BioOne COMPLETE

From Public Forecasts to Occupational Heat Safety Management: Evaluating the Effectiveness of the National Weather Service's Operational Heat Risk Forecasts

Authors: Kim, Yoojun, and Ham, Youngjib

From Public Forecasts to Occupational Heat Safety Management: Evaluating the Effectiveness of the National Weather Service's Operational Heat Risk Forecasts

Yoojun Kim[1] (ID) and Youngjib Ham[1]

[1]Department of Construction Science, Texas A&M University, College Station, TX, USA.

ABSTRACT: Heat poses a major environmental risk to occupational safety, necessitating timely insights into associated risks to safeguard workers. In June 2022, the National Weather Service (NWS) initiated operational wet bulb globe temperature (WBGT) forecasts, offering valuable information for heat risk management. This study evaluates the effectiveness of NWS WBGT forecasts, aiming to identify potential areas of caution and improvements for their application for occupational safety management. To this end, the study examines 1.3 million hourly historical NWS WBGT forecast data, comparing it with observed data from 252 weather stations across the US during the summer of 2023. The results offer key insights, revealing that: (1) the accuracy of NWS WBGT forecasts is influenced more by the times of interest than by the forecast horizons; (2) NWS WBGT forecast accuracy varies across different climates in the US, with air temperature bias being the most influential factor in this inaccuracy; and (3) while NWS WBGT forecasts accurately identify the lowest heat risks (i.e. no heat risk), their performance decreases at higher risk levels, emphasizing the importance of careful interpretation in safety management. These insights offer guidance for more cautious interpretations of NWS WBGT forecasts and lay the foundation for future studies on leveraging operational weather forecasting services in effective heat mitigation strategies.

KEYWORDS: Occupational safety, heat risk, heat exposure, WBGT, safety management

Introduction

Timely evaluation of heat exposure is essential to mitigate one of the most critical environmental concerns affecting workers.[1] Among the myriad of heat indices, such as the Universal Thermal Climate Index (UTCI)[2] and Physiological Equivalent Temperature (PET),[3] the Wet Bulb Globe Temperature (WBGT) index is widely recognized as an international reference for assessing environmental heat risks in occupational settings.[4] Here, heat risk refers to environmental heat stress assessed by WBGT, indicating potential risks for heat-related illnesses among workers. This index incorporates weather variables such as natural wet-bulb temperature (T_{nwb}), air temperature (T_a), and globe temperature (T_g), forming the basis for safety guidelines.[5] Occupational safety institutions like the National Institute for Occupational Safety and Health (NIOSH),[6] the American Conference of Governmental Industrial Hygienists (ACGIH),[7] and the International Organization for Standardization (ISO),[8] provide the WBGT-based heat exposure limits, such as ACGIH's thermal limit values and NIOSH's recommended alert limit and recommended exposure limits. These include ACGIH's thermal limit values and NIOSH's recommended alert limits and recommended exposure limits, along with other heat mitigation strategies, such as appropriate work/rest scheduling to accommodate high heat-sensitive work areas and periods.[6]

Despite the crucial role of the WBGT index in occupational safety management, obtaining timely WBGT information in outdoor workplaces presents practical challenges. It often requires additional weather sensors, such as a black globe thermometer,[9] which introduce logistical and cost constraints onsite along with the need for frequent calibrations.[10] Predicting future occupational heat risks for strategic mitigation planning is even more complex. Such predictions require large-scale numerical simulations and expert adjustments to properly capture global and local weather patterns, alongside daily and seasonal fluctuations.[11] To address these challenges, previous studies have explored the use of weather forecasting services to estimate predictive WBGT values as effective alternatives.[12,13] One notable example is the HEAT-SHIELD project,[14] which leverages European Center for Medium Range Weather Forecasts (ECMWF)[15] to provide daily maximum WBGT forecasts for Europe. Despite these efforts, gaps remain in the application of such forecasts to occupational safety. Specifically, short-term forecast information, such as hourly updates, is crucial for effective safety planning, as maximum values alone do not account for diurnal variations in heat risks. Additionally, the accuracy of these forecasts has not been validated, raising concerns about their reliability for safety management purposes.

In June 2022, the US National Weather Service (NWS) initiated its operational WBGT forecasts, hereafter referred to as NWS WBGT forecasts.[16] This service provides public forecasts on a 2.5 km grid for the contiguous United States, updated hourly for up to 36 hours, every 3 hours for up to 72 hours, and every 6 hours for up to 168 hours, also including Hawaii, Guam,

and Puerto Rico at 3-hour intervals for up to 72 hours and 6-hour intervals for up to 168 hours.[17] Recognizing the practical value of these forecasts, recent studies have assessed their accuracy against observed data, which are measurements collected directly from weather stations. For instance, Ahn et al[18] evaluated the NWS WBGT data from 2018 to 2019 across the US and found average discrepancies ranging from –0.64°C to 1.46°C, while Clark et al[19] examined forecast accuracy for North Carolina during the summers of 2019 to 2021, noting variations in accuracy depending on the level of heat risk and time of day.

Although these evaluations provide valuable insights, a comprehensive understanding of their role in supporting critical safety decisions requires more than just overall accuracy assessments. For example, if the forecasts perform well in non-heat-risk scenarios but falter under heat-sensitive conditions, their reliability for practical applications may be compromised. Similarly, evaluating forecast performance over different time horizons (e.g. 1 hour ahead versus 6 hours ahead) is crucial, as high fidelity in short-term forecasts but poor performance in longer-term scenarios would limit the service's application in occupational safety management. Moreover, these evaluations were conducted before the service became operational in 2022. Since then, WBGT calculation methods have been refined in several areas, including direct and diffused solar radiation, convective heat transfer coefficients, adjustment of solar flux based on cloud cover, and the calculation of natural wet-bulb temperature.[20] However, the accuracies of these updates have not yet been explored.

These gaps in knowledge motivate our study to revisit the evaluation of NWS WBGT forecasts by comparing them with observed weather data, aiming to better understand its role in occupational safety management. To this end, we have compiled an extensive dataset, resulting in 1.3 million hourly data points, including in-situ observations at 252 locations and historical NWS WBGT forecasts during the summer months of 2023 in the US. Our analysis addresses forecast horizons, times of interest, climate types, and the ability to detect occupational heat risk levels to uncover practical insights and identify areas for future research and improvement. The main contributions of this study are threefold. First, it explores hourly NWS WBGT forecast data, improving the granularity across times of interest and forecast horizons. The findings offer empirical insights into the effectiveness of NWS WBGT forecasts according to different heat risk levels. Second, the study evaluates weather biases affecting forecast accuracy across various climate types, providing insights on potential improvement plans to minimize climate-specific biases in future research endeavors and highlighting heat-risk-vulnerable climate types that require more vigilant occupational heat risk monitoring. Third, it provides empirical evidence on the importance of addressing multiple factors in assessing heat-health warning systems, beyond overall accuracy. This underscores the need to incorporate forecast horizons, times of interest, climate types,

and the ability to detect occupational heat risk levels. Overall, this study contributes to improving the role of operational heat risk forecasting services in effectively managing heat exposures for occupational workers.
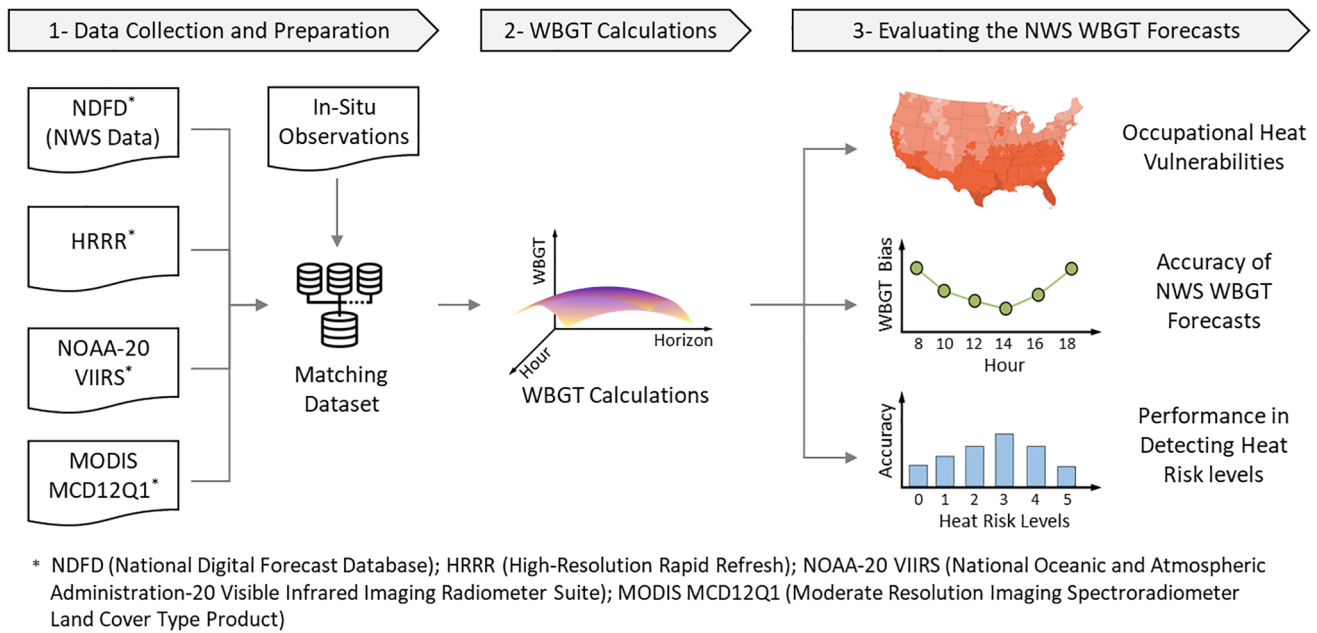
Materials and Methods

Our study evaluates the effectiveness of NWS WBGT forecasts by comparing them to observations from in-situ weather stations across the US. To accomplish this, we integrated data from multiple sources into a unified dataset, aligning observation locations with forecast times to enable a direct comparison, as detailed in the data collection and preprocessing section. This integration process involved collecting hourly data from June 1, 2023, to August 31, 2023, during the hours of 6:00 AM to 8:00 PM, from 252 weather stations in the US. Following data collection, WBGT values were calculated using the NWS WBGT calculation method to analyze the reliability of NWS WBGT forecasts in the context of occupational safety management. This section details the research methodology, outlining the procedures of (1) data collection and preparation, (2) WBGT calculations, and (3) an evaluation of the effectiveness of NWS WBGT forecasts, as illustrated in Figure 1.

Data collection and preprocessing

National digital forecast database (NDFD). The NDFD integrates digital weather forecasts from the weather forecast offices, the river forecast centers, and the national centers for environmental prediction.[11] Our study focused on collecting hourly data on weather variables pertinent to NWS WBGT forecasts from the NDFD, including T_a, RH, cloud cover, and air velocity at 10 m (V_{a10}). The NDFD provides weather forecasts for these variables on an hourly basis for up to 36 hours and every 3 hours for up to 168 hours, with a 2.5 km gridded horizontal resolution.[21] The frequency of updates can vary based on the discretions of the weather forecast offices.[22] The NDFD is publicly available through the National Oceanic and Atmospheric Administration (NOAA)'s data portals, accessible via Hypertext Transfer Protocol (HTTP) or File Transfer Protocol (FTP).[23]

Additional data for NWS WBGT computation. The NWS WBGT computation requires additional data not available in the NDFD.[20] These include (1) surface air pressure, (2) surface albedo, and (3) surface roughness length. For surface air pressure, the NWS algorithm uses the 50th percentile mean sea level pressure from the National Blend of Models (NBM) and converts this to surface air pressure. This study, however, directly sourced surface air pressure from the High-Resolution Rapid Refresh (HRRR) model,[24] streamlining the computation by avoiding the need for conversion. For surface albedo, the NWS algorithm employs the NOAA17 Advanced Very High Resolution Radiometer (AVHRR3), which is no longer publicly available. Alternatively, this study has sourced data

* NDFD (National Digital Forecast Database); HRRR (High-Resolution Rapid Refresh); NOAA-20 VIIRS (National Oceanic and Atmospheric Administration-20 Visible Infrared Imaging Radiometer Suite); MODIS MCD12Q1 (Moderate Resolution Imaging Spectroradiometer Land Cover Type Product)

Figure 1. Overview of studying the effectiveness of NWS WBGT forecasts.

from the Visible Infrared Imaging Radiometer Suite (VIIRS) Level 3 daily gridded land surface albedo from the NOAA-20 satellite.[25] Additionally, land cover type data required for calculating surface roughness length was sourced from the Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type Product (MCD12Q1) version 6.1,[26] then converted to surface roughness length, according to Supplemental Table 1.

In-situ observation data. Observed weather data were collected from 5 weather station networks: Automated Weather Data Network[27]; Agriculture and meteorology[28]; California Irrigation Management Information System[29]; Soil Climate Analysis Network[30]; and WeatherSTEM.[31] Consequently, approximately 1.8 million hourly data for T_a, RH, V_a, Global Horizontal Irradiance (GHI) and their geolocation data (latitude and longitude) were collected from 1508 stations. The data were further refined through 3 rounds of filtering. First, we eliminated hourly data recorded when the sun's elevation was below 3°, according to the NWS WBGT calculations to exclude nighttime observations.[20] Second, we applied quality control measures to remove statistically suspect observations. This involved filtering out hourly weather data outliers for each station's monthly data by identifying observations with z-scores in the top or bottom 0.5% for each weather parameter over the study period. Finally, to avoid an overconcentration of data from specific regions, we randomly selected stations based on their latitudes and longitudes within a 100 km radius. Table 1 and Figure 2 show the number of stations and data points according to the Köppen climates based

Table 1. Number of stations and data points for Köppen climates.

KÖPPEN CLIMATE TYPE	STATION[a]	OBSERVATION[b]
Humid subtropical climate (Cfa)	98	108
Cold desert climate (BSk)	52	53.3
Warm-summer humid continental climate (Dfb)	32	31.9
Hot-summer humid continental climate (Dfa)	26	21.9
Mediterranean climate (Csb)	12	14.6
Hot desert climate (BWh)	7	8.2
Mediterranean climate (Csa)	6	7.2
Cold desert climate (BWk)	4	3.7
Warm-summer continental Mediterranean climate (Dsb)	3	3.7
Subarctic climate with dry winters (Dwa)	3	2.5
Subarctic climate with dry winters (Dwb)	3	2.5
Tropical monsoon climate (Am)	2	2.2
Hot desert climate (BSh)	1	1.2
Oceanic climate (Cfb)	1	1.2
Hot-summer continental Mediterranean climate (Dsa)	1	1.2
Tundra climate (ET)	1	0.8

[a]Number of stations.
[b]Number of hourly data points (in thousands).
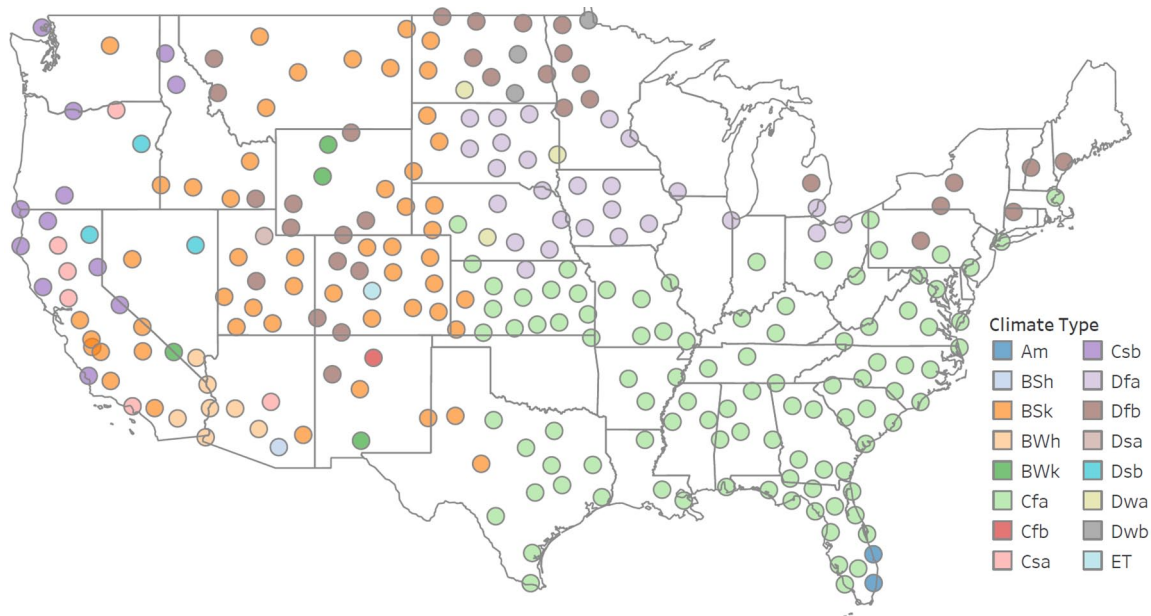
Figure 2. Spatial distributions of in-situ stations based on Köppen climates according to Table 1.

on climate trends from 1981 to 2000.[32] The collected data represent that Humid subtropical climate (Cfa), Cold desert climate (BSk), Warm-summer humid continental climate (Dfb), and Hot-summer humid continental climate (Dfa) are the most frequent Köppen climates in the US. More details on how the Köppen climates are determined can be found in this reference.[32]

Matching dataset. The NDFD data, additional data for NWS WBGT computation, and in-situ observation data were aligned based on their forecast horizons and local standard times (LST). This alignment process involved matching the LST of the in-situ observations with the NDFD data across forecast intervals of 1, 6, 12, 18, and 24 hours. For example, an in-situ observation recorded on July 2nd at 12:00 in the Central Time Zone (UTC-5 hours) was matched with the NDFD data as follows: the 24-hour forecast data was collected from the NDFD data issued on July 1st at 17:00, the 18-hour forecast data from July 1st at 23:00, the 12-hour forecast data from July 2nd at 05:00, the 6-hour forecast data from July 2nd at 11:00, and the 1-hour forecast data from July 2nd at 16:00. This integration effort resulted in an extensive dataset of 1 300 737 hourly data points over the study period. These data points were used to calculate WBGT values, allowing for a direct comparison between the NWS WBGT forecasts and the WBGT values derived from in-situ observations across different local work hours and forecast horizons.

WBGT calculations

The WBGT index is the most widely used thermo-physiological model for occupational heat stress assessments.[33] Its outdoor calculations involve 3 variables: T_{nwb}, T_g, and T_a, as shown in equation (1).[34]

$$WBGT = 0.7T_{nwb} + 0.2T_g + 0.1T_a \tag{1}$$

The NWS employs equation (2) for T_{nwb} calculations in its WBGT computation mechanism.[20]

$$T_{nwb} = T_{wb} + 0.001651 GHI - 0.09555 V_a + 0.13235(T_a - T_{wb}) + 0.20249 \tag{2}$$

Where T_{wb} represents the thermodynamic wet bulb temperature in °C. To ensure consistency in comparing in-situ observations with NWS forecasts, this study uses a formula[35] for T_{wb} calculation, as shown in equation (3).

$$T_{wb} = T_a \tan^{-1}(0.151977(RH + 8.313659)^{0.5}) + \tan^{-1}(T_a + RH) - \tan^{-1}(RH - 1.676331) + 0.00391838 RH^{1.5} \tan^{-1}(0.023101 RH) - 4.686035 \tag{3}$$

For *GHI* calculation, in accordance with the NWS WBGT computation mechanism, the daily maximum of solar radiation flux (GHI_{max}) is calculated based on the Environment Canada weather forecast model.[20] To accommodate the impact of cloud cover on solar radiation flux, this value is subsequently refined using equation (4).

$$GHI = GHI_{max}(1 - 0.75n^{3.4}) \tag{4}$$

Where, *n* represents the cloud cover percentage, expressed as a fraction (0.0-1.0). Subsequently, the hourly solar radiation flux is determined using a Gaussian distribution, under the assumption that the maximum daily solar radiation flux occurs at noon.[36]

For V_a calculation, given that NDFD only provides V_{a10} (i.e. air velocity at 10 m), the equation (5) is further employed, where *l* represents the roughness length in meters, obtained from the MODIS MCD12Q1 dataset.

Table 2. Work/rest ratios and water intake based on WBGT index and work intensities.[6]

HEAT STRESS LEVEL	WBGT INDEX (°C)	EASY WORK (250W)		MODERATE WORK (425W)		HARD WORK (600W)	
		WORK/REST (MIN)	WATER INTAKE (QTH⁻¹)	WORK/REST (MIN)	WATER INTAKE (QTH⁻¹)	WORK/REST (MIN)	WATER INTAKE (QTH⁻¹)
1	25.6-27.7	Unlimited	0.5	Unlimited	0.75	40/20	0.75
2	27.8-29.4	Unlimited	0.5	50/10	0.75	30/30	1
3	29.4-31.1	Unlimited	0.75	40/20	0.75	30/30	1
4	31.1-32.2	Unlimited	0.75	30/30	0.75	20/40	1
5	32.2+	50/10	1	20/40	1	10/50	1

$$V_a = V_{a10} \frac{\log(2/l)}{\log(10/l)} \quad (5)$$

T_g is derived from the equation (6).[37]

$$T_g^{\,4} + CT_g = B + CT_a \quad (6)$$

Once B and C are determined using equations (7) and (8), the remaining value to be determined from the quartic equation is T_g. This value is calculated under the assumption that the real solution to the quartic equation is the one nearest to T_a.[37]

$$B = GHI(\frac{f_{db}}{4\sigma \cos(z)} + f_{dif} \frac{1.2}{\sigma}) + \varepsilon_a T_a^{\,4} \quad (7)$$

$$C = \frac{0.228 \times (3600V_a)^{0.58}}{5.3865 \times 10^{-8}} \quad (8)$$

Where, f_{dif} is the diffuse radiation constant, defined as the sky cover percentage, with a minimum value set at 0.25; f_{db} is the direct radiation constant, calculated as $1 - f_{dif}$; σ represents the Stephan-Boltzmann constant (i.e. $5.6-7 \times 10^{-8}$ W/m²·K⁴); z is zenith angle in degrees, determined using the Solar Position and Intensity (SOLPOS) algorithm[38]; and ε_a is the thermal emissivity calculated as

$$\varepsilon_a = 0.575e_a^{\,(1/7)} \quad (9)$$

Where e_a is the atmospheric vapor pressure, determined by equation (10).[39]

$$e_a = \exp(\frac{17.67(T_d - T_a)}{T_d + 243.5}) \times (1.0007 + 0.00000346P)$$
$$\times 6.112\exp(\frac{17.502T_a}{240.97 + T_a}) \quad (10)$$

Where, P is the surface air pressure in millibars, sourced from the HRRR dataset,[24] and T_d is the dew point temperature in °C, calculated using equation (11).[40]

$$T_d = T_a - (\frac{100 - RH}{5}) \quad (11)$$

The WBGT heat risk levels are evaluated using the following reference criterion outlined in Table 2,[41] originally proposed by the armed services, which targets acclimated average-sized people. This criterion is also used by the Occupational Safety and Health Administration (OSHA).[42] This criterion was chosen for its detailed safety guidelines, including work/rest ratios and water intake recommendations, which offer actionable insights for managing occupational safety.[6] This study investigates the performance of NWS WBGT forecasts, focusing on their ability to accurately assess heat risk levels using this criterion.

Evaluating the effectiveness of NWS WBGT forecasts

The effectiveness of NWS WBGT forecasts is assessed through the comparison of matched WBGT values derived from observations and forecasts across 5 time horizons (1, 6, 12, 18, and 24 hours), focusing on (1) the numerical closeness between observed and forecasted WBGT values and (2) the detection performance of heat risk levels. Forecast accuracy at each hourly observation is determined using root-mean-square error (RMSE) and mean bias error (MBE), based on formulas,[43] according to equations (12) and (13):

$$RMSE = \sqrt{\frac{\sum(\widehat{y_i} - y_i)^2}{n}} \quad (12)$$

$$MBE = \frac{\sum(\widehat{y_i} - y_i)}{n} \quad (13)$$

Where $\widehat{y_i}$ represents the forecasted value for the ith data point, y_i is the observed value for the ith data point, and n is the total number of data points. The performance of NWS WBGT forecasts to identify occupational heat risk levels is measured using a row-normalized confusion matrix, which identifies discrepancies between forecasted and observed WBGT heat risk levels. The confusion matrix is widely used to assess detection performance.[44,45] Given the prevalence of data points at the lowest heat risk level, which exhibit variability across Köppen climates, this study adjusts the confusion
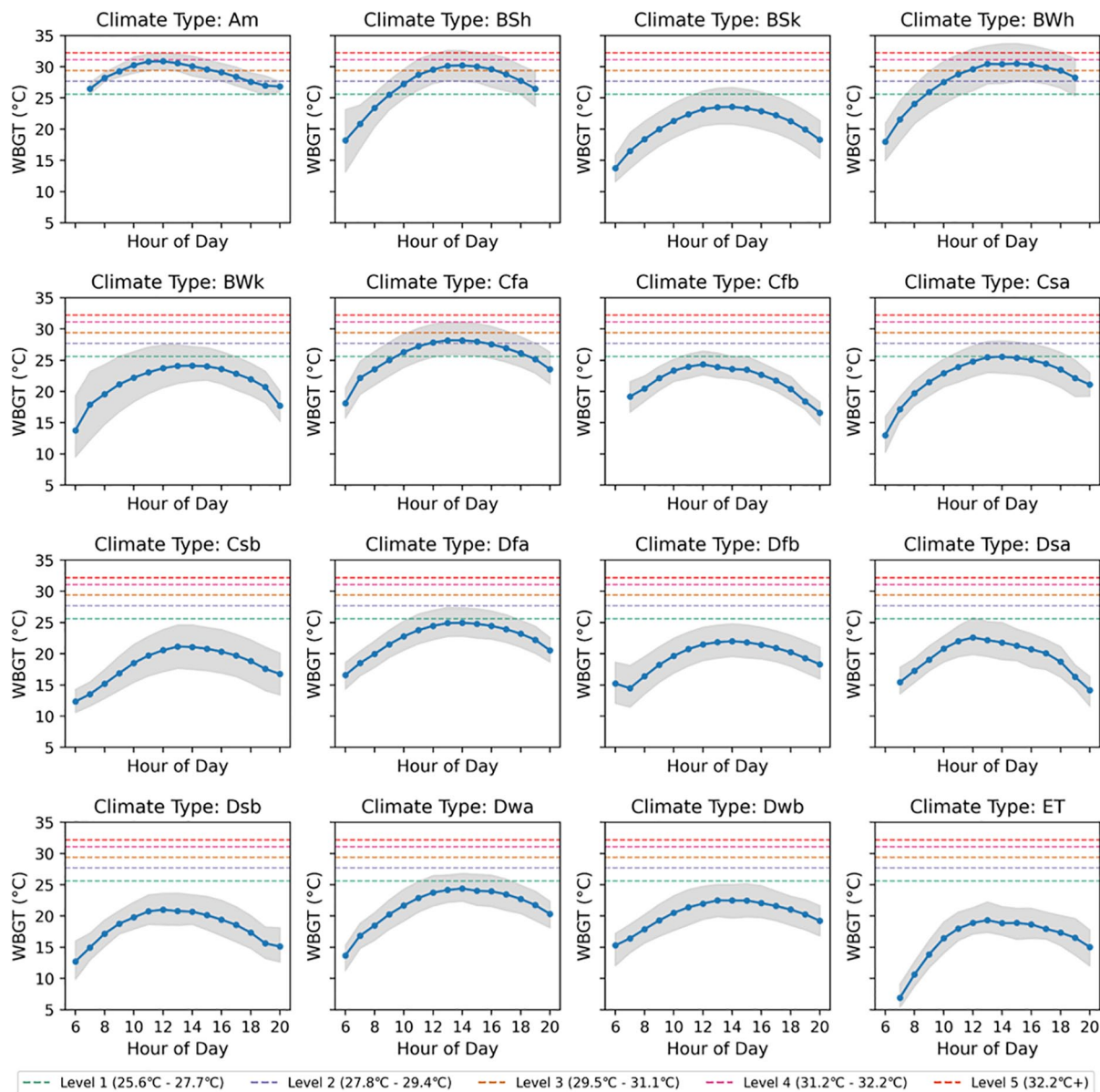
Figure 3. Hourly WBGT variations and interquartile ranges across Köppen climates during local standard work hours (06:00-20:00).

matrix by normalizing each row. This process involves dividing the counts by the row totals and converting the results to percentage form to better recognize the NWS WBGT forecasts performance across the different heat risk levels and Köppen climates.

Results

Occupational heat vulnerabilities based on Köppen climates

Figure 3 features a blue line representing the mean of hourly WBGT values, encompassed by shaded areas indicating the upper and lower quartiles. Furthermore, colored horizontal dotted lines highlight the WBGT levels ranging from Level 1

(25.6 °C-27.7 °C) to Level 5 (over 32.2 °C). Generally, the WBGT trends form a bell-shaped curve, peaking between 12:00 and 14:00 across all climate types, reflecting the typical pattern of occupational heat exposure in the US. Notably, climate types "Csb," "Dfb," "Dsa," "Dsb," "Dwb," and "ET" have upper quartile values that do not surpass Level 1, indicating relatively lower occupational environmental heat risks. Furthermore, climate types "BSk," "BWk," "Cfb," "Csa," "Dfa," and "Dwa" mostly remain WBGT values below Level 1 for the mean and lower quartiles, but their upper quartile may occasionally surpass Level 1. The most pronounced occupational heat risks are observed in climate types "Am," "BSh," "BWh," and "Cfa," with working hours often ranging between Level 1 and Level 5. Additionally, the data reveals variability in the
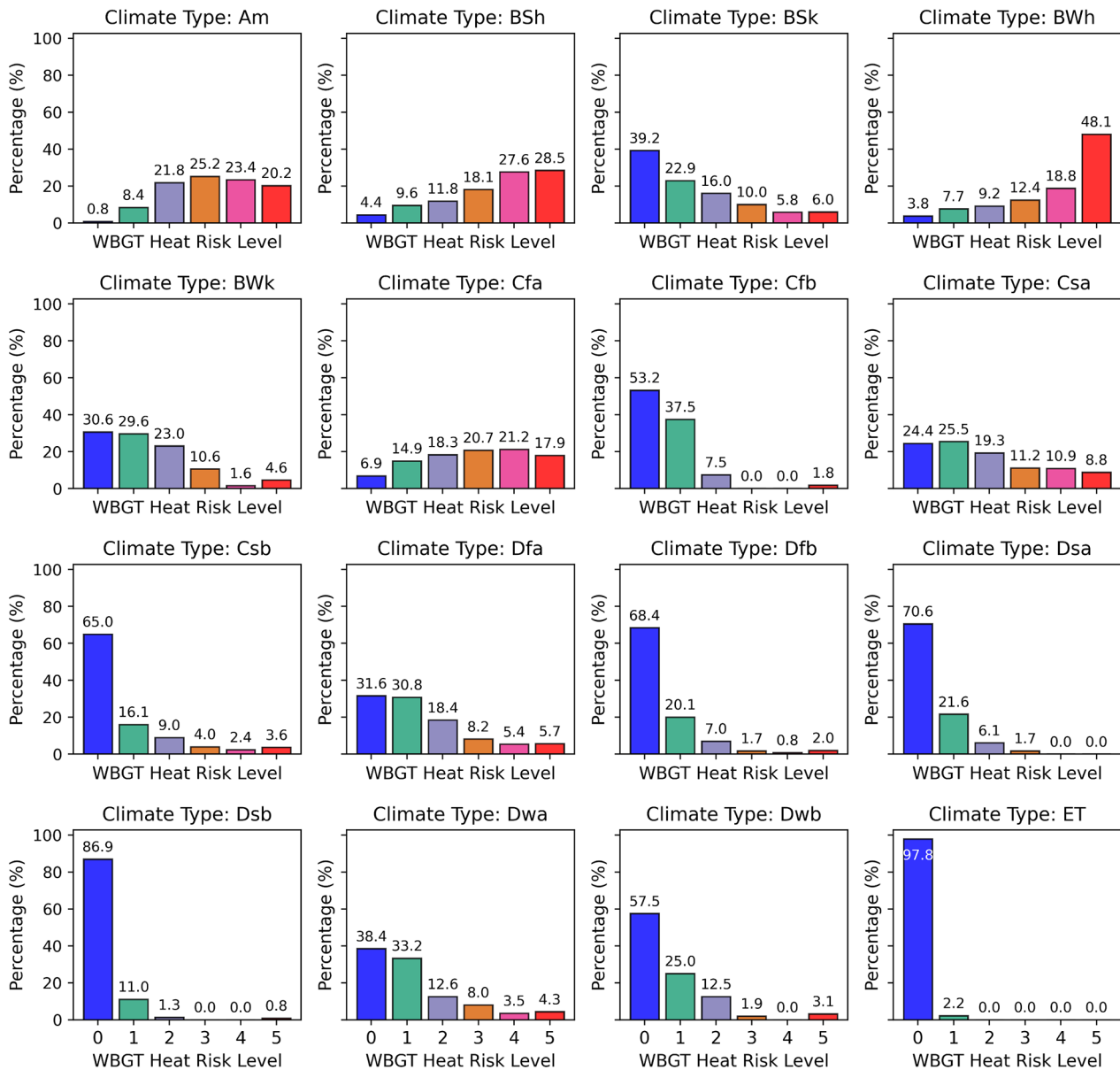
Figure 4. Distribution of WBGT heat risk levels by Köppen climate classification.

interquartile ranges of WBGT across these climates, with "Am" showing a narrower interquartile range, indicating consistent occupational heat risk, while "BSh," "BWh," and "Cfa" display wider ranges, suggesting more daily variations in occupational heat risks. Figure 4 represents the occupational heat risk level distributions across Köppen climates, where each colored bar indicates the proportion of WBGT heat risk levels. In climates "Dsb" and "ET," the majority of heat risk levels are at Level 0 (below 25.6°C), indicating lower occupational heat risks. In contrast, climates "Am," "BSh," "BWh," "Cfa," and "Csa" predominately face heat risks above this threshold, emphasizing distinct occupational heat vulnerabilities. Notably, "Am," "Cfa," and "Csa" show relatively consistent variances in heat risk level distributions, while "BSh" and "BWh" exhibit increasing trends

in percentages across the risk levels with Level 5 (over 32.2°C) being the most prevalent.

Accuracy of NWS WBGT forecasts compared to in-situ observations

Given that WBGT is a function of the weather variables, the accuracy of NWS WBGT forecasts is intrinsically associated with the accuracy of its input weather variables: T_a, RH, V_a, and GHI. Table 3 outlines the MBE and RMSE for these forecasts over 5 time horizons (1, 6, 12, 18, and 24 hours) and 7 local standard times (6:00-20:00 at 2-hour intervals) for the 4 weather variables and WBGT forecasts. The MBE and RMSE for T_a vary from −0.04°C to −1.84°C and from 1.85°C to

Table 3. Comparative analysis of MBE and RMSE metrics for T_a, RH, V_a, GHI, WBGT across five horizons of 1, 6, 12, 18, and 24 hours at 7 local standard time (6:00-20:00) at 2-hour intervals.

VARIABLE	HORIZON	LOCAL STANDARD TIME (MBE/RMSE)							
		6:00	8:00	10:00	12:00	14:00	16:00	18:00	20:00
T_a (°C)	1 h	0.50/2.10	1.32/2.56	0.52/2.03	−0.04/1.85	−0.52/1.94	−0.86/2.17	−1.21/2.45	−1.61/**2.97**
	6 h	0.55/2.14	1.36/2.56	0.46/2.04	−0.13/1.89	−0.64/2.02	−1.00/2.27	−1.37/2.56	−1.80/**3.04**
	12 h	0.56/2.15	1.37/2.58	0.45/2.09	−0.17/1.99	−0.68/2.10	−1.01/2.30	−1.40/2.59	−1.80/**3.06**
	18 h	0.54/2.17	1.37/2.59	0.43/2.10	−0.17/1.99	−0.70/2.14	−1.05/2.40	−1.43/2.68	−1.83/**3.11**
	24 h	0.54/2.17	1.36/2.60	0.43/2.14	−0.23/2.04	−0.72/2.18	−1.05/2.41	−1.43/2.69	−1.84/**3.13**
RH (%)	1 h	0.81/10.05	−3.49/13.22	−0.98/10.69	1.05/9.51	2.41/9.86	3.31/10.62	4.19/11.58	**5.42/14.31**
	6 h	0.86/10.46	−3.46/13.37	−0.82/10.80	1.27/9.66	2.73/10.15	3.91/11.15	4.90/12.15	**6.34/14.93**
	12 h	0.86/10.56	−3.48/13.42	−0.59/11.11	1.52/10.13	2.91/10.13	3.88/11.18	4.87/12.20	**6.30/14.99**
	18 h	1.12/10.72	−3.51/13.53	−0.57/11.12	1.52/10.16	2.97/10.16	4.05/11.57	5.03/12.64	**6.43/15.30**
	24 h	1.10/10.74	−3.48/13.58	−0.50/11.39	1.61/10.37	3.02/10.37	4.01/11.60	5.02/12.64	**6.43/15.35**
V_a (m/s)	1 h	−0.29/2.22	0.02/2.23	0.26/2.38	0.20/2.55	0.14/2.78	−0.02/**2.87**	−0.43/2.81	−0.91/2.59
	6 h	−0.43/2.24	−0.04/2.25	0.19/2.38	0.13/2.56	0.07/2.81	−0.10/**2.91**	−0.56/2.88	−1.07/2.65
	12 h	−0.44/2.27	−0.05/2.27	0.16/2.40	0.07/2.56	0.01/2.82	−0.13/**2.91**	−0.57/2.89	−1.11/2.68
	18 h	−0.48/2.29	−0.09/2.28	0.14/2.41	0.06/2.58	0.01/2.84	−0.20/**2.94**	−0.62/2.91	−1.17/2.70
	24 h	−0.50/2.30	−0.13/2.28	0.12/2.45	0.07/2.61	0.03/2.86	−0.20/**2.95**	−0.62/2.92	−1.20/2.72
GHI (W/m²)	1 h	34.82/89.08	189.33/247.50	210.36/291.94	152.58/283.93	224.51/326.81	**328.94/384.69**	188.58/236.66	35.22/63.67
	6 h	34.08/88.59	188.21/246.90	208.79/291.55	149.80/283.86	222.49/327.02	**327.73/384.05**	188.59/236.74	35.21/63.73
	12 h	34.84/89.20	188.48/247.34	206.73/291.61	145.16/283.62	218.57/325.61	**327.93/384.21**	188.42/236.84	35.62/64.22
	18 h	34.57/89.21	188.14/247.32	207.00/292.55	146.36/285.52	217.54/325.98	**325.77/383.03**	187.73/236.18	35.41/64.03
	24 h	34.28/89.18	187.79/247.49	205.11/292.41	147.66/288.00	217.75/326.89	**326.23/383.64**	188.04/236.47	35.32/63.77
WBGT (°C)	1 h	0.86/2.44	**1.72/2.96**	0.83/2.23	0.15/2.01	0.15/1.99	0.49/2.03	0.01/1.91	−0.72/2.24
	6 h	0.92/2.47	**1.75/2.97**	0.80/2.22	0.10/2.03	0.08/2.01	0.43/2.03	−0.05/1.92	−0.81/2.20
	12 h	0.93/2.48	**1.77/2.99**	0.80/2.26	0.09/2.08	0.06/2.04	0.42/2.03	−0.08/1.93	−0.82/2.22
	18 h	0.93/2.49	**1.77/3.00**	0.80/2.26	0.09/2.08	0.05/2.05	0.40/2.07	−0.09/1.97	−0.83/2.25
	24 h	0.93/2.49	**1.77/3.00**	0.78/2.29	0.04/2.11	0.03/2.08	0.40/2.08	−0.10/1.97	−0.84/2.26

The highest MBE and RMSE values for each variable and horizon are in bold, while the lowest values are underlined.

3.13°C, respectively. For *RH*, these metrics range from −0.50% to 6.43% and from 9.51% to 15.35%, respectively; for V_a, from 0.01 to −1.20 m/s and from 2.22 to 2.95 m/s. T_a and *RH* show the most pronounced errors at 20:00 and the highest accuracy at around noon, whereas V_a shows the best performance at 6:00 and 8:00, deteriorating between 16:00 and 20:00. *GHI* discrepancies are notably higher, with MBE and RMSE ranging from 34.08 to 325.77 W/m² and from 63.67 to 384.69 W/m², respectively. Overall, the MBE and RMSE for WBGT range from 0.01°C to 1.77°C and from 1.91°C to 3.00°C, respectively, with the most significant inaccuracies observed at 8:00 and the smallest between 12:00 and 18:00.

Figure 5 illustrates how these biases in NWS WBGT forecasts vary across Köppen climates. While MBE remains around ±2.0°C and RMSE remains within 3.0°C in most Köppen climates, the climates "Cfb," "Dsa," and "Dsb" exhibit larger errors with a decreasing trend, peaking at 6:00, reaching their minimum around 12:00, and continuing to decrease until 20:00.
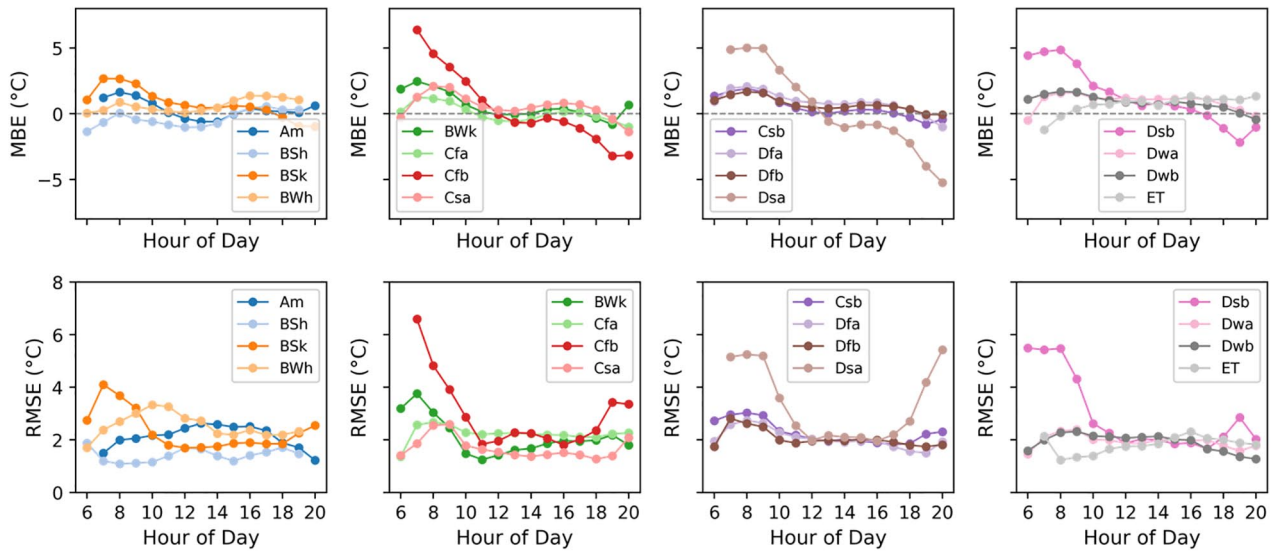
Figure 5. Hourly variations in MBE and RMSE for NWS WBGT forecasts across Köppen climates during local standard work hours (06:00-20:00).

This feature is more evident in RMSE trends, where "Cfb" and "Dsa" display an inverted bell curve, with the best performance around 12:00. In contrast, "Dsb" experiences the largest errors at 6:00, which subsequently decreases but still shows relatively larger errors compared to other climates. Next, this study explores how the identified biases of NWS WBGT forecasts across Köppen climates misrepresent occupational heat risks and identify underlying weather factors that necessitate further improvements in their accuracies.

Performance in detecting occupational heat risk levels

Figure 6 compares observed and forecasted WBGT risk levels across Köppen climates using normalized confusion matrices. Except for the "Am" climate, the NWS WBGT forecasts demonstrate high accuracy in detecting heat risks at Level 0 (below 25.6°C), with success rates of 86.4% for "BWh," and above 90% for other climates. However, for heat risk levels beyond Level 0, performances are generally below 50%. In "BSh," "BWh," and "Cfa" climates, which are vulnerable to work-related heat risks evidenced in Figure 4, the accuracy for predicting higher heat risk levels (equal to or exceeding Level 1) is not as reliable. Forecasts in "BWh," and "Cfa" tend to underestimate risks in these climates, whereas "BSh" forecasts are more likely to overestimate.

We further investigate the causes of certain low performance by analyzing the impact of discrepancies between in-situ observations and NWS weather forecasts (i.e. T_{a_diff}, RH_{diff}, V_{a_diff}, and GHI_{diff}) on the $WBGT_{diff}$ (i.e. the difference between observed and forecasted WBGT values). These analyses utilize beta coefficients derived from regression analyses, with results presented in Figure 7. The symbol "*" denotes statistical significance with a P-value below .05, and the values under each x-axis label in parentheses represent the variance inflation

factor (VIF) values. The R^2 values, ranging from 0.88 to 0.96, show strong linearity in explaining variances in WBGT forecast bias, and VIFs, not exceeding 3.6, suggest minimal multicollinearity among variables.[46] The analysis identifies T_{a_diff} as the most influential factor, with the second most influential factor alternating between RH_{diff} and GHI_{diff}, whereas V_{a_diff} has the least influence. Notably, in climates "Cfb," "Dsa," "Dsb," T_{a_diff} significantly influences $WBGT_{diff}$.

Discussion

This study highlights the importance of comprehensive assessments in occupational heat risk forecasts, specifically focusing on the performance of NWS WBGT forecasts with respect to local time, forecast horizons, and the ability to alarm occupational heat risks across Köppen climates. Utilizing in-situ observations from 252 weather stations and historical NWS WBGT forecasts from the summer of 2023 in the US, the study underscores the necessity for increased vigilance toward occupational heat threats in certain Köppen climates in the US, such as "Am," "BSh," "BWh," "Cfa," and "Csa," where heat risk levels during working hours often exceed Level 0. These climates demand more regular monitoring due to frequent fluctuations in WBGT heat risk levels from Level 1 to Level 5, which can lead to misconceptions about actual risk levels without adequate monitoring.

The study reveals a stronger association between NWS WBGT forecast errors and time of day rather than forecast horizons. Prior research found that the RMSE for NWS WBGT forecasts in North Carolina was 1.3°C from May to September during 2019 to 2021,[19] and another study identified that NWS WBGT forecasts' RMSE varied from –0.64°C to 1.46°C from April to October during 2018 to 2019 in the US.[18] These errors likely stem from underlying inaccuracies in the component weather forecasts, highlighting the limitations of relying on a single metric for forecast
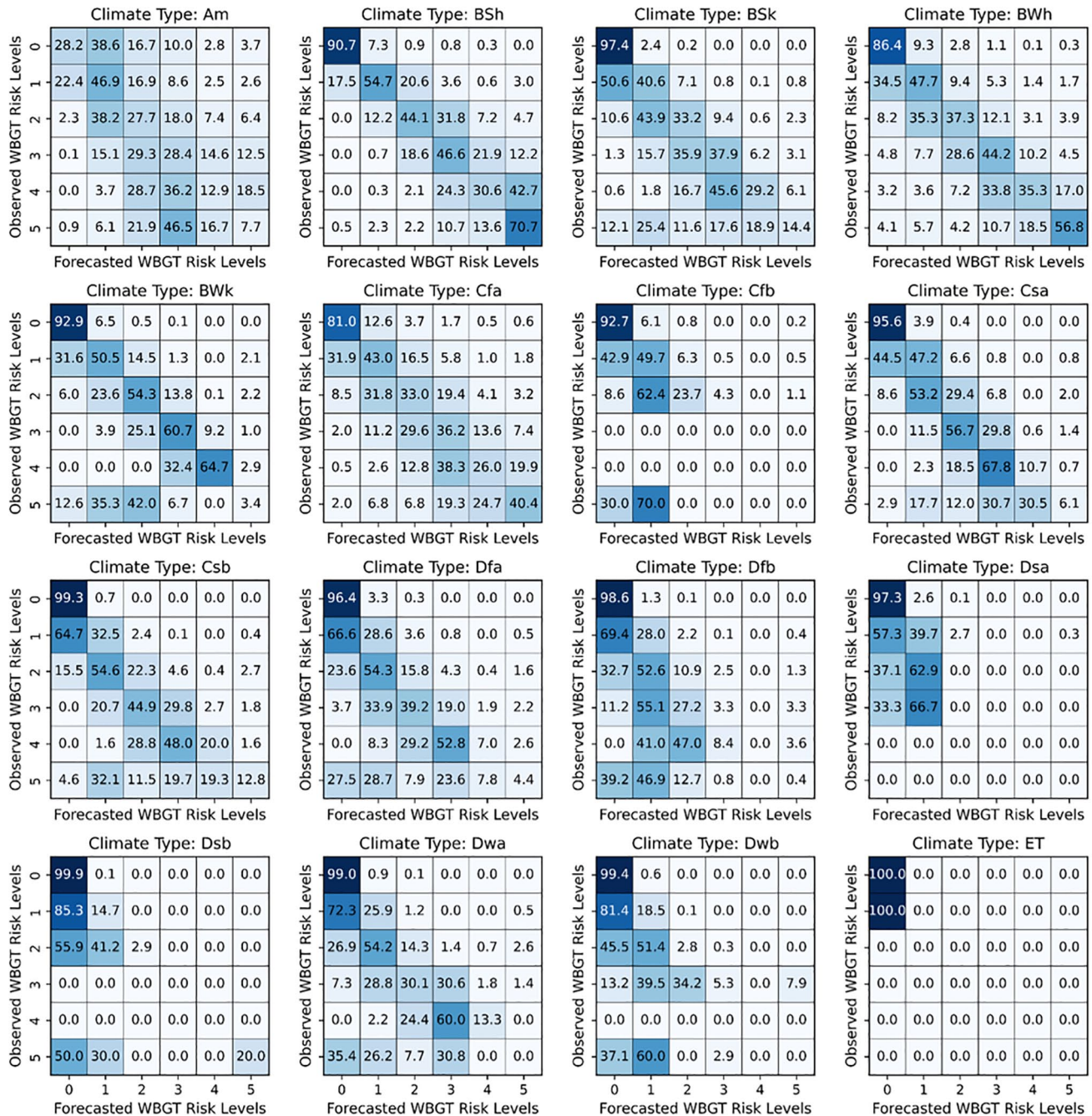
Figure 6. Comparison of observed and forecast WBGT risk levels across Köppen climates using normalized confusion matrices.

evaluation. The analysis in this study further reveals its temporal trends, with specific times of the day and forecast horizons showing increased accuracy or discrepancies for different component weather forecasts, as shown in Table 3. In most Köppen climates, MBE is approximately ±2.0°C and RMSE remains within 3.0°C, as represented in Figure 5. However, the climates "Cfb," "Dsa," and "Dsb" exhibit relatively larger errors, peaking at 6:00, reaching a minimum around 12:00, and continuing to decrease until 20:00. While such notable errors are observed in these climates, it is also noteworthy that the observed time frames with greater errors, such as 6:00 and after 16:00, are most likely

to be under Level 0 heat risks (i.e. no occupational heat risk), as evidenced in Figure 3.

The study also evaluates the effectiveness of NWS WBGT forecasts in identifying occupational heat risk levels. Results reveal consistently low performance in detecting WBGT heat risk levels (equal to or exceeding Level 1) across Köppen climates, with forecast performance correlating more strongly with time of day than with forecast horizons within a 1-day period, as shown in Figure 6. While NWS WBGT forecasts effectively identify the lowest heat risks (i.e. no heat risk), their performance declines at higher risk levels, often below 50%, highlighting the need for cautious interpretation of NWS
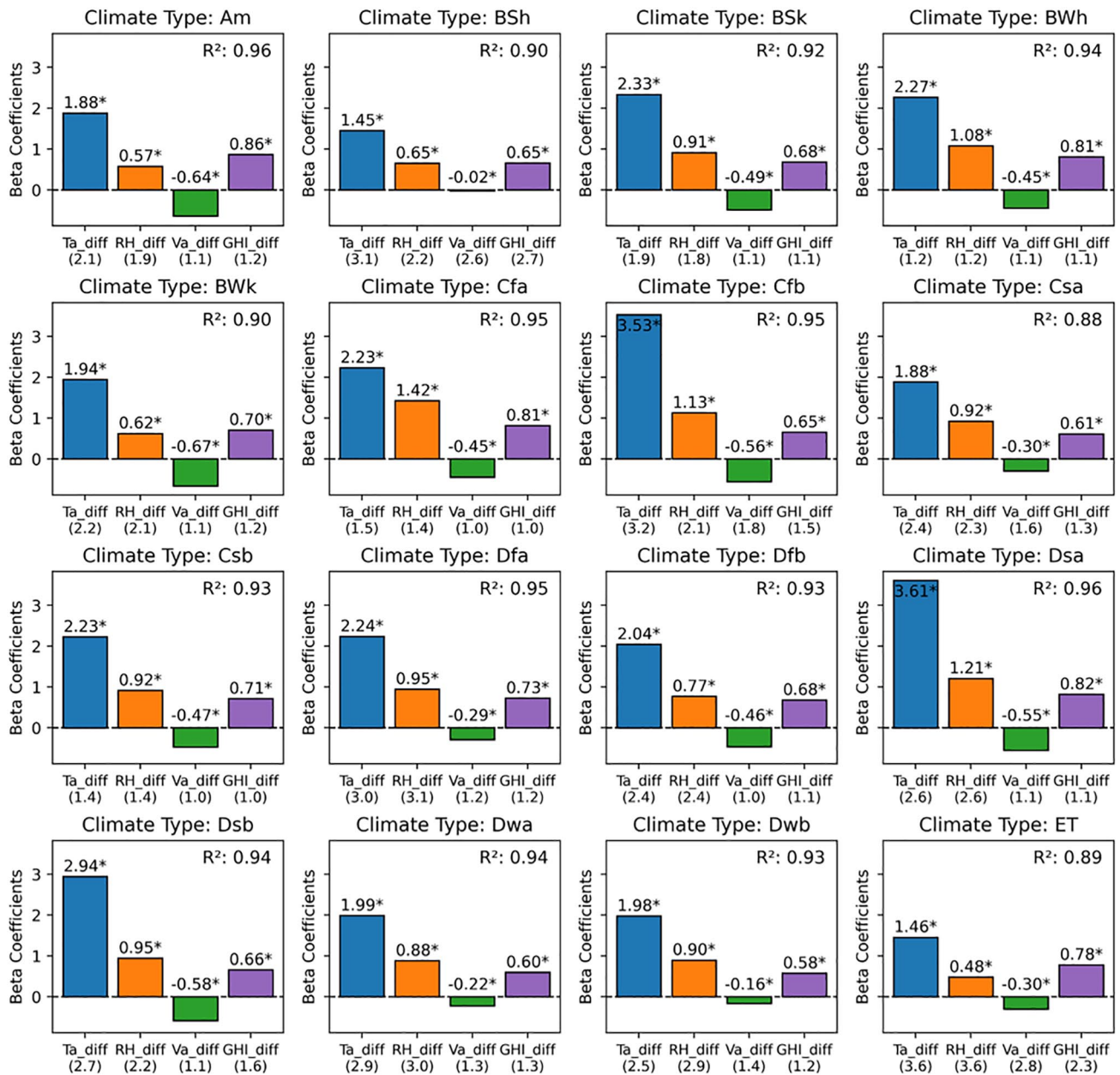
Figure 7. Beta coefficients and VIFs of T_{a_diff}, RH_{diff}, V_{a_diff}, and GHI_{diff} for $WBGT_{diff}$ across Köppen climates.

WBGT forecasts in safety management. For instance, adopting a more conservative interpretation (e.g. include the one level below) of forecasts in "BWh" and "Cfa" climates could be recommended to improve accuracy, potentially exceeding 60% in all instances. The "Am" climate demonstrates lower overall forecast accuracy, indicating a need for more careful forecast interpretation. While forecasts are reliable for detecting Level 0 heat risks, a careful approach is advised for higher heat risk levels, underlining the importance of considering regional climate characteristics in interpreting NWS WBGT forecasts. Without further adjustments, such as post-processing[47] or a conservative interpretation strategy, the NWS WBGT forecasts may not provide a reliable basis for evaluating occupational heat risk levels for safety management. This also

highlights the need for more careful attention to its performance during heat-sensitive periods (e.g. Level 1 to Level 5), where a few degrees of error can affect the implementation of appropriate safety standards, as exemplified in Table 2.

Along with the quantitative results, which can be directly used for practical decision support regarding NWS WBGT forecasts, this study offers several practical implications. For instance, it identifies empirical causes behind the low performance by examining the impact of discrepancies between in-situ observations and NWS weather forecasts, as shown in Figure 7. In all climates, T_{a_diff} significantly influences $WBGT_{diff}$, which is closely related to the associated weather biases and underlying NWS WBGT calculation mechanism, which assigns greater weights to T_a. Another notable finding is that

the second most influential variable is either $GHI_{\_diff}$ or $RH_{\_diff}$, depending on Köppen climates. For practical scenarios, these insights provide climate-specific recommendations for minimizing relevant weather forecast biases and thus effectively improving the reliability of NWS WBGT forecasts, such as using post-processing techniques widely employed to improve weather forecasts.[47] Furthermore, since operational weather forecasts are a common feature in most countries,[48] the methodologies developed in this study for evaluating public weather forecast-driven occupational heat risk forecasts can be readily adapted and applied internationally. For instance, the evaluation framework used in this study could be employed to assess the performance of heat risk forecasts provided by the HEAT-SHIELD project,[14] which encompasses the entire European region. Extending this evaluation approach to different operational forecast systems around the world would yield new insights into the effectiveness of these systems in managing occupational safety. This extended application could help identify specific areas where forecast accuracy needs improvement and develop tailored strategies to mitigate occupational heat risks in diverse climatic conditions.

Despite its contributions, this study has several limitations that open avenues for future research. First, the study primarily uses the NWS's WBGT calculation mechanism for direct comparisons between NWS WBGT forecasts and in-situ observation-based WBGT. Therefore, this analysis does not address the reliability of the NWS's WBGT calculation validated by direct measurements using additional weather sensors, such as a black globe thermometer. Future research also could explore other WBGT estimation methods, such as those proposed by Lemke and Kjellstrom,[49] Bernard,[50] and Liljegren et al[51] to evaluate their accuracy. Second, the ability to identify occupational heat risk levels is closely linked to the standards of occupational heat levels, which may incorporate different risk thresholds. Future studies could benefit from sensitivity analyses exploring different safety guidelines, providing practitioners with practical insights for interpreting NWS WBGT forecasts under various guidelines. For example, OSHA provides recommendations on exposure limits for light work activity at 30°C, moderate work activity at 27.8°C, and heavy work activity at 26.1°C.[6] Revisiting the method of this study to evaluate the performance of weather forecasts under different guidelines would provide more tailored insights for their applications. Lastly, the data used in this study is based on observations from distributed weather stations for the hours of 6:00 AM to 8:00 PM during the summer months of 2023 in the US, resulting in small samples in certain climate types (e.g. BSh, Dsa). Additional data would help address this limitation, allowing for a more comprehensive analysis that incorporates other yearly trends and extends the observation period beyond the current timeframe. Furthermore, the effectiveness of heat risk forecasts is primarily focused on outdoor work environments. Different work environments (e.g. indoors) may need to consider other environment-related conditions (e.g. indoor air

temperature), requiring additional types of data for their applications in safety management.

Conclusion

To properly mitigate the forthcoming heat risks in occupational settings, predictive information on environmental heat is fundamental for developing effective measures to protect the health and safety of workers. Leveraging the benefits of public weather forecasting has the potential to address this need. In this context, this study evaluates the performance of NWS WBGT forecasts. The findings identify areas where heightened caution and improvements could enhance this understanding for proactive occupational safety management. Furthermore, this study discusses its practical implications and extensive research benefits. Overall, this research contributes to enhancing the role of operational heat risk forecasting services, ensuring better management of heat exposures for occupational workers.

Author Contributions

YK conceptualized and designed the study, collected and analyzed the data, and wrote the manuscript. YH provided advice, supervised the study, and reviewed and revised the manuscript.

ORCID iD

Yoojun Kim https://orcid.org/0000-0002-0861-0132

Availability of Data and Materials

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

1. Varghese BM, Hansen A, Bi P, Pisaniello D. Are workers at risk of occupational injuries due to heat exposure? A comprehensive literature review. *Saf Sci.* 2018;110:380-392.
2. Bröde P, Fiala D, Błażejczyk K, et al. Deriving the operational procedure for the Universal Thermal Climate Index (UTCI). *Int J Biometeorol.* 2012;56:481-494.
3. Höppe P. The physiological equivalent temperature: a universal index for the biometeorological assessment of the thermal environment. *Int J Biometeorol.* 1999;43:71-75.
4. Grifoni D, Messeri A, Crisci A, et al. Performances of limited area models for the WORKLIMATE heat-health warning system to protect worker's health and productivity in Italy. *IJERPH.* 2021;18:9940.
5. Fatima SH, Rothmore P, Giles LC, Bi P. Impacts of hot climatic conditions on work, health, and safety in Australia: a case study of policies in practice in the construction industry. *Saf Sci.* 2023;165:106197.

6. Jacklitsch B, Williams W, Musolin K, et al. *Occupational Exposure to Heat and Hot Environments*. Department of Health and Human Services, Centers for Disease Control and Prevention. Published online 2016. https://stacks.cdc.gov/view/cdc/37911

7. ACGIH, Threshold Limit Values (TLVs) and Biological Exposure Indices (BEIs). American Conference of Governmental Industrial Hygienists; 2021:2021.

8. ISO. *ISO 7243:2017 Ergonomics of the Thermal Environment Assessment of Heat Stress Using the WBGT (Wet Bulb Globe Temperature) Index*. 3rd ed. International Organization for Standardization (ISO); 2017.

9. Ghalhari GF, Heidari H, Dehghan SF, Asghari M. Consistency assessment between summer simmer index and other heat stress indices (WBGT and humidex) in Iran's climates. *Urban Clim*. 2022;43:101178.

10. Kim Y, Ham Y. Revealing the impact of heat radiation on construction: a microclimate simulation using meteorological data and geometric modeling. *J Constr Eng Manag*. 2024;150:04024016.

11. Alessi MJ, DeGaetano AT. A comparison of statistical and dynamical downscaling methods for short-term weather forecasts in the US Northeast. *Meteorol Appl*. 2021;28:1976.

12. Leroyer S, Bélair S, Spacek L, Gultepe I. Modelling of radiation-based thermal stress indicators for urban numerical weather prediction. *Urban Clim*. 2018;25:64-81.

13. Kingma BRM, Steenhoff H, Toftum J, et al. ClimApp-integrating personal factors with weather forecasts for individualised warning and guidance on thermal stress. *IJERPH*. 2021;18:11317.

14. Morabito M, Messeri A, Noti P, et al. An occupational heat-health warning system for Europe: the HEAT-SHIELD platform. *IJERPH*. 2019;16:2890.

15. Vitart F, Balsamo G, Buizza R, et al. Sub-seasonal predictions; 2014. Accessed June 19, 2024. https://www.ecmwf.int/sites/default/files/elibrary/2014/12943-sub-seasonal-predictions.pdf

16. NOAA. SCN22-43: wet bulb globe temperature grids in the national digital forecast database will transition to operational status on June 1, 2022. Published online 2022. Accessed March 18, 2024. https://vlab.noaa.gov/web/mdl/ndfd-technical-notices

17. NWS. National Weather Service (NWS) instruction 10-201: National Digital Forecast Database (NDFD) and local database description and specifications. Published online November 21, 2023. http://www.nws.noaa.gov/directives/

18. Ahn Y, Uejio CK, Rennie J, Schmit L. Verifying experimental wet bulb globe temperature hindcasts across the United States. *GeoHealth*. 2022;6:e2021GH000527.

19. Clark J, Konrad CE, Grundstein A. The development and accuracy assessment of wet bulb globe temperature forecasts. *Weather Forecast*. 2024;39:403-419.

20. Boyer T. NDFD wet bulb globe temperature algorithm and software design. Published online 2022. Accessed February 6, 2024. https://vlab.noaa.gov/documents/6609493/0/NDFD+WBGT+Description+Document.pdf/158bd409-4433-4dc0-295a-6deba5c6754d?t=1704311112004

21. NWS. National digital forecast database and local database description and specifications (NWSI 10-201). Published 2019. Accessed February 6, 2024. https://www.weather.gov/media/mdl/ndfd/pd01002001curr.pdf

22. Page WG, Wagenbrenner NS, Butler BW, Forthofer JM, Gibson C. An evaluation of NDFD weather forecasts for wildland fire behavior prediction. *Weather Forecast*. 2018;33:301-315.

23. NOAA. National digital forecast database (NDFD). Published 2023. Accessed February 6, 2024. https://vlab.noaa.gov/web/mdl/ndfd-grid-data

24. NOAA. High-resolution rapid refresh (HRRR). Published 2023. Accessed February 6, 2024. https://vlab.noaa.gov/web/mdl/nbm

25. NOAA. Visible infrared imaging radiometer suite (VIIRS) level 3 daily gridded land surface albedo. Published 2023. Accessed February 6, 2024. https://www.avl.class.noaa.gov/saa/products/search?sub_id=0&datatype_family=JPSS_NGRN&submit.x=26&submit.y=6

26. United States Geological Survey (USGS). Moderate resolution imaging spectroradiometer (MODIS) land cover type product (MCD12Q1). Published 2022. Accessed February 6, 2024. https://lpdaac.usgs.gov/products/mcd12q1v061/

27. High Plains Regional Climatic Center. Automated weather data network (AWDN). Published 2023. Accessed February 6, 2024. https://hprcc.unl.edu/awdn/access/index.php

28. US Bureau of Reclamation. Agriculture and meteorology (AgriMet). Published 2023. Accessed February 6, 2024. https://www.usbr.gov/pn/agrimet/agrimet-map/agrimap.html

29. California Department of Water Resources. California irrigation management information system (CIMIS). Published 2023. Accessed February 6, 2024. https://et.water.ca.gov/Rest/Index

30. US Department of Agriculture. Soil climate analysis network (SCAN). Published 2023. Accessed February 6, 2024. https://www.nrcs.usda.gov/resources/data-and-reports/soil-climate-analysis-network

31. WeatherSTEM. WeatherSTEM (STEM). Published 2023. Accessed February 6, 2024. https://www.weatherstem.com/

32. Chen D, Chen HW. Using the Köppen classification to quantify climate variation and change: an example for 1901–2010. *Environ Dev*. 2013;6:69-79.

33. Kakaei H, Omidi F, Ghasemi R, Sabet MR, Golbabaei F. Changes of WBGT as a heat stress index over the time: a systematic review and meta-analysis. *Urban Clim*. 2019;27:284-292.

34. Dimiceli VE, Piltz SF, Amburn SA. Black globe temperature estimate for the WBGT index. In: Kim HK, Ao SI, Rieger BB eds. *IAENG Transactions on Engineering Technologies*. Vol. 170. Lecture Notes in Electrical Engineering. Springer Netherlands; 2013:323-334.

35. Stull R. Wet-Bulb temperature from relative humidity and air temperature. *J Appl Meteorol Climatol*. 2011;50:2267-2269.

36. Guo M, Zang H, Gao S, et al. Optimal tilt angle and orientation of photovoltaic modules using HS algorithm in different climates of China. *Appl Sci*. 2017;7:1028.

37. Dimiceli V, Piltz S, Amburn S. Estimation of black globe temperature for calculation of the wet bulb globe temperature index. Proceedings of the World Congress on Engineering and Computer Science 2011. Vol 2.; 2011.

38. The National Renewable Energy Laboratory (NREL). NREL's solar position and intensity (SOLPOS 2.0). Published 2001. Accessed March 1, 2024. https://www.nrel.gov/grid/solar-resource/solpos.html

39. Hunter C, Minyard C. Estimating wet bulb globe temperature using standard meteorological measurements. 1999. https://www.osti.gov/servlets/purl/14888-ZH0AwM/native/

40. Lawrence MG. The relationship between relative humidity and the dewpoint temperature in moist air: a simple conversion and Applications. *Bull Am Meteorol Soc*. 2005;86:225-234.

41. US Department of Defense (DOD). Technical manual NEHC-TM-OEM. 6260.6A, prevention and treatment of heat and cold stress injuries. Published online 2007. https://www.med.navy.mil/Navy-and-Marine-Corps-Force-Health-Protection-Command/Environmental-Health/Occupational-and-Environmental-Medicine/Occupational-and-Environmental-Medicine-Division/Technical-Manuals-and-Guidance/

42. National Weather Service. Wet bulb globe temperature. Accessed June 19, 2024. https://www.weather.gov/arx/wbgt

43. Willmott CJ. Some comments on the evaluation of Model Performance. *Bull Am Meteorol Soc*. 1982;63:1309-1313.<

44. Choi Y, Seo S, Hong T, Koo C. A classification model using personal biometric characteristics to identify individuals vulnerable to an extremely hot environment. *J Manag Eng*. 2024;40:04024001.

45. Luo H, Wang M, Wong PKY, Tang J, Cheng JCP. Construction machine pose prediction considering historical motions and activity attributes using gated recurrent unit (GRU). *Autom Constr*. 2021;121:103444.

46. Daoud JI. Multicollinearity and regression analysis. *J Phys Conf Ser*. 2017;949:012009.

47. Vannitsem S, Bremnes JB, Demaeyer J, et al. Statistical postprocessing for weather forecasts: review, challenges, and avenues in a big data world. *Bull Am Meteorol Soc*. 2021;102:E681-E699.

48. Olaniyan EA, Cafaro C, Ogungbenro SB, et al. Performance evaluation of a high-resolution regional model over West Africa for operational use: a case study of August 2017. *Meteorol Appl*. 2022;29:e2080.

49. Lemke B, Kjellstrom T. Calculating workplace WBGT from meteorological data: a tool for climate change assessment. *Ind Health*. 2012;50:267-278.

50. Bernard TE, Pourmoghani M. Prediction of workplace wet bulb global temperature. *Appl Occup Environ Hyg*. 1999;14:126-134.

51. Liljegren JC, Carhart RA, Lawday P, Tschopp S, Sharp R. Modeling the wet bulb globe temperature using standard meteorological measurements. *J Occup Environ Hyg*. 2008;5:645-655.