

The Development and Accuracy Assessment of Wet Bulb Globe Temperature Forecasts

JORDAN CLARK^{a,b}, CHARLES E. KONRAD^{a,c}, AND ANDREW GRUNDSTEIN^d

^a *The University of North Carolina at Chapel Hill, Chapel Hill, North Carolina*

^b *Nicholas Institute for Energy, Environment and Sustainability, Duke University, Durham, North Carolina*

^c *NOAA Southeast Regional Climate Center (SERCC), Chapel Hill, North Carolina*

^d *Department of Geography, The University of Georgia, Athens, Georgia*

(Manuscript received 11 May 2023, in final form 13 November 2023, accepted 15 November 2023)

ABSTRACT: Heat is the leading cause of weather-related death in the United States. Wet bulb globe temperature (WBGT) is a heat stress index commonly used among active populations for activity modification, such as outdoor workers and athletes. Despite widespread use globally, WBGT forecasts have been uncommon in the United States until recent years. This research assesses the accuracy of WBGT forecasts developed by NOAA's Southeast Regional Climate Center (SERCC) and the Carolinas Integrated Sciences and Assessments (CISA). It also details efforts to refine the forecast by accounting for the impact of surface roughness on wind using satellite imagery. Comparisons are made between the SERCC/CISA WBGT forecast and a WBGT forecast modeled after NWS methods. Additionally, both of these forecasts are compared with in situ WBGT measurements (during the summers of 2019–21) and estimates from weather stations to assess forecast accuracy. The SERCC/CISA WBGT forecast was within 0.6°C of observations on average and showed less bias than the forecast based on NWS methods across North Carolina. Importantly, the SERCC/CISA WBGT forecast was more accurate for the most dangerous conditions (WBGT > 31°C), although this resulted in higher false alarms for these extreme conditions compared to the NWS method. In particular, this work improved the forecast for sites more sheltered from wind by better accounting for the influences of land cover on 2-m wind speed. Accurate forecasts are more challenging for sites with complex microclimates. Thus, appropriate caution is necessary when interpreting forecasts and onsite, real-time WBGT measurements remain critical.

SIGNIFICANCE STATEMENT: This research assesses the accuracy of wet bulb globe temperature (WBGT) forecasts. WBGT is a heat stress index that accounts for impacts of air temperature, humidity, wind, and radiation. It is widely used in occupational, athletic, and military settings for heat stress assessment, yet WBGT forecasting in the United States is a relatively new development. These forecasts can be used by decision-makers to better plan activities. We found that WBGT forecasts by NOAA's Southeast Regional Climate Center and Carolinas Integrated Sciences and Assessments were within 0.6°C of observations overall in North Carolina and less biased than forecasts based on methods used by the U.S. National Weather Service, which had larger, colder biases that present potential safety issues in planning.

KEYWORDS: Surface observations; Bias; Automatic weather stations; Forecast verification/skill; Forecasting; Hindcasts

1. Introduction

Exposure to extreme heat leads to more deaths than any other weather event in the United States (CDC 2010; NWS 2020a). 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 (Kovats and Kristie 2006). While the heat index is commonly used in the United States for warning of dangerous conditions, other tools such as the NWS Heat Risk tool are used and are currently being further spread. The heat index (Rothfusz 1990) accounts for the effect of humidity and air temperature,

notably only in the shade. However, other heat stress indices, such as wet bulb globe temperature (WBGT), provide more comprehensive assessments of environmental heat stress (Budd 2008; Hondula et al. 2014).

Prior to 2018/19, WBGT was not a routinely forecast variable in the United States. 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 and CISA 2023). The United States National Weather Service (NWS) released an experimental WBGT forecast in 2019 that was made operational 1 June 2022 (NWS 2019a, 2022). The research presented here assesses the accuracy of the SERCC WBGT forecast and the forecasts made with the methods used by the NWS through comparisons against observed WBGT in North Carolina and estimated WBGT at weather stations in select parts of the CONUS, providing insight on future improvements.

Corresponding author: Jordan Clark, jordan@alumni.unc.edu

DOI: 10.1175/WAF-D-23-0076.1

© 2024 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

Brought to you by NOAA Central Library | Unauthenticated | Downloaded 08/13/24 08:13 PM UTC

a. Wet bulb globe temperature

The U.S. military developed WBGT in the 1950s to reduce the incidence of heat-related casualties at training camps during times of extreme heat (Budd 2008; Yaglou and Minard 1957). WBGT is calculated by adding together three components: 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 shortwave radiation and longwave radiation from the surface of 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. 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 the evaporation of water from this wick mimics the cooling effect of sweat evaporating off of human skin. From these three measures, WBGT is calculated according to the following equation:

$$\text{WBGT} = 0.7\text{NWB} + 0.2T_g + 0.1T_a, \quad (1)$$

where NWB is the natural web bulb temperature, T_g is the black globe temperature, and T_a is the dry bulb temperature (ambient air temperature).

WBGT continues to become more popular for assessing heat stress, including applications in athletics (Casa et al. 2015; Roberts et al. 2021), occupational settings (ACGIH 2017; ISO 1989, 2017; NIOSH 2016; OSHA 2017), and the military (US Department of the Army 2022). Numerous U.S. 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 North Carolina, South Carolina, and Georgia. Ultimately, WBGT is utilized based on a given value's corresponding level of danger, often referred to as flag level. The WBGTs used to define each threshold and associated activity modifications vary depending on setting and application (such as in occupational environments or athletics). An example of common values used for thresholds is 26.7°–29.4°C (green flag), 29.5°–31.0°C (yellow flag), 31.1°–32.1°C (red flag), and 32.2°C+ (black flag). These particular threshold values, which are the thresholds utilized throughout this paper, are utilized in high school athletics in North Carolina (NCHSAA 2016) and are identical to the thresholds used by the U.S. Army (US Department of the Army 2022), except for the green flag, which for the latter begins at 27.8°C instead of 26.7°C. Flag levels denote increasing levels of danger, with green flag requiring limited modifications to activity and black

flag, for example, requiring cancellation of high school football practices in various states.

b. Forecasts of WBGT

Accessible forecasts of WBGT enable advance knowledge of potentially hazardous periods, which has applications in numerous sectors and settings. For example, in athletics and occupational environments, certain workouts or occupational activities may be necessary but also lead to increased heat strain. With the associated activity modification guidelines (such as work/rest ratios) and recommendations on adjustments to clothing or equipment based on WBGTs, forecasts of the index allow for planning of activities to occur on certain days or at certain times of day.

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 comprehensive than the heat index (Budd 2008; Hondula et al. 2014). However, there has been limited research on the accuracy of WBGT forecasts. Although WBGT forecasts have been in use in some parts of the world for many years, such as in Japan (Hoshi and Inaba 2007), these forecasts have not been common in the United States.

In 2018, NOAA's Southeast Regional Climate Center (SERCC) and Carolinas Integrated Sciences and Assessments (CISA) created a WBGT forecast tool, which was made operational in the summer of 2019 (SERCC and CISA 2023). As of 2022, this forecast tool was operational for the eastern two-thirds of the contiguous United States. Similarly, in 2019, the NWS began including WBGT forecasts in an experimental version of their gridded forecast data, the National Digital Forecast Database (NDFD) (NWS 2019a). This forecast recently became operational in 2022 (NWS 2022). However, the SERCC/CISA and NWS WBGT forecasts utilize different methods. One aspect of the research here is to assess the accuracy of these different forecast methodologies.

Complicating the forecasting of WBGT, 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 hundreds of meters (Verkaik et al. 2005). This variability is not well captured by weather models due to their coarser resolution, since the highest-resolution output of widely available and spatially expansive models is 2.5 km. In addition to assessing the accuracy of these WBGT forecasts, this research also assesses the utility of supplementing subforecast 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 Pasquill-Gifford (PG) stability classes for the downscaling of wind from 10 to 2 m (Bowen et al. 1983; Frank et al. 2020; U.S. EPA 2000).

The importance of accurately incorporating the wind at 2 m for WBGT is paramount, given that high (hot) values of WBGT are sensitive to slight changes in environmental conditions, particularly changes in wind speed. 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. For example, under full sun and with a dewpoint of 21.1°C and an air temperature of 30°C, a slight change in wind speed from 1.3 to 0.6 m s⁻¹ can result in a change in WBGT of two flag categories (from yellow to black flag).

The SERCC/CISA forecast tool utilizes both the NWS National Digital Forecast Database (NDFD) and the NWS National Blend of Models (NBM) as inputs to a routine that computes the WBGT forecast (SERCC and CISA 2023). Both products forecast at hourly time steps out to 36 h in the future, but the NDFD hourly forecast changes to a 6-hourly forecast at 72 h past initialization while the NBM forecasts 3-hourly values out to 192 h. Thus, the NDFD is used for 1–69 h and the NBM for 72–120 h for the SERCC/CISA forecast tool. References to the SERCC/CISA WBGT forecast will be hereafter referred to as S-WBGT. The S-WBGT methodology supplemented with surface roughness for downscaling wind speeds will hereafter be referred to as S-WBGT Z.

The forecast analysis here serves as an update to an assessment of forecast accuracy presented in Clark and 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+*) was 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 S-WBGT Z forecast. Two sources of satellite imagery were used to assess if the higher-resolution imagery resulted in more accurate surface roughness values and thus a more accurate WBGT forecast. In addition to looking at forecast bias overall between methods, accuracy of WBGT forecasts using the NBM and NDFD are compared and variations in accuracy across space, weather condition, and time are assessed.

2. Data

a. Weather forecast data

Hourly gridded forecast data for two NWS forecast products (four runs per day: 0000, 0600, 1200, and 1800 UTC) were archived for the summers of 2019–21: National Digital Forecast Database (NDFD) and the National Blend of Models (NBM). The NDFD is a product containing a mosaic of digital weather forecasts from the NWS and National Centers for Environmental Prediction (NCEP) (NWS 2019b). The NBM is an ensemble of guidance, a statistically derived blend from numerous numerical weather prediction models (Craven et al. 2020). Two versions of the NBM were utilized here, version 3.2 (Craven et al. 2020) for the summer of 2019 and version 4.0 (which constitutes 85% of the NBM forecast data) for the summers of 2020 and 2021 (NWS 2020b). While the majority of the analysis focuses on the 24-h forecast, this incorporates a window surrounding this forecast lead time,



FIG. 1. Field work instruments (WBGT meters and weather station). 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 collocated with the WBGT meter.

ranging from 21 to 27 h to ensure forecasts are for relevant times of day and for larger sample sizes of forecast data points.

b. WBGT data

A weather station was collocated with a WBGT meter to provide data quality checks. The weather station was a Davis Instruments Vantage Pro 2 Plus (Fig. 1) that recorded all variables needed to estimate WBGT at 10-s intervals, detailed below. 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) (Fig. 1). The ISO guidelines (Parsons 2006) require that the black globe have a diameter of 0.15 m and the temperature sensor should measure 20°–120°C with accuracies of $\pm 0.5^\circ\text{C}$ for 20°–50°C and $\pm 1^\circ\text{C}$ for temperatures greater than 50°C. The natural wet bulb temperature sensor should be a cylindrical shape, have a diameter of 6 ± 1 mm and length of 30 ± 5 mm, and measure the range 5°–40°C with an accuracy of $\pm 0.5^\circ\text{C}$. The wick situated over the temperature sensor must be white and made of cotton or other water-absorbent material. Last, the dry bulb temperature sensor should be positioned within a radiation shield and measure the range 10°–60°C with a $\pm 1^\circ\text{C}$ accuracy (Parsons 2006).

The instruments were situated such that they were 1.5 m above the ground, on average, with the weather station anemometer situated highest at 2 m. Data were recorded at various intervals (2-, 5-, 10-, 30-, 60-, or 120-s 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 2019–21: the Horace Williams Airport in Chapel Hill, North Carolina, and within suburban environments in Chapel Hill, Durham, and Shelby, North Carolina.

Two methods were used to identify when instruments were in the shadows of buildings or trees during data collection. First, we periodically captured images of the instruments to directly observe any shading. Second, we compared the measured solar radiation from the weather station with estimated clear-sky solar radiation for 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 with no obstacles between the sky and sensors (i.e., no trees). The first 15 min 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). WBGT varies rapidly over small time periods due to slight changes in insolation (e.g., from clouds) and wind speed, which fluctuate rapidly. Thus, to robustly compare with the (instantaneous) hourly time step forecast data, the Observed WBGT was averaged over the course of 20 min (10 min before and after the top of every hour).

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) (Fig. 2). 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). Iowa Environmental Mesonet (IEM) archive (Iowa Environmental Mesonet 2021) 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 2021) was used to download the ECONet weather station data. ECONet weather stations directly measured solar radiation, which is required to estimate WBGT. 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, detailed below. Hourly data for all stations for the heat season (1 May–30 September) 2019–21 were used.

c. 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). Imagery included scenes covering the entire state of North Carolina and parts of South Carolina, Georgia, Alabama, Tennessee, Virginia, Pennsylvania, Illinois, Oklahoma, and Texas. All imagery utilized was captured on three dates:

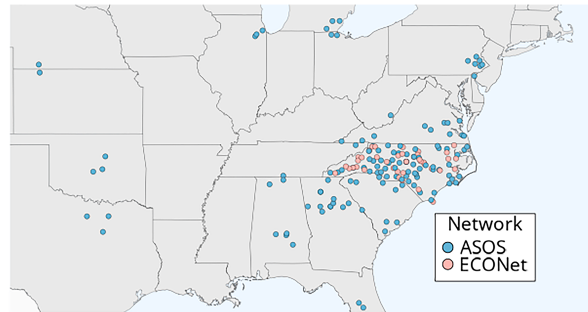


FIG. 2. Map of the weather stations used (by network) for station WBGT.

29 August, 30 August, and 7 September 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 17 August 2019. The 2019 National Land Cover Database (NLCD) was retrieved for the CONUS to provide the land cover information utilized here (Dewitz 2021).

3. 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 and Kjellstrom 2012; Patel et al. 2013). S-WBGT forecasts likewise applied the Liljegren et al. methodology for the NDFD/NBM. The necessary formulas for using the Liljegren et al. (2008) method were provided by the R package “wbgt” (Lieblich and Spector 2017).

To estimate solar radiation for ASOS/AWOS weather stations, the reported cloud cover categories were converted to a percentage cloud cover. These stations report cloud cover at several levels, using the following categories: clear (0%–5%], few (5%–25%], scattered (25%–50%], broken (50%–87%], and overcast (87%–100%] (brackets denote inclusivity) (NOAA et al. 1998). Comparing the observed solar radiation at ECONet weather stations with estimated solar radiation revealed that using the maximum value within each of these ranges produced the most accurate results, that is, 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.

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:

$$\text{Srad} = R_0(1 - 0.75n^{3.4}), \quad (2)$$

where n is the cloud cover fraction (0.0–1.0) and R_0 is the clear-sky direct radiation (W m^{-2}) estimated using (3):

$$R_0 = 990 \sin(\varnothing - 30), \quad (3)$$

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

TABLE 1. The solar radiation delta T (SRDT) method was used to determine Pasquill–Gifford stability classes and the corresponding power-law exponent (drawn from U.S. EPA 2000).

Stability class	Urban exponent	Rural exponent	Stability class	Urban exponent	Rural exponent
A	0.15	0.07	D	0.25	0.15
B	0.15	0.07	E	0.30	0.35
C	0.20	0.10	F	0.30	0.55

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

$$U_Z = U_r \left(\frac{Z}{Z_r} \right)^p, \quad (4)$$

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 (Bowen et al. 1983; Frank et al. 2020; U.S. EPA 2000). The “urban” exponents were used here and are provided in Table 1, since this is implemented in the SERCC/CISA WBGT tool after preliminary field work revealed that the “rural” exponents resulted in 2-m wind speeds that were too fast.

The solar radiation delta T (SRDT) method was used to determine Pasquill–Gifford stability classes and the corresponding power-law exponent (Frank et al. 2020; U.S. EPA 2000). The SRDT method serves as an indicator of atmospheric stability by using observed solar radiation during the day and vertical temperature difference at night (Frank et al. 2020; U.S. EPA 2000). For observations (or forecast data points) with wind speeds of less than 1 m s^{-1} , the wind speed values were increased to 1 m s^{-1} . This decision was based on the sensitivity of the anemometers installed at the weather stations that were used (NOAA et al. 1998).

a. 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 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 and Zhang 2009) since this parameterization (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 with (5):

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}}, \quad (5)$$

where NIR is reflectance in the near-infrared range and RED is reflectance in the red range. Following Zheng et al. (2014) and Markert et al. (2019), the GVF was calculated using (6):

$$\text{GVF} = \frac{\text{NDVI} - \text{NDVI}_{\min}}{\text{NDVI}_{\max} - \text{NDVI}_{\min}}, \quad (6)$$

where NDVI is the NDVI of a given pixel, NDVI_{\max} is the maximum NDVI for a land cover class, and NDVI_{\min} is the NDVI of bare soil (0.01). The land cover classification data were drawn from the 2019 National Land Cover Database (NLCD) (Dewitz 2021). Since the NLCD has a resolution of 30 m, the Sentinel NDVI was rescaled to match this resolution before calculating NDVI_{\max} per land cover class and the GVF.

Finally, surface roughness was calculated with (7) (Markert et al. 2019; Zheng et al. 2014):

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

where GVF is green vegetation fraction, $Z_{0m,\min}$ is the minimum surface roughness in meters, and $Z_{0m,\max}$ is the maximum surface roughness in meters, with the values for the latter two provided in Table 2 (Markert et al. 2019).

To incorporate the influences of surface roughness at varying spatial scales for a given location, a weighted average surface roughness was calculated. The different configurations for weighting each scale for calculating this average are provided in Table A1 in appendix A. This average used the roughness values from the following scales: 30 m (10%), 100 m (25%), 250 m (50%), and 500 m (15%). After deriving the weighted average surface roughness, the wind speeds to be used for S-WBGT Z were downscaled utilizing the following function from van Den Berg (2004):

$$u(z) = u(z_{\text{ref}}) \frac{\ln(Z/Z_0)}{\ln(Z_{\text{ref}}/Z_0)}, \quad (8)$$

TABLE 2. Surface roughness length ranges for the eight land cover classes. Drawn from Markert et al. (2019).

Landcover class	$Z_{0m,\min}$	$Z_{0m,\max}$
Evergreen forest	0.5	0.5
Coniferous forest	0.5	0.5
Mixed forest	0.2	0.5
Water	0.0001	0.0001
Bare	0.01	0.01
Cropland	0.05	0.15
Urban and built up	0.5	0.5
Impervious	0.03	0.03

where Z_0 is surface roughness length in meters (weighted average surface roughness), Z_{ref} is the height of original wind speed (10 m), Z is the height of the downscaled wind speed (2 m), and u is wind speed (van Den Berg 2004). This equation differs from (4) since surface roughness is being utilized in this case. This was calculated for both the forecast data and weather station data. After downscaling the wind speeds, the S-WBGT Z was calculated, with the methods for converting cloud cover to solar radiation identical to the ones described above for S-WBGT.

b. NWS WBGT

As of 2022, WBGT is a forecast parameter in the NDFD (NWS 2022). However, it is not a parameter in the current version of the NBM. Additionally, since the methods used by the NWS in their forecast WBGT product have changed over the duration of this study, the NDFD forecast WBGT values were not used directly for this assessment. Instead, the forecast WBGT was recalculated using the relevant parameters from the NBM/NDFD based on the latest methods used by the NWS. For the forecast WBGT calculated with methods used by the NWS (“NWS WBGT”), all data processing of solar radiation values and wind speed (downscaled without surface roughness) were identical to those described above for S-WBGT. Since the NDFD does not have the shortwave solar radiation parameter found in NBM v4.0, WBGT forecasts calculated from the NBM did not use this field, but instead derived solar radiation from the forecast percentage cloud cover. The NWS WBGT, however, utilizes different methods for estimating the black globe temperature and the natural wet bulb temperature (Boyer 2022). Black globe temperature estimation utilizes a slightly modified version of the methodology from Dimiceli et al. (2011). 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 2022, personal communication).

In 2022, the NWS methodology for estimating the natural wet bulb temperature in their operational, gridded NDFD product was changed (Boyer 2022; T. Boyer 2022, personal communication). However, the methodology employed by the NWS before this change is used here, since it was found to perform notably better in comparison with the field observations collected in this study, which was a modified version of the method developed in Hunter and Minyard (1999):

$$T_n = T_w + 0.00117S - 0.233u + 1.072, \quad (9)$$

where T_n is the natural wet bulb temperature ($^{\circ}\text{C}$), T_w is the psychrometric wet bulb temperature ($^{\circ}\text{C}$), S is solar irradiance (W m^{-2}), and u is 2-m wind speed (m s^{-1}) (Boyer 2022).

c. Analysis

Observed WBGT data were collected throughout the summers of 2019–21 from May to September at the locations detailed above. 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 hour of day and weather condition. It is worth noting that the sample sizes for the observed and station WBGT differ, given that the station WBGT consists of numerous weather stations while observed WBGT draws from field work with a limited number of instruments (Tables C1 and C2 in appendix C). Comparisons were made between the NBM and NDFD WBGT forecasts and at varying forecast lead times for these forecast products. To determine the utility of incorporating surface roughness into the forecast, comparisons were made between the forecast accuracy of the S-WBGT and S-WBGT Z . Only data points between 0600 and 2000 local time (LT) 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 WBGT, 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 Research Applications Laboratory 2015), the following statistics were calculated and compared for each flag level:

1) WBGT FLAG ACCURACY METRICS (BY FLAG LEVEL)

- 1) Percent correct (%): Total of correct forecasts and correct rejections (no event forecast, no event observed) divided by total number of forecasts (Jolliffe and Stephenson 2012).
- 2) Hit rate (%): percentage of correct forecasts (e.g., black flag forecast, black flag observed, hit) (Jolliffe and Stephenson 2012).
- 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 (Jolliffe and Stephenson 2012; NCAR Research Applications Laboratory 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 (Jolliffe and Stephenson 2012).

2) 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 farthest from black flag (Jolliffe and Stephenson 2012; NCAR Research Applications Laboratory 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

TABLE 3. Observed WBGT forecast mean absolute error. S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds. NWS WBGT refers to the WBGT forecast utilizing the NWS methods, applied to forecast data from the NBM and NDFD, respectively, in this table.

Metric	Method	$\geq 29.5^{\circ}\text{C}$	$\geq 32.2^{\circ}\text{C}$
NBM	S-WBGT	0.87	0.73
	S-WBGT Z	1.08	0.65
	NWS WBGT	1.02	1.20
NDFD	S-WBGT	1.19	1.34
	S-WBGT Z	1.02	0.85
	NWS WBGT	1.48	1.97

events and less credit is awarded to correct forecasts of common events (Jolliffe and Stephenson 2012).

- 2) Heidke skill score (HSS): verification measure for categorical forecasts that includes correct random forecasts. Values range from negative infinity to 1. Scores greater than zero mean the forecast does better than random chance (Jolliffe and Stephenson 2012).

4. Results

a. Forecast bias comparisons

The research presented here had two objectives: 1) assessing the accuracy of WBGT forecasts and 2) determining if forecasts can be improved by using surface roughness to downscale wind speeds. Unless otherwise stated, forecast bias refers to the “relative bias” between forecasts and observations, with positive and negative relative bias indicating over and under forecasting, respectively.

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 in the NDFD were more negatively biased than NBM biases (and with higher MAEs) (Table 3), particularly when WBGT was greater than or equal to 32.2°C (Fig. 3).

For both observed and station WBGT, forecast MAE of the three forecast methods were within 0.9° and 0.4°C of one another, respectively, when WBGT was less than 31.1°C . However, at higher values of WBGT ($32.2^{\circ}\text{C}+$), the NWS WBGT median bias and MAE (Fig. 4) increased markedly, as well as the RMSE (Table 4), particularly for station WBGT. Median forecast bias for all methods relative to observed WBGT were positive (warm) when WBGT was less than or equal to 31°C . However, NWS WBGT and S-WBGT had a negative median bias when observed WBGT exceeded 31° and 32.1°C , respectively (Fig. 4). The S-WBGT Z forecast bias remained slightly positive at the highest WBGTs (Fig. 4).

There were two notable differences with station compared to observed WBGT forecast accuracy (Fig. 4). First, the relative bias and MAE for both S-WBGT methods were lower when station WBGT was less than 31.1°C , and S-WBGT Z had a slightly negative bias when station WBGT was equal to or greater than 32.2°C (-0.1°C). Second, NWS WBGT bias was more negative when station WBGT was greater than or equal to 31.1°C compared to observed WBGT, and the median bias for both observed and station WBGT was negative (too cold) across all WBGTs, with median biases of -0.4° and -1.6°C when station WBGT was 26.7° – 29.4°C and 32.2°C or above, respectively (Fig. 4).

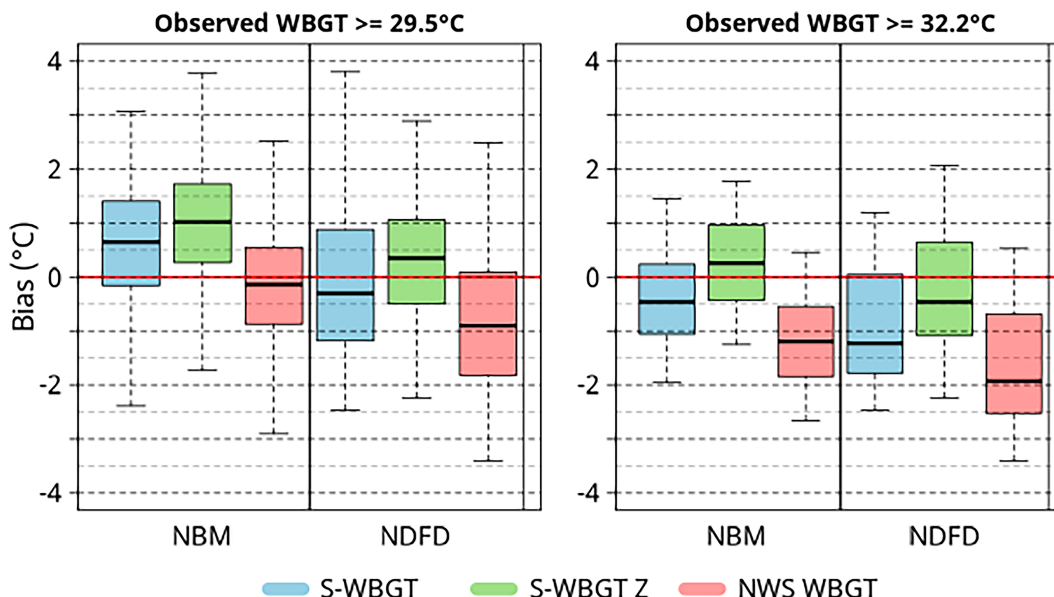


FIG. 3. Observed WBGT forecast bias (24-h forecast): NBM and NDFD. Boxplot whiskers extend up to 1.5 times the interquartile range (e.g., top whisker is $1.5 \times \text{IQR} + \text{third quartile value}$). S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds. NWS WBGT refers to the WBGT forecast utilizing the NWS methods.

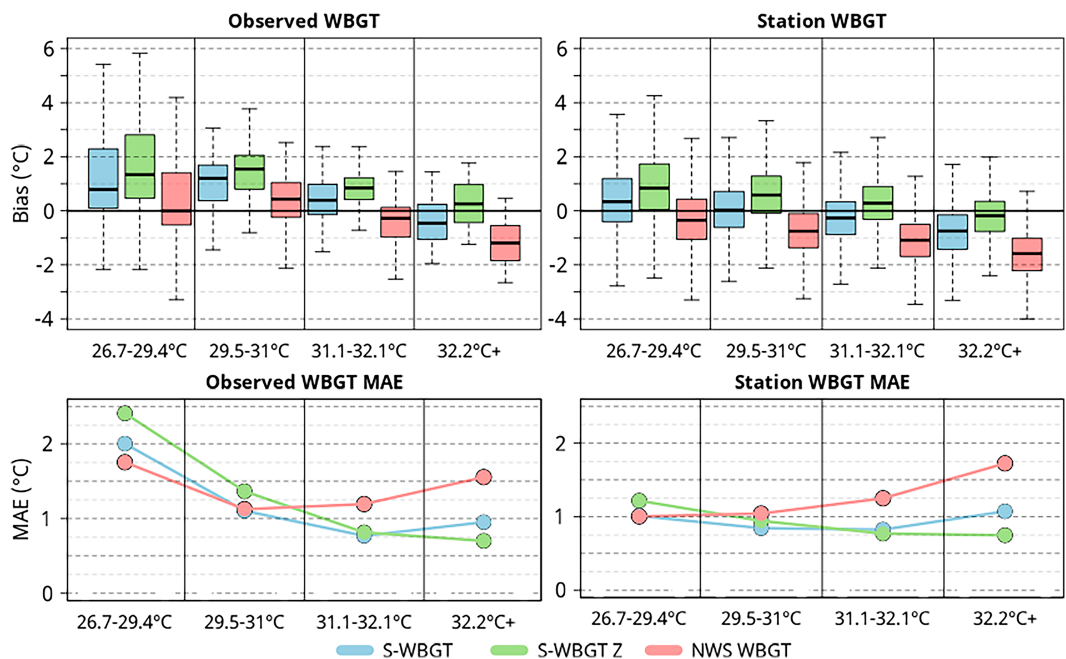


FIG. 4. WBGT relative forecast (top) bias and (bottom) MAE (24-h forecast). Boxplot whiskers extend up to 1.5 times the interquartile range (e.g., top whisker is $1.5 \times \text{IQR} + \text{third quartile value}$). S-WBGT Z is SERCC/CISA forecast with wind speed downscaled to 2 m using PG stability classes (surface roughness). S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds. NWS WBGT refers to the WBGT forecast utilizing the NWS methods.

b. Variations in bias

Assessing the forecast relative bias spatially (average daily maximum station WBGT) demonstrates the high variability across the region and between forecast methods (Fig. 3). Figures displaying WBGT forecast bias for a broader spatial domain are included in Fig. B1 in appendix B.

A majority of stations displayed a negative bias (from -0.8° to -0.1°C) for S-WBGT, with the most negative bias occurring at three stations located along the Appalachian Mountains (Fig. 5). However, S-WBGT Z bias was greater than S-WBGT (0° – 1°C), and positive at all but ten stations,

five of which were along the Appalachian Mountains (Fig. 5). The differences in bias between the two S-WBGT methods ranged from 0.2° to 1.0°C ($\text{WBGT} > 28.9^{\circ}\text{C}$), with the largest bias difference occurring at stations with lower roughness (detailed below). Last, NWS WBGT forecast bias was notably more negative, with the coolest bias along the Appalachian Mountains, southern and western piedmont of North Carolina, and in coastal areas of North Carolina (Fig. 5).

Forecast bias varied across the daylight hours for observed and station WBGT. When observed WBGT was equal to or greater than 29.5°C , both S-WBGT forecasts were positive at

TABLE 4. Mean and standard deviations (SD) of the observed/station WBGT and the three forecast WBGT values (when $\text{WBGT} \geq 26.7^{\circ}\text{C}$). RMSE and variance ratio of observed/station WBGT when $\text{WBGT} \geq 26.7^{\circ}\text{C}$ and $\geq 29.4^{\circ}\text{C}$. Variance ratio is forecast:observed/station. S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds. NWS WBGT refers to the WBGT forecast utilizing the NWS methods.

		Mean ($^{\circ}\text{C}$)	SD ($^{\circ}\text{C}$)	Forecast RMSE ($^{\circ}\text{C}$)		Forecast variance ratio ($^{\circ}\text{C}$)	
		$\geq 26.7^{\circ}\text{C}$		$\geq 26.7^{\circ}\text{C}$	$\geq 29.4^{\circ}\text{C}$	$\geq 26.7^{\circ}\text{C}$	$\geq 29.4^{\circ}\text{C}$
Observed WBGT	Observed WBGT	30.3	1.8				
	S-WBGT	30.9	1.5	1.7	1.3	0.67	1.03
	S-WBGT Z	31.5	1.5	1.9	1.3	0.66	1.05
	NWS WBGT	31.2	1.5	1.7	1.5	0.66	1.01
Station WBGT	Station WBGT	29.1	1.6				
	S-WBGT	29.2	1.7	1.2	1.1	1.17	1.55
	S-WBGT Z	29.7	1.8	1.4	1.1	1.22	1.55
	NWS WBGT	28.6	1.6	1.3	1.4	1.08	1.42

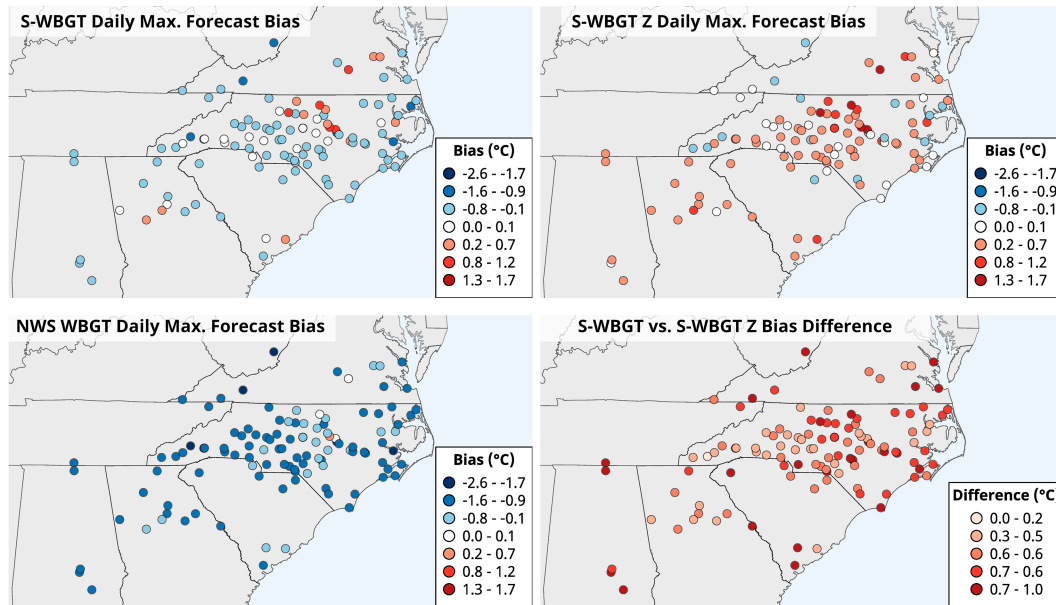


FIG. 5. Station WBGT daily maximum bias (24-h lead time): Regional variations across the Carolinas. S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds. NWS WBGT refers to the WBGT forecast utilizing the NWS methods.

all hours (except for 1100 and 1800 LT), and the NWS WBGT was negatively biased throughout the day, except for 1600 LT (Fig. 6). Compared to 29.5°C, all biases were larger for every hour of day when WBGT was greater than 32.2°C (not pictured). 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 (Fig. 6). The highest median bias for S-WBGT and S-WBGT Z occurred at 1300 and 1400 LT.

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 (Fig. 6).

Furthermore, S-WBGT forecast bias was markedly greater when observed WBGT was less than 31.1°C, particularly when dewpoint temperatures were high, wind speed was relatively high and variable (wide range of values) (0.6–2.2 m s⁻¹), and solar radiation was both low and variable (Fig. 7). When

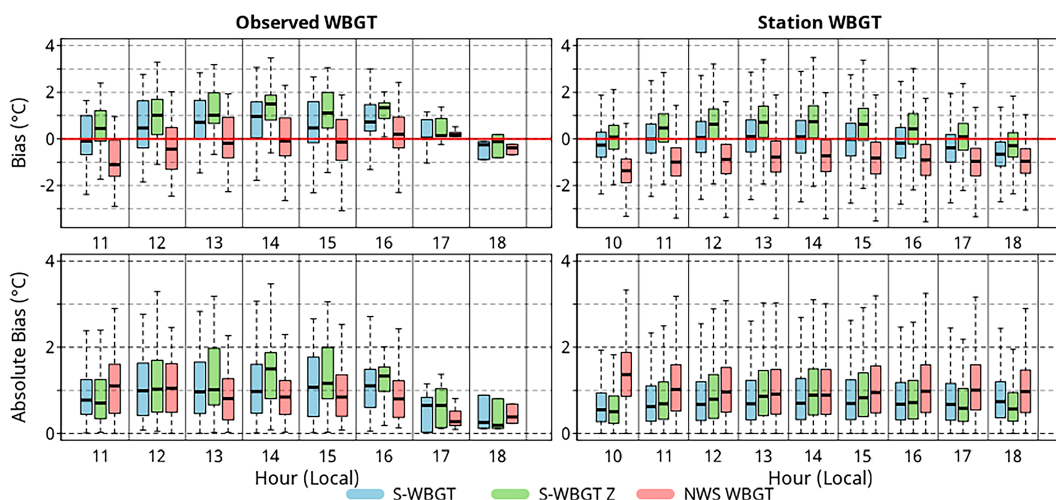


FIG. 6. WBGT forecast bias (24-h lead time) by hour of day (NBM) (WBGT $\geq 29.5^{\circ}\text{C}$). Boxplot whiskers extend up to 1.5 times the interquartile range (e.g., top whisker is $1.5 \times \text{IQR} + \text{third quartile value}$). S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds. NWS WBGT refers to the WBGT forecast utilizing the NWS methods.

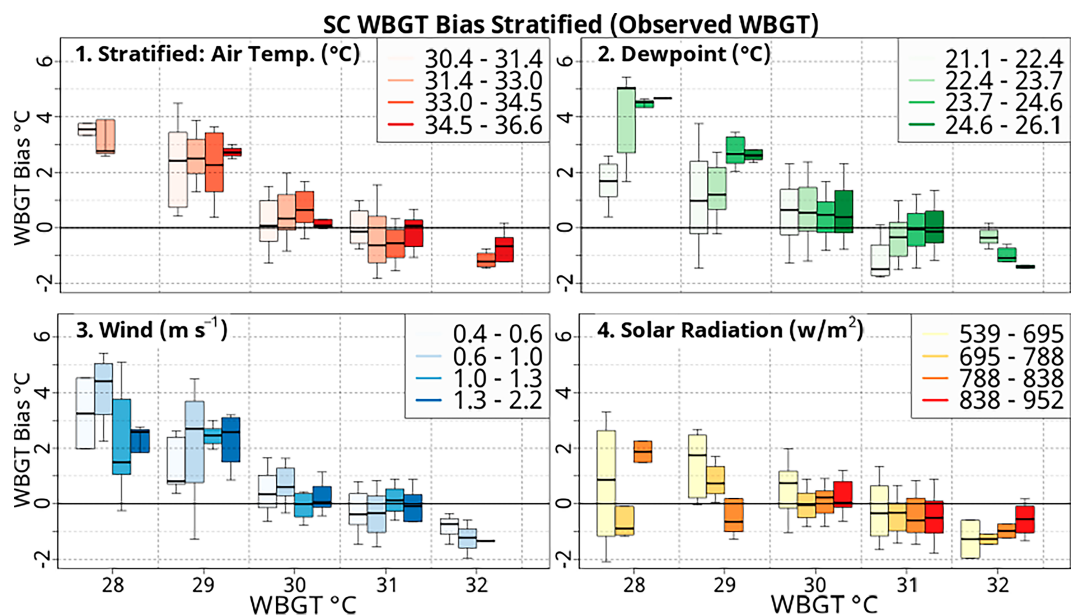


FIG. 7. S-WBGT bias (24-h lead time) stratified by meteorological variables (observed WBGT). Boxplot whiskers extend up to 1.5 times the interquartile range (e.g., top whisker is $1.5 \times \text{IQR} + \text{third quartile value}$). S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds.

observed WBGT was greater than or equal to 30°C, there was less variation in forecast bias across the different strata. Similar patterns were seen when stratifying station WBGT forecast bias (not pictured), but with less variation in bias magnitudes.

c. Forecast WBGT flag accuracy

The forecast WBGT flag accuracy results (corresponding to the flag thresholds discussed above) detailed here refer to the NBM 24-h forecast. Confusion matrices for all results are provided in [Tables C1](#) and [C2](#). The most notable difference between methods when assessing percentage of correct forecasts for each flag was with observed WBGT ([Table 5](#)). Both S-WBGT methods had the highest percent correct for yellow flag. For black flag

observed, S-WBGT Z had a roughly 10% lower percent correct than the SC and NWS WBGT ([Tables 5](#) and [6](#)). The bias scores were similar and indicated under forecasting (bias score < 1) for green and yellow flags (observed) for both S-WBGT methods ([Table 5](#)). For red and black observed flags, both S-WBGT methods over forecast (bias score > 1) and the S-WBGT Z bias notably increased from red (1.07) to black flag (2.11) ([Table 5](#)). Station WBGT flag accuracy was similar; however, S-WBGT Z bias was 1.69 versus 1.07 for red flag ([4](#)) and the NWS WBGT hit rate for black flag dropped from 42% to 18% ([Table 6](#)).

For observed and station WBGT flags, the false alarm ratio, which is the total number of false alarms for a given flag level divided by the total number of forecasts for that flag level

TABLE 5. Forecast WBGT flag accuracy compared to observed WBGT (NBM 24-h forecast). S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the NWS methods.

Metric	Method	Green flag (26.7°–29.4°C)	Yellow flag (29.5°–31.0°C)	Red flag (31.1°–32.1°C)	Black flag (32.2°C+)
Percent correct	S-WBGT	81	70	66	76
	S-WBGT Z	81	72	65	66
	NWS WBGT	78	58	62	77
Hit rate (%)	S-WBGT	26	30	42	84
	S-WBGT Z	25	21	23	94
	NWS WBGT	28	39	39	42
Bias score	S-WBGT	0.33	0.69	1.42	1.53
	S-WBGT Z	0.29	0.45	1.07	2.11
	NWS WBGT	0.57	1.28	1.49	0.67
False alarm ratio	S-WBGT	0.21	0.57	0.70	0.45
	S-WBGT Z	0.14	0.54	0.78	0.55
	NWS WBGT	0.51	0.69	0.74	0.37

TABLE 6. Forecast WBGT flag accuracy compared to station WBGT (NBM 24-h forecast). S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the NWS methods.

Metric	Method	Green flag (26.7°–29.4°C)	Yellow flag (29.5°–31.0°C)	Red flag (31.1°–32.1°C)	Black flag (32.2°C+)
Percent correct	S-WBGT	78	70	84	93
	S-WBGT Z	74	66	81	90
	NWS WBGT	77	68	86	94
Hit rate (%)	S-WBGT	72	55	43	52
	S-WBGT Z	59	49	42	75
	NWS WBGT	86	46	24	18
Bias score	S-WBGT	0.83	1.16	1.32	1.21
	S-WBGT Z	0.61	1.20	1.69	2.24
	NWS WBGT	1.16	0.97	0.71	0.31
False alarm ratio	S-WBGT	0.14	0.52	0.67	0.57
	S-WBGT Z	0.08	0.59	0.75	0.67
	NWS WBGT	0.26	0.53	0.66	0.42

(Jolliffe and Stephenson 2012), was highest for red flags by similar magnitudes across all forecast methods (Tables 5 and 6). The hit rate was notably higher for the S-WBGT methods relative to the NWS WBGT for 1) observed black flag (Table 5) and 2) red and black flags for station WBGT (Table 6). In both cases, the S-WBGT Z had the highest hit rates for black flag (94% and 75% for observed and station black flags, respectively).

Last, there were notable differences in bias between the 24- and 48-h NBM forecast for black flag, with a lower bias for all methods for the 48-h forecast (both observed and station WBGT) (Table 7). However, the hit rate was higher for the 24-h forecast, 92% versus 84% for S-WBGT, 94% versus 77% for S-WBGT Z, and 42% versus 8% for NWS WBGT (Table 7).

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 (Table 8), with the S-WBGT having the

highest scores, followed by S-WBGT Z and the NWS WBGT. At increasing forecast lead times, the GSS for S-WBGT Z remained relatively consistent (Table 9). Contrasting observed WBGT, at increasing forecast lead times, the station WBGT GSS for both S-WBGT methods decreased, with a GSS of 0.31 for S-WBGT, 0.41 for S-WBGT Z, and 0.18 for NWS WBGT for a 72-h forecast (Table 9).

d. Surface roughness

In addition to assessing the accuracy of WBGT forecasts, this research aimed to improve the S-WBGT forecast by incorporating surface roughness into the downscaling of wind speeds from 10 to 2 m. 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-m wind speeds compared to the winds downscaled using the SRDT method and Pasquill–Gifford stability classes (used in S-WBGT). Given the high sensitivity of WBGT to wind speed (e.g., the rapid increase in WBGT under low wind speeds), the improved (minimized) station WBGT bias for S-WBGT Z when WBGT was equal to or greater than 32.2°C can be directly related to an improved 2-m wind speed forecast with surface roughness being used (Fig. 4). Furthermore, the S-WBGT Z forecast bias had lower

TABLE 7. WBGT black flag accuracy (NBM 24- and 48-h forecast). S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the U.S. NWS methods.

Metric	Method	Observed WBGT		Station WBGT	
		24	48	24	48
Bias score	S-WBGT	1.53	0.92	1.21	0.51
	S-WBGT Z	2.11	1.77	2.24	1.43
	NWS WBGT	0.67	0.15	0.31	0.04
Hit rate	S-WBGT	84%	62%	52%	16%
	S-WBGT Z	94%	77%	75%	40%
	NWS WBGT	42%	8%	18%	2%
False alarm ratio	S-WBGT	0.45	0.33	0.04	0.01
	S-WBGT Z	0.55	0.57	0.09	0.04
	NWS WBGT	0.37	0.50	0.01	0.00
Percent correct	S-WBGT	76%	91%	93%	96%
	S-WBGT Z	66%	84%	90%	94%
	NWS WBGT	77%	86%	94%	96%

TABLE 8. Verification scores for WBGT flag forecast by method (NBM 24-h forecast). S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the NWS methods.

	Method	Heidke skill score	Gerrity skill score	Percent correct
Observed WBGT	S-WBGT	0.28	0.40	47
	S-WBGT Z	0.22	0.34	42
	NWS WBGT	0.16	0.32	38
Station WBGT	S-WBGT	0.41	0.55	63
	S-WBGT Z	0.34	0.62	55
	NWS WBGT	0.36	0.36	62

TABLE 9. Gerrity skill scores for NBM WBGT flag forecasts. S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds. NWS WBGT is the WBGT forecast utilizing the NWS methods. Scores for forecast lead times of 24, 48, and 72 h.

	Method	24 h	48 h	72 h
Observed WBGT	S-WBGT	0.40	0.36	0.31
	S-WBGT Z	0.34	0.33	0.34
	NWS WBGT	0.32	0.17	0.11
Station WBGT	S-WBGT	0.55	0.36	0.31
	S-WBGT Z	0.62	0.47	0.41
	NWS WBGT	0.36	0.21	0.18

magnitudes of positive bias increases at and immediately after solar noon compared to the S-WBGT and NWS WBGT (Fig. 6).

The correlation between surface roughness length and differences in the station WBGT between the S-WBGT and S-WBGT Z were statistically significant ($p < 0.01$) with a Pearson correlation coefficient of 0.63 when WBGT was greater than or equal to 31.1°C. As surface roughness increased, differences between the S-WBGT Z bias compared to the bias of other methods also increased (Fig. 8). Thus, S-WBGT Z was a particularly better forecast for sites with rougher surfaces.

Last, weighted average surface roughness with different weightings for the roughness at varying spatial scales (30, 100, 250, and 500 m) (Table A1) was not sensitive to the different configurations of weights for calculating this average (Fig. A1). Variations in the station WBGT calculated using these different surface roughness values and corresponding wind speeds ranged from -0.1° to 0.2°C , but the majority of differences (between the 25th and 75th percentiles) ranged from -0.05° to $+0.05^{\circ}\text{C}$ (Fig. A1). The analysis here thus utilized the surface roughness weighting schema of 30 m (10%), 100 m (25%), 250 m (50%), and 500 m (15%).

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.

5. Discussion and conclusions

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 comprehensive 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, with the NWS forecast becoming

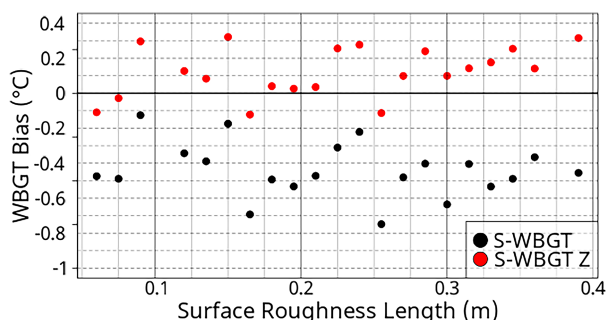


FIG. 8. S-WBGT vs S-WBGT Z average station WBGT forecast bias relative to surface roughness. S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds.

operational in 2022. 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 ASOS and mesonet stations. This research also evaluated efforts to improve the SERCC WBGT forecast by more accurately estimating the wind speed at 2 m based on wind speed at 10 m above the ground.

a. 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 driven primarily by differences in the forecast wind speed between the two forecast products, with the NDFD wind speeds being faster than the NBM (Fig. D1 in appendix D). The NBM is used heavily to inform the NDFD gridded product (Craven et al. 2018). 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. However, the current version of the NBM incorporated adjustments to the wind speed forecast, so it may be that this adjustment results in faster winds, and more negatively (cool) biased WBGT relative to what is seen in this study. Additionally, it is important to reiterate that the method used by the NWS for estimating the natural wet bulb component has changed from what was used here (Boyer 2022). However, this study utilized the modified Hunter and Minyard (1999) method since it was found to be more accurate.

The S-WBGT [which used the Liljegren et al. (2008) methodology for calculating WBGT] was found to be more accurate than the method used by the NWS in most instances, particularly when conditions were dangerous ($\text{WBGT} > 31.1^{\circ}\text{C}$). When WBGT was relatively cooler (i.e., less than 29.4°C), the S-WBGT was more positively (warm) biased than the NWS WBGT. This warm bias could lead to premature or unnecessary cancellations of outdoor activities such as sports practices,

which might pose logistical and economic challenges. However, at higher levels of heat stress and WBGTs, the NWS method underestimated the WBGT while the S-WBGT provided more accurate results, with a lower relative bias, lower MAE, and higher hit rate. A cold bias at extreme WBGTs ($>31.1^{\circ}\text{C}$) could result in insufficient planning and resource allocation for hazardous periods, which is critical since it is as these WBGTs when health is more likely to be impacted by heat exposure.

These findings of higher accuracy with the S-WBGT further support existing research regarding the accuracy of the Liljegren et al. (2008) methodology (Lemke and Kjellstrom 2012; Patel et al. 2013). For red and black flags, S-WBGT was within 0.8°C of the 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 (0.8°C) could still lead to WBGT flag misclassification in some instances, it is sufficiently accurate when compared to the larger error ranges of common WBGT meters and established standards, e.g., from the Japanese National Institute of Occupational Safety and Health (Racinais et al. 2022). Furthermore, in this case, the S-WBGT bias erred on the side of caution (being too warm) under these most thermally stressful conditions, which is preferred here in relation to WBGT being used to safeguard health.

b. Variations in forecast bias

WBGT forecast bias varied across space, weather conditions, and 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 this region 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. Additional downstream implications of this geography arise from adiabatic warming driven by westerly winds in the lee of the mountains.

For the NWS WBGT, the central North Carolina region had the most accurate forecasts on average. 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 North Carolina 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 hypothesized to be the reason behind this pattern for the sparsely forested Midwest and Southern Plains, which have climatologically faster wind speeds.

Distinct variations in forecast bias were seen when stratifying by dewpoint 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 sensitivity of WBGT to wind speed was revealed by the overall range of S-WBGT bias being greater when wind speeds

were low ($<0.6\text{--}1.0\text{ m s}^{-1}$). The high variability in the S-WBGT forecast bias under low and variable solar radiation arises partly from the difficulty in forecasting cloud cover. This cloud cover, in turn, impacts the estimated solar radiation. This is also evident in the diurnal curve of forecast bias, with the bias of all forecast methods becoming increasingly positive toward midday (e.g., the period of highest solar radiation) and then decreasing through the afternoon.

c. 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 analysis revealed that the S-WBGT and S-WBGT Z were superior in forecasting black flag, with higher hit rates. However, this paralleled a tendency to over forecast red and, particularly, black flag, with bias scores greater than one and higher false alarm ratios for black flag than the NWS WBGT: 0.45 (S-WBGT) and 0.55 (S-WBGT Z) compared to 0.37 for the NWS WBGT (Table 5). This pattern was true when assessing Observed and Station WBGT forecast flags. However, with station WBGT, the magnitude of the NWS WBGT under forecasting red and black flag, and S-WBGT and S-WBGT Z over forecasting these flags was slightly higher.

Ultimately this reveals a delicate balance that results in a decision between 1) being more certain that when a black flag is forecast, a black flag will be observed; but if a black flag is not forecast, it still very well could be observed (NWS WBGT) and 2) if a black flag is forecast, it may occur and, if it is not forecast, one has higher confidence in that being true (S-WBGT methods). This latter option is particularly true for locations more sheltered from wind, for which the S-WBGT Z produces a more accurate forecast. Given the use of WBGT in protecting health, erring on the side of overforecasting black flag conditions might be preferable to underforecasting. Even if these conditions do not materialize, activities can be adjusted. However, both under- and overforecasting have logistical implications: for instance, a sudden cancellation of high school practices due to forecasting changes can disrupt parents' work schedules, who then have to pick up their child. Furthermore, the HSS and GSS support the use of the S-WBGT methods, since they both have higher scores and higher percent corrects relative to the NWS WBGT (Table 8). While the GSS for S-WBGT Z and NWS WBGT are very close (0.34 versus 0.32, respectively), it is important to note that the GSS is accounting for the "closeness" of the categorical forecast misses. However, even though the NWS methods resulted in "close" misses when a black flag was observed, the difference in activity modifications and health implications between a red and black flag are important to consider.

d. Surface roughness

Last, 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 measured and forecasted for 10 m above the ground, they

must be downscaled to 2 m to assess human heat stress. This study compared two methods of downscaling (translating) winds: 1) Pasquill–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. This implies that the S-WBGT Z is more reliable in complex terrain (e.g., increased tree cover or urban environments with large structures). Future work could ascertain the roughness level 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, 100 m, etc.), any substantial differences and possible benefits of using the higher resolution imagery are unlikely and significantly less feasible to operationalize.

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. Further research should be conducted to confirm the accuracy of WBGT forecasts in comparison to in situ WBGT across the broader CONUS. 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.

Last, it is important to acknowledge that WBGT estimation involves multiple input variables, each with its own uncertainty. These uncertainties, such as those in downscaling wind speed and estimating solar radiation, can compound and affect the final WBGT values. While our methodology aims to minimize the effects of these uncertainties on the resulting estimations of the natural wet bulb temperature and black globe temperature, they should be considered when interpreting the results.

Overall, WBGT forecasts are more challenging for areas with complex microclimates, particularly microclimates with higher surface roughness (i.e., areas with many trees or structures). The forecasts should be used with caution in such areas. Additional information to complement these forecasts include 1) developing a general understanding of how your

specific microclimate influences WBGT and 2) comparing WBGT readings with the forecast, from which user-based bias corrections could be estimated. WBGT forecasts enable robust planning of outdoor activity; however, measurements of WBGT onsite at the time of activity remain critical.

Acknowledgments. The authors wish to acknowledge the staff at Horace Williams Airport in Chapel Hill, North Carolina, that allowed for WBGT data collection and the North Carolina State Climate Office for access to the weather station data located there.

Data availability statement. The methods for estimating WBGT based on standard meteorological variables are provided in Liljegren et al. (2008) and in the R package “wbgt”. The methods for creating superpixels from collecting images of cloud cover are detailed in the documentation for the R package “SuperpixelImageSegmentation”. Archived data for the National Digital Forecast Database (NDFD) are available through National Centers for Environmental Information (NCEI).

APPENDIX A

Surface Roughness Weighted Average Schema

Table A1 displays seven configurations of calculating the weighted average surface roughness across different spatial scales for a given pixel. Figure A1 displays the difference between station WBGT calculated when using weighted average surface roughness with configuration 1 (Table A1) compared to the six other configurations.

TABLE A1. Surface roughness weighted average schema.

Scale (m)	Weights (%) for each scale						
	1	2	3	4	5	6	7
30	10	10	15	15	25	25	10
100	25	50	35	50	40	25	20
250	50	25	35	20	25	25	40
500	15	15	15	15	10	25	30

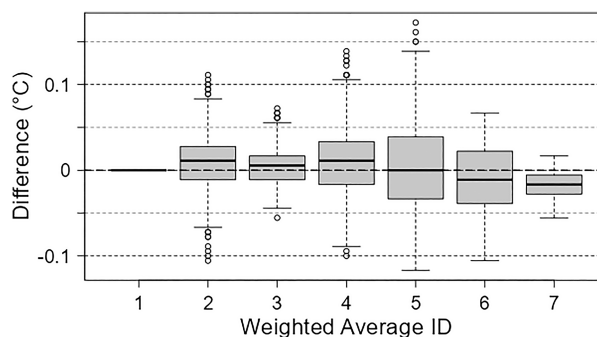


FIG. A1. Station WBGT sensitivity to differences in weighted averages across spatial scales.

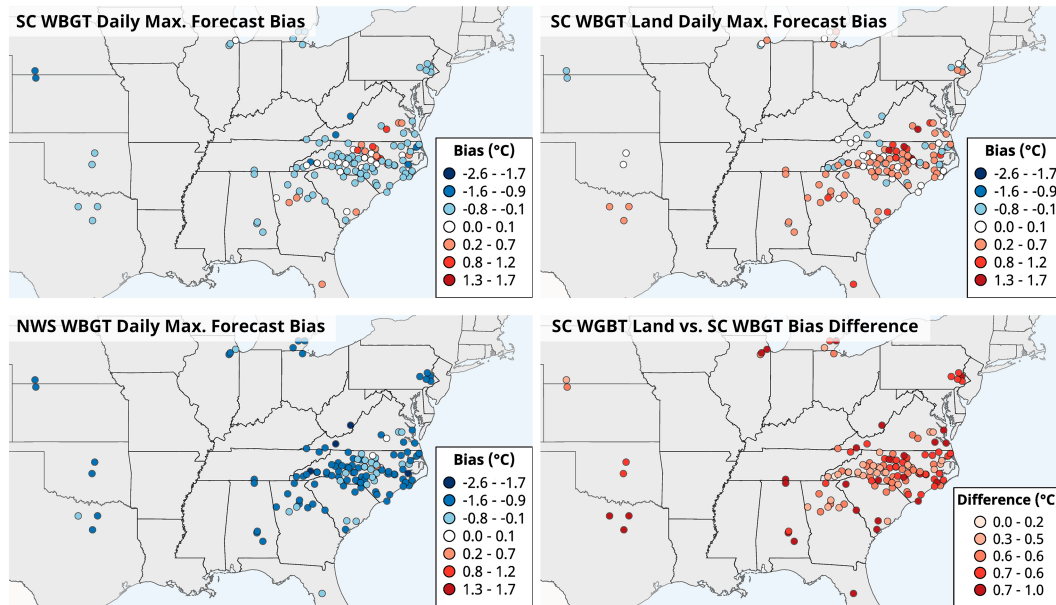


FIG. B1. Station WBGT bias: Regional variations.

APPENDIX B

Station WBGT Bias: Regional Variations

Bias is calculated as forecast WBGT minus station WBGT. Forecast bias is for a 24-h lead time. S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds (see Fig. B1). NWS WBGT refers to the WBGT forecast utilizing the NWS methods.

APPENDIX C

Confusion Matrices for Observed and Station WBGT Flag Forecast Accuracy Assessment

WBGT flags are as follows: 1) green flag (26.7°–29.4°C), 2) yellow flag (29.5°–31.0°C), 3) red flag (31.1°–32.1°C), and

4) black flag (32.2°C+). The assessment is for a forecast lead time of 24 h. S-WBGT (Z) are the forecasts using the Liljegren method, with PG stability classes (surface roughness) to downscale wind speeds (see Tables C1 and C2). NWS WBGT refers to the WBGT forecast utilizing the NWS methods.

TABLE C1. Confusion matrices for observed WBGT flag forecast accuracy assessment.

Forecast Flag	Observed flag											
	S-WBGT				S-WBGT Z				NWS WBGT			
	1	2	3	4	1	2	3	4	1	2	3	4
1	76	8	2	0	64	5	1	0	92	26	5	0
2	51	56	12	3	49	40	7	0	53	72	49	17
3	34	59	59	23	39	55	28	8	24	62	43	53
4	16	49	45	88	29	72	82	106	2	12	21	44

TABLE C2. Confusion matrices for station WBGT flag forecast accuracy assessment.

Forecast Flag	Station flag											
	S-WBGT				S-WBGT Z				NWS WBGT			
	1	2	3	4	1	2	3	4	1	2	3	4
1	38 264	5904	273	30	33 069	2803	82	5	38 264	5904	273	30
2	12 764	15 666	3656	680	17 982	13 858	1858	185	12 764	15 666	3656	680
3	1985	5350	4524	2010	3770	8434	4413	1219	1985	5350	4524	2010
4	406	1404	2080	2889	941	3249	4180	4200	406	1404	2080	2889

APPENDIX D

NBM and NDFD 2-m Wind Comparison

Figure D1 displays a comparison of forecast wind speeds (24-h lead time) between the National Blend of Models (NBM) and National Digital Forecast Database (NDFD). Boxplot whiskers extend up to 1.5 times the interquartile range (e.g., top whisker is $1.5 \times \text{IQR} + \text{third quartile value}$). S-WBGT Z is SERCC/CISA forecast with wind speed downscaled to 2 m using PG stability classes (surface roughness).

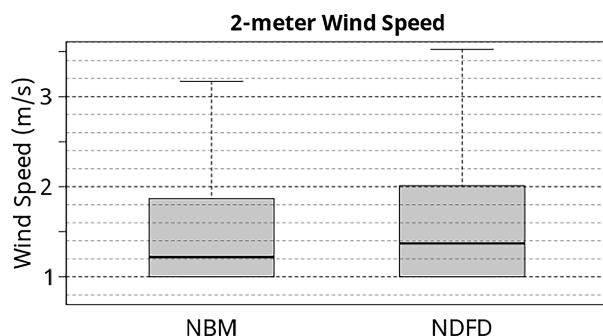


FIG. D1. NBM and NDFD 2-m wind comparison.

REFERENCES

- ACGIH, 2017: *TLVs and BEIs: Threshold Limit Values for Chemical Substances and Physical Agents and Biological Exposure Indices*. American Conference of Governmental Industrial Hygienists, 288 pp.
- Bastiaanssen, W. G. M., M. Meneti, R. A. Feddes, and A. A. M. Holtslag, 1998: A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. *J. Hydrol.*, **212–213**, 198–212, [https://doi.org/10.1016/S0022-1694\(98\)00253-4](https://doi.org/10.1016/S0022-1694(98)00253-4).
- Bowen, B. M., J. M. Dewart, and A. I. Chen, 1983: Stability-class determination: A comparison for one site. *Sixth Symp. on Turbulence and Diffusion*, Boston, MA, Los Alamos National Laboratory, 5 pp., <https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-82-3382>.
- Boyer, T. R., 2022: NDFD wet bulb globe temperature algorithm and software design. NWS Meteorological Development Laboratory, 8 pp., <https://vlab.noaa.gov/documents/6609493/7858379/NDFD+WBGT+Description+Document.pdf/fb89cc3a-0536-111f-f124-e4c93c746ef7?t=1642792547129>.
- Budd, G. M., 2008: Wet-bulb globe temperature (WBGT)—Its history and its limitations. *J. Sci. Med. Sport*, **11**, 20–32, <https://doi.org/10.1016/j.jsams.2007.07.003>.
- Casa, D. J., and Coauthors, 2015: National Athletic Trainers' Association position statement: Exertional heat illnesses. *J. Athletic Train.*, **50**, 986–1000, <https://doi.org/10.4085/1062-6050-50.9.07>.
- CDC, 2010: Climate change and extreme heat events. Centers for Disease Control and Prevention, 20 pp., <https://www.cdc.gov/climateandhealth/pubs/climatechangeandextremeheatevents.pdf>.
- Chen, F., and Y. Zhang, 2009: On the coupling strength between the land surface and the atmosphere: From viewpoint of surface exchange coefficients. *Geophys. Res. Lett.*, **36**, L10404, <https://doi.org/10.1029/2009GL037980>.
- Clark, J., and C. E. Konrad, 2020: Accuracy assessment of experimental wet bulb globe temperature forecasts across North Carolina and the Continental United States. 11 pp., https://www.cisa.sc.edu/Pubs_Presentations_Posters/Reports/2020_Clark%20and%20Konrad_WBGT%20Assessment.pdf.
- Cooper, E., A. Grundstein, A. Rosen, J. Miles, J. Ko, and P. Curry, 2017: An evaluation of portable wet bulb globe temperature monitor accuracy. *J. Athletic Train.*, **52**, 1161–1167, <https://doi.org/10.4085/1062-6050-52.12.18>.
- Copernicus Sentinel data, 2019: Processed by ESA. Accessed 1 April 2021; https://doi.org/10.5270/S2_-742ikth.
- Craven, J. P., and Coauthors, 2018: Overview of National Blend of Models version 3.1. Part I: Capabilities and an outlook for future upgrades. *25th Conf. on Probability and Statistics*, Austin, TX, Amer. Meteor. Soc., 7.3, <https://ams.confex.com/ams/98Annual/webprogram/Paper325347.html>.
- , D. E. Rudack, and P. E. Shafer, 2020: National blend of models: A statistically post-processed multi-model ensemble. *J. Oper. Meteor.*, **8** (1), 1–14, <https://doi.org/10.1519/nwajom.2020.0801>.
- Dewitz, J., 2021: National Land Cover Database (NLCD) 2019 products (version 2.0, June 2021). U.S. Geological Survey, accessed 1 September 2022, <https://doi.org/10.5066/P9KZCM54>.
- Dimiceli, V. E., S. F. Piltz, and S. A. Amburn, 2011: Estimation of black globe temperature for calculation of the wet blub globe temperature index. *Proc. World Congress on Engineering and Computer Science 2011*, San Francisco, CA, Newswood Limited, 9, https://www.iaeng.org/publication/WCECS2011/WCECS2011_pp591-599.pdf.
- Frank, C. W., B. Pospichal, S. Wahl, J. D. Keller, A. Hense, and S. Crewell, 2020: The added value of high resolution regional reanalyses for wind power applications. *Renewable Energy*, **148**, 1094–1109, <https://doi.org/10.1016/j.renene.2019.09.138>.
- Grundstein, A., C. Williams, M. Phan, and E. Cooper, 2015: Regional heat safety thresholds for athletics in the contiguous United States. *Appl. Geogr.*, **56**, 55–60, <https://doi.org/10.1016/j.apgeog.2014.10.014>.
- Hondula, D. M., J. K. Vanos, and S. N. Gosling, 2014: The SSC: A decade of climate-health research and future directions. *Int. J. Biometeor.*, **58**, 109–120, <https://doi.org/10.1007/s00484-012-0619-6>.
- Hoshi, A., and Y. Inaba, 2007: Prediction of heat disorders in Japan. *Global Environ. Res.*, **11**, 45–50, http://www.aires.or.jp/attach.php/6a6f75726e616c5f31312d31656e67/save/0/0/11_1-07.pdf.
- Hunter, C. H., and C. O. Minyard, 1999: Estimating wet bulb globe temperature using standard meteorological measurements. *Proc. Second Conf. on Environmental Applications*, Long Beach, CA.
- Iowa Environmental Mesonet, 2021: ASOS-AWOS-METAR data download. Iowa State University, accessed 1 November 2021, <https://mesonet.agron.iastate.edu/request/download.phtml>.
- ISO, 1989: Hot environments—Estimation of the heat stress on working man, based on the WBGT-index (wet bulb globe temperature). Tech. Doc. ISO7243, 9 pp., <https://cdn.standards.iteh.ai/samples/13895/59f0e814f83345bba03bd2bba1632cb6/ISO-7243-1989.pdf>.
- , 2017: Ergonomics of the thermal environment—Assessment of heat stress using the WBGT (wet bulb globe temperature) index. Tech. Doc. ISO7243:2017, 18 pp., <https://www.iso.org/obp/ui/en/#iso:std:iso:7243:ed-3:vi:en>.
- Jolliffe, I. T., and D. B. Stephenson, Eds., 2012: *Forecast Verification: A Practitioner's Guide in Atmospheric Science*. John Wiley & Sons, 292 pp.

- Kasten, F., and G. Czeplak, 1980: Solar and terrestrial radiation dependent on the amount and type of cloud. *Sol. Energy*, **24**, 177–189, [https://doi.org/10.1016/0038-092X\(80\)90391-6](https://doi.org/10.1016/0038-092X(80)90391-6).
- Kestrel, 2021: Users Guide Kestrel 5400 Heat Stress Tracker. Kestrel, 24 pp., <https://kestrelinstruments.com/mwdownloads/download/link/id/14/>.
- Kopec, R. J., 1977: Response of the wet-bulb-globe-thermometer heat stress index to selected land use surfaces. *Southeast. Geogr.*, **17**, 133–145, <https://doi.org/10.1353/sge.1977.0009>.
- Kovats, R. S., and L. E. Kristie, 2006: Heatwaves and public health in Europe. *Eur. J. Public Health*, **16**, 592–599, <https://doi.org/10.1093/eurpub/ckl049>.
- Lemke, B., and T. Kjellstrom, 2012: Calculating workplace WBGT from meteorological data: A tool for climate change assessment. *Ind. Health*, **50**, 267–278, <https://doi.org/10.2486/indhealth.MS1352>.
- Liebllich, M., and J. Spector, 2017: WBGT: Wet Bulb Globe Temperature, version 1.2. R package, <https://github.com/mdljts/wbgt/blob/master/DESCRIPTION>.
- Liljegren, J. C., R. A. Carhart, P. Lawday, S. Tschopp, and R. Sharp, 2008: Modeling the wet bulb globe temperature using standard meteorological measurements. *J. Occup. Environ. Hyg.*, **5**, 645–655, <https://doi.org/10.1080/15459620802310770>.
- Markert, A., R. Griffin, K. Knupp, A. Molthan, and T. Coleman, 2019: A spatial pattern analysis of land surface roughness heterogeneity and its relationship to the initiation of weak tornadoes. *Earth Interact.*, **23**, <https://doi.org/10.1175/EI-D-18-0010.1>.
- NOAA, Department of Defense, Federal Aviation Administration, and United States Navy, 1998: Automated Surface Observing System (ASOS) user's guide. NOAA Rep. 19981015-073, 89 pp.
- NCAR Research Applications Laboratory, 2015: Verification: Weather Forecast Verification Utilities, version 1.42. R package, <https://CRAN.R-project.org/package=verification>.
- NC CRONOS, 2021: North Carolina State Climate Office. Accessed 1 November 2021, <https://climate.ncsu.edu/cronos/>.
- NCHSAA, 2016: Health and safety. North Carolina High School Athletic Association, accessed 1 July 2021, <https://www.nchsaa.org/wp-content/uploads/2023/07/16-17HealthandSafety.pdf>.
- NIOSH, 2016: *NIOSH Criteria for a Recommended Standard: Occupational Exposure to Heat and Hot Environments*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Institute for Occupational Safety and Health, DHHS (NIOSH), 192 pp.
- NWS, 2019a: Public Information Statement 19-24 updated. National Weather Service Headquarters, Accessed 22 August 2023, https://nsdesk.servicenowservices.com/nws/?id=cards_request_pc&table=x_anno2_cards_prod_catalog&sys_id=59a6184c1ba2e510f874202ce54bcb8c&view=sp.
- , 2019b: National digital forecast database and local database description and specifications. Tech. Rep. NWSI 10-201, 19 pp., <https://www.weather.gov/media/mdl/ndfd/pd01002001curr.pdf>.
- , 2020a: Weather related fatality and injury statistics. National Weather Service, accessed 1 June 2022, <https://www.weather.gov/hazstat/>.
- , 2020b: Public Information Statement 20-16. National Weather Service Headquarters, 3 pp., https://www.weather.gov/media/notification/pdf2/pns20-16nbm_v4.0.pdf.
- , 2022: Service Change Notice 22-43 updated. National Weather Service Headquarters, 2 pp., https://www.weather.gov/media/notification/pdf2/scn22-43_wet_bulb_globe_temp_t2o_aaa.pdf.
- OSHA, 2017: OSHA Instruction Directive TED-01-00-015. OSHA, accessed 1 November 2020, https://www.osha.gov/dts/osta/otm/otm_iii/otm_iii_4.html.
- Parsons, K., 2006: Heat stress standard ISO 7243 and its global application. *Ind. Health*, **44**, 368–379, <https://doi.org/10.2486/indhealth.44.368>.
- Patel, T., S. P. Mullen, and W. R. Santee, 2013: Comparison of methods for estimating wet-bulb globe temperature index from standard meteorological measurements. *Mil. Med.*, **178**, 926–933, <https://doi.org/10.7205/MILMED-D-13-00117>.
- Racinais, S., and Coauthors, 2022: IOC consensus statement on recommendations and regulations for sport events in the heat. *Br. J. Sports Med.*, **57**, 8–25, <https://doi.org/10.1136/bjsports-2022-105942>.
- Roberts, W. O., L. E. Armstrong, M. N. Sawka, S. W. Yeargin, Y. Heled, and F. G. O'Connor, 2021: ACSM expert consensus statement on exertional heat illness: Recognition, management, and return to activity. *Curr. Sports Med. Rep.*, **20**, 470–484, <https://doi.org/10.1249/JSR.0000000000000878>.
- Rothfusz, L. P., 1990: The heat index “equation”. NWS Tech. Doc. SR 90-23, 2 pp.
- SERCC, and CISA, 2023: Wet Bulb Globe Temperature (WBGT) tool. Convergence of Climate-Health-Vulnerabilities, accessed 1 August 2022, <https://convergence.unc.edu/tools/wbgt/>.
- Su, Z., 2002: The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrol. Earth Syst. Sci.*, **6**, 85–100, <https://doi.org/10.5194/hess-6-85-2002>.
- US Department of the Army, 2022: Heat stress control and heat casualty management. Tech. Bulletin Rep. TB MED 507, 81 pp., https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN35159-TB_MED_507-000-WEB-1.pdf.
- US EPA, 2000: Meteorological monitoring guidance for regulatory modeling applications. EPA-454/R-99-005, 171 pp., https://www.epa.gov/sites/production/files/2020-10/documents/mmgrma_0.pdf.
- van den Berg, G. P., 2004: Effects of the wind profile at night on wind turbine sound. *J. Sound Vib.*, **277**, 955–970, <https://doi.org/10.1016/j.jsv.2003.09.050>.
- Verkaik, J. W., A. J. M. Jacobs, A. B. C. Tijn, and J. R. A. Onvlee, 2005: Local wind speed estimation by physical downscaling of weather model forecasts. Accessed 2022, <https://www.knmi.nl/research/publications/local-wind-speed-estimation-by-physical-downscaling-of-weather-model-forecasts>.
- Yaglou, C., and D. Minard, 1957: Control of heat casualties at military training centers. *Arch. Indus. Health*, **16**, 302–305.
- Zeng, X., Z. Wang, and A. Wang, 2012: Surface skin temperature and the interplay between sensible and ground heat fluxes over arid regions. *J. Hydrometeor.*, **13**, 1359–1370, <https://doi.org/10.1175/JHM-D-11-0117.1>.
- Zheng, D., R. van der Velde, Z. Su, M. J. Booij, A. Y. Hoekstra, and J. Wen, 2014: Assessment of roughness length schemes implemented within the Noah land surface model for high-altitude regions. *J. Hydrometeor.*, **15**, 921–937, <https://doi.org/10.1175/JHM-D-13-0102.1>.