BAMS Article

A New Heat Stress Index for Climate Change Assessment

John R. Lanzante[a]

KEYWORDS:
Climate change;
Health;
Societal impacts

ABSTRACT: The heat index (HI), based on Steadman's model of thermoregulation, estimates heat stress on the human body from ambient temperature and humidity. It has been used widely both in applications, such as the issuance of heat advisories by the National Weather Service (NWS) and for research on possible changes in the future due to climate change. However, temperature/humidity combinations that exceed the applicable range of the model are becoming more common due to climate warming. Recent work by Lu and Romps has produced an extended heat index (EHI), which is valid for values outside the range of the original HI. For these values, the HI can underestimate the EHI by a considerable amount. This work utilizes observed data from 15 U.S. weather stations along with bias-adjusted output from a climate model to explore the spatial and temporal aspects of the disparity between the HI and the EHI from the recent past out to the end of the twenty-first century. The underestimate of human heat stress by the HI is found to be the largest for the most extreme cases (5°–10°C), which are also the most impactful. Conditions warranting NWS excessive heat warnings are found to increase dramatically from less than 5% of days historically at most stations to more than 90% in the future at some stations. Although, by design, the scope of this work is limited, it demonstrates the need for the adoption of the EHI for both applications and research.

SIGNIFICANCE STATEMENT: An index based on temperature and humidity was developed over 40 years ago to estimate human heat stress. It has been used to issue heat advisories as well as to study possible changes in heat stress due to climate change. However, since this index was not intended to handle higher temperatures that are becoming more common due to global warming, a new version was recently developed. This study uses past U.S. weather data, as well as future climate projections to study the differences between the original and updated indices. It demonstrates that during heat waves, the original version can underestimate heat stress by a considerable amount and urges the adoption of the new index.

DOI: 10.1175/BAMS-D-24-0030.1
Corresponding author: John Lanzante, John.Lanzante@noaa.gov
Manuscript received 30 January 2024, in final form 12 September 2024, accepted 7 October 2024

AFFILIATION: [a] NOAA/Geophysical Fluid Dynamics Laboratory, Princeton, New Jersey

1. Introduction

In recent years, reports of extreme heat throughout the world are becoming commonplace. The expert opinion of the Intergovernmental Panel on Climate Change (IPCC) has expressed virtual certainty that the intensity, duration, and frequency of heat waves have increased globally (Seneviratne et al. 2021) and that human activities are the primary cause (IPCC 2023). This has been accompanied by increased human mortality and morbidity (IPCC 2023). In the United States, heat has been the leading cause of weather fatalities, eclipsing floods, tornadoes, hurricanes, and lightning (www.weather.gov/hazstat). Furthermore, global warming will continue to increase in the near term and extremes will increase as well (IPCC 2023). There is, however, considerable uncertainty regarding the magnitude of the increases as these depend crucially on future human activities such as emissions of greenhouse gases.

Since the need to adapt to increased heat is inevitable, it is incumbent to consider the range of possible climate trajectories. Projections based on global climate models (GCMs) are the leading tool in such endeavors. To estimate the effect of heat on the human body, a number of different indices have been developed that are based on weather variables including air temperature, humidity, solar radiation, and wind (Baldwin et al. 2023). One of the most popular is the heat index employed operationally by the U.S. National Weather Service (NWS) to issue heat watches, warnings, and advisories. The basis for the heat index is a physiological model (Steadman 1979) based on human thermoregulation assuming optimal choice of clothing, full shade, and unlimited drinking water. Implementation of this model is somewhat complicated as it depends on a system of equations. The widely used NWS implementation [heat index (HI)] is based on a polynomial fit to a table of values produced from these equations (Rothfusz 1990) and depends only on air temperature and relative humidity.

Recent studies (Mukherjee et al. 2021; Dahl et al. 2019) have applied the HI to GCM projections to estimate possible changes in heat stress over the Continental United States (CONUS) during the latter part of the twenty-first century. These studies show an increase in heat stress emanating from the Gulf Coast into the Midwest and Atlantic coastal regions. However, there is a problem with the HI in that it was only intended to be valid over a particular range of temperature and humidity that would be exceeded rarely over Earth in the historical past (Steadman 1979; Lu and Romps 2022). However, the polynomial fit as defined can extrapolate beyond the intended bounds, resulting in unphysical results, the consequences of which are demonstrated by Romps and Lu (2022) and Dahl et al. (2019). As shown by the latter authors, these "out-of-bounds" conditions become increasingly more common in the latter part of the twenty-first century.

Recently, Lu and Romps (2022) have modified the Steadman model, in the form of an extended heat index (EHI) to accommodate the out-of-bounds conditions. The HI model becomes undefined/unphysical when the temperature falls either below or above certain limits considered by Steadman. For lower temperatures, Steadman assumed a fixed reference vapor pressure, which could exceed the saturation pressure. In contrast, for these conditions, the EHI model sets the reference vapor pressure to the heat index's saturation vapor pressure. For even lower temperatures, adjustments are made so that the heat index is very nearly equal to the air temperature, regardless of the relative humidity. For higher temperatures,

the HI becomes undefined because Steadman equates the skin's evaporation rate to the sweating rate, which causes the skin's vapor pressure to exceed its saturation value. The EHI remedy for this is to allow sweat to drip off the skin. The details of the modifications, which are actually much more complicated than briefly summarized above, can be found in Lu and Romps (2022). The EHI has been shown to correspond well with laboratory heat-stress experiments on human subjects (Lu and Romps 2023).

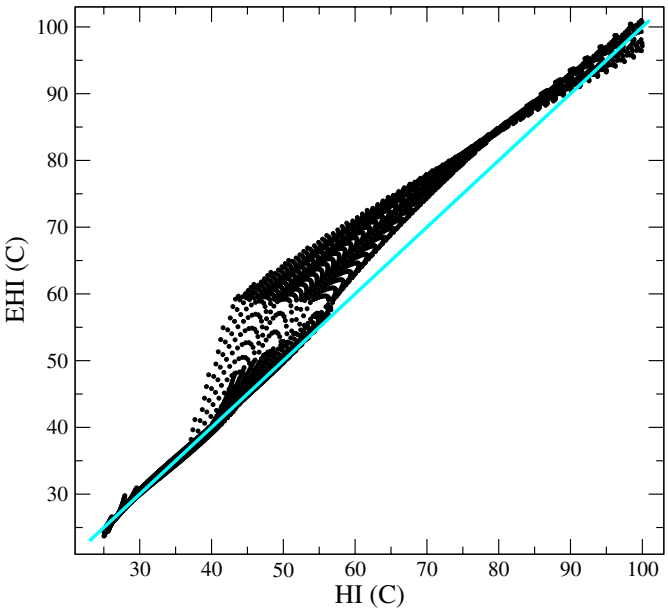

FIG. 1. Scatterplot of HI (°C) (abscissa) vs EHI (°C) (ordinate) for HI ranging from 25° to 100°C produced by all possible combinations of temperature in 0.5-K increments and relative humidity from 0% to 100 in 1% increments. The black points are the (HI, EHI) pairs, while the *Y* = *X* line is cyan.

Since the EHI is valid for all combinations of temperature and humidity, this results in values that differ, sometimes substantially, from those based on the HI via polynomial extrapolation. This can be seen in Fig. 1, which, for illustrative purposes, presents a scatterplot of HI versus EHI over all possible combinations of temperature and humidity, which produce HI from 25° to 100°C. Although the EHI can be less than the HI, over the range of values realizable over the United States during the twenty-first century (as seen later <80°C), the EHI is primarily greater than the HI. The effect of humidity is somewhat more complex than depicted here as this plot presents the range of possible differences. In a given climate regime, only a subset of these points is realizable.

The primary purpose of this work is to make the point that the EHI should be used rather than the HI when considering possible effects of twenty-first century climate change on heat stress over the United States. This is illustrated by a comparison of these two indices at select locations over the United States using both observed data as well as GCM output. These analyses show that the disparity between the HI and the EHI has large geographic and temporal variations and that the use of the HI could potentially grossly underestimate increases in heat stress in some locations for some realizations of future climate. In addition, it is shown that there exists the potential for a dramatic increase in both the frequency and magnitude of heat stress at the end of the twenty-first century at some locations.

2. Data and methods

The observed data used here were obtained from the National Centers for Environmental Information's (NCEI's) Global Surface Summary of the Day (GSOD) database (https://www1.ncdc.noaa.gov/pub/data/gsod/). Time series of daily maximum temperature (tmax) and daily average dewpoint (avgTd) were extracted for 15 U.S. stations (13 CONUS plus Hawaii and the Caribbean). The stations (see Table 1) were selected in order to minimize the number of missing observations while covering different geographic regions/climate zones. Note that throughout this paper, for brevity, the stations are referenced by their two-letter state abbreviation (Table 1).

Only daily summer (i.e., June–August) values for the historical period of 1985–2014 were utilized. The avgTd and tmax were combined to derive an estimated relative humidity (rh), which was then used with tmax to compute values of HI and EHI, representing estimates of

Table 1. Station names, with state abbreviations, latitude, and longitude. Station identifiers are also given as the World Meteorological Organization (WMO) station identifier and International Civil Aeronautical Organization (ICAO) location indicator.

State	Name	ICAO	WMO	Latitude	Longitude
AZ	Phoenix-Mesa Airport	KIWA	722786	33.30	111.67
CA	Long Beach Airport	KLGB	722970	33.81	118.15
CU	Guantanamo Bay Naval Base	MUGM	783670	19.90	75.15
FL	Patrick Air Force Base	KCOF	747950	28.23	80.60
HI	Honolulu Airport	PHNL	911820	21.32	157.94
IL	Chicago O'Hare Airport	KORD	725300	41.96	87.93
LA	New Orleans Airport	KMSY	722310	30.00	90.28
MA	Boston Logan Airport	KBOS	725090	42.31	71.01
MD	Andrews Air Force Base	KADW	745940	38.82	76.87
NC	Raleigh-Durham Airport	KRDU	723060	35.89	78.78
NY	Buffalo Niagara Airport	KBUF	725280	42.94	78.74
OK	Tulsa Airport	KTUL	723560	36.20	95.88
SD	Ellsworth Air Force Base	KRCA	726625	44.15	103.10
UT	Salt Lake City Airport	KSLC	725720	40.77	111.97
WA	Seattle-Tacoma Airport	KSEA	727930	47.45	122.31

the maximum daily values of these heat indices. The computation of the HI has been performed via application of the Rothfusz (1990) formula, and the EHI has been computed via code made available by Romps (https://www.romps.berkeley.edu/papers/pubdata/2020/heatindex/heatindex.f90).

The GCM data are taken from the Geophysical Fluid Dynamics Laboratory's (GFDL's) GFDL-CM4 (Held et al. 2019) and CMIP6 (Eyring et al. 2016) historical run (r1i1p1f1), with a nominal resolution of 100 km, as well as the corresponding SSP585 future run (r1i1p1f1). The land model grid point nearest the GSOD station was used. Summer values of daily maximum temperature (tmax) and daily minimum relative humidity (rhmin) were extracted for three time periods: 1985–2014 (historical), 2031–60 (near future), and 2071–2100 (far future). The values of tmax and rhmin were used to compute values of the HI and EHI as indicated above.

The observed and raw GCM data are not directly comparable for several reasons: 1) The GCM may have biases in temperature and humidity compared to the observations; 2) the nearest GCM grid point does not correspond exactly to the weather station location; 3) the GCM value actually represents some footprint around its grid point; thus, this may introduce complications especially in coastal areas and in areas of complex orography (Dixon et al. 2016; Lanzante et al. 2018); and 4) the HI/EHI is calculated differently for the observations (tmax, avgTd) than the GCM (tmax, rhmin) due to limitations in the values that have been archived.

To deal with these incongruities, the GCM data have been adjusted using a technique, multivariate bias correction (MBC; Cannon 2018), which is a multivariate extension of quantile delta mapping (QDM; Cannon et al. 2015) (https://github.com/cran/MBC/blob/master/R/MBC-QDM.R), commonly used in statistical downscaling (SD) and bias adjustment. Although there are several distinct categories of SD, the distributional methods, of which QDM is one, are especially appropriate here since the interest is in the distribution of values as well as the extremes. QDM has been found to be as good or better than any of a set of common distributional approaches (Lanzante et al. 2018, 2019). MBC extends QDM by incorporating the dependence between variables. By application of MBC, which is trained on the historical observations, we expect to minimize incongruities 1–4. As such, all results reported here are based on the SD of the GCM output via MBC, rather than the raw GCM output. The interested reader is referred to the appendix for more details on MBC and its application here.

3. Results

To begin the exploration of the discrepancy between heat indices, scatterplots of the HI versus the EHI have been constructed for all 15 stations. Three of the more interesting cases are shown here: Fig. 2 (LA, FL, and CU). In these plots, Figs. 2a–c show the results for the historical period for MBC applied to the GCM, while Figs. 2d–f and g–i are for MBC in the near future and far future, respectively. Note that plots for GSOD (not shown) are, by construction, nearly identical to those for MBC in the historical period. That both are very similar is expected as MBC adjusts the GCM values to match the distribution of observed values during the historical period. It then applies the same mapping to adjust the future values.

In a given panel, each star represents a single day. The distance above the $Y = X$ line represents the amount by which the HI underestimates the EHI. That the majority of the points lie close to the line indicates a close correspondence between the HI and EHI. However, at the highest levels of heat stress, large departures are indicated by the red points, which represent differences between the HI and EHI that exceed 6 standard deviations. Note that the bi-weight standard deviation (Lanzante 1996) was used in lieu of the traditional standard deviation. The reason is that the bi-weight, unlike the traditional metric, is not much influenced by the outlying values and instead is characteristic of the bulk of the values that lie close to the line.

Examining Fig. 2 for LA, it is seen that during the historical period, there are only three notable discrepancies at the high end of the values. While in the near future, there is a moderate increase in extreme differences; in the far future, there is a huge increase, and most of the differences exceed 5°C. Similar behavior is seen for FL, although somewhat more large differences are found in the historical period. The results for CU are qualitatively similar but with fewer exceedances in the historical period.

A rendering of the geographical pattern of the extreme differences (red points in Fig. 2) is shown in Fig. 3, which displays bars whose height is proportional to the percentage of time that extreme positive differences between the EHI and HI occur. The four bars, from left to right, correspond to GSOD historical, MBC historical, MBC near future, and MBC far future, respectively. For all stations, the extreme differences are rare in the historical period. For a few of the stations, extreme differences start to become noticeable in the near future and for about a third extreme differences become very frequent in the far future.

It is not surprising that the spatial variations shown in Fig. 3 are similar to those found by Romps and Lu (2022) for large historical departures between the EHI

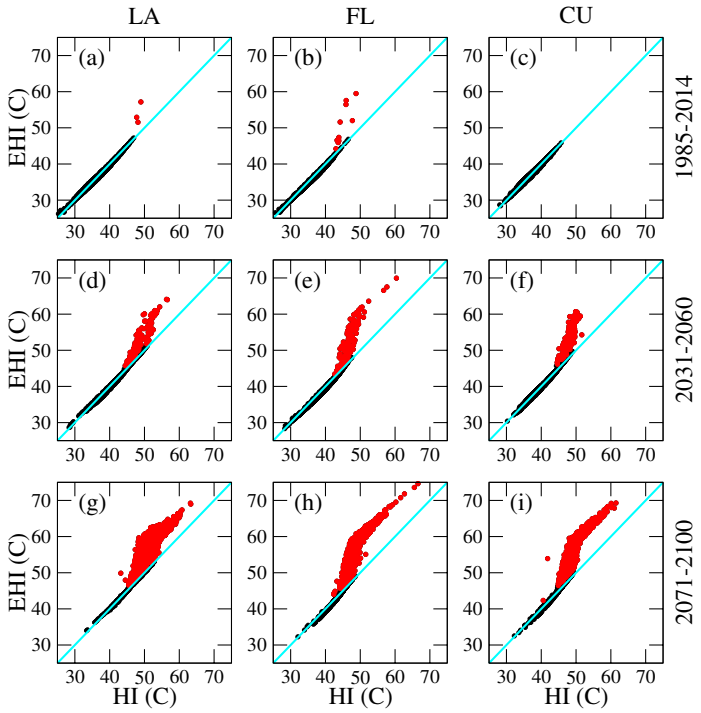

Fig. 2. Scatterplots of summertime daily maximum HI (°C) (abscissa) vs EHI (°C) (ordinate) for (a),(d),(g) LA, (b),(e),(h) FL, and (c),(f),(i) CU. (a)–(c) Based on GCM output that has been adjusted via MBC for the historical time period (1985–2014). (d)–(f) As in (a)–(c), but for the near future (2031–60). (g)–(i) As in (d)–(f), but for the far future (2071–2100). Future values are based on the SSP585 scenario. The $Y = X$ line is shown in cyan. Each point (black or red) represents a summer day during the time period. Points that fall more than six standard deviations based on the bi-weight standard deviation (Lanzante 1996) from the line are given in red.

and HI, for HI out-of-bounds occurrences by Dahl et al. (2019) and for future increases in the HI by Dahl et al. (2019) and Mukherjee et al. (2021). Areas where the EHI exceeds the HI by a large amount correspond reasonably well to areas where heat stress is expected to increase the most. These are areas that are able to tap moisture from subtropical waters, extending from the Caribbean and Gulf of Mexico northward into the CONUS. Even though there are locations for which large differences between the EHI and HI are relatively rare, this does not obviate the need for adoption of the EHI. Romps and Lu (2022) examined cases for which the EHI exceeded its historical 99.9th percentile and found large underestimates of the EHI by the HI. Although these are rare cases in the current climate, the associated events can have large societal impacts in terms of human morbidity and mortality as they corresponded to prominent historical heat waves. These are precisely the instances in which underestimates of heat stress can be the most consequential.

Having examined spatial and temporal patterns of the discrepancy, it is of interest to examine the magnitudes by which the EHI exceeds the HI (corresponding to the red points in Fig. 2). These differences are shown in the form of boxplots (Fig. 4) for 10 of the 15 stations—the remaining stations had too few cases for meaningful results. For most stations, future differences are commonly 3°–8°C with some instances approaching or exceeding 10°C. Even in the historical past, discrepancies of 3°–10°C could be realized for

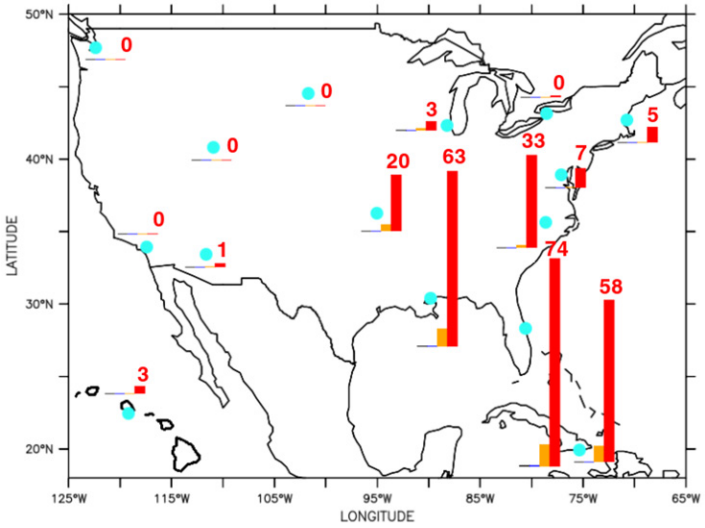

Fig. 3. Bars whose height is proportional to the percentage (%) of extreme positive differences of summertime daily maximum (EHI – EH) for each station and time period. The station location is given by a cyan circle with four bars from left to right: black (GSOD 1985–2014), blue (MBC 1985–2014), orange (MBC 2031–60), and red (MBC 2071–2100). Future values are based on the SSP585 scenario. The red numbers (%) correspond to the height of the red bars. The extremes correspond to the red points displayed in red in Fig. 2, as well as those for the other stations (not shown).

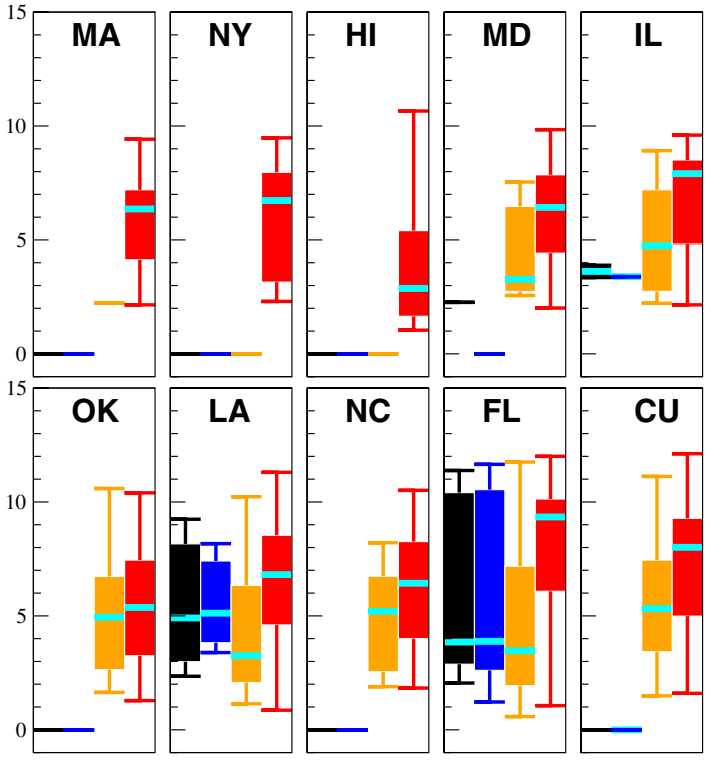

Fig. 4. Boxplots of the magnitudes (°C) of summertime daily maximum of the EHI – HI differences for the extreme cases depicted in Fig. 3 with the colors and ordering from left to right same as in Fig. 3. Each solid box extends from the lower to upper quartile with whiskers depicting the minimum and maximum values; the median is show in cyan. The state abbreviations at the top of each panel correspond to those given in Table 1.

the most impacted stations (FL, LA, IL). The NWS issues various levels of heat advisories when the HI is expected to exceed certain thresholds (varying across the nation but often 100° or 105°F).

Next, histograms of the heat indices for select stations are presented to show distributional changes. Figure 5 shows historical (black and cyan curves) and future (orange and red curves) distributions of the EHI for (a) IL, (b) NC, (c) LA, and (d) CU. By construction, in the historical period, the distributions of the observations (not shown) and MBC are virtually identical. The relative GCM cold bias in the EHI, which is fully adjusted by MBC, is not necessarily indicative of a GCM deficiency. The sign of this difference is consistent with that expected (Diaconescu et al. 2023) due to the difference in how the EHI was computed (rhmin from GSOD and avgTd from the GCM). There is a progression in MBC, with the distributions shifting rightward, from historical to near future to far future. For IL and NC by the time of the far future, approximately half or more of the values are "novel" in that they have never occurred during the historical period. In addition, in the near and particularly in the far future, the right tail becomes elongated encompassing some very extreme values. For LA and CU, which are located closer to the subtropical moisture sources, the behavior is more extreme than for IL and NC. For the near future, roughly half of the values are novel, whereas for the far future, the vast majority are novel. If this scenario were realized, humans living in these regions would commonly experience summer heat stress that their recent ancestors had never felt.

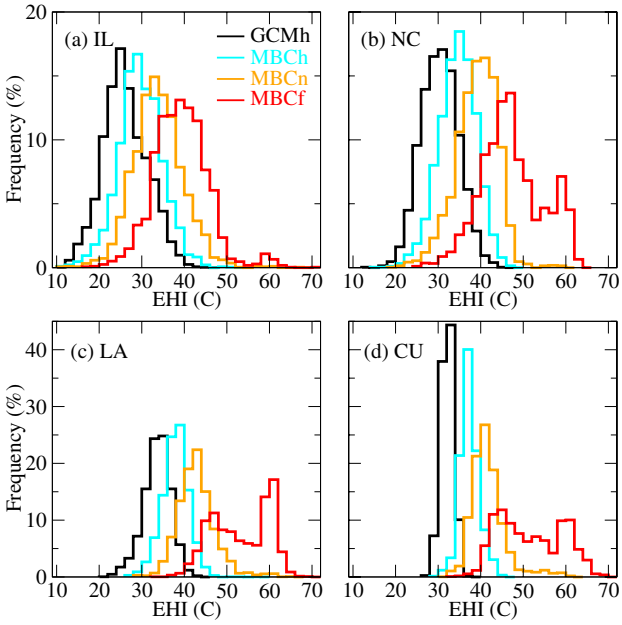

FIG. 5. Histograms of summertime daily maximum EHI (°C) for (a) IL (b) NC, (c) LA, and (d) CU. GCM historical distributions are in black, MBC historical distributions are in cyan, MBC near-future distributions are in orange, and MBC far-future distributions are in red. Future values are based on the SSP585 scenario.

The final set of histograms (Fig. 6) depict the differences in distributions between the HI and the EHI for the same four stations shown in Fig. 5. While Fig. 5 shows remarkable future changes in heat stress, Fig. 6 shows how the most extreme changes are underestimated by the HI as compared to the EHI, as the right tail of the EHI is elongated compared to that of the HI. The implication for the assessment of impacts on human health could be quite severe.

Since it has been shown that the EHI can often exceed the HI by a considerable amount, a natural question to ask is what impact would it have if the NWS utilized the EHI instead of the HI in issuing heat advisories? While the gut reaction is that one would expect more frequent advisories, the answer is not what it would seem. While the criteria used by the NWS vary by location, often a heat index exceeding 105°–110°F will trigger a public alert. As an exercise, suppose we consider a value of 41°C, which is near the low end of this range as our criteria. Looking more closely at Fig. 1, we see that there is not much difference between the HI and EHI near 41°C. Only at higher values does the EHI depart more consistently from the HI.

To better quantify this, 2 × 2 contingency tables have been constructed for each station and time period. The average frequencies over the 15 stations are given in Table 2. The two left-most columns give the average frequencies for which neither index would trigger an alert (nn) or

both would (yy). The sum of these two frequencies represents the percentage of time that they agree in terms of alerts. For all time periods, the agreement is 98% or greater. Thus, somewhat counterintuitively, the EHI does not appear to provide any advantage for the issuance of heat advisories. However, if in the future as the magnitude of extreme heat-stress events increases, the NWS adopts a higher-level criterion as a trigger for alerts, it is possible (as seen in Fig. 1) that use of the EHI would lead to a greater frequency of advisories. Note that since in practice, the NWS incorporates additional criteria such as a duration of 2 days for the exceedances (www.weather.gov/safety/heat-ww), the exceedances in Table 2 may slightly overestimate the frequencies of advisories.

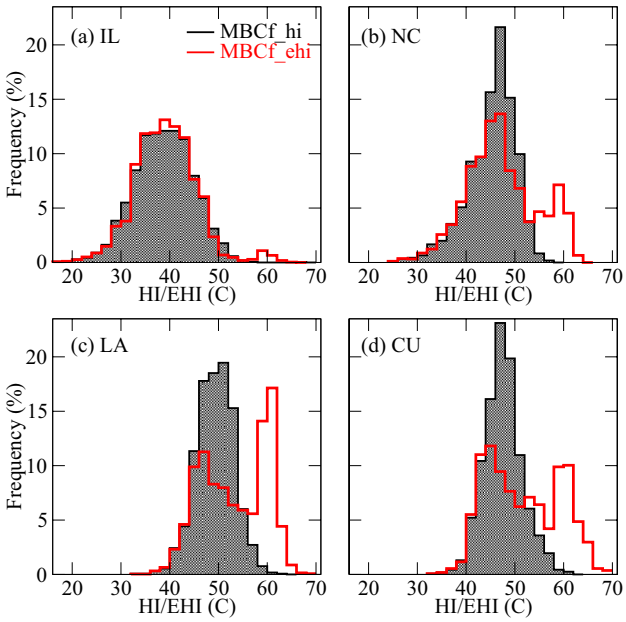

Fig. 6. Histograms of MBC far future for HI (°C) (black, gray filled) and EHI (°C) (red) for (a) IL, (b) NC, (c) LA, and (d) CU. Future values are based on the SSP585 scenario.

Perhaps, the most remarkable aspect of Table 2 is the incredible increase in the frequency of heat advisory-level stress (yy) in the far future. To explore this in more detail Fig. 7 shows the yy percentages by station for the different time periods. There are some more subtle details by which stations differ in their progression over time (the explanations for which would require more in-depth analysis and may be GCM dependent). However, the overarching conclusion is that heat advisory conditions are projected to become much more common at all locations under the SSP585 scenario. For about 40% of the stations, heat advisories would be issued 75% or more of all summer days and a few would have advisory-level conditions almost every day. About half of the stations, which have rarely reached advisory conditions in the current climate, could expect to have advisory-level conditions 10%–50% of the time.

4. Discussion

This work has explored some of the differences between the widely used NWS implementation (Rothfusz 1990) of the HI originally derived by Steadman (1979) and an updated (Lu and Romps 2022) version (EHI), which extends the Steadman model into the regime of higher temperatures. At the time when Steadman derived his physiological model, this regime was considered so rare as to not merit serious effort. In the decades that have passed, climate change has rendered that assumption dubious and motivated the development of the EHI.

This work has compared these two metrics, designed to estimate heat stress on the human body, over a representative network of 15 U.S. stations. In conjunction with the output from a

Table 2. Averages over the 15 stations from 2 × 2 contingency tables comparing the frequency (%) for which NWS-like daily heat advisories would be issued based on the summertime daily maximum HI and EHI. An advisory is triggered if the heat index reaches or exceeds 41°C. The column nn (yy) is the frequency for which the summertime daily maximum of neither (both) HI nor (and) EHI reach 41°C on a given day, while yn (ny) is the frequency for which HI (EHI) meets the criteria but EHI (HI) does not. Future values are based on the SSP585 scenario.

	nn	yy	yn	ny
GSOD 1985–2014	94.6	5.0	0.3	0.1
MBC 1985–2014	94.7	4.9	0.3	0.1
MBC 2031–60	74.4	24.2	1.3	0.2
MBC 2071–2100	45.0	54.2	0.6	0.3

leading GCM, which has been adjusted for biases compared to station observations, comparisons have been made for both the recent past and the future extending to the end of the twenty-first century under the SSP585 scenario.

The most important finding of this work is the large difference between the EHI and the HI for some locations in the future. There are distinct geographical patterns in both the projected changes in heat stress and the differences between the two metrics. Regions that more effectively tap subtropical moisture show a more impactful increase in heat stress as well as larger differences between the two metrics. As previously shown by Romps and Lu (2022), the traditional HI is liable to underestimate the most extreme heat stress by a potentially large amount. Here, it is found that in the future, in the more humid locations, the EHI exceeds the HI by an extreme amount in a fifth to as much as three quarters of all days. While the typical magnitude of these extreme differences is 5°C at times, it can exceed 10°C. Such differences in the HI imply important differences in impacts on human health. The impact of these differences is magnified by the dramatic increase in the frequency of high heat-stress days, for example, as measured by the conditions, which would warrant the issuance of an excessive heat warning by the NWS. While for most locations, this occurs on less than 5% of summertime days in the historical period; in the future, this may occur at least a third of the time at most stations and about 90% or more of the time at about a third of the locations.

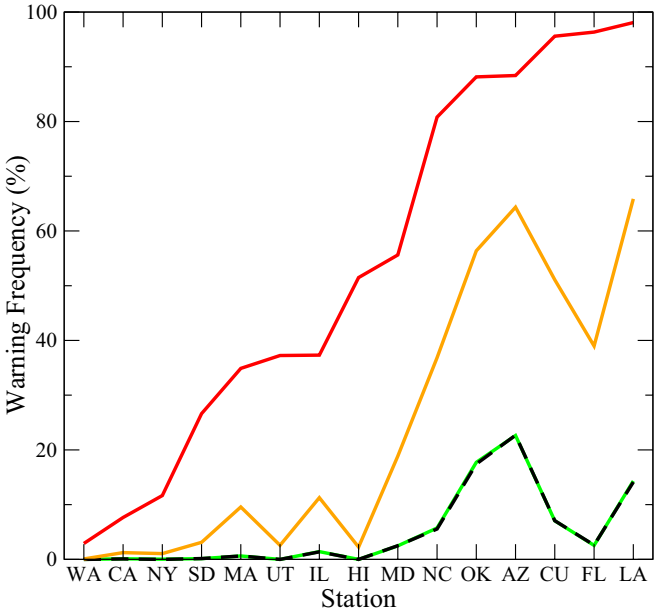

FIG. 7. The frequency (ordinate; % of days) for which both HI and EHI summertime daily maxima reach or exceed 41°C by station (abscissa) and time period (different curves). Green (dashed black) are for GSOD (MBC) for 1985–2014, while orange (red) is for 2031–60 (2071–2100) for MBC. Future values are based on the SSP585 scenario.

It is important to note that this work is by design a demonstration project and is by no means intended to provide a definitive answer as to what heat stress the U.S. population will be subjected to in the future. It is not exhaustive, rather illustrative. However, it does provide one plausible scenario. There are a number of caveats that limit the conclusions with regard to the future as this work considers only a single:

1) GCM
2) Bias adjustment method
3) Forcing scenario
4) GCM realization (i.e., ensemble member)

With regard to caveat 1, the amount of anthropogenic warming varies widely among GCMs. By one metric, climate sensitivity, the GFDL model ranks as intermediate (Scafetta 2023). Furthermore, it has been shown to be credible in some key aspects compared to observations (Held et al. 2019). Nevertheless, there is no definitive answer as to which model is "right" and a range of solutions are considered plausible.

With regard to caveat 2, there are a plethora of statistical downscaling/bias-adjustment methods. The chosen MBC method (Cannon 2018), as a distributional approach, is particularly appropriate. QDM, the basis for MBC, has been shown to perform well compared to

other methods in its class in "perfect model" evaluations (Lanzante et al. 2019), which test the ability to perform under conditions of climate change.

With regard to caveat 3, the SSP585 future forcing scenario that has been employed here is the "worst-case" CMIP6 scenario. Nevertheless, it is considered a plausible outcome (Schwalm et al. 2020) in a highly uncertain future that depends crucially on human mitigation, which to date has not kept up with past pledges to reduce greenhouse gas emissions.

With regard to caveat 4, the natural or internal climate variability plays a role in how climate change progresses even in the face of strong anthropogenic forcing. A single GCM realization does not fully reflect the range of possibilities. Furthermore, in so much as CMIP6 coupled-climate model simulations are unconstrained, i.e., free-running, there is potentially some degree of inherent mismatch between the past observations and the GCM historical run with regard to internal variability. As such this could have an effect on the historical "training" of MBC, which would be reflected in the bias adjustment in the future.

An additional caveat 5 regards the manner in which the heat index values were computed from daily data. Generally speaking, one might consider the daily maximum heat index, which could be reasonably estimated from hourly observations of temperature and humidity, as most relevant to human health as it represents the maximum stress. Due to data constraints, the GCM data were not available at hourly time resolution, as is often the case. For the GCM, tmax and rhmin were used with the implicit assumption that the minimum relative humidity occurs at the time of maximum temperature. This assumption would hold if the specific humidity was constant over the course of the day. However, particularly during the warm season, one might expect a slight increase in specific humidity during the afternoon via enhanced evapotranspiration driven by diurnal heating. In addition, an airmass change (e.g., via advection, frontal passage, and sea breeze) could also lead to a higher relative humidity at the time of the maximum temperature. On any day when rhmin does not occur at the time of tmax, the estimated heat index from these values will be lower than that based on hourly or higher frequency temporal resolution data, leading to a cold bias in the estimated heat index.

There is a question as to the impact of this cold bias. On the one hand, Diaconescu et al. (2023) found that the distributions of the Humidex heat-stress index calculated from tmax and rhmin were very similar to those calculated directly from hourly data over a network of 319 Canadian stations. Ongoing work by GFDL's Empirical Statistical Downscaling team (https://www.gfdl.noaa.gov/esd_eval/) is currently addressing this issue. Preliminary findings[1] suggest that while the differences are greater than 1°C only about 5% of the time over a network of stations in the Northeastern United States, nevertheless, the tasmax/rhmin approximation could lead to ~20% underestimate of the number of days that the National Weather Service would issue a heat advisory.

[1] See video (46:25–1:02:00) of Keith Dixon's presentation at the 104th annual meeting of the American Meteorological Society at https://ams.confex.com/ams/104ANNUAL/meetingapp.cgi/Session/65953.

While these findings motivate further study on this important topic, the impact of this on the conclusions of this study is mitigated here by the application of MBC to the heat indices rendering equal treatment of the observed and model data by transforming the GCM values toward those of the observations. Since the observational approach assumed a constant dewpoint over the course of the day, differences from the "true" 24-h maximum heat index could be either positive or negative. This more likely leads to overall cancellation of bias in the aggregate as opposed to use of tmax/rhmin whose bias is inherently one-sided (i.e., negative). Nevertheless, the use of daily summary values for both observations and the GCM represents an additional source of uncertainty.

In conclusion, in consideration of (1)–(5), it is clear that an informed assessment of potential future changes in heat stress requires a much more comprehensive analysis that

considers the various uncertainties that exist. However, given that all of the choices made here are reasonable and plausible suggests that the EHI should be used instead of the HI in future work involving climate change. Furthermore, even in the historical and near-future time periods, the EHI has been shown to provide more urgency for extreme events. Given that those tend to be the most societally impactful, suggests for the adoption of the EHI, particularly in some geographical locations. Even to the extent that the SSP585 scenario could overestimate the magnitude of late twenty-first century impacts, the spatial patterns of change found here may prove to be realistic via the time-shift paradigm (Tebaldi and Knutti 2018; Lee et al. 2021).

Guidance such as provided here could also be incorporated by the recently launched National Integrated Heat Health Information System (NIHHIS) (www.heat.gov), which is a collaboration of a number of government entities. The NIHHIS was founded by the National Oceanic and Atmospheric Administration's (NOAA) Climate Program Office (CPO), and National Weather Service (NWS), in partnership with the Centers for Disease Control and Prevention (CDC).

The findings of this project provide impetus for more comprehensive analyses of future heat stress. The Empirical Statistical Downscaling (ESD) team at GFDL (https://www.gfdl.noaa.gov/esd_eval) is currently embarked on such work and is examining not only the difference between the HI and the EHI but other choices made in estimating heat stress. In addition, while the code made available by Romps (https://www.romps.berkeley.edu/papers/pubs-2020-heatindex.html) represents a very valuable contribution to the community, application to large datasets, such as those from GCM output, can prove to be computationally costly. Some accurate shortcuts that greatly reduce computational cost have been developed (Lanzante 2024).

Acknowledgments. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF. Thanks also to Tom Delworth and Tom Knutson for very useful comments on an earlier draft of this manuscript as well as to David Romps, an anonymous reviewer, and Editor Art DeGaetano during the course of the review process.

Data availability statement. The GSOD data are available at https://www.ncei.noaa.gov/access/search/data-search/global-summary-of-the-day/, while the GFDL GCM data can be accessed at https://esgf-node.llnl.gov/projects/cmip6/.

APPENDIX

In this work, statistical downscaling has been applied in order to downscale/bias-adjust GCM values with respect to station observations. This process utilizes a station observation and the nearest GCM gridpoint value. In essence, the method is "trained" via the relationship between the cumulative distribution functions (CDFs) of the historical station observations and the historical GCM output values. This mapping is then applied to adjust future GCM values.

In determining how to bias adjust the heat indices employed in this work, a fundamental complication is the fact that the heat index is computed from two independent quantities: temperature and humidity. As such, there are three philosophically distinct approaches to consider:

1) Apply separate univariate adjustment (here QDM) to temperature and humidity and then use these to compute the heat index

2) Compute the heat index from temperature and humidity and then apply univariate (here QDM) adjustment to the heat index
3) Apply a multivariate (here bivariate MBC) adjustment to temperature and humidity and then use these to compute the heat index.

A key consideration is the dependency between temperature and humidity. For a given location and climate state, there are certain combinations of these two variables that may occur and others than will not occur. Since approach 1 does not account for this, it is not a viable approach. In approach 2, there will be some implicit accounting for the dependency since only certain values of the heat index are realizable and since the univariate method QDM, by virtue of its distributional nature, varies the amount of adjustment as a function of position within the distribution. The most appropriate choice is approach 3 since it explicitly accounts for the dependency between the two variables. An additional consideration for approach 3 is the choice of humidity variable. In this case, dewpoint proves to be a natural choice as it is a temperature variable like daily maximum temperature. Note that while MBC (see below) provides a strong constraint on the covariability of maximum temperature and dewpoint, the possibility of an unphysical result in which the latter exceeds the former exists. However, this does not occur here most likely because the separation between the two is typically fairly large.

The results reported in this work are based on approach 3; however, all of the results (not shown) have been repeated using approach 2. Both approaches yield qualitatively consistent results and in many instances are quantitatively similar. The results are most similar for the historical period. During the future, the most noteworthy difference is a sizeable increase based on approach 3 versus approach 2 in the frequency of extreme differences (Fig. 3) at the stations closest to the subtropical moisture sources, with moderate decreases for eastern stations farther from the sources. However, typical magnitudes of the differences (Fig. 4) are fairly similar as are the exceedance frequencies (Table 2 and Fig. 7).

The MBC algorithm operates to produce a mapping between two multivariate distributions: a source (here the GCM) and a target (here GSOD). The source distribution is adjusted iteratively by applying a random orthogonal rotation to the datasets followed by quantile mapping (QDM) of the rotated marginal distributions and then inversely rotating the resulting data. The rotation provides linear combinations of the original variables rather than the original variables separately in an attempt to preserve the relations between the variables. Iteration ceases when the source and target distributions are sufficiently similar. As a final step, QDM is applied separately to the original variables and these values are reordered according to the ranks of iteratively derived solution. This final step insures the trend-preserving property of QDM while maintaining the dependence structure of the iterative algorithm. The remainder of this appendix describes the QDM algorithm.

Conceptually, the QDM method is an extension of the commonly used simple bias adjustment (SBA) approach. For example, SBA would take the difference in means between historical observations and historical model (i.e., GCM) values and add this difference to each future model value. The QDM method, which is one from a class of distributional methods, performs bias adjustment in a similar fashion except that the bias adjustment factor varies by position within the statistical distribution.

Mathematically QDM can be expressed as

$$D_f(x) = x + F_{\mathrm{Oh}}^{-1}\left[F_{\mathrm{Mf}}(x)\right] - F_{\mathrm{Mh}}^{-1}\left[F_{\mathrm{Mf}}(x)\right], \tag{A1}$$

where x is the future GCM value to be adjusted and $D_f(x)$ is the downscaled/bias-adjusted value for the future. In this case, the argument x refers to the heat index (either HI or EHI).

Here, F is the CDF and F^{-1} is its inverse (or equivalently a quantile function). Subscripts Oh, Mh, and Mf refer to observations-historical, model-historical, and model-future, respectively.

In a sense, the adjustment procedure utilizes the future GCM value (x) as a "first guess" and then amends it based on the difference between the second and third terms on the rhs of (A1). The innermost terms on the rhs of (A1), i.e., "$F_{\mathrm{Mf}}(x)$," are the CDF of the future model values evaluated at the future value to be bias-adjusted. Numerically, this is a percentile; thus, it indicates its relative position within its distribution.

For example, suppose x lies in the upper tail of the Mf distribution at the 90th percentile. The second (third) term on the rhs evaluates the Oh (Mh) quantile function at this same percentile. Thus, the bias-adjusted value is the difference between the 90th percentile values of the Oh and Mh distributions added to the value to be adjusted (i.e., x, the 90th percentile value of the Mf distribution). Note that by way of this construction, it is possible to adjust novel values. That is, model values in the future that are out of the bounds of the model historical distribution will always be defined.

In this paper, as in many studies, bias adjustment is applied to both the historical and future time periods. When applied to the future, there is an implicit assumption of "stationarity" that the mapping derived in the historical period [i.e., the difference between the second and third terms on the rhs of (A1)] is valid in the future. When applied to the historical period (i.e., replace all "f" subscripts with "h"), the third term on the rhs of (A1) reduces to "x," which cancels with the first term on the rhs of (A1), yielding

$$D_h(x) = F_{\mathrm{Oh}}^{-1}\left[F_{\mathrm{Mh}}(x)\right]. \tag{A2}$$

Here, the percentile of the x value in the Mh distribution is entered into the quantile function of the Oh distribution. In our example, the 90th percentile value from the historical model distribution would be replaced by the 90th percentile value from the historical observational distribution. As such, this ensures that the distribution of bias-adjusted values very nearly matches those of the observations. This expression (A2) is sometimes referred to as "quantile mapping" although since this expression is used inconsistently in the literature, some may use it to describe other variants of distributional mapping.

References

Baldwin, J. W., T. Benmarhnia, K. L. Ebi, O. Jay, N. J. Lutsko, and J. K. Vanos, 2023: Humidity's role in heat-related health outcomes: A heated debate. *Environ. Health Perspect.*, **131**, 055001, https://doi.org/10.1289/EHP11807.

Cannon, A. J., 2018: Multivariate quantile mapping bias correction: An *N*-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dyn.*, **50**, 31–49, https://doi.org/10.1007/s00382-017-3580-6.

——, S. R. Sobie, and T. Q. Murdock, 2015: Bias correction of GCM precipitation by quantile mapping: How well do methods preserve changes in quantiles and extremes? *J. Climate*, **28**, 6938–6959, https://doi.org/10.1175/JCLI-D-14-00754.1.

Dahl, K., R. Licker, J. T. Abatzoglou, and J. Declet-Barreto, 2019: Increased frequency of and population exposure to extreme heat index days in the United States during the 21st century. *Environ. Res. Commun.*, **1**, 075002, https://doi.org/10.1088/2515-7620/ab27cf.

Diaconescu, E., H. Sankare, K. Chow, T. Q. Murdock, and A. J. Cannon, 2023: A short note on the use of daily climate data to calculate Humidex heat-stress indices. *Int. J. Climatol.*, **43**, 837–849, https://doi.org/10.1002/joc.7833.

Dixon, K. W., J. R. Lanzante, M. J. Nath, K. Hayhoe, A. Stoner, A. Radhakrishnan, V. Balaji, and C. F. Gaitan, 2016: Evaluating the stationarity assumption in statistically downscaled climate projections: Is past performance an indicator of future results? *Climatic Change*, **135**, 395–408, https://doi.org/10.1007/s10584-016-1598-0.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937–1958, https://doi.org/10.5194/gmd-9-1937-2016.

Held, I. M., and Coauthors, 2019: Structure and performance of GFDL's CM4.0 climate model. *J. Adv. Model. Earth Syst.*, **11**, 3691–3727, https://doi.org/10.1029/2019MS001829.

IPCC, 2023: Summary for policymakers. *Climate Change 2023: Synthesis Report*, H. Lee and J. Romero, Eds., IPCC, 1–34, https://doi.org/10.59327/IPCC/AR6-9789291691647.

Lanzante, J. R., 1996: Resistant, robust and non-parametric techniques for the analysis of climate data: Theory and examples, including applications to historical radiosonde station data. *Int. J. Climatol.*, **16**, 1197–1226, https://doi.org/10.1002/(SICI)1097-0088(199611)16:11<1197::AID-JOC89>3.0.CO;2-L.

——, 2024: Efficient and accurate shortcuts for calculating the Extended Heat Index. *J. Appl. Meteor. Climatol.*, https://doi.org/10.1175/JAMC-D-24-0081.1, in press.

——, K. W. Dixon, M. J. Nath, C. E. Whitlock, and D. Adams-Smith, 2018: Some pitfalls in statistical downscaling of future climate. *Bull. Amer. Meteor. Soc.*, **99**, 791–803, https://doi.org/10.1175/BAMS-D-17-0046.1.

——, D. Adams-Smith, K. W. Dixon, M. J. Nath, and C. E. Whitlock, 2019: Evaluation of some distributional downscaling methods as applied to daily maximum temperature with emphasis on extremes. *Int. J. Climatol.*, **40**, 1571–1585, https://doi.org/10.1002/joc.6288.

Lee, J.-Y., and Coauthors, 2021: Future global climate: Scenario-based projections and near-term information. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 553–672, https://doi.org/10.1017/9781009157896.006.

Lu, Y.-C., and D. M. Romps, 2022: Extending the heat index. *J. Appl. Meteor. Climatol.*, **61**, 1367–1383, https://doi.org/10.1175/JAMC-D-22-0021.1.

——, and ——, 2023: Predicting fatal heat and humidity using the heat index model. *J. Appl. Physiol.*, **134**, 649–656, https://doi.org/10.1152/japplphysiol.00417.2022.

Mukherjee, S., A. K. Mishra, M. E. Mann, and C. Raymond, 2021: Anthropogenic warming and population growth may double US heat stress by the late 21st century. *Earth's Future*, **9**, e2020EF001886, https://doi.org/10.1029/2020EF001886.

Romps, D. M., and Y.-C. Lu, 2022: Chronically underestimated: A reassessment of US heat waves using the extended heat index. *Environ. Res. Lett.*, **17**, 094017, https://doi.org/10.1088/1748-9326/ac8945.

Rothfusz, L. P., 1990: The heat index "equation" (or, more than you ever wanted to know about heat index). NWS Southern Region Headquarters Tech. Attachment SR 90-23, 2 pp., https://www.weather.gov/media/ffc/ta_htindx.PDF.

Scafetta, N., 2023: CMIP6 GCM validation based on ECS and TCR ranking for 21st century temperature projections and risk assessment. *Atmosphere*, **14**, 345, https://doi.org/10.3390/atmos14020345.

Schwalm, C. R., S. Glendon, and P. B. Duffy, 2020: RCP8.5 tracks cumulative CO_2 emissions. *Proc. Natl. Acad. Sci. USA*, **117**, 19 656–19 657, https://www.pnas.org/doi/10.1073/pnas.2007117117.

Seneviratne, S. I., and Coauthors, 2021: Weather and climate extreme events in a changing climate. *Climate Change 2021: The Physical Science Basis*, V. Masson-Delmotte et al., Eds., Cambridge University Press, 1513–1766, https://doi.org/10.1017/9781009157896.013.

Steadman, R. G., 1979: The assessment of sultriness. Part I: A temperature-humidity index based on human physiology and clothing science. *J. Appl. Meteor.*, **18**, 861–873, https://doi.org/10.1175/1520-0450(1979)018<0861:TAOSPI>2.0.CO;2.

Tebaldi, C., and R. Knutti, 2018: Evaluating the accuracy of climate change pattern emulation for low warming targets. *Environ. Res. Lett.*, **13**, 055006, https://doi.org/10.1088/1748-9326/aabef2.