WBGT from weather station observations, wind speed should be measured at 2 meters above the ground. Since most weather stations measure wind 10 meters above the surface, it must be translated into a 2-meter wind speed using a vertical wind profile and logarithmic transformation. An example of this concept can be seen in Figure 3.2. Utilizing wind speeds at 2 meters is important since this is the general height of humans and thus the height of the wind that impacts body temperature, and since the wind speed at 10 meters is consistently faster. Given the sensitivity of WBGT to wind speed, accuracy with this downscaling is critical.
Variations in surface type, e.g., grass versus a tennis court, produce varying degrees of thermal stress since different surfaces absorb and reflect varying amounts of radiation, which determines the amount of heat absorbed by the surface and its temperature (Oke, 1987). This temperature, often called skin temperature, is therefore proportional to the amount of longwave radiation emitted from the surface and incident on the human body. Since skin temperature is largely determined by the amount of direct solar radiation incident on the surface, which is a function of cloud cover and atmospheric transmissivity, the same surface can have vastly different temperatures based on slight changes in radiation. The main factors influencing incident radiation, especially over small spatial and temporal scales, are land cover (shade from vegetation) and cloud cover (US EPA, 2000). Like wind speed, changes in radiation contribute to large and sudden swings in WBGT at the same location over the course of only minutes. Thus, to most reliably capture the overall thermal stress on the body during a period, WBGT should be averaged over a set time interval, with literature recommending a variety of intervals, such as a 5-minute (Thorsson et al., 2007), 30-minute (Kopec, 1977), or 60-minute average (OSHA, 2017).

The third environmental variable contributing to microclimatic variations in WBGT is humidity. Thermally, changes in humidity determine the rate of evaporation of sweat off human
skin, and thus the rate at which the body can cool itself through evaporative cooling. Like wind speed and radiation, land cover is critical since nearby bodies of water, specifically the evaporation of that water, significantly contribute to local humidity levels (Oke, 1987). Further, soil characteristics that vary across different soil types, such as retention capacity and drainage, result in varying amounts of water evaporating from the soil and thus local differences in humidity. Differences in soil moisture are yet another factor influencing skin temperature. Wet soil has a greater latent and smaller sensible heat flux than dry soil, meaning dry soil will be warmer than wet soil under the same amount of incident radiation (Oke, 1987).

Lastly, while these environmental variables are important to consider separately as determiners of microclimatic variability, these variables act in concert to define uniquely hot or cool microclimates. Generally, the amount of radiation a surface receives influences skin temperature and local moisture flux from bodies of water and the soil. As the surface is continually heated, the air at the surface warms and rises. The degree of surface warming influences the amount of air being warmed and the speed at which it rises, generating wind. As this rising air is replaced by the cooler and drier air above, local humidity levels are changed, and together with the wind generated from the radiative heating of the surface, the overall heat stress of a particular location and microclimate is determined.

Existing research has highlighted the variations in WBGT as a function of surface type (Grundstein & Cooper, 2020; Kosaka et al., 2018; Pryor et al., 2017), the wide range of variations in temperature across surfaces (Kopec, 1977), and the range of impacts of the overall radiative environment to human body temperature (Hardin & Vanos, 2018). Additional research has also espoused the challenges and inaccuracies associated with using WBGT and other heat stress indices estimated from a weather station far from the site of interest, emphasizing the importance of real-time, on-site measurements (Grundstein et al., 2022; Pryor et al., 2017).
The research presented here expands upon this existing literature in several ways. Compared to prior work, it examines a much larger sampling of days across microclimates in an array of athletic and suburban environments, particularly days with high thermal stress (WBGT > 88°F–90°F). Further, a full-size WBGT meter that conforms to standards outlined by the International Organization for Standardization is utilized to assess the accuracy of WBGT estimates and WBGT measurements made by other devices. Given the importance of accurately measuring and estimating WBGT since it is used to safeguard health, the bias of the Liljegren et al. (2008) method for estimating WBGT and one of the more commonly used and robust WBGT meters, the Kestrel 5400, is compared to ground truth WBGT observations. Additionally, WBGT measurements and skin temperature readings are made and compared over various surfaces and within distinct microclimates (e.g., grass, tennis courts, forests, sun, and shade).

This research further contributes to existing literature by evaluating the accuracy of different methods for estimating clear-sky radiation and the subsequent modification of that clear-sky radiation by percentage cloud cover to attain an estimate of incident radiation at the surface. This is important since WBGT is not commonly measured at weather stations and since most weather stations either have no solar radiation measurement or only provide observed cloud cover amounts. Assessing the accuracy of estimating solar radiation, accounting for cloud cover percentage, has important implications for creating forecasts of WBGT since these typically rely on a forecast cloud cover value.

**Research Questions**

The following research questions were addressed:

1) What is the overall accuracy of the Liljegren et al. (2008) methodology for estimating WBGT? Does the assumption by this method that the skin temperature is equivalent to the air temperature introduce significant bias?
2) What is the accuracy of the Kestrel 5400 WBGT compared to ground truth, and how do the measurements by these devices compare to estimated WBGT from a weather station?

3) How does skin (surface) temperature across different surface types relate to variations in WBGT (e.g., shade vs. sun; tennis court vs. grass)? How much overall difference is there in WBGT measured over these surfaces?

4) How does the accuracy of different commonly used methods for estimating clear-sky radiation compare to one another? Can they be improved?
   a. What is the accuracy of modifying clear-sky radiation by percentage cloud cover?

Data

Observed WBGT Measurements

WBGT data were recorded alongside a weather station with three different instruments: two Kestrel 5400 Heat Stress Tracker devices and a custom research unit manufactured by Kestrel to meet the specifications for a WBGT meter as outlined by the International Organization for Standardization, first used in Cooper et al. (2017) (Figure 3.3). The custom WBGT meter measures the three components of WBGT separately. The device contains a thermometer inside a naturally aspirated radiation shield for recording the dry bulb temperature. The natural wet bulb temperature is measured by a temperature probe wrapped in wicking material resting above a water reservoir, with the bottom of the wick submerged to keep the wick constantly wet. The globe temperature is measured with a matte black globe thermometer with a diameter of 0.15 meters and a temperature sensor inside (Cooper et al., 2017). As with the other instruments, the WBGT meter was designed to record data 1.5–2 meters above ground level. Lastly, dependent upon the location and the frequency of access to the location, the WBGT meter recorded data at multiple intervals (2-, 5-, 10-, 30-, 60-, or 120-second intervals). The WBGT measured by this device will hereafter be referred to as “Observed WBGT”.

45
The Kestrel 5400 devices directly measure black globe temperature, although the diameter of the globe (2.5 cm) is significantly smaller than the standard black globe thermometer (0.15 m) (Figure 3.3). The manufacturer adjusts the measurements of this smaller globe to account for this size difference. The devices do not directly record natural wet bulb temperature but instead offer approximations of this component, due to the challenges with the natural wet bulb thermometer requiring a wetted wick and thus frequent access to water.

The Kestrel 5400s were mounted atop tripods designed for the devices with the Kestrel Vane Mount holding the actual device, which allows the device to rotate freely according to wind direction. In addition to the WBGT variables, the Kestrel 5400 devices also record air temperature, dew point temperature, relative humidity, barometric pressure, and wind speed. The Kestrel 5400s were set to record data at 1.5 meters above the ground at various intervals (2-, 5-, 10-, and 30-second intervals), depending upon the location and ability to access that location regularly, detailed below.
Note. The WBGT meter (right) with a black globe thermometer, natural wet bulb temperature probe situated in a water reservoir, and dry bulb sensor in a radiation shield. A weather station (left) and a Kestrel 5400 (middle) were also co-located with the WBGT meter.

Weather Station Measurements

The weather station utilized to collect standard meteorological variables was a Davis Instruments Vantage Pro 2 Plus (Figure 3.3). The station was co-located with the WBGT meter to provide data quality checks. This weather station made measurements of the following variables with a temporal resolution of ten seconds: air temperature, dew point temperature, barometric pressure, relative humidity, solar radiation, and wind speed. The weather station instruments were mounted on a tripod to situate the station 1.5 meters above ground level. Routine calibration for relative humidity was conducted by placing the temperature and humidity sensor (Sensirion SHT31) in humidity chambers, calibrating to two points, with a sodium chloride slurry and magnesium chloride slurry for calibrating to 75% and 33%, respectively.
Additionally, a temperature probe was anchored adjacent to the ground to record skin temperature measurements, also at a temporal resolution of ten seconds. This temperature probe was manufactured by Davis Instruments (SKU 6475). Lastly, attached at the top of the weather station tripod was an IP camera positioned to take photos of the sky conditions at an interval of 10 seconds (Figure 3.3). The exact camera model ultimately utilized is not pictured in Figure 3.3. The camera was manufactured by Amcrest (model number: IP8M-2496EB-V2) with a resolution of 8 megapixels (3840x2160) and a wide viewing angle of 125 degrees. Cloud cover images were only obtained for the summer of 2021 in Durham and Shelby, NC.

**Data Collection Locations**

Measurements were taken at several locations throughout the summers of 2019-2021 in North Carolina: 1) three high schools in the Triangle Region of NC (Cedar Ridge High School in Hillsborough, NC, Green Level High School in Cary, NC, and Wake Forest High School in Wake Forest, NC), including measurements at different practice and playing fields across these campuses and tennis courts, 2) the Horace Williams Airport in Chapel Hill, NC, and 3) within suburban environments in Chapel Hill, NC, Durham, NC, and Shelby, NC.

**Methods**

**Observed WBGT**

The observed WBGT data (i.e., the individual measures of natural wet bulb temperature, black globe temperature, and dry bulb temperature) were consistently cross-referenced with Station WBGT from the weather station co-located with the WBGT meter. This was done to ensure no errant data points since the cables connecting the instruments to the data loggers sometimes became loose. On these rare occasions, this resulted in readings ranging from 50%-70% of what they should have been (e.g., the natural wet bulb temperature was supposed to be around 80°F, but the instrument suddenly reported it as 62°F for a brief period). Further, the dry bulb sensor was
compared to the dry bulb readings from the weather station to ensure consistent dry bulb temperature readings. To ensure that the instruments were properly situated with respect to sun and shade, measurement times were cross referenced with regularly captured images of the observation area and with pyranometer measurements of solar radiation from the co-located weather station.

When the WBGT meter was put in place, the first 15 minutes of data were discarded, since the instruments require time to equilibrate with the environment (Kestrel, 2021). Routine checks were conducted to ensure the water reservoir for the natural wet bulb temperature remained sufficiently full to keep the wick consistently moist. If the water level was insufficient to keep the wick moist, the natural wet bulb temperature data collected between then and the last time it was checked (the prior evening) were discarded. When additional water was required, distilled water kept in the shade at ambient air temperature levels was used. After refilling the water reservoir, the observations for the next 15 minutes were discarded to ensure the wick had sufficient time to dampen and the temperature probe equilibrate.

Due to the rapid variability of WBGT over short time intervals and for better comparison between the three devices (Observed WBGT, Kestrel 5400, and weather station), minute-level averages were calculated from the collected data by each device, from which 15-minute moving averages were derived. The 15-minute average was chosen as a compromise between other recommended averaging periods for WBGT (5- and 30-minute averages) and following Liljegren et al. (2008).

**Station WBGT**

WBGT was estimated from the weather station observations using the method developed in Liljegren et al. (2008), which has consistently been found as highly accurate (Lemke & Kjellstrom,
The estimated weather station WBGT will be hereafter referred to as “Station WBGT”.

**Clear-Sky Radiation Estimates and Image Analysis**

Estimates of clear-sky radiation from three methods were compared to measured solar radiation: 1) Kasten & Czeplak, 1980, 2) Ryan-Stolzenbach model (Ryan & Stolzenbach, 1972), and 3) Bras model (Bras, 1990). These three models will hereafter be referred to as 1) Kasten, 2) Ryan, and 3) Bras.

The Ryan and Bras models for estimating clear-sky radiation have customizable parameters to account for the impact of atmospheric transmissivity on the incident radiation at the surface, 0.0–1.0 for the Ryan model and an integer between 1–5 for the Bras model, with 5 being the most turbid. The default parameters were utilized, specifically setting the atmospheric transmissivity coefficient (ATC) to 0.8 in the Ryan model and the corresponding variable in the Bras model to 3. In addition to estimates using the default parameters, these parameters were changed to determine if more accurate estimates could be made based on the observed solar radiation by the weather station. First, however, cloud-free periods and the percentage cloud cover were identified.

To estimate the percentage cloud cover for the weather station and WBGT observations, each image corresponding to the date and times of those observations was segmented and classified, from which the total pixels of the image classified as clouds were divided by the total number of pixels in the image. The images were segmented using the simple linear iterative clustering (SLIC) method combined with the affinity propagation (AP) clustering algorithm (Zhou, 2015). Superpixels using the SLIC method were constructed and then clustered based on similarity with the AP clustering algorithm, which does not require predefining the number of clusters (Zhu et al., 2017). This was completed in R using the packages “SuperpixelImageSegmentation”
(Mouselimis, 2022b; Zhou, 2015) and “OpenImageR” (Achanta et al., 2010; Achanta et al., 2012; Buchner, 2013; Haghighat et al., 2015; Haghighat, 2015; Machine, 2012; Mouselimis, 2022a).

After the image segmentation and clustering, the color representing each cluster, which was determined by averaging the color (RGB values) across all pixels in the cluster, was converted to HSV values in R. The HSV values corresponding to clouds and clear sky were determined manually from a sample of images. Once these characteristics were determined, all images were processed. To ensure adequate performance, random selections of five images per day were selected and manually reviewed to determine if the configuration was correctly segmenting and classifying clusters as clouds or clear sky. Some clusters identified in the image had very few total pixels. When these clusters were 1) unable to be classified based on the pre-determined characteristics of clouds and clear sky and 2) exceeded 10% of the total pixels in the image, the cloud cover percentage derived from that image was not included in the analysis.

Although images were captured every 10 seconds, only images every 30 seconds were utilized from June 20th, 2021 - September 11th, 2021. Lastly, only images between the hours of 9:30 and 16:00 local time were used, due to issues with classifying the images when the sun angle was lower and since this was the period in which all instruments were consistently and fully exposed to the sun (or whatever the current sky conditions were). The percentage cloud cover was then calculated by dividing the total number of pixels within each cluster denoted as clouds by the total number of pictures in the entire image. Out of all images of sky conditions taken, 83,859 images were able to be used. 3,036 of these images were discarded due to the percentage of pixels unable to be classified being greater than 10%.

To assess the accuracy of modifying clear-sky radiation by percentage cloud cover to obtain an estimate of the solar radiation incident at the surface, the clear-sky radiation value
corresponding to the timestamp of each image was modified by the calculated percentage cloud
cover using (2):

\[ S_{rad} = R_0 \times (1 - 0.75n^{3.4}), \]  

(2)

where \( n \) is the cloud cover fraction (0.0–1.0) and \( R_0 \) is the clear-sky direct radiation (\( \text{w/m}^2 \))
detailed above) (Kasten & Czeplak, 1980).

**Analysis**

The accuracy of the Liljegren WBGT estimation and the WBGT measured by the Kestrel 5400
were assessed by comparing them to the Observed WBGT. The overall accuracy and the accuracy at
1°F WBGT increments were calculated. In addition to the general and relative biases, the accuracy
of the Liljegren method and Kestrel 5400 device were assessed using the three following accuracy
verification metrics:

1) **Hit Rate (%):** percentage of correct WBGT measurements or estimates (black flag observed,
black flag estimated, hit) (US NWS, 2019b).

2) **False Alarm Ratio:** total number of false alarms for a given WBGT divided by the total
number of observations at that WBGT (US NWS, 2019b).

3) **Bias Score:** measure indicating the direction of bias (positive/negative) in addition to the
magnitude (ratio of the frequency of measuring/estimating a WBGT to the frequency of
observations at that WBGT). Values greater than 1 correspond to positive (warm) biases.
Values less than 1 correspond to negative (cool) biases (NCAR, 2015).

Comparisons of the WBGT in a variety of locations and shade conditions were assessed,
including: 1) full sun versus full shade, 2) full sun versus shaded forest, 3) shaded field versus
shaded forest, and 4) tennis court versus grassy field. These comparisons only utilize the Kestrel
5400 WBGT since there were two devices, and concurrent measurements allowed for the most
robust comparisons. To assess the statistical significance of the differences in WBGT as measured
by the Kestrel 5400 and estimated with the Liljegren methodology, and differences across the
different surface types, Wilcoxon rank sum tests with continuity correction were utilized to compare
means. Statistical significance was evaluated at the 95% confidence level.

The skin temperature probe was co-located with the other field work instruments to
determine the impact of skin temperature on the accuracy of the estimation and Kestrel 5400 values. Since the Liljegren methodology assumes that the skin temperature is equivalent to the air
temperature for estimating the black globe temperature, comparisons of the bias for the black
globe temperature and WBGT overall were assessed as a function of the difference in air and skin
temperature.

Lastly, evaluating the accuracy of the three methods for estimating clear-sky radiation used
here first required identifying cloud-free periods. The derived cloud cover percentage from each
image and observed solar radiation readings were averaged out to a 5-minute interval to suppress
differences arising from any slight mismatches in timesteps of the images and observations. The
percentage cloud cover at a 5-minute interval was then determined to be cloud-free if it had less
than 5% cloud cover. For these periods, the estimates of clear-sky radiation from each of the three
methods were calculated and then directly compared to the corresponding 5-minute averaged
solar radiation observation. Finally, the accuracy assessment of modifying clear-sky radiation by
percentage cloud cover (2) followed those same steps, except all images were used instead of only
the cloud-free images (cloud cover values 0%-100%).

Results

Overview

First, the results for assessing the overall biases and patterns of the Station WBGT and
Kestrel 5400 WBGT will be presented, compared to Observed WBGT, followed by the assessments
of their accuracy with commonly used verification metrics. Second, skin temperature readings and
WBGT differences across surfaces and shade conditions will be detailed. Lastly, the results detail the accuracy of the methods for estimating clear-sky radiation and the ensuing modification of that clear-sky radiation by percentage cloud cover, estimated from the images of sky conditions.

**Biases in the Station and Kestrel 5400 WBGT**

Comparisons with the observed WBGT revealed that both the Station and Kestrel 5400 WBGT display a positive (warm) median bias (Figure 3.4). The Kestrel 5400 WBGT median bias was +0.6°F (95% CI: 0.6°F–0.7°F) higher than the Station WBGT (Figure 3.4).

**Figure 3.4. WBGT Biases by Device and by 2°F Increments of WBGT.**

*Note. Differences between the Station WBGT and Kestrel 5400 WBGT (left) with Observed WBGT, and differences between the Station and Kestrel 5400 WBGT (left). Difference between the Station WBGT bias as a function of the Observed WBGT, incrementing at intervals of 2°F (right).*
Table 3.2. *WBGT Thresholds: Wilcoxon Difference of Means.*

<table>
<thead>
<tr>
<th>WBGT Threshold</th>
<th>Station vs. Observed</th>
<th>Kestrel 5400 vs. Observed</th>
<th>Kestrel vs. Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.4°F (0.3°F – 0.5°F)</td>
<td>1.2°F (1.1°F – 1.3°F)</td>
<td>0.6°F (0.6°F – 0.7°F)</td>
</tr>
<tr>
<td>80</td>
<td>0.4°F (0.4°F – 0.5°F)</td>
<td>1.3°F (1.2°F – 1.3°F)</td>
<td>0.7°F (0.6°F – 0.8°F)</td>
</tr>
<tr>
<td>85</td>
<td>0.5°F (0.4°F – 0.5°F)</td>
<td>1.4°F (1.3°F – 1.4°F)</td>
<td>0.8°F (0.7°F – 0.9°F)</td>
</tr>
<tr>
<td>88</td>
<td>0.4°F (0.4°F – 0.5°F)</td>
<td>1.5°F (1.4°F – 1.6°F)</td>
<td>0.9°F (0.9°F – 1.0°F)</td>
</tr>
<tr>
<td>90</td>
<td>0.4°F (0.4°F – 0.5°F)</td>
<td>1.6°F (1.5°F – 1.7°F)</td>
<td>1.1°F (1.1°F – 1.2°F)</td>
</tr>
<tr>
<td>92</td>
<td>-0.1°F (-0.2°F – -0.1°F)</td>
<td>1.5°F (1.4°F – 1.5°F)</td>
<td>1.3°F (1.2°F – 1.3°F)</td>
</tr>
</tbody>
</table>

*Note.* Wilcoxon Difference of Means of means results for 1) Station WBGT, 2) Kestrel 5400 WBGT, and 3) Kestrel 5400 WBGT relative to Station WBGT across all WBGTs, and for when WBGT was equal to or greater than 80°F, 85°F, 88°F, 90°F, and 92°F. 95% confidence intervals are included in parentheses.

The assessment revealed a distinct pattern in the Kestrel 5400 WBGT bias (Figure 3.5). As observed WBGT increased, the Kestrel 5400 had an increasingly positive (warm) bias, with mean biases of +1.3°F (1.2°F–1.3°F) at 80°F and +1.6°F (1.5°F–1.7°F) at 90°F (Figure 3.5) (Table 3.2). At WBGTs greater than 86°F, the median and also lower 75th percentile of the Kestrel 5400 WBGT bias were positive (Figure 3.5). Contrasting this, the Station WBGT did not display a distinct pattern of changing bias as WBGT increased, with its median bias remaining between +0.1°F–+0.8°F for all WBGTs between 78°F–92°F (Figure 3.4). At extreme WBGTs of 94°F, the Station WBGT bias had a slight negative (cool) median bias of -0.7°F (Figure 3.4).
Note. Bias of the Kestrel 5400 WBGT relative to 1) Observed WBGT (left) and 2) the Station WBGT, with the x-axis indicating the Observed WBGT at 2°F increments.

The bias of the Kestrel 5400 WBGT relative to the Station WBGT became increasingly positive as Observed WBGT increased (Figure 3.5). At WBGTs less than or equal to 84°F, the Kestrel 5400 bias was slightly negative relative to the Station WBGT (Figure 3.5). However, as WBGT increased, this bias became increasingly positive (warm), with the Kestrel 5400 WBGT median bias relative to the Station WBGT being +1°F and +1.9°F, respectively, at Observed WBGTs of 90°F and 94°F (Figure 3.5). All comparisons of means between the Kestrel 5400 WBGT and Station WBGT were statistically significant (Table 3.2).

A disaggregation of the biases of the Station and Kestrel 5400 WBGT reveals that the natural wet bulb temperature for both was positively (warm) biased (Table 3.3). The magnitude of bias was lowest for the Station, with a bias of 1.3°F, compared to the Kestrel 5400, which had a bias of 2.6°F (Table 3.3). Contrasting this, the Station and Kestrel 5400 black globe temperatures were negatively (cool) biased. As was the case with the natural wet bulb temperature, the station
estimation of black globe temperature was more accurate than the Kestrel 5400, with biases of -2.6°F and -3.9°F, respectively (Table 3.3).

Table 3.3. Natural Wet Bulb and Black Globe Temperature Difference in Means.

<table>
<thead>
<tr>
<th></th>
<th>Station vs. Observed</th>
<th>Kestrel 5400 vs. Observed</th>
<th>Kestrel 5400 vs. Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natural Wet Bulb</td>
<td>1.3°F (1.3°F - 1.3°F)</td>
<td>2.6°F (2.6°F - 2.7°F)</td>
<td>1.3°F (1.2°F - 1.3°F)</td>
</tr>
<tr>
<td>Black Globe Temperature</td>
<td>-2.6°F (-2.8°F - 2.5°F)</td>
<td>-3.9°F (-4.1°F - 3.7°F)</td>
<td>-1.7°F (-1.9°F - 1.5°F)</td>
</tr>
</tbody>
</table>

*Note.* Wilcoxon difference of means to assess differences of the natural wet bulb temperature and black globe temperature for the Station and the Kestrel 5400 measurements. 95% confidence intervals are included in parentheses.

**WBGT Accuracy Metrics**

Analyses of the hit rate, false alarm ratio, and bias scores further revealed the differences between the Station WBGT and Kestrel 5400 WBGT (Figure 3.6), with the Station WBGT better matching the Observed WBGT.

The hit rate was comparable until WBGT exceeded 85°F, at which point the Kestrel 5400 WBGT hit rate continuously decreased, while the Station WBGT hit rate was consistent from 87°F–90°F and then increased slightly at WBGTs higher than 90°F (Figure 3.6). Likewise, the false alarm ratio was similar between the two until WBGT exceeded 85°F. AT WBGTs greater than or equal to 89°F, the Kestrel 5400 false alarm ratio ranged from 0.9–0.99, while the Station WBGT false alarm ratio ranged between 0.7–0.8 (Figure 3.6).
Figure 3.6. *WBGT Accuracy Metrics.*

Note. The hit rate, false alarm ratio, and bias score for the Station and Kestrel 5400 WBGT at 1°F increments. Shading represents the corresponding WBGT flag level of the values (including the standards that have 92°F as the threshold for black flag).

Lastly, the Kestrel 5400 and Station WBGT bias scores at WBGTs less than 85°F were similar, being slightly cool-biased (Figure 3.6). When Observed WBGT was between 86°F–88°F, the Kestrel 5400 WBGT had a cooler bias than the Station WBGT (Figure 3.6). Overall, the Station WBGT bias scores at dangerously high WBGTs of 90°F–94°F were notably lower than the Kestrel 5400 WBGT bias scores, although both were positively biased. The largest difference was at extreme WBGTs of 94°F, with the bias score of the Kestrel 5400 being 2.8 compared to 1.75 with the Station WBGT (Figure 3.6).
**WBGT Variations: Surface Types and Shade**

**Skin Temperature**

The observed skin temperatures on bare ground, alive grass, and dormant grass differed with respect to both their mean and ranges (Figure 3.7). Overall, the range of observed skin temperatures was greatest for bare ground (69°F–140°F), followed by dormant grass (75°F–130°F), and then alive grass (76°F–121°F) (Figure 3.7). The skin temperature of bare ground was the highest, with a mean of 104°F, and 25th and 75th percentile values of 94°F and 113°F, respectively (Figure 3.7). Alive grass had a mean temperature of 92°F, while dormant grass had a mean of 96°F (Figure 3.7). Both grass types had similar 25th percentiles of roughly 88°F–89°F. However, the 75th percentile was higher for dormant grass (108°F compared to 99°F) (Figure 3.7).

**Figure 3.7. Observed Skin Temperatures.**

The mean bias of the Station black globe temperature (utilizing the Liljegren et al. (2008) method) was positive (warm) when the air temperature was warmer than the skin temperature, and negatively (cool) biased when skin temperature was warmer than air temperature (Figure 3.8). The patterns in bias between the Station WBGT and Kestrel 5400 WBGT were similar when assessed as a function of the difference between the skin and air temperature, with both being positive (warm) when skin temperature was less than air temperature. Also, both WBGTs had a warm bias when skin temperature was greater than air temperature. When skin temperature was 10°F–40°F
higher than air temperature, the median bias of the Station WBGT ranged between 0.1°F–1°F, compared to median biases for the Kestrel 5400 ranging from 0.8°F–1.2°F (Figure 3.8).

**Figure 3.8. Black Globe Temperature Biases Relative to Difference between Skin and Air Temperature.**

![Figure 3.8](image)

*Note.* Comparison of the Station versus observed black globe temperature (top), and the Station and Kestrel 5400 WBGT biases compared to Observed WBGT (bottom) as a function of the difference between skin and air temperature (skin minus air temperature).

**Comparison across surfaces and shade conditions**

Statistically significant differences were identified in the means of the WBGT (measured with the Kestrel 5400) devices over a tennis court and neighboring grassy field, with the tennis court being 1.6°F (0.5°F–2.5°F) warmer (Table 3.4).
Table 3.4. Differences in WBGT between Locations.

<table>
<thead>
<tr>
<th>Location 1</th>
<th>Location 2</th>
<th>Difference in WBGT Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun</td>
<td>Shade</td>
<td>6.8°F (6.1°F–7.4°F)</td>
</tr>
<tr>
<td>Sunned field</td>
<td>Shaded Forest</td>
<td>9.7°F (9.3°F–10.1°F)</td>
</tr>
<tr>
<td>Shaded field</td>
<td>Shaded Forest</td>
<td>0.8°F (0.3°F–1.7°F)</td>
</tr>
<tr>
<td>Tennis Court</td>
<td>Grass Field</td>
<td>1.6°F (0.5°F–2.5°F)</td>
</tr>
</tbody>
</table>

Note. Difference in means of WBGT measured concurrently by the two Kestrel 5400 devices in different locations and shade conditions. The 95% confidence interval is included in parentheses. Positive values indicate the values at “Location 1” were higher than those at “Location 2”.

Overall, the WBGT measured in the sun was around 6.7°F–6.8°F warmer than WBGT in the shade (excluding shaded observations in the forest) (Figure 3.9). However, the difference between WBGT measured in the sun and WBGT measured in a shaded forest was greater, with a difference in WBGT of 9.7°F (9.3°F–10.1°F) (Table 3.4).

Figure 3.9. WBGT Sun versus WBGT Shade.

As Observed WBGT increased, the differences between the WBGT in the sun and shade steadily increased, with the median difference being +1.8°F at 80°F, +5°F at 86°F, and +8.7°F at 92°F.
(Figure 3.9). A statistically significant difference between the means of the WBGT in the sun and shade were found, with a difference of 6.8°F (6.1°F–7.4°F) (Table 3.4).

The difference in WBGT measured over 1) a grassy field that had been irradiated periodically throughout the day and 2) bare ground within a forest with a thick canopy and consistent shade was statistically significant. The WBGT measured over the shaded grassy field was 0.8°F (0.3°F–1.7°F) warmer than the WBGT measured in the forested shade (Table 3.4).

**Estimating Clear-Sky Radiation**

The methods for estimating clear-sky radiation and modifying that radiation by percentage cloud cover all performed well, but there was some variability in bias. Comparing the estimated radiation from each of the three methods for estimating clear-sky radiation, including modifying those estimates by percentage cloud cover, showed that all methods were within +/- 75 w/m², excluding the Ryan model with the default ATC set to 0.8 (Figure 3.10). Under clear sky or nearly clear sky (when cloud cover percentage was less than or equal to 5%), the Kasten method showed the largest bias (median of +72 w/m²), while the Ryan method had the lowest when ATC was set to 0.65 (median of -15 w/m²) (Figure 3.10). Further, the Ryan method was slightly negative (too low), while the other two methods were positively biased (too high) (Figure 3.10). As with the Ryan method, the default settings in the Bras method for atmospheric turbidity resulted in estimated solar radiation being too high. Adjusting this by changing the value representing this turbidity from 3 to 4 improved results, but the Ryan method (0.65 ATC) remained more accurate (Figure 3.10).
Figure 3.10. *Clear-sky Radiation Estimates.*

**Accuracy of Clear-Sky Radiation Estimates**

<table>
<thead>
<tr>
<th>Cloud Cover Percentage &lt;= 5%</th>
<th>With Cloud Cover Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Comparison of estimated clear-sky radiation and observed solar radiation from the 1) Kasten, 2) Ryan (with default ATC of 0.8 and 0.65), and 3) Bras Methods (with default ATC setting of 3 and 4) under cloud-free sky (left). Comparison of estimated radiation and observed solar radiation after modifying by cloud cover percentage (right).

The biases with all three methods for estimated solar radiation derived from modifying clear-sky radiation by percentage cloud cover are shown in Figure 3.10. The pattern in bias between the methods is similar to the accuracy of the clear-sky radiation discussed above, with the Kasten method having the largest bias and the Ryan method having the smallest bias (Figure 3.10). However, the mean biases of all methods were positive (too high) compared to observations, with means of 80 w/m² (Kasten), 7 w/m² (Ryan ATC 0.65), and 65 w/m² (Bras, default parameters) (Figure 3.10).

**Discussion and Conclusion**

This research addressed four specific questions. The first two questions were concerned with discerning and comparing the accuracy of 1) estimated WBGT from a weather station using the Liljegren et al. (2008) methodology and 2) the WBGT as measured by a Kestrel 5400 WBGT.
meter. The Liljegren estimation methodology had a lower bias than the Kestrel 5400 when compared to Observed WBGT, with overall biases being +0.4°F and +1.2°F warmer than observed WBGT, respectively (Figure 3.4). Furthermore, the bias of the Liljegren estimation does not vary significantly as WBGT increases, which was the case for the Kestrel 5400 (Figure 3.4). This pattern of increasing bias at higher WBGTs for the Kestrel 5400 is hypothesized to be related to the required approximations the device uses to adjust for the small, non-standard black globe and the “waterless” natural wet bulb temperature.

The Station (Liljegren) WBGT and Kestrel 5400 WBGT hit rates, bias scores, and false alarm ratios were similar when WBGT was between 70°F–85°F. As had already been determined, the Kestrel 5400 WBGT being more positively biased than the Observed WBGT resulted in relatively low hit rates at the most dangerous WBGTs and high numbers of false alarms.

Additionally, the Liljegren methodology assumes that the skin (surface) temperature is equivalent to the air temperature in the equation for estimating the black globe temperature. Given the large differences between air and skin temperature that are often measured, particularly over surfaces such as bare ground or tennis courts, this research question sought to gauge the sensitivity of the black globe estimation to this assumption and what bias it may introduce. Based on comparing the Liljegren WBGT biases when skin and air temperature were equivalent, or the skin temperature was up to 45°F warmer, little difference was revealed (Figure 3.8). Thus, this assumption was not found to be a significant contributor to the overall bias of the Station WBGT, further evident given that the Station WBGT is positively biased, yet the station black globe temperature is negatively biased, as skin temperatures become increasingly warmer than the air temperature.

The third research question evaluated the differences in WBGT between measurement locations, e.g., sun versus shade, and how WBGT varied as a function of surface type and the skin
temperature of those surfaces. In addition to sun versus shade comparisons, measurements of WBGT from within a forested area and over tennis courts were compared.

Overall, the three surfaces investigated in this work (bare ground, alive grass, and dormant grass) had markedly different skin temperatures (Figure 3.7). The bare ground temperatures were the highest, largely due to the relatively lower albedo of this surface, which allowed more radiation to be absorbed. The skin temperature of dormant grass was higher than alive grass, which is similarly related to albedo (Figure 3.7). However, this difference is also related to the relatively higher latent heat flux of the alive grass (with the dormant grass having a higher sensible heat flux) since the alive grass is transpiring. The grass being alive is also indicative of moister soil and increased latent heat flux.

A statistically significant difference was found in the sun vs. the shade (Table 3.4), with the WBGT in the sun being 6.8°F (6.1°F–7.4°F) warmer, and these differences increased as WBGT increased. From an average difference of around 1°F–1.5°F when WBGT was 80°F–82°F, when WBGT was equal or greater to 90°F and 94°F, the WBGT in the sun was on average 8°F and 9.1°F warmer than concurrent shaded WBGT measurements, respectively. The largest differences at the most extreme WBGTS are driven by the increased differences in the radiative components of the environment between sun and shade that influence heat stress and WBGT.

The difference between WBGT in the sun (grassy field) and WBGT in a shaded forest was statistically significant, with the sunned grassy field being 9.7°F (9.3°F–10.1°F) warmer than the shaded forest WBGT. Additionally, the shaded WBGT measurements made over a grassy field versus shaded measurements from within a forest were +0.8°F (0.3°F–1.7°F) warmer than the concurrent measurements in the forest. These differences are hypothesized to be at least partially a result of the amount of radiation the respective surfaces had received throughout the day. The location in the grassy field was in direct sunlight for approximately 4–5 hours on average before
becoming shaded by trees, after which the measurements were taken. However, the measurement location in the forest was selected based on the full coverage of shade provided by a thick tree canopy. Thus, while both sets of measurements were made concurrently in the shade, the WBGTs over the surface that had been irradiated earlier in the day were warmer than the WBGTs over the surface that had been shaded throughout the day.

Lastly, WBGT measurements made over a tennis court were +1.6°F (0.5°F–2.5°F) warmer than concurrent measurements made over a neighboring grassy field, with roughly 200 feet between the sites. This differs from the results in Grundstein & Cooper (2020) and Kopec (1977), which showed minimal difference in WBGT between a hard surface (tennis court) and grass. This difference is hypothesized to result from differences in the weather conditions in the studies, as the mean WBGT observed in Grundstein & Cooper (2020) over all surfaces was less than 82.4°F while the mean WBGT of concurrent observations on a tennis court and grass was 90°F and 88.6°F, respectively, in this study.

The implications of these differences are important because the conditions in this study were significantly more thermally stressful, and it is at these WBGTs that health effects could be ameliorated with awareness of the differences across surfaces. Grundstein & Cooper (2020) and Kopec (1977) note that competing factors likely contributed to their finding of little difference between a tennis court and a grassy field (e.g., higher dew point temperatures and lower skin temperatures over the grass and the reverse over the tennis court). Thus, additional research could confirm if these competing factors are less impactful at the most extreme heat stress levels and provide further clarity on differences between tennis courts and grass overall.

Finally, the fourth question addressed in this research compared the accuracy of different methods for estimating clear-sky radiation and the accuracy of modifying that clear-sky radiation by percentage cloud cover. The three methods assessed were: 1. Kasten model (Kasten & Czeplak,
1980), 2. Ryan-Stolzenbach model (Ryan & Stolzenbach, 1972), and 3. Bras model (Bras, 1990). Between these methods, with the default parameters (atmospheric transmissivity) in the Ryan and Bras models, the Bras model had the lowest bias in estimating clear-sky radiation. However, setting the ATC to 0.65 instead of the default 0.8 in the Ryan model resulted in clear-sky radiation estimates that outperformed all other methods tested, and outperformed the resulting change in bias for the Bras model when modifying its ATC (Figure 3.10). Importantly, the use of varied settings of the ATC should be validated further. Changing the ATC setting too drastically from the default would not be recommended, especially in a different climate or season, but awareness that there are substantive differences between methods is important as it may help explain unexpected differences in the results of a given study.

The modification of clear-sky radiation by the calculated percentage cloud cover from the images of sky conditions performed well in comparison to observed solar radiation, with median differences of within +/- 75 w/m². The differences between methods, in this case, mirrored the differences with estimating clear-sky radiation. While the Ryan method with an ATC of 0.65 underestimated clear-sky radiation, this method with cloud cover was slightly positive (too high) by 6.9 w/m² (-17.7–31.6 w/m²), but overall, not different from zero to a statistically significant degree.

This study had limitations that are important to note. First, the results and comparisons made here may differ in other climates and regions, but similarities in other humid, sub-tropical locations would be expected. Second, the image analysis to classify and calculate percentage cloud cover could be improved in future research by more robust image classification, which is increasingly possible with the advancements in machine learning algorithms. Future research should also address the varying character of cloud types (e.g., height and density) as that certainly impacts the radiation ultimately incident at the surface. However, this would require a
reformulation of modifying clear-sky radiation estimates by cloud cover, since the method used here simply allows for one value to represent the total sky conditions (2).

Overall, the research presented here further confirmed prior research highlighting the relative accuracy of the Liljegren et al. (2008) methodology for estimating WBGT. The Kestrel 5400 WBGT meter had a higher bias, particularly when WBGT was greater than or equal to 88°F (+1.5°F–1.7°F), which is 0.2°F–0.4°F greater than the stated accuracy of the device by the manufacturer (1.3°F) (Kestrel, 2020). Relatively small bias has major implications when it comes to WBGT, given that the difference between flag levels is only 2°F–5°F. The bias here resulted in the device “erring on the side of caution”, but if the reverse were true, the device would indicate conditions were safe for activity when conditions could be quite threatening to health. Additionally, with the knowledge of positive bias in the device, as was the case here, there is concern about doubts in the device readings and subsequent user-based bias correction that could erroneously conclude that the WBGT is lower than it is.

Awareness of the increasingly large difference between the sun and shade WBGT as WBGT in the sun increases provides further basis for encouraging activity to be moved to the shade on the most thermally stressful days, if possible, and emphasizes the importance of on-site measurements. This also should encourage future planning of outdoor athletic facilities to factor in shade access more heavily for the hottest times of the day. Lastly, the identification of a more accurate method for estimating clear-sky radiation, relative to the observations assessed here in this study, and confirmation of the accuracy of modifying that radiation by percentage cloud cover, provides useful insight for future efforts to estimate WBGT in real-time from weather stations that do not measure solar radiation. This finding also emboldens efforts to forecast WBGT that rely on forecast cloud cover values, ultimately reducing uncertainty with estimating WBGT from estimated variables themselves.
CHAPTER 4: THE DEVELOPMENT AND ACCURACY ASSESSMENT OF WET BULB GLOBE TEMPERATURE FORECASTS

Exposure to extreme heat leads to more deaths than any other weather event in the United States (CDC, 2010; NOAA’s National Weather Service, 2020). Therefore, efforts to safeguard health contemporarily are extremely important and will remain critical given the implications of a warming world due to climate change.

To prevent adverse health outcomes resulting from heat exposure, systems have been developed that provide warnings to the public during particularly dangerous periods of hot weather. These systems, often referred to as heat-health warning systems (HHWS), are based on weather forecasts of variables, such as air temperature or the Heat Index, the latter being the most common metric to assess heat stress in the United States (Kovats & Kristie, 2006). While the Heat Index accounts for the effect of humidity and air temperature, notably only in the shade, other heat stress indices, such as wet bulb globe temperature (WBGT), provide more robust assessments of environmental heat stress (Budd, 2008; Hondula et al., 2014).

Prior to 2018–2019, WBGT was not a routinely forecast variable. In 2018, NOAA’s Southeast Regional Climate Center (SERCC) and Carolinas Integrated Sciences and Assessments (CISA) developed a WBGT forecast tool, which was operationalized in the summer of 2019 (SERCC & CISA, 2023). The US NWS released an experimental WBGT forecast in 2019 that was operationalized in 2021 (NOAA, 2021). The research presented here assesses the accuracy of these WBGT forecasts and provides insight on future improvements.

Background
**Wet bulb globe temperature (WBGT)**

To reduce military casualties at training camps during times of extreme heat, the US military developed WBGT in the 1950s (Budd, 2008; Yaglou & Minard, 1957). WBGT is calculated by adding together three components (1): dry-bulb temperature, natural wet bulb temperature, and black globe temperature (Budd, 2008). The black globe temperature is measured using a black globe thermometer. The temperature probe is suspended inside the black globe, and the globe itself is unshielded from radiation. The black globe temperature is an indicator of the temperature due to radiative forcing incident on human skin, including both direct and diffuse short-wave radiation and long-wave radiation from the surface of the Earth (Kopec, 1977; Liljegren et al., 2008). The temperature of the black globe is also influenced by wind speed. The second term, the dry-bulb temperature, is a standard measure of ambient air temperature, with the temperature sensor located inside of a radiation shield that is naturally ventilated (Liljegren et al., 2008). Unlike the commonly measured and estimated psychrometric wet bulb temperature, which is similarly located in a radiation shield like the dry bulb thermometer, the natural wet bulb thermometer is measured unshielded from radiation (Liljegren et al., 2008). There is a wet wick that is wrapped around the bulb of the thermometer, and based on environmental influences such as wind speed, the evaporation of water from this wick mimics the cooling effect of sweat evaporating off of human skin. Several environmental variables influence the natural wet bulb temperature, including wind speed, relative humidity, and solar radiation (Budd, 2008; Kopec, 1977; Liljegren et al., 2008; Yaglou & Minard, 1957). From these three measures, WBGT is calculated according to the following equation:

\[
WBGT = 0.7 \times NWB + 0.2 \times Tg + 0.1 \times Ta,
\]

where \(NW B\) is the natural wet bulb temperature, \(Tg\) is the black globe temperature, and \(Ta\) is the dry bulb temperature (ambient air temperature).
Ultimately, WBGT is utilized based on a given value’s corresponding level of danger, often referred to as flag level (Table 4.1). For each level, or flag, there are associated guidelines meant to help keep a worker or athlete cool by dictating things such as the amount of water required for adequate hydration and the duration of rest breaks (ISO, 1989, 2017). While there is variation amongst organizations in the specific WBGTs used to define each flag level, the most common thresholds are given in Table 4.1.

Table 4.1. WBGT Safety Guidelines from the North Carolina High School Athletic Association.

<table>
<thead>
<tr>
<th>WBGT (*°F)</th>
<th>Athletic Activity Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 80</td>
<td>Unlimited activity with primary cautions for new or unconditioned athletes or extreme exertion; schedule mandatory rest/water breaks (5 min water/rest break every 30 min).</td>
</tr>
<tr>
<td>80–84.9</td>
<td>Normal practice for athletes; closely monitor new or unconditioned athletes and all athletes during extreme exertion. Schedule mandatory rest/water breaks (5 min water/rest break every 25 min).</td>
</tr>
<tr>
<td>85–87.9</td>
<td>New or unconditioned athletes should have reduced intensity practice and modifications in clothing. Well-conditioned athletes should have more frequent rest breaks and hydration as well as cautious monitoring for symptoms of heat illness. Schedule frequent mandatory rest/water breaks (5 min water/rest break every 20 min). Have cold or ice immersion pool on site for practice.</td>
</tr>
<tr>
<td>88–89.9</td>
<td>All athletes must be under constant observation and supervision. Remove pads and equipment. Schedule frequent mandatory rest/water breaks (5 min water/rest break every 15 min). Have cold or ice immersion pool on site for practice.</td>
</tr>
<tr>
<td>90+</td>
<td>SUSPEND PRACTICE/MUST INCLUDE MANDATORY BREAKS AS DIRECTED BY GAME-DAY ADMINISTRATOR DURING CONTEST.</td>
</tr>
</tbody>
</table>

Note. Adapted from Raleigh NWS WFO (2017).

WBGT is recommended for occupational environments by the United States Occupational Safety and Health Administration (OSHA, 2017). Outside of the United States, WBGT is the recommended method for determining heat stress by the International Organization for Standardization (ISO, 2017). Additionally, numerous US states now require measurements of WBGT to determine if it is safe for athletic practice outdoors (Grundstein et al., 2015; NCHSAA, 2016), such as in the states of North Carolina, South Carolina, Georgia, Florida, Massachusetts, and New Jersey.

While WBGT has been increasing in popularity for assessing heat stress, there remain several challenges with WBGT. WBGT meters that measure each of the three components (dry bulb,
black globe temperature, and natural wet bulb temperature) are the most accurate. However, these devices are 200%–400% more expensive than devices that offer approximations of the natural wet bulb temperature and black globe temperature. While cheaper, the devices providing such approximation often have notably higher biases in measurements of WBGT (Budd, 2008; Cooper et al., 2017). Additionally, neither natural wet bulb thermometers nor black globe thermometers are common instruments installed at weather stations, and thus real-time WBGT values cannot be assessed at most stations. However, numerous models have been developed that estimate WBGT based on standard meteorological variables recorded at weather stations. These methods vary in accuracy. The methodology presented in Liljegren et al. (2008) has been found to be most accurate in several studies (Lemke & Kjellstrom, 2012; Patel et al., 2013).

High (hot) values of WBGT are sensitive to slight changes in environmental conditions, particularly changes in wind speed. While atmospheric stability influences the degree to which wind speed higher in the atmosphere mixes down to lower levels closer to the ground, local wind speed at the surface is also influenced by surrounding trees and buildings (US EPA, 2000). There is less difference in wind speed across altitudes above trees and buildings compared to the differences in wind speed at the surface relative to wind 30–50 feet above the surface (Oke, 1987). The cause of this is attributable mainly to objects that extend up from the ground interacting with wind at higher altitudes. This interaction, or source of friction, leads to slower wind speeds at the surface. The amount of friction, also called surface roughness, thus ultimately determines the degree to which wind decelerates as altitude decreases towards ground level. With the sensitivity of WBGT (and the human body) to small differences in wind speed, variations in surface roughness can lead to vastly different WBGTs at the ground between nearby locations. An example of this sensitivity and the implications therein is seen in Figure 4.1, which shows there can be a difference of up to two WBGT flag levels when wind speeds decrease from 3 mph to 1.5 mph.
Figure 4.1. *Relationship between Station WBGT and wind speed.*

![Graph showing the relationship between WBGT and wind speed for different cloud cover conditions.](image)

*Note.* WBGT estimated using Liljegren et al. (2008) method. Relationship depicted for three categories of cloud cover, with an air temperature of 86°F (30°C) and dew point of 70°F (21.1°C).

In order to estimate WBGT from weather station observations, wind speeds at 10 meters above the surface must be downscaled to 2 meters, using a vertical wind profile and logarithmic transformation. An example of this concept can be seen in Figure 4.2.

Using wind speed at 2 meters and accurately downscaling wind to this height is important since 1) wind speed at 10 meters above the surface is faster than wind speed at 2 meters above the surface, 2) WBGT is incredibly sensitive to small changes in wind speed, and 3) humans are 1.5–2 meters tall on average.

Figure 4.2. *Variation in wind speed as a function of height above ground.*
Forecasts of WBGT

Accessible forecasts of WBGT enable strategic planning, given its increasing utilization and the required use of the index to assess heat stress. Additionally, with forecasts of WBGT available, this lays the groundwork for future use of WBGT by other sensitive groups, since it is recognized to be more robust than the Heat Index (Budd, 2008; Hondula et al., 2014). However, there has been limited research on the accuracy of WBGT forecasts and these forecasts are not common. One example of a WBGT forecast can be seen in Japan, where the Japan Meteorological Agency has an operationalized WBGT forecast and utilizes it as a basis for heat warnings, much like the use of the Heat Index in the United States (Hoshi & Inaba, 2007). In 2019, the US NWS began including WBGT forecasts in an experimental version of their gridded forecast data, the National Digital Forecast Database (NDFD) (US NWS, 2019a). This forecast was recently operationalized in 2021 (NOAA, 2021).

In 2018, NOAA's Southeast Regional Climate Center (SERCC) and Carolinas Integrated Sciences and Assessments (CISA) created a WBGT forecast tool, which was operationalized in the summer of 2019 (SERCC & CISA, 2023). As of 2022, this forecast tool was operational for the eastern two-thirds of the contiguous United States. The SERCC/CISA (SC WBGT) forecast tool and the NWS operational WBGT forecast, however, utilize different methods for estimating WBGT. One aspect of the research here is to assess the accuracy of these different forecast methodologies.
The degree to which any weather model accurately captures the heterogeneity of land types and microclimates within a given pixel of its model output is limited, as there can be wide variation in WBGT and the microclimatic variables influencing WBGT across 100s of meters (Verkaik et al., 1998). This variability fails to be fully captured by weather models due to inadequate resolution, since the highest resolution model output is at 2.5 km. In addition to assessing the accuracy of these WBGT forecasts, this research also assesses the utility of supplementing sub-forecast grid scale surface roughness information to better tune the downscaling of wind speeds for calculating the WBGT forecast. The current SERCC/CISA forecast tool employs Pasquil-Gifford stability classes for the downscaling of wind from 10 to 2 meters.

The SERCC/CISA forecast tool utilizes both the US NWS National Digital Forecast Database (NDFD) and the US NWS National Blend of Models (NBM) in creating the WBGT forecast (SERCC & CISA, 2023). Both models forecast at hourly timesteps out to 36 hours in the future, but the NDFD hourly forecast changes to a 6-hourly forecast at 72 hours past initialization while the NBM forecasts 3-hourly values out to 192 hours. Thus, the NDFD is used for 1–69 hours and the NBM for 72–120 hours for the SERCC/CISA forecast tool. References to the SERCC/CISA WBGT forecast will be hereafter referred to as SC WBGT. The SC WBGT methodology supplemented with surface roughness for downscaling wind speeds will hereafter be referred to as SC WBGT Land.

The forecast analysis here serves as an update to an assessment of forecast accuracy presented in Clark & Konrad (2020). For assessing accuracy, the WBGT forecasts will be compared with 1) Observed WBGT measurements collected with a WBGT meter that is compliant with standards established by the International Organization for Standardization and 2) WBGT estimated from weather stations. Satellite imagery (Sentinel 2A/2B and Landsat 8 ETM+) were used to calculate the normalized difference vegetation index (NDVI) and green vegetation fraction (GVF)
to then estimate surface roughness across portions of the forecast domain for the SC WBGT Land forecast.

**Research Questions**

The following research questions were addressed:

1. What is the accuracy of the WBGT forecast and how does this accuracy vary across methodologies for estimating WBGT?
   a. What is the accuracy of SC WBGT when using the NBM vs. the NDFD?
   b. What is the accuracy of the current NWS WBGT forecast and how does it compare to the accuracy of SC WBGT?
   c. Is there a spatial pattern to forecast errors (e.g., mountain valleys or closed landscapes having higher error)?
   d. How does WBGT forecast bias vary as a function of hour of day?

2. For downscaling wind speeds from 10 meters to 2 meters, does the incorporation of gridded surface roughness data improve the WBGT forecasts, compared to when the winds are downscaled using Pasquill-Gifford Stability Classes?
   a. For generating gridded surface roughness, does the finer resolution of Sentinel satellite imagery (10 m) offer significant improvement relative to Landsat (30 m)?

**Data**

**Weather Forecast Data**

Hourly gridded forecast data for two US NWS forecast models (four runs per day: 0Z, 6Z, 12Z, 18Z) were archived for the summers of 2019-2021: the National Blend of Models (NBM) and National Digital Forecast Database (NDFD). Both models have a spatial resolution of 2.5 km. The NBM data were downloaded at regular intervals from the NOAA Operational Model Archive and
Distribution System (Rutledge et al., 2006). Both operational and experimental NDFD data were downloaded from US NWS FTP servers.

The following variables from both forecast models were used: 2-meter air temperature, 2-meter dewpoint temperature, 10-meter wind speed, and total cloud cover percentage.

**Weather Station Measurements**

A weather station was co-located with a WBGT meter to provide data quality checks. The weather station was a Davis Instruments Vantage Pro 2 Plus (Figure 4.3) that recorded all variables needed to estimate WBGT. Data were recorded at 10-second intervals for the following variables: air temperature, relative humidity, dew point temperature, barometric pressure, wind speed, and solar radiation. All instruments were placed 1.5 meters above the ground. The humidity sensor (Sensirion SHT31) was regularly calibrated to ensure accuracy by placing the sensor inside humidity chambers that included either a sodium chloride slurry for calibrating the humidity sensor to 33%, or a magnesium chloride slurry for calibrating the humidity sensor to 75%.

**Observed WBGT Measurements**

WBGT data were recorded with a WBGT meter designed to meet the specifications for a WBGT meter as outlined by the International Organization for Standardization, first used in Cooper et al. (2017) (Figure 4.3). The device contains a dry bulb thermometer inside a standard, naturally ventilated radiation shield. The black globe temperature is measured with a matte black globe that has a diameter of 0.15 meters (Cooper et al., 2017). Lastly, the natural wet bulb temperature is measured using a thermometer suspended above a water reservoir. A wick was placed over the thermometer, with the end of the wick submerged in the water reservoir to keep the wick wet. The instruments were situated such that they were 1.5 meters above the ground. Data were recorded at various intervals (2-, 5-, 10-, 30-, 60-, or 120-second intervals), depending on the location of the meter and thus the frequency at which it could be accessed.
Measurements were taken at several locations throughout the summers of 2019-2021: three high schools in the Triangle Region of NC, the Horace Williams Airport in Chapel Hill, NC, and within suburban environments in Chapel Hill, NC, Durham, NC, and Shelby, NC.

**Figure 4.3. Field Work Instruments (WBGT meters and weather station).**

*Note.* The WBGT meter (right) with a black globe thermometer, natural wet bulb temperature probe situated in a water reservoir, and dry bulb sensor in a radiation shield. A weather station (left) and a Kestrel 5400 (middle) were also co-located with the WBGT meter.

**Station WBGT**

Station WBGT was estimated using weather stations from three networks. These three networks include the Automated Surface Observing System (ASOS) network, Automated Weather Observing System (AWOS) network, and the North Carolina Environment and Climate Observing Network (ECONet), which is maintained by the North Carolina State Climate Office. In total, 169 stations were used (130 stations from ASOS/AWOS and 39 stations from the ECONet) (Figure 4.4). The 130 ASOS/AWOS stations were mainly in North Carolina. Stations in other states were selected.
based on the availability of Sentinel satellite imagery (detailed below). The Iowa Environmental Mesonet archive (IEM) was used to download the weather station data for stations on the ASOS and AWOS networks. The Climate Retrieval Observations Network of the Southeast Database housed at the North Carolina State Climate Office (NC CRONOS) was used to download the ECONet weather station data.

Meteorological variables from the hourly weather station data included: 2-meter air temperature, relative humidity, dew point temperature, 10-meter wind speed, and barometric pressure. Stations that were a part of the ECONet weather station network also measured solar radiation. However, for ASOS/AWOS weather stations, total cloud cover is reported in lieu of solar radiation. The reported cloud cover was used to estimate the observed solar radiation for the stations on this network. Hourly data for all stations for the heat season (May 1 – September 30) 2019–2021 were used.

**Figure 4.4. Map of the weather stations used (by network) for Station WBGT.**
Satellite Imagery and Land Cover

Satellite imagery from Sentinel-2A and Sentinel-2B were downloaded from the Copernicus Open Access Hub, specifically the Level-1C product (Copernicus Sentinel data, 2019). This imagery has a spatial resolution of 10 meters. Imagery included scenes covering the entire state of North Carolina and parts of South Carolina, Georgia, Alabama, Tennessee, Virginia, Pennsylvania, Illinois, Oklahoma, and Texas. The distribution of stations can be seen in Figure 4.4. All imagery utilized was captured on three dates: August 29th, August 30th, and September 7th, 2019. In addition to the Sentinel imagery, Landsat 8 ETM+ Level-2 imagery was also obtained for the state of NC, with a collection date of August 17th, 2019.

The 2019 National Land Cover Database (NLCD) was retrieved for the CONUS to provide the land cover information utilized here (Dewitz, 2021).

Methods

Station WBGT across the ASOS/AWOS network was estimated using the Liljegren et al. (2008) methodology, which has been found to be most accurate at estimating WBGT (Lemke & Kjellstrom, 2012; Patel et al., 2013). SC WBGT forecasts likewise applied the Liljegren et al. 2008 methodology for the NDFD/NBM forecast models. The necessary formulas for using the Liljegren et al. 2008 method were provided by the R package “wbgt” (Lieblich & Spector, 2017).

Solar radiation was estimated for the ASOS/AWOS weather stations since these stations do not directly measure this variable. To estimate solar radiation, the reported cloud cover was utilized. Likewise, the forecast model data provided cloud cover, thus solar radiation was also estimated to calculate the WBGT forecast. The ASOS/AWOS stations report cloud cover at several levels, using the following categories: clear (0%–5%), few (5%–10%), scattered (25%–50%), broken (50%–87%), and overcast (87%–100%) (National Oceanic and Atmospheric Administration et al., 1998). Comparing the observed solar radiation values with estimated solar radiation values at
ECONet weather stations revealed that using the maximum value within each of the aforementioned ranges produced the most accurate results, i.e., clear (5%), few (10%), scattered (50%), broken (87%), and overcast (100%). Since ASOS/AWOS stations measure clouds at multiple levels, the layer with the largest amount of cloud cover was used to derive the percentage cloud cover variable. For example, if an observation reported scattered (50%) at level 1, few (10%) at level 2, and broken (87%) at level 3, the cloud cover at level 3 was utilized, which is 87% in this example.

With 1) the percentage cloud cover at ASOS/AWOS stations and 2) the cloud cover provided in the forecast data, the clear-sky direct radiation value (based on the time and location of a station observation or forecast data point) was modified by the percentage cloud cover to estimate solar radiation with the following equation:

\[
S_{rad} = R_0 \ast (1 - 0.75n^{3.4}),
\]

where \(n\) is the cloud cover fraction (0.0–1.0) and \(R_0\) is the clear-sky direct radiation (w/m²) estimated using (3):

\[
R_0 = 990 \ast \sin(\emptyset - 30),
\]

where \(\emptyset\) is solar elevation angle (Kasten & Czeplak, 1980).

The forecast data and observed weather station data reported wind speed at 10 meters. Except for the SC WBGT Land (detailed below), these measurements were logarithmically downscaled from 10 meters to 2 meters using the following function:

\[
U_z = U_r \left(\frac{Z}{Z_r}\right)^p,
\]

where \(U_z\) is the mean wind speed at height \(Z\) above ground, \(U_r\) is the wind speed at the reference height \(Z_r\), and \(p\) is the power-law exponent (US EPA, 2000). The exponents are provided in Appendix 2.1. The “Urban” exponents were utilized here. The Solar Radiation Dela-T (SRDT) method was used to determine Pasquill-Gifford Stability classes and the corresponding power-law exponent.
The SRDT method serves as an indicator of atmospheric stability by using observed solar radiation during the day and vertical temperature difference at night (US EPA, 2000). For observations (or forecast data points) with wind speeds of less than 1 meter per second, the wind speed value was increased to 1 meter per second. This decision was based on the sensitivity of the anemometers installed at the weather stations that were used (Heuser et al., n.d.; National Oceanic and Atmospheric Administration et al., 1998).

**Surface Roughness and Wind Speed**

A variety of methods have been developed for characterizing surface roughness across different land cover and land use types. Several characteristics of the land surface have been used to assess surface roughness length, including NDVI (Bastiaanssen W.G.M. et al., 1998; Markert et al., 2019), GVF (Markert et al., 2019; Zeng et al., 2012), and leaf area index (LAI) (Su, 2002; Zheng et al., 2014). The methodology utilized here drew upon the method developed for the Noah Land Surface Model (LSM) version 3.4.1 (Chen & Zhang, 2009) since this parametrization (7) was found to be most consistent across different land cover types and climate patterns (Markert et al., 2019; Zheng et al., 2014).

The first step was to atmospherically correct the Sentinel satellite imagery using the Sentinel Application Platform (SNAP) software and the plugin Sen2Cor. The Landsat imagery was already atmospherically corrected. For both the Sentinel and Landsat imagery, the NDVI was calculated using the following equation:

\[
NDVI = \frac{NIR - RED}{NIR + RED}, \tag{5}
\]

where \(NIR\) is reflectance in the near infrared range (Sentinel band 8; Landsat band 5) and \(RED\) is reflectance in the red range (Sentinel/Landsat band 4). Following Zheng et al. (2014) and Markert et al. (2019), the GVF was calculated using (6):
\[ GVF = \frac{NDVI_{\text{max}} - NDVI_{\text{min}}}{NDVI_{\text{max}} - NDVI_{\text{min}}}, \quad (6) \]

where \( NDVI \) is the NDVI of a given pixel, \( NDVI_{\text{max}} \) is the maximum NDVI for a land cover class, and \( NDVI_{\text{min}} \) is the NDVI of bare soil (0.01). The land cover classification data was drawn from the NLCD 2019. Since the NLCD has a resolution of 30 meters, the Sentinel NDVI was rescaled to match this resolution before calculating \( NDVI_{\text{max}} \) per land cover class and the GVF.

Once the GVF was calculated, the surface roughness, \( Z_{0m} \), was calculated from the following equation (Markert et al., 2019; Zheng et al., 2014):

\[ Z_{0m} = (1 - GVF) \times Z_{0m,\text{min}} + GVF \times Z_{0m,\text{max}}, \quad (7) \]

where \( GVF \) is green vegetation fraction, \( Z_{0m,\text{min}} \) is the minimum surface roughness in meters, and \( Z_{0m,\text{max}} \) is the maximum surface roughness in meters, with the values for the latter two provided in Table 4.2 (Markert et al., 2019).

<table>
<thead>
<tr>
<th>Land Class</th>
<th>( Z_{0m,\text{min}} )</th>
<th>( Z_{0m,\text{max}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evergreen forest</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Coniferous Forest</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Mixed forest</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Water</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Bare</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>Croplands</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td>Urban and built up</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Impervious</td>
<td>0.03</td>
<td>0.03</td>
</tr>
</tbody>
</table>

*Note.* Drawn from Markert et al. (2019).

Given the high pixel-to-pixel variability in surface roughness commonly found at local scales, solely utilizing the derived surface roughness values at 30 meters would disregard important land surface information surrounding a location (i.e., neighboring pixels) that impact wind speed. Thus, a weighted average of surface roughness values at the following spatial resolutions was calculated for each point for both Sentinel and Landsat imagery: 30 m, 100 m, 250
m, and 500 m. Seven total configurations for the weighting of each spatial resolution for calculating this weighted average were examined to test the sensitivity of the resulting WBGT forecast (Appendix 4.1).

After deriving the weighted average surface roughness, the wind speeds to be used for SC WBGT Land were downscaled utilizing the following function from van Den Berg (2004),

\[ u(z) = u(z_{ref}) \times \frac{\ln(Z/Z_0)}{\ln(Z_{ref}/Z_0)} \]

where \( Z_0 \) is surface roughness length in meters (weighted average surface roughness), \( Z_{ref} \) is the height of original wind speed (10 meters), \( Z \) is the height of the downscaled wind speed (2 meters), and \( u \) is wind speed (van Den Berg, 2004). This was calculated for both the forecast data and weather station data. After downscaling the wind speeds, the SC WBGT Land was calculated, with the methods for converting cloud cover to solar radiation identical to the ones described above for SC WBGT.

**NWS WBGT**

For the WBGT calculated using the NWS methodology (“NWS WBGT”), all data processing regarding solar radiation values and wind speed (downscaled without surface roughness) were identical to those described above for SC WBGT. The NWS WBGT, however, utilizes different methods for estimating the black globe temperature and the natural wet bulb temperature. Black globe temperature estimation utilizes a slightly modified version of the methodology from Dimiceli et al. (2011), which was originally developed by the Tulsa, Oklahoma NWS office. These modifications include: 1) direct beam radiation capped at 0.75 (instead of 1.0) and 2) the convective heat transfer coefficient set to 0.228 (instead of 0.315) for daytime and 0 at night (T. Boyer, personal communication, October 3, 2022).
In 2022, the NWS methodology for estimating the natural wet bulb temperature was changed (T. Boyer, personal communication, October 3, 2022). However, the methodology employed by the NWS before this change is used here, since it was found to perform notably better. The new method results in cooler values for the natural wet bulb temperature on the order of several degrees Fahrenheit, using the following equation:

\[
T_n = T_w + 0.001651S - 0.09555u + 0.13235T_{wd} + 0.20249,
\]

where \( T_n \) is the natural wet bulb temperature (°F), \( T_w \) is the wet bulb temperature (°F), \( S \) is solar irradiance (\( \text{w/m}^2 \)), \( u \) is wind speed (m/s), and \( T_{wd} \) is the wet bulb depression (°F) (the difference between the dry bulb temperature and the wet bulb temperature). The prior method utilized a modification of the method developed in Hunter & Minyard (1999), which was derived for hot and dry conditions in the US (Lemke & Kjellstrom, 2012). The original Hunter & Minyard (1999) method for estimating the natural wet bulb temperature is:

\[
T_n = T_w + 0.021S - 0.42u + 1.93,
\]

where \( T_n \) is the natural wet bulb temperature (°F), \( T_w \) is the wet bulb temperature (°F), \( S \) is solar irradiance (\( \text{w/m}^2 \)), and \( u \) is wind speed (m/s). The modified formula used by the NWS (and here to calculate NWS WBGT) is provided in (11):

\[
T_n = T_w + 0.0021S - 0.42u + 1.93,
\]

**Observed WBGT**

The dry bulb temperature measured for the Observed WBGT was cross referenced with the dry bulb temperature measured by the co-located weather station for validation. Additionally, the estimated natural wet bulb temperature and black globe temperature from the co-located weather station were compared with observations of these variables on the WBGT meter to filter out instances where the instruments briefly became disconnected and thus reported incorrect data.
Regularly captured images of the instruments during data collection were used to determine the times during which these instruments were shaded, in addition to using comparison of measured solar radiation on the weather station and estimated clear-sky solar radiation based on the location and time. This was completed to ensure the accuracy of the Observed WBGT being utilized, in that shaded WBGT would skew the forecast bias assessment since the forecasts were not made for shaded WBGT.

All data were collected over grassy surfaces. The first 15 minutes of data for a given instance of data collection were discarded. This allowed adequate time for the instrumentation to equilibrate to the environment after being stored indoors or in the shade (Kestrel, 2021). The water reservoir for the natural wet bulb thermometer was regularly checked to ensure the water level was high enough to maintain moisture on the wick. In the event the water level was not high enough, all observations between that time and the time the reservoir was last checked were excluded from this analysis. Distilled water stored in the shade (thus at the ambient air temperature) was used to fill the water reservoir. Similarly to the equilibration of instruments when first setting them up at a location, the 15 minutes immediately following refilling the reservoir were excluded from analysis.

WBGT varies rapidly over small time periods (tens of seconds to a few minutes) due to fluctuations in conditions, such as cloud cover and wind gusts. To ensure robust comparison with the hourly forecast data, the forecast WBGT was compared to the Observed WBGT averaged over the course of twenty minutes (10 minutes before and after the top of every hour).

**Analysis**

Forecast accuracy for Observed WBGT was defined as the forecast WBGT minus the Observed WBGT. For Station WBGT (WBGT estimated at ASOS/AWOS and ECONet stations), the
accuracy was defined as the forecast WBGT minus the Station WBGT, as estimated by the Liljegren et al. (2008) methodology.

In addition to assessing the overall hourly forecast accuracy for each method compared to both Observed WBGT and Station WBGT, the forecast accuracy was stratified by weather conditions and hour of day. Comparisons were made between the NBM and NDFD WBGT forecasts and at varying forecast lead times for these models. To determine the utility of incorporating surface roughness into the forecast, comparisons were made between the forecast accuracy of the SC WBGT and SC WBGT Land. Only data points between 6 AM and 8 PM local time were included since this is the period during which heat stress and WBGT are highest.

Since WBGT is used based on the corresponding flag level for a given value, the accuracy of forecasting the flag levels associated with the forecast, Observed WBGT, and Station WBGT were also assessed. Contingency tables for WBGT flag accuracy were created and verification statistics calculated for each forecast method. Using the R package ‘verification’ (NCAR, 2015), the following statistics were calculated and compared for each flag level: percent correct, hit rate, bias score, and false alarm ratio. Additionally, two verification measures were used to assess the forecasts overall (not by flag level): Gerrity Skill Score (GSS) and Heidke Skill Score (HSS).

**WBGT Flag Accuracy Metrics (by flag level):**

1) **Percent Correct (%):** Total of correct forecasts (hits + correct rejections) divided by total number of forecasts (US NWS, 2019b).

2) **Hit Rate (%):** percentage of correct forecasts (hit = forecasted X and X observed) (e.g., black flag forecast, black flag observed, hit) (US NWS, 2019b).

3) **Bias Score:** measure indicating the direction of bias (positive/negative) in addition to the magnitude (ratio of frequency of forecasting a flag level to the frequency of observations at
that flag level). Values greater than 1 correspond to positive (warm) biases. Values less than 1 correspond to negative (cool) biases (NCAR, 2015).

4) False Alarm Ratio: total number of false alarms (event forecast, but not observed) for a given flag level divided by the total number of forecasts for that flag level (US NWS, 2019b).

**WBGT Flag Accuracy Metrics (overall assessment):**

1) Gerrity Skill Score (GSS): verification measure for categorical forecasts that accounts for the ‘closeness’ of categories in its assessment, e.g., green flag being closer to yellow flag but furthest from black flag (NCAR, 2015). Values range from -1 to 1, with 1 being a perfect forecast. In a Gerrity Skill Score analysis of accuracy, greater credit is given to correct (and almost correct) forecasts of rare events and less credit is awarded to correct forecasts of common events (Jolliffe & Stephenson, 2012).

2) Heidke skill score (HSS): verification measure for categorical forecasts that includes correct random forecasts. Values range from negative infinity to 1. Positive scores mean the forecast does better than random chance (US NWS, 2019b).

**Results**

The research presented here was focused on two overall questions and will be detailed in this order: 1) assessing the accuracy of WBGT forecasts and 2) determining if forecasts can be improved by using surface roughness to downscale wind speeds.

**Overview: WBGT Forecast Accuracy**

First, the overall forecast accuracy results will be presented, including comparisons between forecasts using the NBM and NDFD weather forecasts. Throughout the period of this study (2019-2021), the NBM was upgraded. However, no noteworthy differences were found in the results presented below due to these changes. Unless otherwise stated, the specific accuracy discussed outside these comparisons refers to the WBGT forecasts based on the NBM. Differences
in the accuracy when comparing the forecast to two sources of ground truth WBGT data will be
detailed: 1) Observed WBGT (direct measures of WBGT in the field) and 2) Station WBGT (WBGT
estimated at ASOS/AWOS and ECONet weather stations). Additionally, the WBGT forecast accuracy
results will include the following:

1) the assessment of both hourly and daily maximum WBGT forecast accuracy,
2) variations in bias across space,
3) comparisons of bias when stratified by hour of day and weather conditions to determine
   patterns in the biases, and
4) accuracy of forecasting WBGT flag levels.

**WBGT Forecast Accuracy**

WBGT forecast bias compared against both Observed and Station WBGT varied as a
function of location, hour of day, when stratified by weather conditions (e.g., temperature,
humidity, wind speed), and also based on the WBGT estimation methodology. Overall, WBGT
forecast biases between the NBM and NDFD were similar for a 24-hour forecast (Appendix 4.2.1).
However, when Observed WBGT was equal to or greater than 85°F and 90°F, the NDFD bias was
higher and importantly, had a notably negative (cold) bias relative to the NBM (Figure 4.5). The 24-
and 48-hour forecast biases were similar, but the forecast bias was more negative (cold) for the 72-
hour forecast (Appendix 4.2.1).
**Figure 4.5.** *WBGT Forecast Bias: NBM and NDFD.*

Note. NBM and NDFD Bias for Observed WBGT for when WBGT was equal to or greater than 85°F (left) and 90°F (right) (24-hour forecast). SC WBGT and SC WBGT Land are the WBGT forecasts using the Liljegren methodology, with the SC WBGT using PG Stability classes and SC WBGT Land using surface roughness to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the US NWS methods.

Forecast bias was positive (warm) when WBGT was between 80°F–88°F, with bias becoming less positive as WBGT increased (Figure 4.6). The NWS WBGT bias was consistently the most negative. Negative median biases were seen with the SC WBGT (-0.8°F) and NWS WBGT (-2.1°F) when WBGT was equal to or above 90°F; however, the SC WBGT Land bias was +0.4°F (Figure 4.6).
Figure 4.6. WBGT Forecast Bias (Relative and Absolute).

Note. Relative bias (top) and absolute bias (bottom) for Observed (left) and Station WBGT (right), for when WBGT was between 80°F–85°F, 85°F–88°F, 88°F–90°F, and equal to or greater than 90°F. Forecast WBGT was from the NBM. SC WBGT and SC WBGT Land are the WBGT forecasts using the Liljegren methodology, with the SC WBGT using PG Stability classes and SC WBGT Land using surface roughness to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the US NWS methods.

For both Observed and Station WBGT forecast accuracy, median absolute biases between the three forecast methods were within 1°F of each other for when WBGT was 80°F–85°F and 85°F–88°F. However, at higher values of WBGT (90°F+), the NWS WBGT absolute bias increased markedly (Figure 4.6), particularly for Station WBGT. There were two notable differences with Station compared to Observed WBGT forecast accuracy (Figure 4.6). First, both SC WBGT methods had lower bias when WBGT was less than 88°F, and SC WBGT Land had a slightly negative bias when WBGT was equal to or greater than 90°F (-0.2°F). Second, NWS WBGT bias was larger when WBGT was greater than or equal to 88°F, and its median bias was negative (too cold) across all WBGTs,
with median biases of -0.8°F and -2.8°F when Station WBGT was 80°F–85°F and 90°F or above, respectively (Figure 4.6).

The accuracy of forecasting daily maximum Observed WBGT was similar to the hourly forecast accuracy (Figure 4.7). However, when the daily maximum WBGT was less than 88°F, the magnitude of bias was lower, particularly for the two SC WBGT forecast methods. NWS WBGT was negatively biased (too cold), other than when WBGT was 85°F–88°F.

**Figure 4.7. Daily Maximum WBGT Forecast Bias.**

<table>
<thead>
<tr>
<th>WBGT °F</th>
<th>Observed WBGT</th>
<th>Station WBGT</th>
</tr>
</thead>
<tbody>
<tr>
<td>80°–85°F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>85°–88°F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>88°–90°F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥90°F</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** Daily maximum forecast relative (top) and absolute bias (bottom) for Observed (left) and Station WBGT (right), for when daily maximum WBGT was between 80°F–85°F, 85°F–88°F, 88°F–90°F, and equal to or greater than 90°F. WBGT ranges correspond to common ranges of WBGT flag levels. Forecast WBGT was from the NBM. SC WBGT and SC WBGT Land are the WBGT forecasts using the Liljegren methodology, with the SC WBGT using PG Stability classes and SC WBGT Land using surface roughness to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the US NWS methods.
Variations Across Space

Assessing the forecast bias spatially (average daily maximum WBGT) demonstrates the high variability across the region and between forecast methods (Figures 8–10).

Figure 4.8. WBGT Bias: Regional Variations.

Note. SC WBGT (Top) and SC WBGT Land (Bottom) Forecast Bias (NBM) for Daily Maximum WBGT for ASOS/AWOS and ECONet weather stations when WBGT was equal to or greater than 85°F. The comparison of forecast WBGT at these stations is relative to the estimated WBGT at these stations using the Liljegren et al. (2008) methodology. Negative bias means the forecast was too cold. SC WBGT and SC WBGT Land are the WBGT forecasts using the Liljegren methodology, with the SC WBGT using PG Stability classes and SC WBGT Land using surface roughness to downscale wind speeds.
Figure 4.9. SC WBGT vs. SC WBGT Land Forecast Bias.

**Difference in Forecast Bias between SC WBGT and SC WBGT Land**

Note. SC WBGT bias is subtracted by SC WBGT Land bias (NBM). SC WBGT and SC WBGT Land are the WBGT forecasts using the Liljegren methodology, with the SC WBGT using PG Stability classes and SC WBGT Land using surface roughness to downscale wind speeds.

A majority of stations displayed a negative bias (-1.9°F–0°F) for SC WBGT, with the most negative bias occurring at three stations located along the Appalachian Mountains (Figure 4.8). However, SC WBGT Land bias was greater than SC WBGT, and positive at all but ten stations, five of which were along the Appalachian Mountains (Figures 4.8–4.9). The differences in bias between the two SC WBGT methods ranged from 0°F–1.8°F (WBGT > 84°F), with the largest bias difference occurring at stations with lower roughness (detailed below). Lastly, NWS WBGT forecast bias was significantly more negative, with the coolest bias along the Appalachian Mountains, southern and western piedmont of NC, and in coastal areas of NC (Figure 4.10).
Figure 4.10. NWS WBGT Bias: Regional Differences.

NWS WBGT Daily Max. Forecast Bias

Note. NWS WBGT Forecast Bias (based on NBM) for Daily Maximum WBGT for ASOS/AWOS and ECONet weather stations when WBGT was equal to or greater than 85°F. Negative bias means the forecast was too cold. NWS WBGT is the WBGT as calculated by the US NWS.

Variations by Hour of Day and Weather Conditions

Forecast bias varied across the daylight hours for Observed and Station WBGT. When Observed WBGT was equal to or greater than 85°F, both SC WBGT forecasts were positive at all hours (except for 1100 and 1800 hours), and the NWS WBGT was negatively biased throughout the day, except for 1600 hours (Figure 4.11). All biases were larger for every hour of day when WBGT was greater than 90°F (Appendix 4.2.2). Like Observed WBGT, Station WBGT forecast bias followed a diurnal pattern, with the biases for all methods becoming more positive (warm) around and immediately after solar noon (Figure 4.11). The highest median bias for SC WBGT and SC WBGT Land occurred at 1300 and 1400 hours. Similar to Observed WBGT, early afternoon was when the NWS WBGT median bias was closest to zero, but the bias remained negative (cool) throughout all hours of the day (Figure 4.11).
Figure 4.11. *WBGT Forecast Bias by Hour of Day.*

**Forecast Bias by Hour of Day (NBM) (WBGT >= 85°F)**

Note. WBGT Forecast bias by hour of day for the three WBGT forecast methods (WBGT >= 85°F) for Observed WBGT (left) and Station WBGT (right). SC WBGT and SC WBGT Land are the WBGT forecasts using the Liljegren methodology, with the SC WBGT using PG Stability classes and SC WBGT Land using surface roughness to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the US NWS methods.

Furthermore, SC WBGT forecast bias was markedly greater when WBGT was less than 88°F, particularly when dew point temperatures were high, wind speed was relatively high and variable (wide range of values) (1.4–4.8 mph), and solar radiation was both low and variable (Figure 4.12). Similar patterns were seen when stratifying Station WBGT forecast bias (Appendix 4.3), but with less variation in bias magnitudes. Across all WBGTs, Station WBGT forecast bias was more negative when wind speeds were slower (less than 4.1 mph) (Appendix 4.3). When dew point temperatures were above 74.9°F and WBGT was less than 86°F, forecast bias was more positive than at lower dew point temperatures. This pattern reversed when WBGT was higher than 86°F, with bias more negative than at lower dew point temperatures. Lastly, when WBGT was above 88°F, SC WBGT bias was more negative when solar radiation was low (median biases of -2.5°F vs. 1.1°F WBGT at higher solar radiation values) (Appendix 4.3).
Figure 4.12. *SC WBGT Bias Stratified by Meteorological Variables for Observed WBGT.*

Noteworthy, SC WBGT Bias for Observed WBGT, stratified by quantiles (0.3, 0.5, 0.75, 0.9, 1) of air temperature, dew point temperature, wind speed, and solar radiation. WBGT (x-axis) rounded to every second nearest integer (NBM 24-hour forecast). SC WBGT is the SERCC-CISA WBGT with wind speed downscaled to 2 meters using PG Stability Classes.

**Forecast WBGT Flag Accuracy**

The forecast WBGT flag accuracy results detailed here refer to the NBM 24-hour forecast.

The most notable difference between methods when assessing “percent correct” was with Observed WBGT (Table 4.3). Both SC WBGT methods had the highest percent correct for yellow flag. For black flag observed, SC WBGT Land had a roughly 10% lower percent correct than the SC and NWS WBGT (Tables 4.3–4.4). The bias scores were similar and indicated underforecasting for green and yellow flags (Observed) for both SC WBGT methods (Table 4.3). For red and black observed flags, both SC WBGT methods overforecast and the SC WBGT Land bias notably increased from red (1.07) to black flag (2.11) (Table 4.3). Station WBGT Flag accuracy was similar; however, SC WBGT Land bias was 1.69 versus 1.07 for red flag (Table 4.4) and the NWS WBGT hit rate for black flag dropped from 42% to 18% (Table 4.4).
Table 4.3. Forecast WBGT Flag Accuracy Compared to Observed WBGT.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Method</th>
<th>Green Flag</th>
<th>Yellow Flag</th>
<th>Red Flag</th>
<th>Black Flag</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SC WBGT</td>
<td>81%</td>
<td>70%</td>
<td>66%</td>
<td>76%</td>
</tr>
<tr>
<td>Percent Correct (%)</td>
<td>SC WBGT Land</td>
<td>81%</td>
<td>72%</td>
<td>65%</td>
<td>66%</td>
</tr>
<tr>
<td></td>
<td>NWS WBGT</td>
<td>78%</td>
<td>58%</td>
<td>62%</td>
<td>77%</td>
</tr>
<tr>
<td>Hit Rate (%)</td>
<td>SC WBGT</td>
<td>26%</td>
<td>30%</td>
<td>42%</td>
<td>84%</td>
</tr>
<tr>
<td></td>
<td>SC WBGT Land</td>
<td>25%</td>
<td>21%</td>
<td>23%</td>
<td>94%</td>
</tr>
<tr>
<td></td>
<td>NWS WBGT</td>
<td>28%</td>
<td>39%</td>
<td>39%</td>
<td>42%</td>
</tr>
<tr>
<td>Bias Score</td>
<td>SC WBGT</td>
<td>0.33</td>
<td>0.69</td>
<td>1.42</td>
<td>1.53</td>
</tr>
<tr>
<td></td>
<td>SC WBGT Land</td>
<td>0.29</td>
<td>0.45</td>
<td>1.07</td>
<td>2.11</td>
</tr>
<tr>
<td></td>
<td>NWS WBGT</td>
<td>0.57</td>
<td>1.28</td>
<td>1.49</td>
<td>0.67</td>
</tr>
<tr>
<td>False Alarm Ratio</td>
<td>SC WBGT</td>
<td>0.21</td>
<td>0.57</td>
<td>0.70</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>SC WBGT Land</td>
<td>0.14</td>
<td>0.54</td>
<td>0.78</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>NWS WBGT</td>
<td>0.51</td>
<td>0.69</td>
<td>0.74</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Note. SC WBGT and SC WBGT Land are the WBGT forecasts using the Liljegren methodology, with the SC WBGT using PG Stability classes and SC WBGT Land using surface roughness to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the US NWS methods.

Table 4.4. Forecast WBGT Flag Accuracy Compared to Station WBGT.

<table>
<thead>
<tr>
<th>Metric</th>
<th>Method</th>
<th>Green Flag</th>
<th>Yellow Flag</th>
<th>Red Flag</th>
<th>Black Flag</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Correct (%)</td>
<td>SC WBGT</td>
<td>78%</td>
<td>70%</td>
<td>84%</td>
<td>93%</td>
</tr>
<tr>
<td></td>
<td>SC WBGT Land</td>
<td>74%</td>
<td>66%</td>
<td>81%</td>
<td>90%</td>
</tr>
<tr>
<td></td>
<td>NWS WBGT</td>
<td>77%</td>
<td>68%</td>
<td>86%</td>
<td>94%</td>
</tr>
<tr>
<td>Hit Rate (%)</td>
<td>SC WBGT</td>
<td>72%</td>
<td>55%</td>
<td>43%</td>
<td>52%</td>
</tr>
<tr>
<td></td>
<td>SC WBGT Land</td>
<td>59%</td>
<td>49%</td>
<td>42%</td>
<td>75%</td>
</tr>
<tr>
<td></td>
<td>NWS WBGT</td>
<td>86%</td>
<td>46%</td>
<td>24%</td>
<td>18%</td>
</tr>
<tr>
<td>Bias Score</td>
<td>SC WBGT</td>
<td>0.83</td>
<td>1.16</td>
<td>1.32</td>
<td>1.21</td>
</tr>
<tr>
<td></td>
<td>SC WBGT Land</td>
<td>0.61</td>
<td>1.20</td>
<td>1.69</td>
<td>2.24</td>
</tr>
<tr>
<td></td>
<td>NWS WBGT</td>
<td>1.16</td>
<td>0.97</td>
<td>0.71</td>
<td>0.31</td>
</tr>
<tr>
<td>False Alarm Ratio</td>
<td>SC WBGT</td>
<td>0.14</td>
<td>0.52</td>
<td>0.67</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>SC WBGT Land</td>
<td>0.08</td>
<td>0.59</td>
<td>0.75</td>
<td>0.67</td>
</tr>
<tr>
<td></td>
<td>NWS WBGT</td>
<td>0.26</td>
<td>0.53</td>
<td>0.66</td>
<td>0.42</td>
</tr>
</tbody>
</table>

Note. Station WBGT is estimated WBGT at weather stations using the Liljegren et al. (2008) methodology. SC WBGT and SC WBGT Land are the WBGT forecasts using the Liljegren methodology, with the SC WBGT using PG Stability classes and SC WBGT Land using surface roughness to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the US NWS methods.
For Observed and Station WBGT flags, the false alarm ratio was highest for red flags by similar magnitudes across all forecast methods (Tables 4.3–4.4). The hit rate was notably higher for the SC WBGT methods relative to the NWS WBGT for 1) Observed black flag (Table 4.3) and 2) red and black flags for Station WBGT (Table 4.4). In both cases, the SC WBGT Land had the highest hit rates for black flag (94% and 75% for Observed and Station black flags, respectively).

Lastly, there were notable differences in bias between the 24- and 48-hour NBM forecast, with a lower bias for all methods for the 48-hour forecast (both Observed and Station WBGT) (Appendix 4.4). However, the hit rate was higher for the 24-hour forecast, 92% vs. 84% for SC WBGT, 94% vs. 77% for SC WBGT Land, and 42% vs. 8% for NWS WBGT (Appendix 4.4).

**Table 4.5. Verification Scores for WBGT Flag Forecast by Estimation Method.**

<table>
<thead>
<tr>
<th>WBGT Flag Verification Scores (NBM 24 hour forecast)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td><strong>Observed WBGT</strong></td>
</tr>
<tr>
<td>SC WBGT</td>
</tr>
<tr>
<td>SC WBGT Land</td>
</tr>
<tr>
<td>NWS WBGT</td>
</tr>
<tr>
<td><strong>Station WBGT</strong></td>
</tr>
<tr>
<td>SC WBGT</td>
</tr>
<tr>
<td>SC WBGT Land</td>
</tr>
<tr>
<td>NWS WBGT</td>
</tr>
</tbody>
</table>

*Note. SC WBGT and SC WBGT Land are the WBGT forecasts using the Liljegren methodology, with the SC WBGT using PG Stability classes and SC WBGT Land using surface roughness to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the US NWS methods.*

In addition to assessing the forecast accuracy for each flag level, two metrics were chosen to summarize the accuracy of each forecast method overall. For Observed WBGT, the Gerrity Skill Score (GSS) and Heidke Skill Score (HSS) were relatively similar across methods (Figure 4.5), with the SC WBGT having the highest scores, followed by SC WBGT Land and the NWS WBGT. At increasing forecast lead times, the GSS for SC WBGT Land remained relatively consistent (Table 4.6). Contrasting Observed WBGT, at increasing forecast lead times, the Station WBGT GSS for both SC
WBGT methods decreased, with a GSS of 0.31 for SC WBGT, 0.41 for SC WBGT Land, and 0.18 for NWS WBGT for a 72-hour forecast (Table 4.6).

**Table 4.6. Gerrity Skill Scores for NBM WBGT Flag Forecasts.**

<table>
<thead>
<tr>
<th>Method</th>
<th>24-hour</th>
<th>48-hour</th>
<th>72-hour</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Observed WBGT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC WBGT</td>
<td>0.40</td>
<td>0.36</td>
<td>0.31</td>
</tr>
<tr>
<td>SC WBGT Land</td>
<td>0.34</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td>NWS WBGT</td>
<td>0.32</td>
<td>0.17</td>
<td>0.11</td>
</tr>
<tr>
<td><strong>Station WBGT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SC WBGT</td>
<td>0.55</td>
<td>0.36</td>
<td>0.31</td>
</tr>
<tr>
<td>SC WBGT Land</td>
<td>0.62</td>
<td>0.47</td>
<td>0.41</td>
</tr>
<tr>
<td>NWS WBGT</td>
<td>0.36</td>
<td>0.21</td>
<td>0.18</td>
</tr>
</tbody>
</table>

*Note. SC WBGT and SC WBGT Land are the WBGT forecasts using the Liljegren methodology, with the SC WBGT using PG Stability classes and SC WBGT Land using surface roughness to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the US NWS methods.*

**Surface Roughness**

**Identified Improvements in Accuracy**

In addition to assessing the accuracy of WBGT forecasts, this research aimed to improve the SC WBGT forecast by incorporating surface roughness into the downscaling of wind speeds from 10 to 2 meters. Using surface roughness to downscale wind speeds had a distinguishable impact on the forecast and resulting bias. Overall, using surface roughness in downscaling resulted in slower 2-meter wind speeds compared to the winds downscaled using the SRDT method and Pasquil-Gifford Stability classes (used in SC WBGT). Given the high sensitivity of WBGT to wind speed (e.g., the rapid increase in WBGT under low wind speeds seen in Figure 4.2), the improved (minimized) bias for SC WBGT Land when WBGT was equal to or greater than 90°F can be directly related to an improved 2-meter wind speed forecast with surface roughness being used (Figure 4.6). Furthermore, the SC WBGT Land forecast bias had lower magnitudes of positive bias increases at and immediately after solar noon compared to the SC WBGT and NWS WBGT (Figure 4.11).
Using a simple linear model to assess the relationship between surface roughness and forecast bias revealed surface roughness as a statistically significant predictor for the bias differences between the SC WBGT Land relative to both the SC WBGT and NWS WBGT (Figure 4.13). As surface roughness increased, differences between the SC WBGT Land bias compared to the bias of other methods also increased (Figure 4.13). Thus, SC WBGT Land was a particularly better forecast for sites with rougher surfaces.

**Figure 4.13. SC WBGT vs. SC WBGT Land Bias relative to Surface Roughness.**

![Average Forecast Bias Across Roughness Lengths](image)

*Note.* Comparison of the differences in forecast WBGT bias between the SC WBGT and SC WBGT Land as a function of surface roughness. Points are average forecast bias for all stations at a given surface roughness value (at 0.25 m roughness length increments). SC WBGT and SC WBGT Land are the WBGT forecasts using the Liljegren methodology, with the SC WBGT using PG Stability classes and SC WBGT Land using surface roughness to downscale wind speeds.

**Weighted Average Surface Roughness.**

Lastly, weighted average surface roughness with different weightings for the roughness at varying spatial scales (30 m, 100 m, 250 m, and 500 m) (Appendix 4.1.1) was not sensitive to the different configurations of weights for calculating this average (Appendix 4.1.2). Variations in the WBGT calculated using these different surface roughness values and corresponding wind speeds ranged from -0.2°F–0.3°F, but the majority of differences (between the 25th and 75th percentiles) ranged from -0.07°F–0.06°F (Appendix 4.1.2). The analysis here thus utilized the surface roughness weighting schema of 30 m (10%), 100 m (25%), 250 m (50%), and 500 m (15%).
**Sentinel vs. Landsat Imagery**

The second aspect of the investigation into surface roughness sought to determine if the finer resolution of Sentinel 2A/2B satellite imagery (10 m) offered significant improvement relative to Landsat 8 ETM+ (30 m). Comparisons revealed no differences between the roughness values derived from these two sources, except for negligible differences in forested areas where there were slight differences in NDVI. Given the limited differences between the two and despite the challenges associated with higher resolution data (e.g., storage space, data processing speed, etc.), the analysis here proceeded with using the Sentinel imagery as it was already processed.

**Discussion and Conclusion**

Wet bulb globe temperature (WBGT) is a heat stress index that is increasingly utilized for safeguarding health, such as in athletics and workplaces, since it is robust in accounting for environmental variables influencing heat stress. However, forecasts of WBGT have not been standard, with efforts to create these forecasts in the United States having only been undertaken recently. Furthermore, validation of the accuracy of these forecasts has been limited spatially, with respect to the number of sites with ground truth measurements, and temporally, by limited number of days sampled for verification. The research presented here assessed the accuracy of WBGT forecasts relative to 1) Observed WBGT with an ISO-compliant WBGT meter and 2) estimated WBGT at weather stations. This research also evaluated efforts to improve these forecasts by more accurately estimating the wind speed at 2 meters based on wind speed at 10 meters above the ground.

**Overall WBGT Forecast Bias**

Comparisons between WBGT forecasts using the NBM and NDFD revealed some differences, particularly at higher WBGTs, with the NDFD having a more negative (cold) bias. It is hypothesized that these differences are primarily driven by differences in the forecast wind speed
between the two models, with the NDFD wind speeds being faster than the NBM (Appendix 4.5). Although the NDFD is derived from the NBM, individual NWS weather forecast offices modify NBM output as they see fit in the creation of the official NDFD forecast. Thus, it is evident that the tendency for wind speeds to be increased from the baseline wind speeds forecast in the NBM has a significant impact on the performance of the NDFD WBGT forecast, particularly when WBGT is high.

The SC WBGT (which used the Liljegren et al. (2008) methodology for calculating WBGT) was found to be more accurate than the method used by the US NWS in most instances, particularly when conditions were dangerous. This pattern was consistent when assessing hourly forecast accuracy as well as the accuracy of forecasting daily maximum WBGT. When WBGT was relatively cooler (i.e., less than 85°F), the SC WBGT was more positively (warm) biased than the NWS WBGT. However, at higher levels of heat stress and WBGTs, the NWS method underestimated the WBGT while the SC WBGT provided more accurate results.

These findings of higher accuracy with the SC WBGT further support existing research regarding the accuracy of the Liljegren et al. (2008) methodology (Lemke & Kjellstrom, 2012; Patel et al., 2013). For red and black flags, SC WBGT was within 1.5°F of Observed WBGT. Accuracy at such high values is most critical given the use of WBGT in making decisions about the safety of outdoor activity. While this error range (1.5°F) could still lead to WBGT flag misclassification in some instances, it is sufficiently accurate when compared to the larger error range of common WBGT meters and, for example, based on standards from the Japanese National Institute of Occupational Safety and Health (Racinais et al., 2022). Furthermore, in this case, the SC WBGT bias erred on the side of caution (being too warm) under these most thermally stressful conditions, which is preferred over a dangerous cold bias since WBGT is ultimately used to safeguard health.
**Variations in Forecast Bias**

WBGT forecast bias varied across space, weather conditions, and based on the hour of the day. Finding that the coldest biases for all methods was located in the Appalachian Mountains was unsurprising due to the intricate influences of terrain in that area, which are not accounted for in this work. Additionally, weather forecasts for the Mountains are more complex during the summer with respect to cloud cover (and thus solar radiation), as terrain-induced, daytime thunderstorms are challenging to predict. Consequently, the cool outflows and debris cloud fields from these storms can potentially cause large forecast errors. This is particularly true for the daily maximum WBGT, which is more likely to be impacted than a given hourly observation.

For the NWS WBGT, three other regions stand out: central NC, the Midwest, and the Southern Plains. On average, the NWS WBGT bias was most accurate in these regions, but was still negative. It is hypothesized that this is related to the effect of wind speed on the differences in the estimated natural wet bulb temperature and black globe temperature from the different methodologies, with there being less difference at higher wind speeds. While central NC is not a region with low surface roughness overall, the stations where the NWS WBGT performs better are stations where surface roughness is low, relative to neighboring stations. This is especially true for the sparsely forested Midwest and Southern Plains, which have climatologically faster wind speeds.

Distinct variations in forecast bias were seen when stratifying by dew point temperature, wind speed, and solar radiation. Rapid variability in these variables (together and separately) leads to high variability in WBGT, thus increasing the likelihood of errors for the one top-of-the-hour forecast value. The importance of wind speed was highlighted by the overall range of SC WBGT bias being greater when wind speeds were low (< 1.4 mph). The SC WBGT forecast bias being most variable when solar radiation was low and variable can be at least partially attributed to the challenge with forecasting cloud cover and the resulting impact on the estimated solar radiation.
from this cloud cover (Figure 4.12). This is also evident in the diurnal curve of forecast bias, with the bias of all forecast methods becoming increasingly positive towards midday (e.g., the period of highest solar radiation) and then reversing through the afternoon (Figure 4.12).

**Forecasting WBGT Flags**

As was the case when assessing WBGT forecast bias at each temperature value, there were noteworthy differences in accuracy between the methods for forecasting WBGT flag levels. The forecast WBGT flag accuracy analysis revealed that the SC WBGT and SC WBGT Land were superior in forecasting black flag, with higher hit rates. The high percentage of correct black flag forecasts by NWS WBGT can be misleading if taken at face value (e.g., 94% for Station WBGT) (Table 4.4). While this means that when the NWS WBGT was forecasting a black flag, a black flag was very likely to occur, there was a consistent tendency to under forecast black flag. Similarly, the high black flag hit rates for both SC WBGT methods partially result from an tendency to over forecast black flags, with the SC WBGT and SC WBGT Land having a 45% and 55% false alarm ratio for observed black flags, respectively (Table 4.3). The narrow range defining red flag (88°F–90°F) presents a challenge with assessing the categorical forecast accuracy, as the assessment can be deceptively skewed poorly even if the bias was only 1°F–2°F. This is evident when comparing the HSS and GSS, since the GSS is higher as it accounts for the ‘closeness’ of the categorical forecast misses (Tables 4.5–4.6).

**Surface Roughness**

Lastly, given the paramount influence of wind speed on WBGT, this research addressed how the influence of land cover and associated surface roughness impact current efforts to forecast wind speed and WBGT. Since wind speeds are faster at 10 meters above the surface, to accurately assess the wind speeds critical for human heat stress, wind at 2 meters above the surface must be estimated. This downscaling (translating) of wind speed can be completed in several ways, with the research here comparing two methods: 1) Pasquil-Gifford Stability classes and 2) surface roughness
values (derived from high-resolution satellite imagery). Incorporating more granular land cover information and surface roughness improved the WBGT forecasts. Importantly, however, this improvement was not uniform. The use of surface roughness for downscaling wind speeds results in more positively (warm) biased WBGT at sites with low surface roughness relative to sites with higher roughness, providing a point of investigation for future research. Future work could ascertain the roughness at and below which the use of surface roughness does not improve the WBGT forecast. Exploring other methods for estimating surface roughness at a reasonable scale may also prove beneficial, including incorporating the influence of differences in surface roughness from different directions (e.g., for a given point, wind from the north travels over dense forest and thus is slowed more by roughness than winds from the south that travel over open fields).

Additionally, there were negligible differences between the surface roughness values derived from the Sentinel and Landsat imagery. It is hypothesized that this is largely due to the use of the NCLD 2019 as the landcover data with which the vegetation indices from the images were paired. If land cover were classified directly from the Sentinel and Landsat imagery themselves, there might have been more of a difference. However, given the challenge of accurately classifying land cover over broad spatial areas and the need to consider surface roughness values at varying scales (e.g., 30 m, 100 m, etc.), any substantial differences and possible benefits of using the higher resolution imagery are unlikely and significantly less feasible to operationalize.

Estimating WBGT remains a challenge due to the complexity of the index since it responds to numerous environmental variables. This research continues to emphasize the importance of selecting the best methodologies for estimating WBGT and demonstrates the ability to forecast WBGT accurately, particularly when the conditions are dangerous. As more organizations and entities begin using WBGT, accurate WBGT forecasts will continue to increase in value as they enable robust planning for outdoor activity and early warning of particularly hazardous periods.
CHAPTER 5: CONCLUSIONS

Exposure to heat presents a major threat to human health, with heat being the leading cause of weather-related death in the United States (CDC, 2010; NOAA's National Weather Service, 2020). In this era of anthropogenic climate change, the threat to human health posed by extreme temperatures contemporarily is further compounded by the looming future threat. Coffel et al. (2018) identified the Southeastern United States as one of four regions projected to see the largest increase in humidity by 2070–2080, with days of highest humidity coinciding with an increase in air temperature of 3°C–4°C, increasing heat stress levels to very dangerous levels. Enhancing our current ability to measure heat stress provides the foundation for future efforts to safeguard health in a warming world. In addition to assessing how well heat stress indices predict morbidity, especially with respect to the utility of using WBGT for this prediction, this research also evaluates the ability to forecast WBGT several days into the future, therein investigating variations in WBGT and heat stress across small distances and different surfaces.

The primary objectives of this dissertation were to:

1. Investigate and compare the relationships of three heat stress indices (air temperature, the Heat Index, and WBGT) with morbidity across North Carolina.

2. Characterize variations in WBGT across small distances (10s to 100s of meters) and over different surfaces, including ascertaining the accuracy of WBGT estimates.

3. Determine if WBGT can be forecast reliably and if improvements can be made to the forecast by incorporating land cover data that are at a higher spatial resolution than the forecast grid itself.
The following research questions were investigated to meet these objectives:

1. How does the use of WBGT to predict different morbidities compare to using air temperature and the Heat Index for this prediction? Do relationships between heat and morbidity vary based on morbidity type, age groups, or geographic location within North Carolina?

2. How does WBGT vary across small distances (10s to 100s of meters), over different surfaces (e.g., tennis courts), and in differing shade conditions (e.g., a shaded forest vs. shaded grass field)? What is the accuracy of estimating WBGT with the methodology developed in Liljegren et al. (2008), and how does this compare with a commonly used WBGT meter?

3. What is the accuracy of forecasting WBGT and how does this compare to other forecasts of the index? Can forecasts be improved by incorporating the varying influence of land cover on wind speed (i.e., variations in surface roughness)?

Each of these questions were addressed in each chapter of the dissertation.

In Chapter 2, emergency department visits from 2007-2016 for North Carolina and two sources of weather data (weather stations and the ERA5-Land) were utilized to assess heat-morbidity relationships. Daily statistics for three heat stress indicators (air temperature, the Heat Index, and WBGT) were calculated for each county. Generalized Additive Models were used to model the relationships between morbidity and each heat stress indicator, accounting for the underlying trends, seasonality, holidays, and differences in geography across the state.

This work determined that meteorological data from weather stations were more robust than the use of the ERA5-Land data in modeling these relationships. After comparing the models for each morbidity and heat stress index combination, daily maximum WBGT (non-lagged, one-day maximum) was found to be a better predictor of morbidity to a statistically significant degree.
HRI and mental health morbidity were found to have more characteristic relationships with increasing temperature (i.e., a “j-shaped” curve) relative to the other morbidities investigated (cardiovascular and all-cause morbidity). Across age groups, the WBGT thresholds at which the largest increases in HRI occurred were similar. However, the WBGT thresholds for all other morbidities were lower (at cooler values) for older age groups, which would be expected given the population’s higher vulnerability to heat. While the magnitudes of increases in morbidity across age groups were similar in most instances, larger increases were found at the identified thresholds for those aged 20–34 and 35–54. This is believed to be related to occupational exposure, with a hypothesis that WBGT may better correlate with this type of exposure and since this age group has more outdoor workers relative to other age groups. Regionally, rural regions were found to have a higher heat-morbidity burden than urban regions for HRI, specifically in the rural coastal plain of NC. However, urban regions had the highest heat-morbidity burden at extremely high WBGTs (greater than 91°F) for all-cause morbidity. Overall, rural regions experienced larger morbidity increases at lower (cooler) values compared to urban regions, which is hypothesized to relate to the higher proportion of outdoor workers in these regions.

Lastly, the research presented in Chapter 2 found the WBGT thresholds for red and black flag (88°F and 90°F, respectively) corresponded to the largest increases in HRI and mental health morbidity for all ages, and all-cause morbidity for those aged 13–54. The largest increases in morbidity when using the Heat Index were at values of 103°F and 108°F, notably lower than existing thresholds the US NWS uses in issuing heat advisories and warnings (although other variables, such as minimum temperature and consecutive days of heat, are also weighed in this decision-making process).

In Chapter 3, the accuracy of WBGT estimates and variations in WBGT over varying distances, within differing microclimates, and across several surface types were assessed. WBGT
data and weather station data were collected in a variety of locations and microclimates, specifically using Kestrel 5400 devices and a WBGT meter that was built to meet standards set by the International Organization for Standardization. Overall, the Liljegren et al. (2008) methodology for estimating WBGT was found to be more accurate than the WBGT as measured by the Kestrel 5400, with biases being +0.4°F and +1.2°F warmer than observed, respectively. The bias of the Kestrel 5400 was also found to increase as WBGT increased. Additionally, the method from Liljegren et al. (2008) for estimating the black globe temperature was not biased due to the underlying assumption that the air temperature is equivalent to the skin temperature.

Along with weather station and WBGT measurements, skin temperatures of bare ground, alive grass, and dormant (senesced) grass were compared, with the skin temperature of bare ground found to be the highest, followed by dormant grass and alive grass. These findings ultimately relate to the differences in albedo between these surfaces as well as the differences in latent versus sensible heat flux.

Large differences were found when comparing WBGT measured over different surfaces and shade conditions. The WBGT measured in the sun was 6.8°F (6.1°F–7.4°F) warmer than shaded WBGT. Furthermore, as WBGT in the sun increased, the difference between the sun and shade WBGT increased markedly. WBGT differences were also found in comparing measurements over a grassy field and a shaded forest, with the grassy field being 9.7°F (9.3°F–10.1°F) warmer. Shaded forest measurements were found to be higher than WBGT measured in a shaded grass field by +0.8°F (0.3°F–1.7°F). This is hypothesized to result from the differences in the amount of radiation the respective surfaces had received throughout the day, since the field had been in direct sun 4–5 hours before becoming shaded. Lastly, WBGT measured over a tennis court was found to be +1.6°F (0.5°F–2.5°F) warmer than WBGT over a neighboring grassy field.
The last component of Chapter 3 compared methods for estimating clear-sky radiation and the accuracy of then modifying that radiation by percentage cloud cover. This is important since few weather stations measure solar radiation, and all airport weather stations in the US measure cloud cover fraction. The “Ryan model” developed in Ryan & Stolzenbach (1972) was found to be most accurate when setting the atmospheric transmissivity coefficient to 0.65. The modification of clear-sky radiation by the calculated percentage cloud cover from the images of sky conditions resulted in median differences with observed solar radiation of +/- 75 w/m².

Chapter 4 presented an accuracy assessment of a WBGT forecast, which was developed through a partnership between NOAA’s Southeast Regional Climate Center (SERCC) and Carolinas Integrated Sciences and Assessments (CISA), referred to as “SC WBGT”. To assess this accuracy, WBGT forecast values were compared to both Observed WBGT (measurements made with the WBGT meter compliant with International Organization for Standardization guidelines used in Chapter 3) and Station WBGT (WBGT estimated from weather stations). The accuracy of the SC WBGT forecast was also compared to the WBGT forecasts made by the US NWS. The US NWS WBGT forecast is the only other available source for forecast WBGT in the US to our knowledge at the time of writing. The second part of Chapter 4 utilized high-resolution satellite imagery from Sentinel 2A/2B and Landsat 8 ETM+, along with the 2019 National Land Cover Database, to derive estimates of surface roughness at a fine scale (30 meters). The WBGT forecast utilizing these surface roughness values to downscale wind speed is referred to as “SC WBGT Land”.

Compared to Observed WBGT, the SC WBGT forecast bias was positive (warm) when WBGT was 80°F–88°F, with bias becoming less positive as WBGT increased. The NWS WBGT was consistently more negative (cool) than the SC WBGT. When WBGT was equal to or above 90°F, both the SC WBGT and NWS WBGT had negative (cool) median biases: -0.8°F and -2.1°F, respectively.
Regional differences in forecast bias were revealed (based on Station WBGT). A majority of stations displayed a negative bias (-1.9°F–0°F) for SC WBGT, with the most negative bias occurring at three stations located along the Appalachian Mountains. Likewise, the NWS WBGT bias was most negative (cool) in this region, and more negatively biased than the SC WBGT at most stations. Also, assessing the WBGT forecast bias based on hour of day revealed the bias was maximized around solar noon, i.e., when the sun is highest in the sky.

Since WBGT is used based on associating a given reading with established flag levels and corresponding recommendations to safeguard health, the accuracy of forecasting each WBGT flag level was assessed. The SC WBGT and SC WBGT Land were superior in forecasting black flag, with high hit rates. This is partially a result of these forecasts over forecasting black flags, with the SC WBGT and SC WBGT Land having a 45% and 55% false alarm ratio for observed black flags, respectively.

The last component of this chapter revealed no differences in the surface roughness values derived from the Sentinel and Landsat imagery, despite the differences in spatial resolution. The use of surface roughness in downscaling wind speeds resulted in slower 2-meter winds relative to the downscaling that used Pasquil-Gifford Stability classes (used in SC WBGT). Thus, the SC WBGT Land was consistently slightly higher than the SC WBGT. The important differences in forecast bias between the SC WBGT and SC WBGT Land (relative to Observed WBGT) are 1) the SC WBGT land had lower increases in bias at solar noon, given the solar radiation values were not used in downscaling wind speeds as was the case with SC WBGT and 2) the SC WBGT Land bias was lower (and slightly positive) than the SC WBGT (which was slightly negative) for when WBGT was equal to or greater than 90°F.

The research presented in this dissertation addressed numerous gaps in literature and complimented a panoply of research on WBGT and heat stress more broadly. Identifying that
WBGT is a robust predictor of morbidity creates a foundation for using WBGT in future heat-health warning systems. With respect to research on North Carolina specifically, this work further affirmed the higher heat-morbidity burden, particularly at cooler values, in rural areas relative to urban areas (Kovach et al., 2015; Sugg et al., 2016). This is significant as it contrasts other findings that focus on and highlight vulnerability to heat in urban areas (Hondula et al., 2013; Ketterer & Matzarakis, 2014; Pálgy et al., 2005).

This research also addressed an existing gap in knowledge related to the utility of WBGT flag levels (i.e., do the standard WBGT values used as thresholds for defining the flag categories correspond with distinct increases in danger?). Verifying these thresholds corresponded with distinct increases in morbidity further supports the use of WBGT and these thresholds in safeguarding health. Similarly, the research finding that morbidity increases at lower values of the Heat Index than what is typically used for issuing heat advisories and warnings supports lowering the Heat Index thresholds.

The WBGT and weather station data collected here spanned significantly longer time periods than those used in numerous studies with similar objectives (Cooper et al., 2017; Dimiceli et al., 2011; Kopec, 1977; Lemke & Kjellstrom, 2012; Liljegren et al., 2008). The research finding that the WBGT was +1.6°F (0.5°F–2.5°F) warmer over a tennis court compared to a neighboring grass field provides a valuable contrast to existing work. These results differ from those in Grundstein & Cooper (2020), which found limited differences in WBGT over these two surfaces. It is hypothesized that a major contributor to these differences is that the WBGTs measured and compared in this study were roughly 8°F warmer. Importantly, the implication is that there may be an even greater difference in WBGT above these surfaces under more thermally stressful conditions, where health effects could be lessened with awareness of such differences across surfaces.
Lastly, this research provided a useful validation of WBGT forecasts. It highlights the importance of choosing appropriate methods for estimating WBGT by demonstrating that the methods in Liljegren et al. (2008) are more accurate. Other less accurate methods present a variety of issues, including instilling a lack of faith in the forecast, which largely nullifies the benefits of the forecast overall, and providing dangerous underestimates of the forecast WBGT under the most thermally stressful conditions. Additionally, this work determined the accuracy of commonly used methods for estimating clear-sky radiation and modifying that radiation by percentage cloud cover. It also demonstrated the utility of using surface roughness at a sub-forecast grid scale in WBGT forecasts, particularly for sites with higher surface roughness. This finding is important since it is at these sites where the use of surface roughness can correct a possibly dangerous cold bias in the forecast WBGT.

Numerous opportunities exist for future research. Further comparison of the use of weather station vs. gridded data (such as the ERA5-Land) would help delineate the best source of weather data to use in future studies on heat-morbidity relationships. This is important since the finding here that weather station data outperformed the ERA5-Land was surprising, and since gridded data is often seen as superior and is increasingly utilized in similar studies. While this research demonstrated WBGT can predict morbidity at a population-level, individual-level analysis of changes in heat stress at existing WBGT thresholds would provide valuable context on the utility of the index, especially given the increasing number of organizations modifying these threshold values.

The comparison between WBGT over tennis courts and a neighboring grass field urges further investigation since the findings here differ from existing research. Future work should consider larger sample sizes over many days and sites when comparing tennis courts and grassy surfaces. Lastly, as efforts to improve WBGT forecasts continue, and assuredly the number of WBGT
forecast sources increases, validating these forecasts will remain an important component, and remain challenging given the limited access to the most robust WBGT meters. Awareness of a given device’s biases when undertaking these forecast validations is critical. Additional validation would be most useful in humid subtropical climates like North Carolina, perhaps with some focus on urban areas and sites with widely varying surface roughness, since WBGT has such high spatial variability. Validating these forecasts in other climates is also important, as WBGT is being increasingly used in the eastern two-thirds of the continental US.
### APPENDIX 2.1: POWER-LAW EXPONENTS USED FOR DOWNSCALING WIND SPEED.

<table>
<thead>
<tr>
<th>Stability Class</th>
<th>Urban Exponent</th>
<th>Rural Exponent</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.15</td>
<td>0.07</td>
</tr>
<tr>
<td>B</td>
<td>0.15</td>
<td>0.07</td>
</tr>
<tr>
<td>C</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>D</td>
<td>0.25</td>
<td>0.15</td>
</tr>
<tr>
<td>E</td>
<td>0.30</td>
<td>0.35</td>
</tr>
<tr>
<td>F</td>
<td>0.30</td>
<td>0.55</td>
</tr>
</tbody>
</table>

*Note.* Table displaying exponents used for downscaling wind speeds with Pasquill-Gifford Stability Classes. The “Urban Exponent” was utilized. Table drawn from US EPA (2000).
APPENDIX 2.2: REGIONAL THRESHOLDS FOR LARGEST MORBIDITY INCREASES.

<table>
<thead>
<tr>
<th>Morbidity</th>
<th>WBGT (°F)</th>
<th>Mountains</th>
<th>Rural Piedmont</th>
<th>Urban Piedmont</th>
<th>Urban Coastal Plain</th>
<th>Rural Coastal Plain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HRI</td>
<td>88</td>
<td>1.16 (1.14-1.21)</td>
<td>1.11 (1.11-1.11)</td>
<td>1.18 (1.18-1.19)</td>
<td>1.13 (1.13-1.14)</td>
<td>1.16 (1.16-1.16)</td>
</tr>
<tr>
<td></td>
<td>89</td>
<td>1.19 (1.18-1.21)</td>
<td>1.12 (1.12-1.12)</td>
<td>1.19 (1.19-1.19)</td>
<td>1.14 (1.14-1.14)</td>
<td>1.17 (1.17-1.17)</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>1.21 (1.13-1.24)</td>
<td>1.13 (1.12-1.13)</td>
<td>1.20 (1.19-1.20)</td>
<td>1.15 (1.14-1.16)</td>
<td>1.18 (1.17-1.18)</td>
</tr>
<tr>
<td></td>
<td>91</td>
<td>1.24 (0.98-1.33)</td>
<td>1.13 (1.13-1.14)</td>
<td>1.20 (1.18-1.22)</td>
<td>1.16 (1.13-1.19)</td>
<td>1.18 (1.16-1.20)</td>
</tr>
<tr>
<td>Mental</td>
<td>86</td>
<td>1.06 (1.06-1.06)</td>
<td>1.00 (1.00-1.00)</td>
<td>1.00 (1.00-1.00)</td>
<td>1.04 (1.04-1.04)</td>
<td>1.03 (1.03-1.03)</td>
</tr>
<tr>
<td></td>
<td>89</td>
<td>1.10 (1.10-1.10)</td>
<td>1.00 (1.00-1.00)</td>
<td>1.01 (1.01-1.01)</td>
<td>1.04 (1.04-1.04)</td>
<td>1.04 (1.03-1.04)</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>1.12 (1.11-1.12)</td>
<td>1.00 (1.00-1.00)</td>
<td>1.01 (1.00-1.01)</td>
<td>1.04 (1.03-1.04)</td>
<td>1.04 (1.03-1.04)</td>
</tr>
<tr>
<td></td>
<td>91</td>
<td>1.13 (1.11-1.15)</td>
<td>1.00 (1.00-1.01)</td>
<td>1.01 (1.00-1.01)</td>
<td>1.04 (1.03-1.04)</td>
<td>1.04 (1.03-1.04)</td>
</tr>
<tr>
<td>All Cause</td>
<td>86</td>
<td>1.02 (1.02-1.02)</td>
<td>0.99 (0.99-0.99)</td>
<td>1.01 (1.01-1.01)</td>
<td>1.02 (1.02-1.02)</td>
<td>1.01 (1.01-1.01)</td>
</tr>
<tr>
<td></td>
<td>90</td>
<td>1.01 (1.00-1.01)</td>
<td>0.99 (0.99-0.99)</td>
<td>1.02 (1.02-1.02)</td>
<td>1.01 (1.01-1.01)</td>
<td>1.00 (1.00-1.00)</td>
</tr>
<tr>
<td></td>
<td>92</td>
<td>1.00 (0.99-1.01)</td>
<td>0.99 (0.99-0.99)</td>
<td>1.03 (1.03-1.03)</td>
<td>1.01 (1.01-1.01)</td>
<td>1.00 (0.99-1.00)</td>
</tr>
<tr>
<td></td>
<td>94</td>
<td>1.00 (0.98-1.01)</td>
<td>0.99 (0.98-0.99)</td>
<td>1.04 (1.04-1.04)</td>
<td>1.01 (1.00-1.01)</td>
<td>0.99 (0.99-1.00)</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>84</td>
<td>1.02 (1.02-1.02)</td>
<td>1.00 (1.00-1.00)</td>
<td>1.00 (1.00-1.00)</td>
<td>1.02 (1.02-1.02)</td>
<td>1.01 (1.01-1.01)</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>1.02 (1.01-1.02)</td>
<td>1.00 (1.00-1.00)</td>
<td>0.99 (0.99-1.00)</td>
<td>1.02 (1.02-1.02)</td>
<td>1.01 (1.01-1.01)</td>
</tr>
<tr>
<td></td>
<td>89</td>
<td>1.00 (1.00-1.00)</td>
<td>0.99 (0.99-0.99)</td>
<td>0.97 (0.97-0.97)</td>
<td>1.02 (1.02-1.02)</td>
<td>1.00 (1.00-1.00)</td>
</tr>
</tbody>
</table>

Note. Table displaying the identified thresholds (bolded values) for each region at which the largest morbidity increases occurred. 95% confidence intervals are included in parentheses.
APPENDIX 4.1

APPENDIX 4.1.1: SURFACE ROUGHNESS WEIGHTED AVERAGE SCHEMA.

Surface Roughness Weighted Average:
weights (%) for each scale tested

<table>
<thead>
<tr>
<th>Scale (m)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>10</td>
<td>10</td>
<td>15</td>
<td>15</td>
<td>25</td>
<td>25</td>
<td>10</td>
</tr>
<tr>
<td>100</td>
<td>25</td>
<td>50</td>
<td>35</td>
<td>50</td>
<td>40</td>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>250</td>
<td>50</td>
<td>25</td>
<td>35</td>
<td>20</td>
<td>25</td>
<td>25</td>
<td>40</td>
</tr>
<tr>
<td>500</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>10</td>
<td>25</td>
<td>30</td>
</tr>
</tbody>
</table>

Note. Table displaying seven configurations of calculating the weighted average surface roughness across different spatial scales for a given pixel.

APPENDIX 4.1.2: WBGT SENSITIVITY TO DIFFERENCES IN WEIGHTED AVERAGES ACROSS SCALES.

WBGT Sensitivity to Weighted Averages of Surface Roughness across Spatial Scales

Note. Plot displays difference between WBGT calculated when using weighted average surface roughness with configuration 1 (Appendix 4.1.1) compared to the six other configurations.
APPENDIX 4.2.

APPENDIX 4.2.1: WBGT FORECAST BIASES.

Observed WBGT Forecast Bias
NBM and NDFD (24-hour forecast)  
WBGT Bias by Forecast Hour (NBM)

Note. WBGT Forecast relative bias for the NBM and NDFD 24-hour forecast (left) and NBM for forecast lead times of 24, 48, and 72 hours (right) (when Observed WBGT was greater than or equal to 80°F).

APPENDIX 4.2.2: WBGT FORECAST BIASES BY HOUR OF DAY.

Forecast Bias by Hour of Day (NBM) (WBGT >= 90°F)

Note. WBGT forecast bias (1-day forecast) by hour of day for when WBGT was equal to or greater than 90°F for Observed WBGT (left) and Station WBGT (right).
APPENDIX 4.3. SC WBGT BIAS STRATIFIED BY METEOROLOGICAL VARIABLES FOR STATION WBGT.

**SC WBGT Bias Stratified (Station WBGT)**

1. Stratified by: Air Temp. (°F)
   - 92.9 - 95.8
   - 95.8 - 98.7
   - 98.7 - 101.4
   - 101.4 - 108.0

2. Dewpoint (°F)
   - 58.3 - 71.0
   - 71.0 - 73.4
   - 73.4 - 75.4
   - 75.4 - 87.0

3. Wind (mph)
   - 3.1 - 4.1
   - 4.1 - 6.9
   - 6.9 - 9.0
   - 9.0 - 27.1

4. Solar Radiation (w/m2)
   - 492 - 631
   - 631 - 777
   - 777 - 865
   - 865 - 1137

*Note.* SC WBGT Bias for Station WBGT, stratified by quantiles (0.3, 0.5, 0.75, 0.9, 1) of air temperature, dew point temperature, wind speed, and solar radiation. WBGT (x-axis) rounded to every second nearest integer (NBM 24-hour forecast). SC WBGT is the SERCC-CISA WBGT with wind speed downscaled to 2 meters using PG Stability Classes.
APPENDIX 4.4: WBGT BLACK FLAG ACCURACY METRICS (NBM 24- AND 48-HOUR FORECASTS).

<table>
<thead>
<tr>
<th>Metric</th>
<th>Method</th>
<th>Observed WBGT</th>
<th>Station WBGT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>24</td>
<td>48</td>
</tr>
<tr>
<td>Bias</td>
<td>SC WBGT</td>
<td>1.53</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>SC WBGT Land</td>
<td>2.11</td>
<td>1.77</td>
</tr>
<tr>
<td></td>
<td>NWS WBGT</td>
<td>0.67</td>
<td>0.15</td>
</tr>
<tr>
<td>Hit Rate (%)</td>
<td>SC WBGT</td>
<td>84%</td>
<td>62%</td>
</tr>
<tr>
<td></td>
<td>SC WBGT Land</td>
<td>94%</td>
<td>77%</td>
</tr>
<tr>
<td></td>
<td>NWS WBGT</td>
<td>42%</td>
<td>8%</td>
</tr>
<tr>
<td>False Alarm Ratio</td>
<td>SC WBGT</td>
<td>0.45</td>
<td>0.33</td>
</tr>
<tr>
<td></td>
<td>SC WBGT Land</td>
<td>0.55</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>NWS WBGT</td>
<td>0.37</td>
<td>0.50</td>
</tr>
<tr>
<td>Percent Correct (%)</td>
<td>SC WBGT</td>
<td>76%</td>
<td>91%</td>
</tr>
<tr>
<td></td>
<td>SC WBGT Land</td>
<td>66%</td>
<td>84%</td>
</tr>
<tr>
<td></td>
<td>NWS WBGT</td>
<td>77%</td>
<td>86%</td>
</tr>
</tbody>
</table>
APPENDIX 4.5. NBM AND NDFD 2-METER WIND COMPARISON.

NBM and NDFD 2-meter wind speed comparison
REFERENCES


CDC. (2010). *Climate Change and Extreme Heat Events*.


Copernicus Sentinel data [2019], processed by ESA.


Guillory, A. (2020). *ERA5: How to calculate wind speed and wind direction from u and v components of the wind?*


Iowa Environmental Mesonet (IEM). *ASOS-AWOS-METAR Data Download*. [Data set]. https://mesonet.agron.iastate.edu/request/download.phtml


NC CRONOS. North Carolina State Climate Office. Raleigh, NC. [Data set]. https://climate.ncsu.edu/cronos/


NOAA. (2017). *Station Pressure.*


North Carolina Disease Event Tracking and Epidemiologic Collection Tool (NC DETECT). [Data set]. Available at: http://www.ncdetect.org


