Challenging the universality of heatwave definitions: gridded temperature discrepancies across climate regions

Erin Coughlan de Perez[1,2] · Julie Arrighi[2] · Joalane Marunye[3]

Abstract

As global studies of climate change depict increasingly dire outcomes of extreme heat, there is an urgent need to understand the appropriateness of heatwave definitions and temperature datasets in different parts of the world. We carry out an intercomparison of the CHIRTS gridded station-satellite temperature dataset with three reanalysis products, ERA5, NCEP-DOE Reanalysis 2, and MERRA2, to assess biases in the absolute value of extreme heat events and the distribution of extreme events. We find close agreement between all four datasets in the magnitude and distribution of extreme temperatures, with a cold bias in the reanalyses over mountainous areas. However, there is little to no agreement between datasets on the timing of extreme heat events in the tropics, and the datasets do not even agree on which month is the hottest month climatologically in these regions. Second, we compare how the four datasets represent the frequency and timing of extreme heat events, using two different types of heatwave definitions: 5-day duration-based extremes and extreme temperature-humidity combinations (heat index). In the case of 5-day heatwaves, there are almost zero events recorded historically in tropical regions. In contrast, high absolute values of the heat index are most common in dry climates, likely due to the dominance of high temperature spikes in these regions, and high heat index events also occur in temperate and tropical regions. There is little agreement between datasets, however, on when these extreme heat index events have happened historically in the tropics. Given these results, we highlight the need for locally developed heatwave metrics for different parts of the world, and we urge against the use of a single heatwave definition in global studies. We also recommend that any studies assessing heat-health relationships in tropical regions beware of the lack of agreement between observational and reanalysis datasets and compare results from different gridded dataset products to estimate uncertainty in heat-health relationships.

Keywords Heatwaves · Reanalysis · Tropics · Humidity · Climate

✉ Erin Coughlan de Perez
   erin.coughlan@tufts.edu

1   Feinstein International Center, Friedman School of Nutrition Science and Policy, Tufts University, Boston, MA, USA

2   Red Cross Red Crescent Climate Centre, The Hague, The Netherlands

3   Department of Geography and Environmental Science, National University of Lesotho, Roma, Lesotho

1 Introduction

The increasing frequency and intensity of extreme heat is the most universal outcome of climate change across the world, and most of Earth's land area has already observed increasing temperatures (IPCC 2022). Although extreme heat is an increasing concern, there is little information on the relative accuracy and usefulness of historical temperature datasets to characterize the behavior and change in extreme heat in different parts of the world. This is particularly true in the tropics, which have been the subject of fewer research studies on heat and heatwaves (Campbell et al. 2018).

In the extratropics, there is a large and growing literature on the dangers of heatwaves, with a ballooning evidence base that links observed extreme temperatures with societal outcomes (Marx et al. 2021). These studies often design bespoke heatwave indices to capture health outcomes, and these can include a set of consecutive days of temperature above a threshold as well as simultaneous heat-humidity indices (Smith et al. 2013). A meta-analysis of heatwave definitions in 2016 found a similar variety, with different combinations of maximum, minimum, and mean temperatures and durations found to relate to health outcomes in different parts of the world, as well as a variety of temperature plus humidity definitions, such as the heat index and apparent temperatures (Xu et al. 2016). Researchers working on that study recommended locally developed heatwave definitions to provide the most relevant indication of potential health outcomes.

Studies that have examined the relative frequency of different heatwave definitions in climate datasets have found that different definitions happen with higher frequency in different regions, and therefore, it cannot be assumed that a heatwave definition with a certain frequency in Canada would occur with the same frequency in Brazil. In a heatwave intercomparison over the USA, for example, Smith et al. (2013) found that the heat index using temperature and humidity was more frequently observed in the southwestern humid part of the country, while definitions focused on simple maximum temperatures were more frequent in other (often drier) regions. That study was a simple climatology of heatwave definitions, not linked to impacts in each region. Heo et al. (2019) did a similar comparison across South Korea, finding different frequencies of each definition in different regions of the country; they found that web-bulb globe temperature had the strongest link with health outcomes. Existing climate change projections of increases in humid heatwaves might also be underestimates, due to models' inability to represent convective precipitation that influences these events (Birch et al. 2022).

WMO-WHO Heat Health Guidance identifies two broad types of heatwaves: dry heatwaves in areas with a variable summer season and moist heatwaves that are associated with warm and humid conditions and nocturnal cloud cover that prevents cooling at night (World Health Organization 2018). This bifurcation of heatwaves into persistent (duration-based) and humid heatwaves was also used by Domeisen et al. (2023) in their characterization of the predictability of heatwaves. However, many studies characterizing heatwave risk globally use a single heatwave definition for the entire world (e.g., Chambers 2020).

Most heat-health research elucidates the relationship between temperature and outcomes in specific locations across the extratropics, with a high number of studies in Europe, North America, and Australia. This includes an emphasis on land-atmosphere feedbacks and periods of blocking and subsidence that influence the increase in hot extremes in these regions (e.g., Jeong et al. 2022; Salathé Jr et al. 2023; Wang et al. 2023). There is little research or documentation of heat impacts in Africa and many other tropical regions of the world (Harrington and Otto 2020). Because a global intercomparison of heatwave datasets

and definitions has never been done, there is no information on which of these definitions occur with which frequency in tropical and understudied regions. In fact, studies that have attempted to calculate the frequency of global heatwaves using multi-day heatwave definitions have found few if any heatwaves in the tropics (e.g., Coughlan de Perez et al. 2018).

Even though there is a lack of heat-health studies in the tropics, many major studies highlight the tropics as an area of concern for extreme heat under climate change. Coffel et al. found that the tropics saw a greater number of exceedance of historical annual maximums in the future with climate change, compared to the rest of the world. However, they acknowledge that this is likely due to the low variability of temperatures in the tropics (Coffel et al. 2017).

It has become accepted practice to simply extrapolate studies about extreme temperature from the extratropics to the tropics, without any inquiry as to whether this is appropriate. For example, Mora et al. (2017) carried out a large review of heatwave definitions and posited a universal deadly temperature-humidity relationship, but almost all of the studies reviewed came from the extratropics. The authors then present their main conclusion with a focus on the tropics, finding that their "universal" extreme heat definition will see the highest increases in tropical regions, and therefore, those regions are at high risk. At the time of writing, this paper had accumulated almost 1000 citations, and the results were featured prominently in an IPCC synthesis report key figure (reproduced here as Fig. 1), with a disclaimer that the data used to derive the heat-health relationship was mainly from the extratropics.

In order to better understand heat-health relationships in different climate zones of the world, we need further studies that connect climate extremes with human health outcomes. While there are several studies using station-based temperature records to derive heat-health relationships in tropical regions, such as Faye et al. (2021), many efforts to understand heat in tropical regions rely heavily on merged station-satellite gridded products and reanalysis data. However, there has been no intercomparison of these datasets for extreme heat, to understand how they differ in representing the most commonly used extreme heat indices.

Existing reanalysis comparisons tend to focus on average values, and Long et al. 2017 found that monthly mean temperatures over latitude zones were generally in agreement among recent reanalyses (CFSR, MERRA, ERA-Interim, JRA-55, and MERRA-2). There have been a few reanalysis intercomparisons for other types of extreme events, including wind (Torralba et al. 2017), the hydrological cycle (Lorenz and Kunstmann 2012), and atmospheric rivers (Collow et al. 2022), but no global intercomparison for extreme heat. A recent study on heat extremes excluded regions in which there was a lack
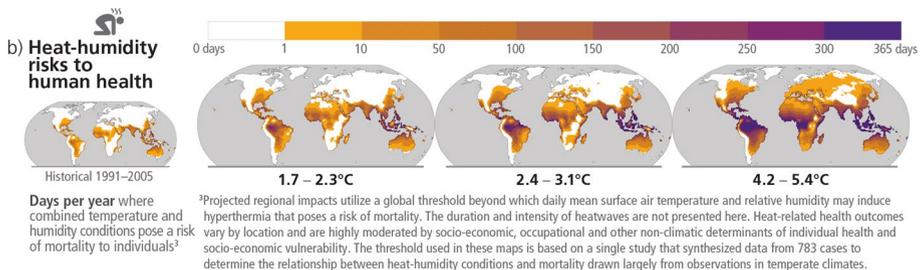

b) **Heat-humidity risks to human health**

0 days 1 10 50 100 150 200 250 300 365 days

Historical 1991–2005   1.7 – 2.3°C   2.4 – 3.1°C   4.2 – 5.4°C

Days per year where combined temperature and humidity conditions pose a risk of mortality to individuals[3]

[3]Projected regional impacts utilize a global threshold beyond which daily mean surface air temperature and relative humidity may induce hyperthermia that poses a risk of mortality. The duration and intensity of heatwaves are not presented here. Heat-related health outcomes vary by location and are highly moderated by socio-economic, occupational and other non-climatic determinants of individual health and socio-economic vulnerability. The threshold used in these maps is based on a single study that synthesized data from 783 cases to determine the relationship between heat-humidity conditions and mortality drawn largely from observations in temperate climates.

Fig. 1 Reproduction of the IPCC synthesis report summary for policymakers figure SPM3b (IPCC 2023)

of consistency between the ERA5 and JRA55 reanalyses for extreme heat events, which included much of Africa (Thompson et al. 2023).

Therefore, the goal of this study is to carry out an intercomparison of extreme heat definitions in several observational and reanalysis datasets. First, we will compare how the reanalysis datasets represent the absolute value and shape of the distribution of extreme temperatures. Secondly, we will compare how the datasets represent the timing of extreme heat events, using two different types of heatwave definitions: long-duration extreme temperatures and extreme temperature-humidity combinations. This intercomparison will enable researchers to understand the inherent limitations of our gridded temperature datasets in different regions of the world and set the stage for future research on the relationship between extreme heat and health outcomes in understudied regions.

2 Methods

We will carry out an intercomparison of extreme heat in four commonly used gridded historical temperature datasets. The Climate Hazards Center Infrared Temperature with Stations (CHIRTS) dataset is a merged station-satellite product widely used in data-sparse regions (Funk et al. 2019). This was developed by merging satellite and station data to generate a dataset of monthly maximum temperatures around the world. The dataset was then combined with ERA5 reanalysis data to create a timeseries of daily temperatures that align with the monthly maximums (Funk et al. 2019). Daily maximum temperature is available from 1983 to 2016 at a native resolution of 0.05°.

The other datasets analyzed here are reanalysis products, derived using a physically consistent climate model to create gridded datasets of many climate variables, including temperatures, aligned with the observational record. The first is ERA5 Land, which is a global reanalysis dataset at 9-km resolution over land areas, based on ERA5 (Hersbach et al. 2020; Muñoz-Sabater et al. 2021). The data is archived on an hourly basis at a resolution of 0.1°.

The other two reanalyses are MERRA2 and NCEP-DOE Reanalysis 2. MERRA2 is run at 0.5° latitude × 0.625° longitude resolution and archived every 6 h. The NCEP2 reanalysis uses a gaussian grid of 1.875 × 1.904129 and is also archived every 6 h.

For this analysis, all datasets were regridded to the native resolution of the MERRA2 dataset (0.5° lat × 0.625° lon) using bilinear interpolation. We used the full global extent of the CHIRTS data, which is all longitudes −180 to 180 and latitudes −60 to 70. For all datasets, we calculated the daily maximum values (Tmax) and annual maximum of daily maximum temperature (TXx) for the years 1983–2015, which are the years available in the CHIRTS dataset. We also calculated daily anomalies by subtracting the average value for that day of the year from all days in each time series.

To estimate the number of long-duration heatwaves, we use the WMO heatwave definition, which is "five or more consecutive days of prolonged heat in which the daily maximum temperature is higher than the average maximum temperature by 5°C (9 °F) or more." The average maximum temperature is calculated as a daily threshold.

To estimate the number of humid heatwaves, we use the heat index (HI). The U.S. National Oceanic and Atmospheric Administration defines a heat index of 32–41°C as "extreme caution," the range of 41–54 °C as "danger," and over 54 °C as "extreme danger" with imminent risk of heat stroke (NOAA, The National Weather Service n.d.). Relative humidity is available for ERA5 and dewpoint temperature is available for the NCEP

reanalysis, and therefore, we use both of those reanalyses to calculate humid heatwaves at each timestep of the dataset and then derive the maximum and minimum HI value per day. Note that MERRA2 also produces dewpoint temperatures, but a validation check on this field identified errors such that we did not calculate humid heatwaves in that reanalysis dataset.

The heat index is a commonly used heatwave measure that converts temperature and humidity into an "apparent temperature" that adjusts the temperature based on the impact of humidity. This is similar in nature to many heat-humidity thresholds and related to several physiological thresholds that incorporate heat, humidity, windspeeds, and sunlight (Anderson et al. 2013).

For the reanalysis intercomparison, we calculate statistical descriptors for the Tmax values (averages, standard deviations) and then calculate the temporal correlation in Tmax between the datasets. Lastly, we compare the frequency and timing of the long-duration heatwaves and the humid heatwaves between datasets. To characterize outcomes by climate region, we aggregate results using the Koppen-Geiger climatic zones (Kottek et al. 2006; Bryant et al. 2017).

3 Results

3.1 Intercomparison of Tmax averages

The CHIRTS and three reanalysis datasets show a strong agreement in the average of the daily maximum temperatures, with few differences between the reanalyses and CHIRTS (Fig. 2). The largest differences are seen over mountain ranges, with the reanalyses displaying a cool bias.

All of the datasets also agree strongly on the overall standard deviation of Tmax (Fig. 3). In the Arctic, the reanalyses overestimate the standard deviation relative to CHIRTS, and
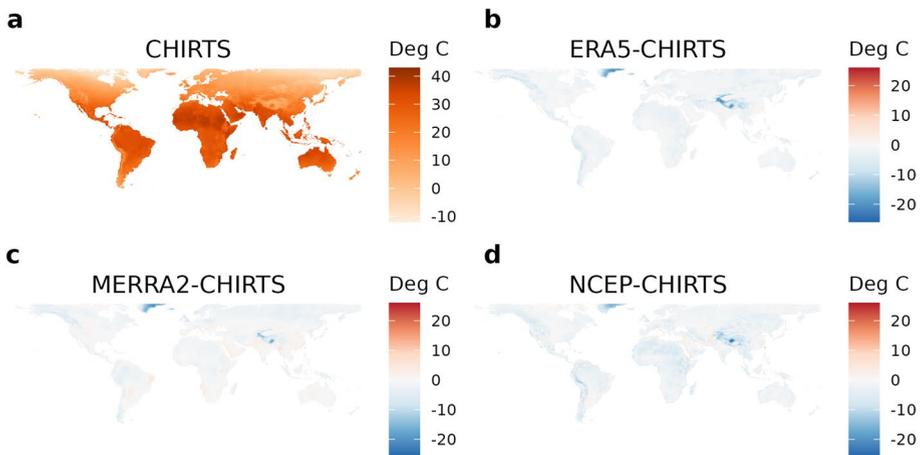
Fig. 2 Average of daily maximum temperatures 1983–2015. Absolute values are shown for **a** CHIRTS dataset, and the difference between the reanalysis and CHIRTS is shown for **b** ERA5, **c** MERRA2, and **d** NCEP
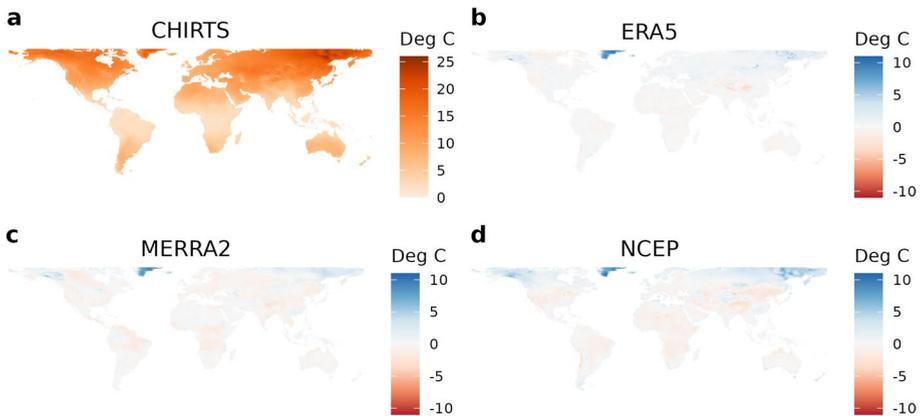
Fig. 3 Standard deviation of daily maximum temperatures 1983–2015. Absolute values are shown for **a** CHIRTS dataset, and the difference between the reanalysis and CHIRTS is shown for **b** ERA5, **c** MERRA2, and **d** NCEP

over most of the remaining land areas the reanalyses have a slight underestimation of the standard deviation in Tmax.

All datasets agree that there is a very small standard deviation of Tmax in the tropics. The "length of the tail" of Tmax or the difference between the 99.9th and 99.5th percentile (see Supplementary Figure 1) is between 0 and 2° in the tropics, while it can reach 6° in some parts of the extratropics.

3.2 Intercomparison of the timings of Tmax and TXx

Despite the strong agreement in the magnitude and statistical distribution of Tmax values globally, CHIRTS and the three reanalysis datasets show a strong disagreement in the timing of the hottest events over much of the world. Figure 4 depicts the simple Pearson correlation in the timeseries of Tmax and TXx for 1983–2015. Because ERA5 is used to create the daily data in CHIRTS based on monthly maximums, these two datasets show the strongest agreements in their daily correlations. However, even in Fig. 4a, the correlation of Tmax between CHIRTS and ERA5, the tropics are less correlated than the extratropics. This is further pronounced in the correlations between other datasets, with almost no association between the Tmax data in CHIRTS and MERRA2 or NCEP over much of the tropics.

The extratropics demonstrate higher correlations between daily anomalies than the tropics, with high disagreement also found in the Himalayan and Andean mountain ranges. Figure 4b, d, f also depicts the correlation between annual temperature maximums, TXx, in these datasets. The lack of correlation in the tropics, even for the CHIRTS-ERA5 datasets, becomes even further pronounced.

However, we find indications that the agreement between datasets has improved over time. Daily maximum temperature correlations between CHIRTS and the three reanalysis have improved over most geographies when comparing results from the first half of the time period (1983–1999) and the second half of the time period (2000–2015). Improvements are most notable in the correlation between CHIRTS and NCEP, especially in the Himalayas. See Supplementary Information Figures SI-4 and SI-5 for details.
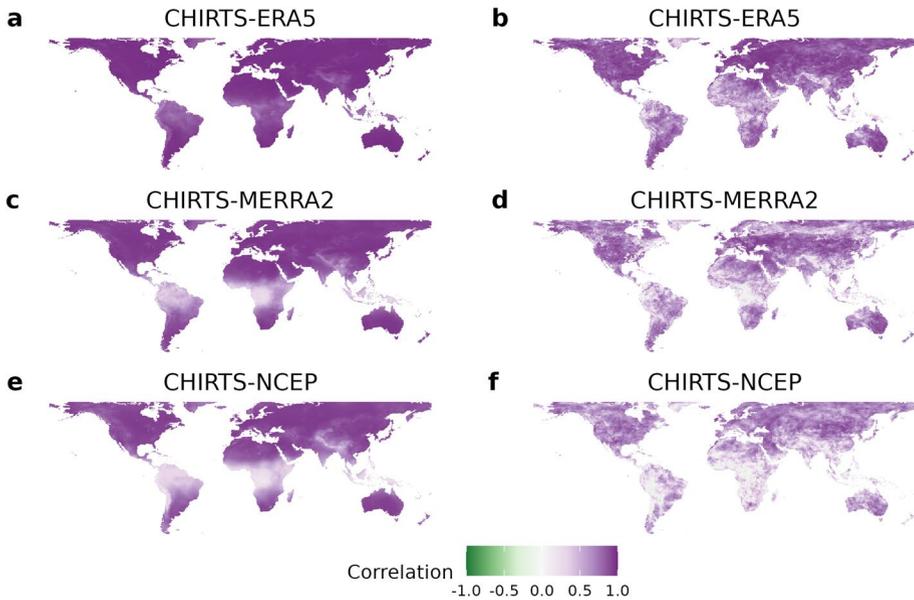
Fig. 4 Correlations of daily and annual maximum temperature among the historical datasets, 1983–2015. **a**, **c**, **e** The correlation of daily Tmax anomalies between CHIRTS and each of the reanalysis datasets. **b**, **d**, **f** The correlation between annual maximum temperatures between CHIRTS and each of the reanalysis datasets

This disagreement about extreme temperatures in tropical regions is further demonstrated in the different annual climatologies of each dataset. We calculated the monthly average of Tmax for the time period of 1983–2015 and identified the month with the highest average Tmax value. Figure 5 shows the absolute value of the difference, in months, between the hottest month in each dataset. So for example, if ERA5 indicated that August was the hottest month in one location and CHIRTS indicated that June was the hottest
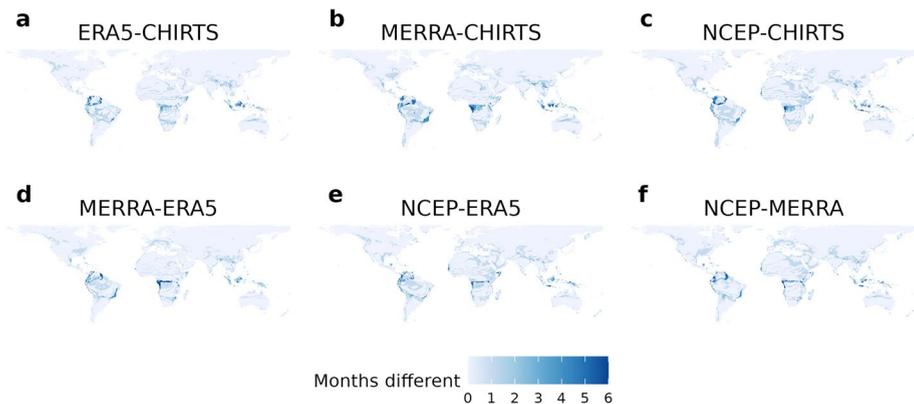
Fig. 5 Difference in monthly climatologies of Tmax (1983–2015). Each map depicts the number of months difference between the hottest month in the two datasets being compared

month, this would be a difference of 2 months. While there is strong agreement again in the extratropics on the hottest month of the year, there is strong disagreement in the tropics between all datasets, with some datasets even differing by 6 months in their climatologically hottest month of the year.

CHIRTS is derived in part from ERA5 data, and therefore, it is not surprising that Fig. 4a shows the highest correlations in daily Tmax anomalies between these datasets. However, Figs. 4b and 5a demonstrate discrepancies between these datasets in the correlation of annual maximum temperatures and the yearly climatologies in some regions. CHIRTS uses satellite data to calibrate the absolute value temperature data, while relying on ERA5 for the temporal sequencing of events. Discrepancies between the datasets could happen in regions with low standard deviation in temperature (see Fig. 3a), where small changes in absolute values of temperature can affect the sequencing of the hottest years or months, or be regions in which CHIRTS and ERA5 have large absolute value differences, such as the Himalayan region highlighted in Fig. 2b.

3.3 Intercomparison of heatwave events

Lastly, we compare the timing and frequency of heatwave events in these datasets, using both the long-duration definition (Fig. 6) and the heat-humidity definition (Fig. 8). The WMO heatwave definition, of 5 consecutive days in which Tmax is greater than the average Tmax by more than 5°C, occurs frequently in the extratropics. Over the time period of 1983–2015, northern regions of the USA, Europe, and Asia experienced more than 100 of these events. However, most of the tropical regions of the world did not register even a single heatwave using this definition in that time period (Fig. 6).

All observational datasets agree on the general spatial pattern of heatwave frequency, and these general spatial patterns are consistent for shorter events. For
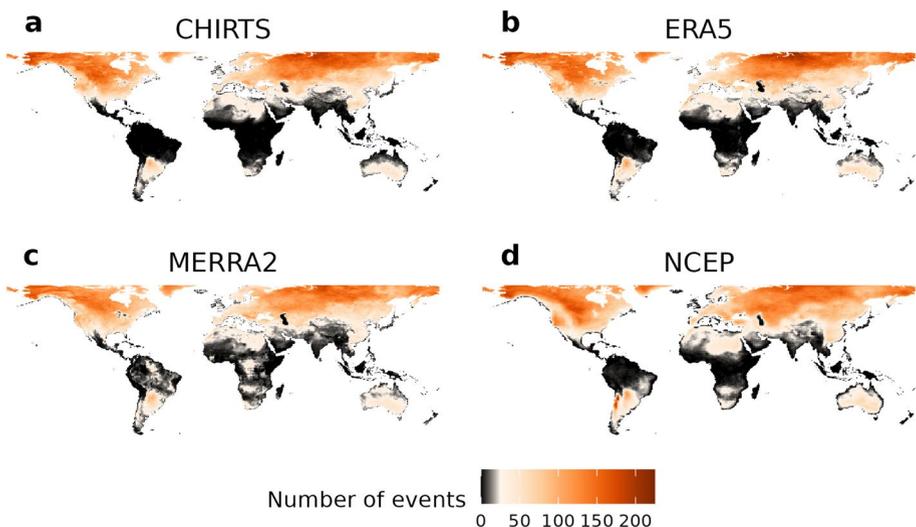

Fig. 6 Number of heatwaves between 1983 and 2015 using the WMO definition of 5 consecutive days of temperature anomalies of 5° or greater. Black indicates regions that have no heatwaves on record using that definition

example, if the duration is reduced to only 2 days (see Supplementary Information Figure SI-2), more events are observed while retaining this general spatial pattern. For events of longer duration and smaller magnitude, CHIRTS and ERA5 produce similar spatial patterns with few events observed, and the MERRA2 and NCEP reanalyses produce very few events anywhere in the world (Supplementary Information Figure SI-3).

For the extreme heat and humidity type of heatwave, we use the heat index, which calculates an "apparent temperature" by adjusting the air temperature based on the relative humidity in the ERA5 reanalysis and the NCEP reanalysis. The absolute values of the 99th percentile of the daily maximum heat index are unrealistic in some parts of the world in the NCEP reanalysis (Fig. 7b), and there is stronger agreement between ERA5 and NCEP in the 99.9th percentile of the daily minimum values of the heat index (Fig. 6c, d).

Areas of the world that register the highest absolute values of the HI are broadly consistent between the two models, including Pakistan, northern Australia, parts of the Sahel, the Arabian Peninsula, and central South America. These are also the regions that show the highest frequency of exceeding dangerous heat index thresholds as their maximum daily value and have registered days in which the heat index does not fall below dangerous thresholds, such as 41.

As found in the intercomparison of extreme temperatures, there is little consistency between datasets in the timing of extreme heat index events in the tropics. Figure 8 depicts the simple Pearson correlation of daily heat index anomalies between ERA5 and NCEP reanalyses, which show strong correlations in the extratropics and weak correlations in the tropics. Correlations tend to be stronger for daily minimum heat index anomalies (Fig. 8d) than daily maximum heat index anomalies (Fig. 8c). Within each model, however, the correlation between daily maximum and minimum heat index anomalies is weak in the tropics and strong in much of the extratropics (Fig. 8a, b), indicating that neither model is finding persistent high heat index values over the course of a day in the tropics. The annual maximum heat index event is also poorly correlated in much of the world, with the highest correlations in the northern extratropics.
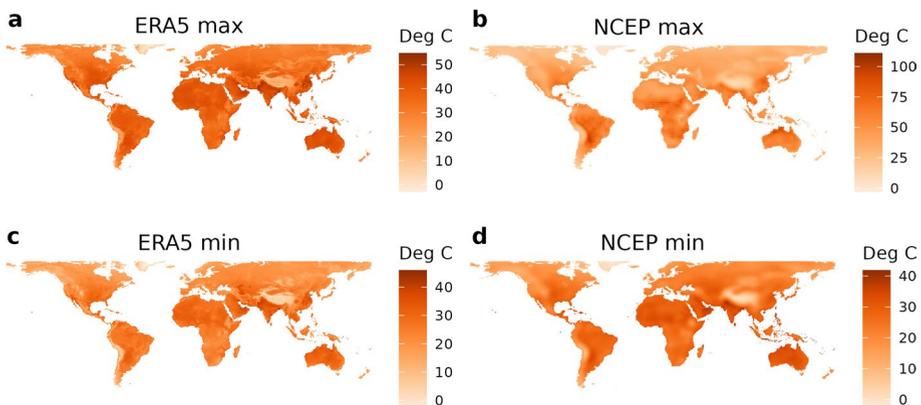

Fig. 7 99.9th percentile of the daily max and min heat index for the ERA5 and NCEP reanalyses. The 99.9th percentile of the daily maximum value of HI is plotted in the top row for the **a** ERA5 and **b** NCEP reanalysis and for the daily minimum value of HI in the bottom row for **c** ERA5 and **d** NCEP. Note that the scales are different across the figures
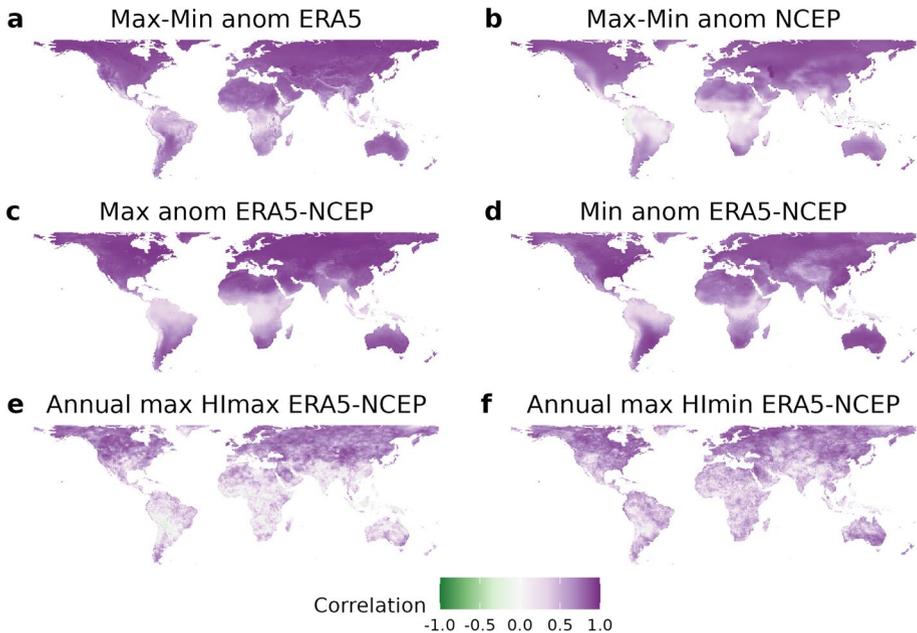
Fig. 8 Correlations of the heat index over time in the ERA and NCEP datasets. The first row depicts correlations of the maximum and minimum daily HI within each model: ERA (**a**) and NCEP (**b**). The second row depicts the correlations between ERA and NCEP for **c** anomalies of daily maximum HI and **d** anomalies of daily minimum HI. The third row depicts the correlations of annual maximums between ERA and NCEP for **e** daily maximum HI and **f** daily minimum HI

The regional differences in results are most apparent when aggregated across climate zones. Figure 9 depicts the average correlations in Figs. 4e and 8c across the Koppen-Geiger climate zones. Tropical regions have correlations averaging less than 0.5, while other regions of the world are higher than 0.75. Discrepancies by climate zone are also apparent in the count of heatwave events using the two definitions explored here. Using the WMO definition of 5 consecutive days of high temperatures, continental and polar regions of the world experience 80 and 90 events on average over the 1983–2015 period, while tropical regions experience almost no such events.

This is the opposite for extreme heat-humidity thresholds. Using a threshold value of 41°C (danger) for the heat index, regions classified as having a dry climate register an average of almost 300 events in the time period 1980–2021, while almost 0 events are recorded on average in the continental and polar climate regions. The ability for dry climate zones to reach high values of the heat index is likely due to the land surface feedbacks that allow for high temperature spikes in these regions, and the contribution of humidity to this index is likely irrelevant in those events. On the other hand, there are approximately 100 events on average in tropical and temperate regions. This pattern is similar for daily minimum heat index values above 32°C.
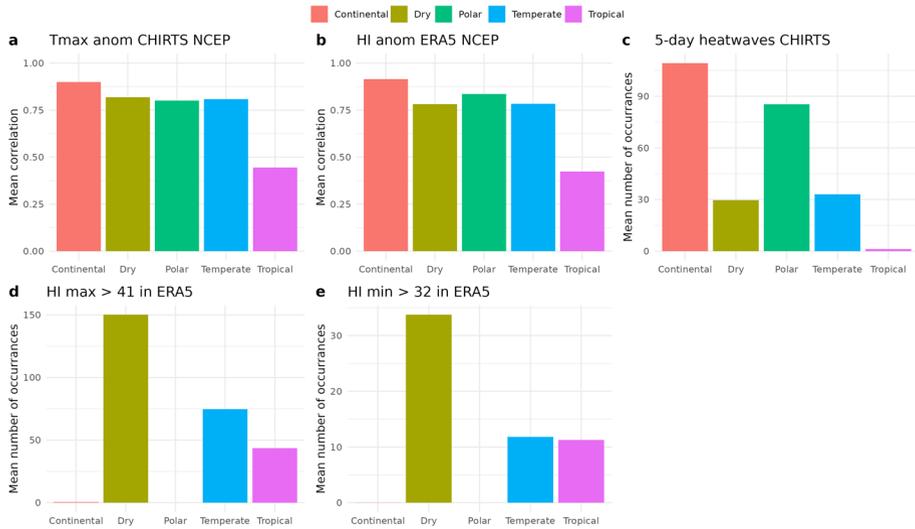
Fig. 9 Aggregations of extreme temperature results by Koppen climate classification regions. Average correlation by region of daily Tmax anomalies CHIRTS-NCEP (**a**) and average correlation by region of daily heat index anomalies ERA5-NCEP. Average number of occurrences of different extreme heat metrics by region: WMO-defined 5-day heatwave events (**c**), daily maximum heat index greater than the "danger" threshold of 41°C, and daily minimum heat index greater than the "extreme caution" threshold of 32°C

4 Discussion and conclusions

The four observational datasets examined here have similar values for daily maximum temperatures globally, although the three reanalyses demonstrate a cold bias over mountainous areas relative to the CHIRTS merged station-satellite temperature data. The shape of the distribution and magnitudes of extreme temperature events are also well aligned between all four datasets.

The timing of extreme temperature events, however, is consistent between datasets in the extratropics and entirely different in the tropics. This is likely due to the small variability of temperature in the tropics relative to the extratropics. There is little correlation between datasets for anomalously hot daily temperatures, annual maximum temperatures, daily maximum heat index, or daily minimum heat index. This indicates that the processes responsible for any anomalously hot events in the tropics are not well represented in the reanalyses.

When it comes to heatwave definitions, the WMO definition of a multi-day heatwave is an unusual event in the extratropics but never occurs in most of the tropics. This is likely due to the fact that multi-day extremes are generated by atmospheric blocking patterns that do not occur in the tropics, and they are often exacerbated by dry land surface feedbacks that do not occur in wetter regions of the world. There is a low correlation between daily maximum and daily minimum heat index values in the tropics as well, indicating that persistence of high heat index values is also not common, and the highest daily maximum heat index values are not happening on the same days as the highest daily minimum heat index values.

Therefore, any attempt to generalize heatwave definitions from the extratropics to the tropics should be avoided, because the dynamics of extreme heat and the driving mechanisms are likely to be dramatically different in the tropics. Multi-day persistent thresholds, such as the WMO heatwave definition, do not occur in much of the tropical world, and absolute heat/humidity thresholds are exceeded so often that they might not be considered extreme events.

We recommend that future research on extreme heat consider different heatwave definitions for different climate zones globally, selecting appropriate definitions for different Koppen-Geiger climate zones. In "continental" and "polar" climate regions, multi-day persistent heatwaves occur while high HI values do not, and therefore, impacts are most likely to be associated with multi-day definitions or high anomalies of the HI (rather than absolute values). In these regions, both CHIRTS and the reanalysis datasets would be useful to characterize these events. "Dry" climate regions can see the highest high absolute values of temperature, and therefore, multi-day events or single-day very high temperature or HI events are likely to be relevant. In "temperate" regions, multi-day heatwaves and high heat index values both occur, and it would be prudent to explore the types of impacts that are associated with each type of event in these regions. In "tropical" regions, there are few, if any, instances of multi-day heat events, and therefore, high HI values are likely to be most relevant for impacts in these regions. Because of the lack of agreement between datasets, it would be prudent to triangulate temperature information with local weather station records. This study was limited to the intercomparison of global gridded temperature datasets, and further studies comparing results with station data can indicate relative accuracy against stations in tropical regions.

This points to the need for locally developed heatwave metrics for different parts of the world, identifying which types of extreme heat, duration, and humidity are associated with different types of societal outcomes in that location. This study did not review or analyze data on health outcomes of these extreme heat metrics, which we recommend as a subject for further research in understudied regions. Thiaw et al. (2022) provide an outline for stakeholder engagement for the development of heat risk forecasts in the Sahel. However, the low consistency between datasets in the timing of anomalously hot events in the tropics should also be a cause for concern in heat-health research. Because the datasets do not agree on the timing of specific hot events, any studies attempting to associate these temperature records with health or other societal outcomes are likely to find null relationships or spurious relationships. Station data can be used in these cases, but in cases with no station data, we recommend that researchers compare results for different gridded dataset products to estimate the uncertainty in the heat-health relationships.

Data availability The datasets analyzed in the current study are available in the following locations:

- CHIRTS: https://www.chc.ucsb.edu/data/chirtsdaily
- ERA5-Land: https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=overview
- NCEP/DOE Reanalysis2: https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html
- MERRA2: https://disc.gsfc.nasa.gov/

Declarations

Competing interests The authors declare no competing interests.

Disclaimer The opinions expressed herein are those of the authors and do not necessarily reflect the views of the U.S. Agency for International Development.

References

Anderson GB, Bell ML, Peng RD (2013) Methods to calculate the heat index as an exposure metric in environmental health research. Environ Health Perspect 121(10):1111–1119

Birch CE, Jackson LS, Finney DL, Marsham JM, Stratton RA, Tucker S et al (2022) Future changes in African heatwaves and their drivers at the convective scale. J Climate 35(18):5981–6006

Bryant C, Wheeler NR, Rubel F, French RH (2017) kgc: Koeppen-Geiger climatic zones. R package version 1.0.0.2. https://cran.r-project.org/package=kgc. Accessed 25 Sept 2023

Campbell S, Remenyi TA, White CJ, Johnston FH (2018) Heatwave and health impact research: a global review. Health Place 53:210–218. https://doi.org/10.1016/j.healthplace.2018.08.017

Chambers J (2020) Global and cross-country analysis of exposure of vulnerable populations to heatwaves from 1980 to 2018. Clim Change 163(1):539–558

Coffel ED, Horton RM, De Sherbinin A (2017) Temperature and humidity based projections of a rapid rise in global heat stress exposure during the 21st century. Environ Res Lett 13(1):014001

Collow AM, Shields CA, Guan B, Kim S, Lora JM, McClenny EE et al (2022) An overview of ARTMIP's Tier 2 Reanalysis Intercomparison: uncertainty in the detection of atmospheric rivers and their associated precipitation. J Geophys Res Atmos 127(8):e2021JD036155

Coughlan de Perez E, Van Aalst M, Bischiniotis K, Mason S, Nissan H, Pappenberger F et al (2018) Global predictability of temperature extremes. Environ Res Lett 13(5):054017

Domeisen DI, Eltahir EA, Fischer EM, Knutti R, Perkins-Kirkpatrick SE, Schär C et al (2023) Prediction and projection of heatwaves. Nature Reviews Earth & Environment 4(1):36–50

Faye M, Dème A, Diongue AK, Diouf I (2021) Impact of different heat wave definitions on daily mortality in Bandafassi. Senegal PloS one 16(4):e0249199

Funk C, Peterson P, Peterson S, Shukla S, Davenport F, Michaelsen J et al (2019) A high-resolution 1983–2016 T max climate data record based on infrared temperatures and stations by the Climate Hazard Center. J Climate 32(17):5639–5658

IPCC (2022) Summary for policymakers. In: Pörtner H-O, Roberts DC, Poloczanska ES, Mintenbeck K, Tignor M, Alegría A, Craig M, Langsdorf S, Löschke S, Möller V, Okem A (eds) Climate change 2022: impacts, adaptation, and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change

IPCC (2023) Summary for policymakers. In: Core Writing Team, Lee H, Romero J (eds) Climate change 2023: synthesis report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, pp 1–34. https://doi.org/10.59327/IPCC/AR6-9789291691647.001

Jeong DI, Cannon AJ, Yu B (2022) Influences of atmospheric blocking on North American summer heat-waves in a changing climate: a comparison of two Canadian Earth system model large ensembles. Clim Change 172(1-2):5

Harrington LJ, Otto FE (2020) Reconciling theory with the reality of African heatwaves. Nature Climate Change 10(9):796–798

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J et al (2020) The ERA5 global reanalysis. Q J Roy Meteorol Soc 146(730):1999–2049

Heo S, Bell ML, Lee JT (2019) Comparison of health risks by heat wave definition: applicability of wet-bulb globe temperature for heat wave criteria. Environ Res 168:158–170

Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. Meteorol Z 15:259–263. https://doi.org/10.1127/0941-2948/2006/0130

Long CS, Fujiwara M, Davis S, Mitchell DM, Wright CJ (2017) Climatology and interannual variability of dynamic variables in multiple reanalyses evaluated by the SPARC Reanalysis Intercomparison Project (S-RIP). Atmospheric Chemistry and Physics 17(23):14593–14629

Lorenz C, Kunstmann H (2012) The hydrological cycle in three state-of-the-art reanalyses: intercomparison and performance analysis. J Hydrometeorol 13(5):1397–1420

Marx W, Haunschild R, Bornmann L (2021) Heat waves: a hot topic in climate change research. Theor Appl Climatol 146(1-2):781–800

Mora C, Dousset B, Caldwell IR, Powell FE, Geronimo RC, Bielecki CR et al (2017) Global risk of deadly heat. Nat Clim Change 7(7):501–506

Muñoz-Sabater J, Dutra E, Agustí-Panareda A, Albergel C, Arduini G, Balsamo G, ... Thépaut JN (2021) ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. Earth Syst Sci Data 13(9):4349–4383

NOAA, The National Weather Service (n.d.) "What is the heat index?" https://www.weather.gov/ama/heatindex. Accessed 25 Sept 2023

Salathé EP Jr, Beggs A, McJunkin C, Sandhu S (2023) The relative warming rates of heat events and median days in the Pacific Northwest from observations and a regional climate model. J Climate:1–24

Smith TT, Zaitchik BF, Gohlke JM (2013) Heat waves in the United States: definitions, patterns and trends. Clim Change 118:811–825

Thiaw WM, Bekele E, Diouf SN, Dewitt DG, Ndiaye O, Ngom Ndiaye MK, Ndiaye PN, Diene N, Diouf M, Diaw A, Diop S, Badj F, Diouf A (2022) Toward experimental heat–health early warning in Africa. Bull Am Meteorol Soc 103(8):E1843–E1860. https://doi.org/10.1175/bams-d-20-0140.1

Thompson V, Mitchell D, Hegerl GC, Collins M, Leach NJ, Slingo JM (2023) The most at-risk regions in the world for high-impact heatwaves. Nat Commun 14(1):2152

Torralba V, Doblas-Reyes FJ, Gonzalez-Reviriego N (2017) Uncertainty in recent near-surface wind speed trends: a global reanalysis intercomparison. Environ Res Lett 12(11):114019

Wang P, Yang Y, Xue D, Qu Y, Tang J, Leung LR, Liao H (2023) Increasing compound hazards of tropical cyclones and heatwaves over southeastern coast of China under climate warming. J Climate 36(7):2243–2257

World Health Organization (2018) Heat and health. https://www.who.int/news-room/fact-sheets/detail/climate-change-heat-and-health. Accessed 01 May 2023

Xu Z, FitzGerald G, Guo Y, Jalaludin B, Tong S (2016) Impact of heatwave on mortality under different heatwave definitions: a systematic review and meta-analysis. Environ Int 89:193–203