RESEARCH ARTICLE

Key Points:

- Extreme temperature events are detected with Gaussian Mixture Models to follow a multimodal rather than a unimodal distribution
- 10-year temperature extremes will occur 13.6 times more frequently under 3.0°C future warming
- Colder days are getting warmer faster than hotter days in high latitudes, whereas it is the opposite for many regions in low latitudes

Author Contributions:

Conceptualization: Aytaç Paçal, Birgit Hassler, Katja Weigel, M. Levent Kurnaz, Veronika Eyring
Formal analysis: Aytaç Paçal
Funding acquisition: Veronika Eyring
Investigation: Aytaç Paçal, Birgit Hassler
Methodology: Aytaç Paçal, Birgit Hassler, Katja Weigel, M. Levent Kurnaz, Michael F. Wehner, Veronika Eyring
Project Administration: Veronika Eyring
Software: Aytaç Paçal
Supervision: Birgit Hassler, Katja Weigel, Veronika Eyring
Validation: Aytaç Paçal
Visualization: Aytaç Paçal

Detecting Extreme Temperature Events Using Gaussian Mixture Models

Aytaç Paçal[1,2] ![ORCID], Birgit Hassler[1] ![ORCID], Katja Weigel[1,2] ![ORCID], M. Levent Kurnaz[3] ![ORCID], Michael F. Wehner[4] ![ORCID], and Veronika Eyring[1,2] ![ORCID]

[1]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany, [2]University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany, [3]Center for Climate Change and Policy Studies, Boğaziçi University, Istanbul, Turkey, [4]Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

Abstract Extreme temperature events have traditionally been detected assuming a unimodal distribution of temperature data. We found that surface temperature data can be described more accurately with a multimodal rather than a unimodal distribution. Here, we applied Gaussian Mixture Models (GMM) to daily near-surface maximum air temperature data from the historical and future Coupled Model Intercomparison Project Phase 6 (CMIP6) simulations for 46 land regions defined by the Intergovernmental Panel on Climate Change. Using the multimodal distribution, we found that temperature extremes, defined based on daily data in the warmest mode of the GMM distributions, are getting more frequent in all regions. Globally, a 10-year extreme temperature event relative to 1985–2014 conditions will occur 13.6 times more frequently in the future under 3.0°C of global warming levels (GWL). The frequency increase can be even higher in tropical regions, such that 10-year extreme temperature events will occur almost twice a week. Additionally, we analyzed the change in future temperature distributions under different GWL and found that the hot temperatures are increasing faster than cold temperatures in low latitudes, while the cold temperatures are increasing faster than the hot temperatures in high latitudes. The smallest changes in temperature distribution can be found in tropical regions, where the annual temperature range is small. Our method captures the differences in geographical regions and shows that the frequency of extreme events will be even higher than reported in previous studies.

Plain Language Summary Extreme temperature events are unusual weather conditions with exceptionally low or high temperatures. Traditionally, the temperature range was determined by assuming a single distribution, which describes the frequency of temperatures at a given climate using their mean and variability. This single distribution was then used to detect extreme weather events. In this study, we found that temperature data from reanalyses and climate models can be more accurately described using a mixture of multiple Gaussian distributions. We used the information from this mixture of Gaussians to determine the cold and hot extremes of the distributions. We analyzed their change in a future climate and found that hot temperature extremes are getting more frequent in all analyzed regions at a rate that is even higher than found in previous studies. For example, a global 10-year event will occur 13.6 times more frequently under 3.0°C of global warming. Furthermore, our results show that the temperatures of hot days will increase faster than the temperature of cold days in equatorial regions, while the opposite will occur in polar regions. Extreme hot temperatures will be the new normal in highly populated regions such as the Mediterranean basin.

1. Introduction

Increasing levels of atmospheric carbon dioxide (CO_2) concentration unequivocally transformed the earth's climate (IPCC, 2021). This surplus of CO_2 in the atmosphere contributes to the greenhouse effect, and by increasing the mean and the variability of global temperatures, it amplifies the risk of high-impact temperature extremes (Baker et al., 2018). The effects of anthropogenic global warming led to the emergence of heat extremes that would not have occurred previously (Robinson et al., 2021). This means that unprecedented heat extremes like the 2010 Russian heatwave or the 2021 Western North America heatwave would have likely not happened without the warming effect (Christidis et al., 2015; Rahmstorf & Coumou, 2011; Thompson et al., 2022). The latter was found to be a remarkable four standard deviations away from the mean (Thompson et al., 2022). The Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) concluded that human influence on the climate system is unequivocal (Eyring et al., 2021) and *virtually certain* to be the main driver of the changes in

Writing – original draft: Aytaç Paçal, Birgit Hassler, Katja Weigel
Writing – review & editing: Aytaç Paçal, Birgit Hassler, Katja Weigel, M. Levent Kurnaz, Michael F. Wehner, Veronika Eyring

hot and cold extremes (Seneviratne et al., 2021). It introduced more frequent and intense hot extremes since the 1950s on land areas while a decrease in cold extremes is observed (IPCC, 2021). Several studies found that the duration, frequency, and intensity of extreme events will increase, and extreme events will be introduced at new locations (Kharin et al., 2013; Mallick et al., 2022; Perkins-Kirkpatrick & Lewis, 2020; Pfleiderer et al., 2019; Rahmstorf & Coumou, 2011; Raymond et al., 2020; Seneviratne et al., 2012, 2021; Sillmann, Kharin, Zhang, et al., 2013; Sillmann, Kharin, Zwiers, et al., 2013; Vogel, Hauser, & Seneviratne, 2020). As the number of occurrences of heat extremes like the 2003 European heatwave and their duration increase, the socio-economic burden of climate change poses a threat to societies (Demiroglu et al., 2020; García-León et al., 2021; Meehl & Tebaldi, 2004; Perera et al., 2020; Robine et al., 2008; Seneviratne et al., 2021).

The warming of the climate causes different changes in different regions. Tropics, polar regions and the Middle East and North Africa (MENA) region, are hot spots of notable climate trend shifts (Hao et al., 2018; Y. Zhang et al., 2022). Iyakaremye et al. (2022) have shown that an abrupt shift in the daily maximum temperatures occurred in Africa in the last two decades compared to the previous 20 years, which introduced more frequent and intense hot days. Moreover, regions in Africa will face a higher increase in temperatures compared to the rest of the globe. Iyakaremye et al. (2021) found that the annual maximum of daily maximum temperatures over Africa is expected to increase by 1.6/2.2°C in the future, while global temperatures are projected to rise by 1.5/2.0°C during the same period. In the MENA region, the frequency and intensity of heatwaves will highly increase by the end of the century under a business-as-usual pathway scenario, which will affect about half of the MENA population (Lelieveld et al., 2016; Ozturk et al., 2021; Zittis et al., 2021). The number of occurrences of exceptionally hot summers, which have 2–4°C hotter temperatures than the long-term average, has also increased from a single event between 1951 and 1980 to five events between 2001 and 2010 in Central and Eastern Europe, where the 2010 heatwave was the hottest and longest event with the largest geographical extent that ever occurred over Europe (Guerreiro et al., 2018; Twardosz & Kossowska-Cezak, 2013). Similarly, other studies also found that the temperature extremes in Europe will increase 20-fold at the end of the century, compared to 1961–1990 (Barriopedro et al., 2011; Nikulin et al., 2011; Schär et al., 2004). Over the Americas, the dry and hot extremes showed an increase both in frequency and spatial scope over the past 122 years (Alizadeh et al., 2020; Cai et al., 2014).

Correctly characterizing the temperature distributions to analyze extreme events is a still-continuing issue as extremes are by definition rare events, and several studies showed that the assumption of distributions or a stationary climate often underestimates the observed heat records (Benestad, 2004; Schär et al., 2004; Anderson & Kostinski, 2010; Fischer & Schär, 2010; Barriopedro et al., 2011; C. Li et al., 2019; Loikith & Neelin, 2019). Thompson et al. (2022) characterized extreme events by calculating a daily extreme index which is the difference between the daily maximum temperature and mean daily maximum temperature divided by the standard deviation. With the assumption of a normal distribution, they found that the 2021 North American heatwave was one of the most extreme events with 4 standard deviations from the mean. Moreover, the authors projected that 20% of the weather risk attribution forecast regions (Stone, 2019) will experience extreme events that are four standard deviations from the means in the future. Other studies found that hot summers will be the norm, that is, mean temperatures exceed the temperature of the historically hottest summer, within the next 1–2 decades (Lewis et al., 2017; Mueller et al., 2016; Vogel, Hauser, & Seneviratne, 2020; Vogel, Zscheischler, et al., 2020).

Common indices to monitor and analyze climate extremes that are used in the climate community at the moment, such as ETCCDI (the Expert Team on Climate Change Detection and Indices), are mostly based on daily mean near-surface air temperature or daily maximum near-surface air temperature (X. Zhang et al., 2011; Alexander et al., 2006). Two standard approaches to detect extreme events are the percentile-over-threshold (POT) and the block maxima method. The block maxima method groups data into an equal length of blocks, for example, month, season, or year, and use the maximum temperature value of each block to fit the data. The POT method defines a threshold, for example, percentiles, and uses all temperature values above this threshold in the analysis. Choosing the percentiles for defining extremes is not trivial as the temperature extremes have a strong seasonality and temporal dependence (Huang et al., 2016). The block maxima method is more commonly used in climate studies because of its simplicity with monthly, seasonal or annual block periods for fitting generalized extreme value (GEV) distribution to temperature and precipitation extremes (Ben Alaya et al., 2020; IPCC, 2021; Kharin et al., 2013; C. Li et al., 2021; Paciorek et al., 2018; Wang et al., 2016; Wehner et al., 2018). The block maxima method, however, does not use all available data, as calculating a single maximum value from a block period throws out the rest of the data. To be approximated by the GEV distribution, the blocks are assumed to

be long enough and "max-stable," which means that if you take the maximum of a group of values selected from a specific GEV distribution, the result will be GEV distributed with the same shape parameter (Ben Alaya et al., 2020; Huang et al., 2016). However, these assumptions might not be valid for all possible use cases or all possible variables. For example, GEV is not the best fit for shorter block lengths as the fit improves with increasing block size (Ben Alaya et al., 2020; Wang et al., 2016). Ben Alaya et al. (2020) argued that the identically distributed random variables assumption of extreme value theory might be problematic for extreme precipitation events. They considered a mixture of GEV distributions to fit precipitation data to demonstrate that the mixture distribution could be a potential explanation for the instability of annual maxima. Kollu et al. (2012) tested wind speed characteristics using mixture probability distribution functions (PDF). They found that conventional PDFs are inadequate to describe wind speed distributions compared to the mixture distributions that they used in the study. A mixture of Gaussians was used by Shin et al. (2022) to describe the distribution of the daily thermal comfort index in South Korea, an index that has a strong seasonality. Ice surface temperature data follows a clear multimodal distribution, according to Clarkson et al. (2022). They also found that a unimodal distribution fit is particularly poor at modeling the tail probabilities. Probability distributions with one and two components are called unimodal and bimodal, respectively, whereas distributions with multiple (two or more) components are called multimodal distributions.

The temperature distributions are expected to move toward warmer temperatures and to change their shape with changing means and standard deviations (IPCC, 2021). Also, the assumption of distribution might not be correct for all geographical regions as daily weather variables show a distinct non-Gaussianity (Kodra & Ganguly, 2014; Linz et al., 2018; Perron & Sura, 2013; Sardeshmukh et al., 2015; Tamarin-Brodsky et al., 2019; E. M. Volodin & Yurova, 2012). Furthermore, several studies found that daily mean, daily maximum and real forecast data of 2m temperatures show bimodal features (Bertossa et al., 2021; Cho & Jeong, 2016; Donat & Alexander, 2012; Grace, 1995; Wilks, 2002). These changes, shifts and bimodalities in the temperature distributions affect the probabilities in the tails. As extreme events are rare events that lie in the tails of a distribution, correctly describing the tails is very important for extreme event detection. Even though the block maxima method is widely used in studies which used block sizes large enough to converge asymptotically to GEV distributions, a GEV distribution is not well suited to describe extreme value data when the bimodality is apparent or block sizes are short (Ben Alaya et al., 2020; Knoben et al., 2019; Sardeshmukh et al., 2015; Wang et al., 2016). Therefore, the properties of the entire probability distribution, that is, mean, standard deviation and shape, are needed to get the tail properties right (Sardeshmukh et al., 2015). A distribution can be described by not only the mean and the standard deviation but also skewness and kurtosis. Donat and Alexander (2012) found that daily minimum and maximum temperatures have significantly shifted toward higher values and skewed toward the hotter part of the distribution. They highlighted that the changes in extremes are related not only to the means but also to other parameters of the daily temperature distribution. Sardeshmukh and Sura (2009) found a parabolic relationship between kurtosis and skewness that cause the non-Gaussianity of the observed daily weather anomalies. Similarly, Tamarin-Brodsky et al. (2022) used a mixture model with three Gaussians to describe the PDF of near-surface atmospheric temperature to analyze the relationship between kurtosis and skewness, as they are important to explain how the tails of the distribution change. They found that two- and three-Gaussian models are useful to explain the relationship between kurtosis and skewness.

In the study presented here, our approach is to utilize the entire temperature distribution to detect extreme events. We implemented Gaussian Mixture Models (GMM), which describe the probability distribution function of data points as a mixture of Gaussian distributions. We determined the number of Gaussian components in the temperature distribution of each grid cell of 46 land regions defined by the IPCC using daily near-surface maximum air temperature data from the historical and future Coupled Model Intercomparison Project Phase 6 (CMIP6) simulations. This choice was supported by previous studies which found distinct bimodality in daily weather variables (Bertossa et al., 2021; Cho & Jeong, 2016; Donat & Alexander, 2012; Grace, 1995; Wilks, 2002) and was verified by applying the same analysis to the European Centre for Medium-Range Weather Forecasts Reanalysis 5th Generation (ECMWF-ERA5) data for the same historical time period (1985–2014). The parameters from the determined distribution components, namely mean, standard deviation and weight, were used to calculate the change in the return period of extreme temperature events between the historical and future periods determined by using global warming levels (GWL). In a stationary climate, the return period of an event describes the average time between the occurrences of a certain event of a defined size. In this study, we analyzed 1-year, 5-year, 10-year and 20-year events, where an n-year event has an occurrence probability of $1/n$ as the climate is not

stationary. Hence, these event magnitudes change as time progresses, where an *n*-year event means that the event in question would be expected to occur once in every *n* years. We only calculated return periods equal to or less than the available future data period to prevent overestimating the return periods of extreme events, since GMM distributions are not bounded. Section 2 presents the climate data and warming levels used in this study, as well as the analyzed regions, and explains the methodology of detecting extreme event return periods by using GMM. Section 3 shows our results obtained using the GMM method for all analyzed IPCC land regions, and Section 4 finalizes the paper with a summary and discussion.

2. Data and Methodology

2.1. Climate Data

For this study, we used multi-year daily near-surface maximum temperatures from the Coupled Model Intercomparison Project Phase 6 (CMIP6), and for which both the historical simulations and the simulations for Shared Socioeconomic Pathways (SSPs) 1–2.6, 2–4.5, 3–7.0, and 5–8.5 scenarios were available (Eyring et al., 2016; O'Neill et al., 2014, 2016). Additionally, the ECMWF-ERA5 data set was included for the 30-year time period (1985–2014) (Hersbach et al., 2018). Table 1 shows the list of models and their resolutions. The 30-year time period from 1985 to 2014 from historical simulations is used as the base to calculate the return values of extreme temperature events, that is, 1-year, 5-year, 10-year and 20-year events. The GWL, as introduced in the IPCC AR6 report, are used to assess the changes in future climate in line with the warming levels defined in the Paris Agreement which are compared to the pre-industrial period (IPCC, 2021). The future period for each model is defined as a 20-year period between 2015 and 2100 when the central year of the running window of the global daily near-surface temperature mean of that model first exceeds 1.5°C, 2°C, 3°C, and 4°C relative to 1850–1900 global daily near-surface mean temperatures. We used the same GWL periods defined for and used in IPCC (Hauser et al., 2022; IPCC, 2021), similarly to Hajat et al. (2022) and Ribeiro et al. (2022). Therefore, we obtained the start and end years of 20-year GWL periods for each CMIP6 simulation from Hauser et al. (2022). Here, we used a longer historical base period (30 years) compared to future GWL periods (20 years) for the analysis to obtain more robust results. This decision was made based on the fact that GMM distributions have no bounds. Therefore, we focused our analysis solely on return periods shorter than our base period. By limiting our analysis to shorter return periods, we can mitigate the biases and outliers that may occur beyond the limits of the data sets. As some data sets did not exceed certain warming levels, they were excluded from the analysis (e.g., NOR-ESM2-MM was not used in calculations for 4°C warming under SPP5-8.5, as it did not exceed this level). Figure 1 shows the historical and future GWL periods for each CMIP6 model used in this study.

We extracted daily maximum near-surface air temperature for 30-year historical and 20-year future periods under GWL for each SSP individually for 46 IPCC land regions that are shown in Figure 1 (Iturbide et al., 2020). All data extraction and preprocessing in this study were performed by using the Earth System Model Evaluation Tool (ESMValTool) version 2.5.0, which is an open-source software package for analyzing and evaluating model simulations (Eyring et al., 2020; Lauer et al., 2020; Righi et al., 2020; Weigel et al., 2021). We extracted the daily maximum near-surface air temperature from each model for each region using shapefiles provided by IPCC (Iturbide et al., 2020), converted units from Kelvin to Celsius, and created a single spatiotemporal Network Common Data Form file for each region. The data were then ready to be used in the diagnostic script written in Python where the extreme events and their return periods were analyzed.

2.2. Return Period Analyses

For the return period calculation of extreme temperature events, that is, 1-, 5-, 10-, and 20-year events, we defined a temperature threshold for an event by calculating the standard deviation distance of the event temperature from the mean temperature in the past, that is, how many standard deviations away the event temperature *was* from the mean. We then applied this temperature threshold value to the future period but calculated its standard deviation distance from the mean using the parameters from the future distribution, that is, how many standard deviations away the event temperature *will be* from the mean. To test the underlying distribution shape of the daily near-surface maximum temperature distribution, we first analyzed data from individual grid cells of each climate model. We found that daily maximum near-surface air temperature data in climate grid cells usually do not follow a unimodal distribution, but rather follow a bimodal distribution, a probability distribution composed of two components.

Table 1
Reanalysis Data and CMIP6 Models Used in This Study to Detect Extreme Temperature Events

Model	Variant	Resolution (km)	Reference
ECMWF-ERA5	Reanalysis	25	Hersbach et al. (2018)
ACCESS-CM2	r1i1p1f1	250	Dix et al. (2019)
ACCESS-ESM1-5	r1i1p1f1	250	Ziehn et al. (2019)
AWI-CM-1-1-MR	r1i1p1f1	100	Semmler et al. (2018)
BCC-CSM2-MR	r1i1p1f1	100	Wu et al. (2018)
CanESM5	r1i1p1f1	500	Swart et al. (2019)
CNRM-CM6-1	r1i1p1f2	250	Voldoire (2018)
CNRM-CM6-1-HR	r1i1p1f2	50	Voldoire (2019)
CNRM-ESM2-1	r1i1p1f2	250	Seferian (2018)
EC-Earth3	r1i1p1f1	100	EC-Earth Consortium (EC-Earth) (2019a)
EC-Earth3-CC	r1i1p1f1	100	EC-Earth Consortium (EC-Earth) (2021)
EC-Earth3-Veg	r1i1p1f1	100	EC-Earth Consortium (EC-Earth) (2019b)
EC-Earth3-Veg-LR	r1i1p1f1	250	EC-Earth Consortium (EC-Earth) (2020)
FGOALS-g3	r1i1p1f1	250	L. Li (2019)
GFDL-ESM4	r1i1p1f1	100	Krasting et al. (2018)
HadGEM3-GC31-LL	r1i1p1f3	250	Ridley et al. (2019a)
HadGEM3-GC31-MM	r1i1p1f3	100	Ridley et al. (2019b)
INM-CM4-8	r1i1p1f1	100	von et al. (2019)
INM-CM5-0	r1i1p1f1	100	E. Volodin et al. (2019)
IPSL-CM6A-LR	r1i1p1f1	250	Boucher et al. (2018)
KACE-1-0-G	r1i1p1f1	250	Byun et al. (2019)
MIROC6	r1i1p1f1	250	Tatebe and Watanabe (2018)
MIROC-ES2L	r1i1p1f2	500	Hajima et al. (2019)
MPI-ESM1-2-HR	r1i1p1f1	100	Jungclaus et al. (2019)
MPI-ESM1-2-LR	r1i1p1f1	250	Wieners et al. (2019)
MRI-ESM2-0	r1i1p1f1	100	Yukimoto et al. (2019)
NESM3	r1i1p1f1	250	Cao and Wang (2019)
NorESM2-LM	r1i1p1f1	250	Seland et al. (2019)
NorESM2-MM	r1i1p1f1	100	Bentsen et al. (2019)
UKESM1-0-LL	r1i1p1f2	250	Tang et al. (2019)

Note. Climate models with spatial resolutions ranging from 50 to 500 km were used in the analyses. The first available ensemble members were chosen. The Reanalysis data set that has a resolution of 25 km was regridded to 100 km and used for evaluating modality.

To calculate the return periods of extreme events, we modeled the probability distribution of multi-year (30 years for the historical base period and 20 years for future GWL periods) daily near-surface maximum temperature data from a grid cell as mixtures of multiple Gaussian distributions, rather than a single Gaussian distribution. GMM is a probabilistic model that describes the data points in a population as a mixture of Gaussian distributions with unknown parameters which are the mean, standard deviation and weight of each Gaussian component, five parameters in total for a bimodal distribution. With this approach, we were able to analyze the change in the distribution, and accurately model the tails of the data compared to a unimodal distribution. When a unimodal distribution is fit to multi-year data with bimodality, it is likely that the resulting distribution will have a larger standard deviation to encompass both modes. This large standard deviation between a unimodal distribution fit to bimodal data can have significant implications for analyses, as the larger standard deviations of unimodal distributions tend to push the extreme events further away from what would be observed if the bimodality were properly accounted for. In other words, when one tries to calculate the threshold of an event as *n* sigma distance from
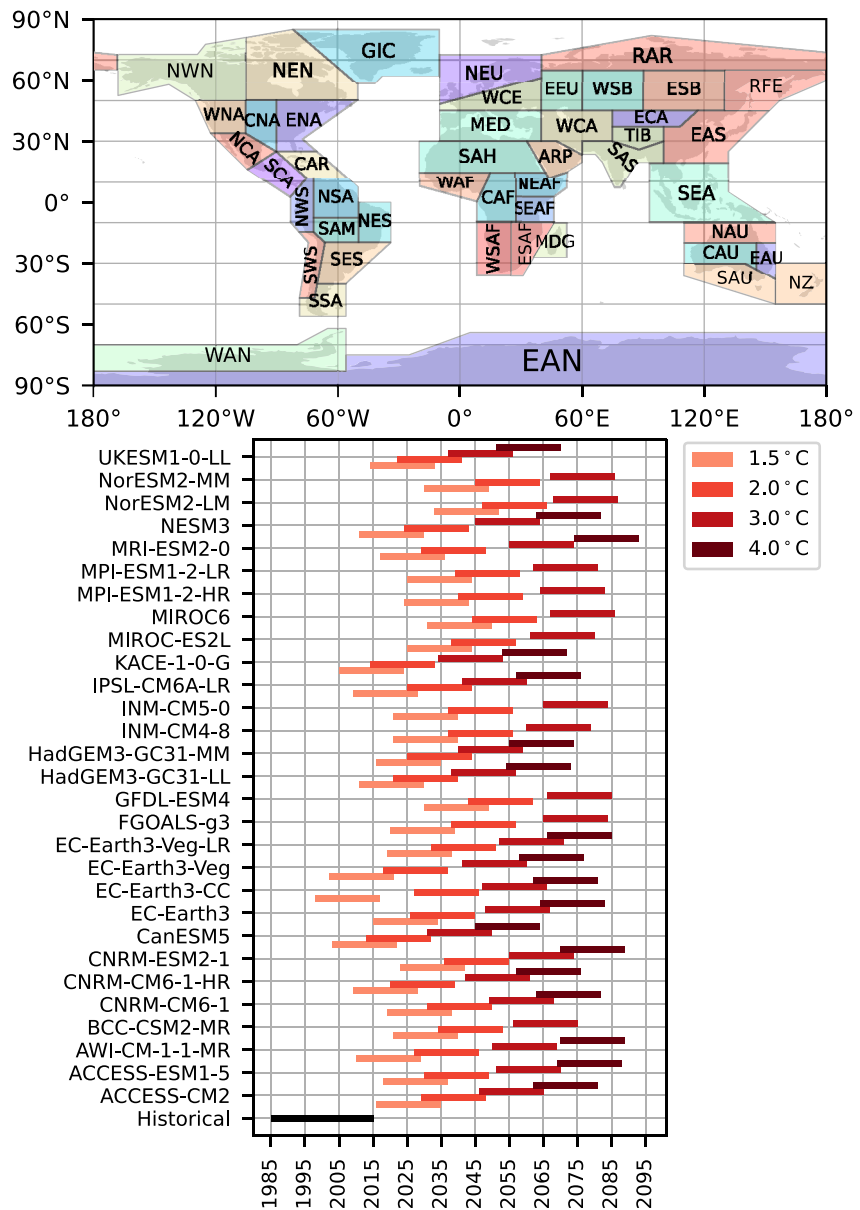
Figure 1. (top) We used 46 land regions defined in Iturbide et al. (2020). See Table S2 for region definitions. (bottom) Future periods of the CMIP6 models when the central year of the 20-year running window exceeds global warming levels (GWL) relative to 1850–1900 base for the SSP5-8.5 scenario are extracted using the data from Hauser et al. (2022). The colors in the graph go from light to dark, each color representing a different level of warming 1.5, 2, 3, and 4°C. These levels are expected to be exceeded around 2026, 2040, 2060, and 2070 respectively. The 30-year historical base period is indicated in gray. Note that different models have different time periods when they exceed the GWL. Future periods for other Shared Socioeconomic Pathway scenarios are presented in Table S3.

the mean, this threshold might be well beyond the maximum value of the distribution. Furthermore, a unimodal distribution fit will affect the measures of central tendency when bimodality exists in data (von Hippel, 2005). An example goodness-of-fit test for normal distribution, GEV distribution with different shape parameters and GMM distributions on the daily maximum temperature data from a random grid cell is presented in Supplementary Material Section 1. We used an unsupervised machine-learning package, the "GaussianMixture" package from open-sourced machine-learning library Scikit-learn, to compute the unknown parameters of the Gaussian components in a mixture that generates all observed data points (Pedregosa et al., 2011). We applied this package to the daily maximum near-surface air temperature data in each grid cell of the CMIP6 models. The "GaussianMixture" package first randomly assigns values to the component parameters and then uses the expectation-maximization

algorithm (EM) to converge their values. EM algorithm fits GMM to data by alternating between two steps, Expectation (E) and Maximization (M). In the E step, it randomly assumes components and calculates the probability of each point to be generated by that component. In the M step, the parameters are tweaked to maximize the likelihood found in the first step. It also uses the Bayesian Information Criteria (BIC) score, which is used to estimate the goodness-of-fit of a distribution and which accounts for both the likelihood function and the number of parameters. Then, the probability distribution function of the mixture model that was fit to multi-year daily near-surface temperature can be written as a linear summation of multiple Gaussian components:

$$p(x) = \sum_{k=1}^{K} \omega_k \mathcal{N}(x|\mu_k, \sigma_k) \tag{1}$$

$$\mathcal{N}(x|\mu_k, \sigma_k) = \frac{1}{\sigma_k \sqrt{2\pi}} \exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right) \tag{2}$$

$$\sum_{k=1}^{K} \omega_k = 1 \tag{3}$$

where K is the number of Gaussian components in the mixture. μ_k, σ_k, and ω_k are the mean, the standard deviation and the weight of the kth component, respectively. Implementing GMM to evaluate multi-year raw daily maximum temperatures allows us to investigate the long-term characteristics of the individual components. This method does not consider the temporal changes within one period, as they can be assumed to be negligible compared to the changes between different time periods. As shown in other studies, mean temperatures are increasing all over the globe (Eyring et al., 2020; IPCC, 2021; Robinson et al., 2021). Using the raw temperatures, we can analyze how the convergence or divergence of the peaks of the different Gaussian components affect the extremes compared to the used historical periods. In our analysis, we have disregarded three or more Gaussian components. This choice was supported by the value of the BIC score and the fact that increasing the number of components tends to cause overfitting, even though BIC scores penalize adding more parameters. In some cases, the BIC scores for the components showed close results for more than three components (see Figure S2 in Supporting Information S1). For instance, the lowest BIC score was reached for a mixture with seven Gaussian components for the distribution of temperatures in a grid cell. However, the highest change in BIC scores occurred when switching from one component to two components. Consequently, we used the gradient of BIC scores rather than using the lowest score. We selected the number of Gaussian components where the highest gradient change occurs in the BIC scores as the best fit. To further prevent overfitting, we also applied the following unimodality test after estimating the BIC scores: If the BIC score returned a bimodal distribution, then the parameters of the mixture distribution components were used for the unimodality test. As shown in Equation 4, if the difference between the means of Gaussian components was less than or equal to twice the minimum of standard deviations, then unimodal distribution was assumed, otherwise, the bimodal distribution fit for the data was kept. It is worth noting that this procedure had a tendency to favor fitting a unimodal distribution. However, after all these tests and checks, the majority of grid cells showed a clear bimodal distribution. For a bimodal distribution, hereafter we referred to the right (left) Gaussian component as "hot (cold) Gaussian" as shown in Figure 2).

$$|\mu_1 - \mu_2| \le 2\min(\sigma_1, \sigma_2) \tag{4}$$

First, we grouped grid cells of a region depending on their modality, either unimodal or bimodal, for each CMIP6 model, and calculated the percentages of grid modalities among all grid cells of a region for each CMIP6 model. We then determined the multi-model mean percentages of grid cell modalities of a region as shown in Figure 3. Additionally, we calculated the global multi-model mean percentage of grid cell modalities using all regions and CMIP6 models. We found that globally 89.58% of all grid cells follow a bimodal distribution in the historical period as shown in the white box in the upper center part of Figure 3. Furthermore, we analyzed the ECMWF-ERA5 data set for the same historical time period (1985–2014) to confirm whether bimodality is also found in data other than model simulations. We regridded the ECMWF-ERA5 data from a 25-km grid to a coarser 1-degree 100-km grid using the nearest neighbor method to have a similar resolution as many CMIP6 data sets. The ECMWF-ERA5 reanalysis data set shows similar results to the CMIP6 models: Globally 86.95% of all grid cells in the ECMWF-ERA5 reanalysis data set follow a bimodal distribution as shown in the white box in the upper center part of Figure 4, while only 13.05% of them follow a unimodal distribution. The temperature
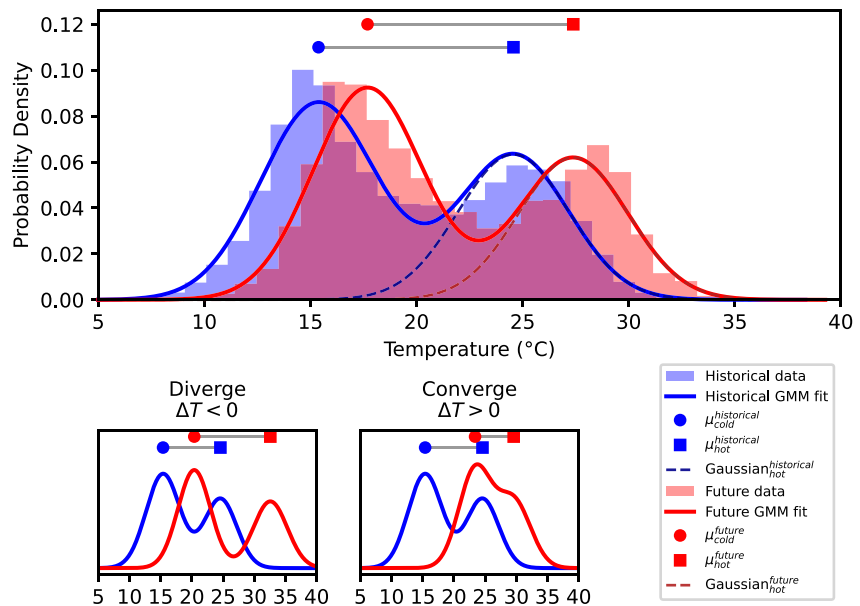
Figure 2. Exemplary bimodal distribution of daily maximum temperatures from a grid cell for the historical 30-year period of 1985–2014 (blue) and future 20-year global warming levels period (red). Blue and red lines show the corresponding Gaussian Mixture Models fit for the historical and future periods, respectively. The shape of the distribution is determined by the parameters of each Gaussian component, which are the means, standard deviations and weights. Here, the means of cold and hot Gaussian peaks are shown with blue (red) dots and squares for the historical (future) period, respectively. The hot Gaussian component used in the analysis is shown with a dashed blue (red) line for the historical (future) period. The bottom two plots show what convergence and divergence of the peaks mean based on the ΔT value.

distributions in the ERA5 and CMIP6 data sets show predominantly similar patterns across various regions, although certain exceptions are observed, particularly in South America. These differences can most likely be attributed to several factors. First, resampling of ERA5 data from a 25 km grid to a coarser 1-degree grid introduces a smoothing effect on the data, which would increase the unimodal grid cells. Additionally, biases in
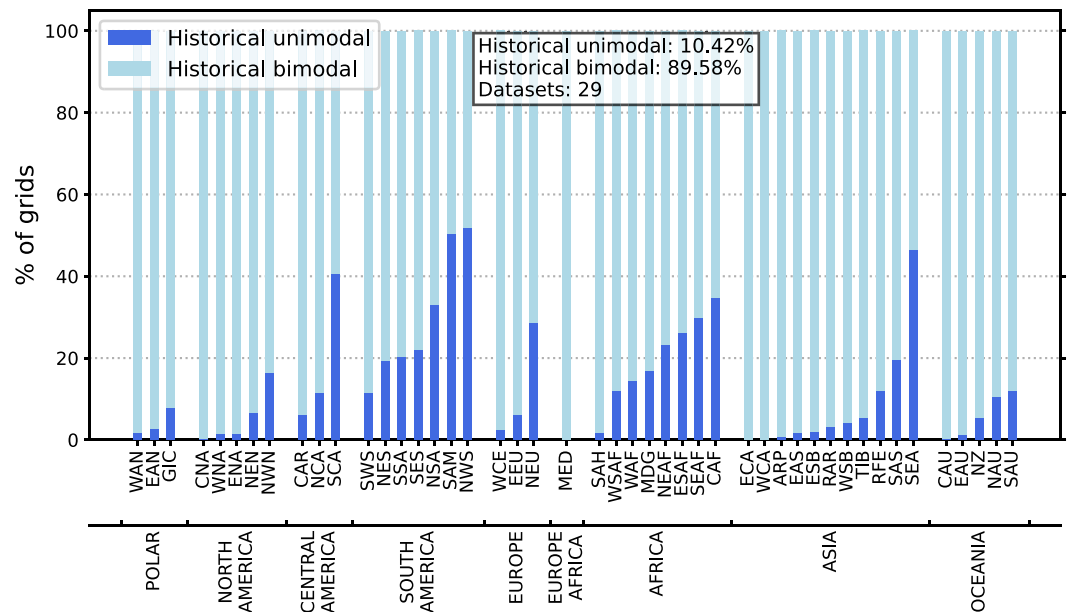

Figure 3. Multi-model mean percentages of grid modalities for the historical period in study regions grouped by continents. Dark and light blue bars show the percentage of grid cells with unimodal or bimodal distribution, respectively, for the historical period of 29 CMIP6 simulations.
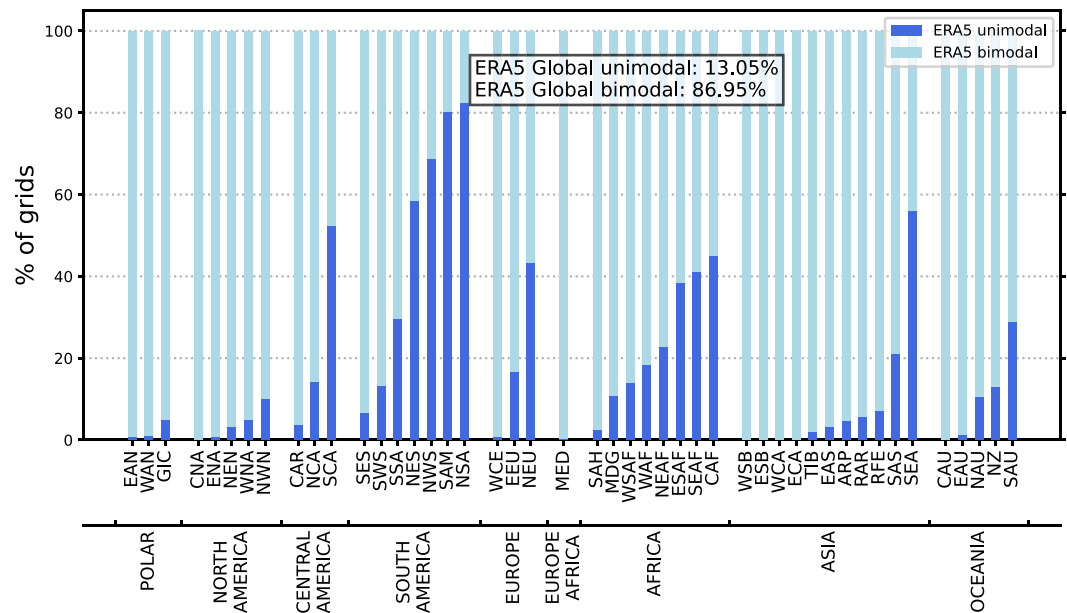
Figure 4. Same as Figure 3 but for ECMWF-ERA5 reanalysis data set.

surface temperature in CMIP6 data set. also contribute to the observed variations from ERA5 (Bock et al., 2020). Nevertheless, as we aim to evaluate the shape of temperature distributions, we did not apply a bias correction and used raw multi-year daily temperature data from CMIP6 models for our analysis.

Then, the parameters of the hot Gaussian component, $\mu_{hot}^{historical}$, $\sigma_{hot}^{historical}$ and $\omega_{hot}^{historical}$, were used to calculate the change in return periods. We only analyzed 1-year, 5-year, 10-year and 20-year events, as GMM are unbounded. One should be careful while calculating the return periods using GMM, as the unbounded tails of the Gaussian component could overestimate the probabilities of longer return periods. Therefore, return periods equal to or less than the analysis period were calculated using GMM. The change in return periods is calculated first in each grid cell of a region and then averaged together to produce regional results for each CMIP6 simulation.

For normally distributed data, the expected percentage of the population inside the $\mu \pm d\sigma$ range is defined as

$$E(\mu \pm d\sigma) = \text{erf}\left(\frac{d}{\sqrt{2}}\right) \tag{5}$$

where erf is the error function and d is the standard deviation distance. The approximate expected frequency, f, outside this range is then defined as the return period of an extreme.

$$1 \ in \ \frac{1}{1 - \text{erf}\left(\frac{d}{\sqrt{2}}\right)} \ days \tag{6}$$

The return period of an event describes the average time between the occurrences of a certain event of a defined size, where an $n - year$ event has an occurrence probability of $1/n$ as the climate is not stationary, where "*year*" is defined as the number of days covered by the hot Gaussian component. The reason for this definition is that the entire probability distribution is composed of both the cold and warm periods of a year, however, our data set consists of daily maximum temperature data spanning 30 (or 20) years, totaling 10,950 (or 7,300) days. Since our analysis specifically aims to identify extreme values using parameters from the hot Gaussian component, we need to consider the number of data points generated by this component as the definition of a "year." We determine the length of a "year" by dividing the data points falling under the hot Gaussian component by the length of the analysis period as shown in Equation 7. For example, we can assume that a symmetrical bimodal distribution results in ~180 days of cold weather and ~180 days of hot weather in a normal 365-day calendar year. For such a symmetric case, a 10-year event would then be a temperature event in 1,800 days (10 years×180 $\frac{days}{year}$). Since we

cannot assume a symmetric distribution for grid cells of each model, we calculated the number of days covered by the hot Gaussian component using the component weights and data set size.

Let $\mathcal{D}$ denote the number of days in L years. Then, a "*year*" in the historical period, $\left| \mathcal{N} \left(\mu_{hot}^{historical}, \sigma_{hot}^{historical} \right) \right|$ is defined as

$$\left| \mathcal{N} \left(\mu_{hot}^{historical}, \sigma_{hot}^{historical} \right) \right| = \frac{\omega_{hot}^{historical} \mathcal{D}}{L} \tag{7}$$

where $\mu_{hot}^{historical}$ is the mean, $\sigma_{hot}^{historical}$ is the standard deviation and $\omega_{hot}^{historical}$ is the weight of hot Gaussian component. The expected frequency of n-year events in the historical period, $f_n^{historical}$, is then calculated by using the length of a year,

$$f_n^{historical} = n \times \left| \mathcal{N} \left(\mu_{hot}^{historical}, \sigma_{hot}^{historical} \right) \right| \quad n = 1, 5, 10, 20 \tag{8}$$

The standard deviation distance of range, $d_n^{historical}$, for an extreme event in the historical period can be calculated by using Equation 6,

$$d_n^{historical} = \mathrm{erf}^{-1} \left(1 - \frac{1}{f_n^{historical}} \right) \sqrt{2} \tag{9}$$

where erf^{-1} is inverse error function. Now, we can calculate a temperature threshold, $\tau_n^{historical}$, for an n-*year* event in the historical period.

$$\tau_n^{historical} = \mu_{hot}^{historical} + d_n^{historical} \sigma_{hot}^{historical} \tag{10}$$

Using this temperature threshold from the historical period, we calculate the standard deviation distance of the temperature threshold of n-*year* event in the future, d_n^{future}, by using the mean μ_{hot}^{future}, and standard deviation σ_{hot}^{future} from the hot Gaussian component of the future distribution.

$$d_n^{future} = \frac{\tau_n^{historical} - \mu_{hot}^{future}}{\sigma_{hot}^{future}} \tag{11}$$

$$f_{\dot{n}}^{future} = \frac{1}{1 - \mathrm{erf} \left(\frac{d_n^{future}}{\sqrt{2}} \right)} \tag{12}$$

Finally, the new value of the return period in the future $\dot{n}$, that is, $\dot{n}$-year event, is calculated by using Equation 8

$$\dot{n} = \frac{f_{\dot{n}}^{future}}{\left| \mathcal{N} \left(\mu_{hot}^{future}, \sigma_{hot}^{future} \right) \right|} \tag{13}$$

where $\left| \mathcal{N} \left(\mu_{hot}^{future}, \sigma_{hot}^{future} \right) \right|$ is length of a "*year*" in the future period.

With this method, we can also analyze if and how much the Gaussian components will shift in the future relative to the historical period. We defined ΔT, as the change in the difference between the means of cold and hot Gaussian components as shown in Equation 14:

$$\Delta T = \delta T_{cold} - \delta T_{hot} \tag{14}$$

$$\delta T_{cold} = \mu_{cold}^{future} - \mu_{cold}^{historical} \tag{15}$$

$$\delta T_{hot} = \mu_{hot}^{future} - \mu_{hot}^{historical} \tag{16}$$

In Figure 2, this change in hot and cold Gaussian means is schematically illustrated. Assuming the future means of Gaussian components are higher than the historical periods, δT_{cold} and δT_{hot} will always be positive. Therefore, a negative ΔT means that the peaks are diverging in the future: the hot Gaussian moves toward warmer temperatures faster than the cold Gaussian, which increases the frequency of hot extremes and induces an overall warmer climate. A positive ΔT means that the peaks are converging: the cold Gaussian moves closer to the hot Gaussian, which increases the number of days with warmer temperatures in the colder mode.
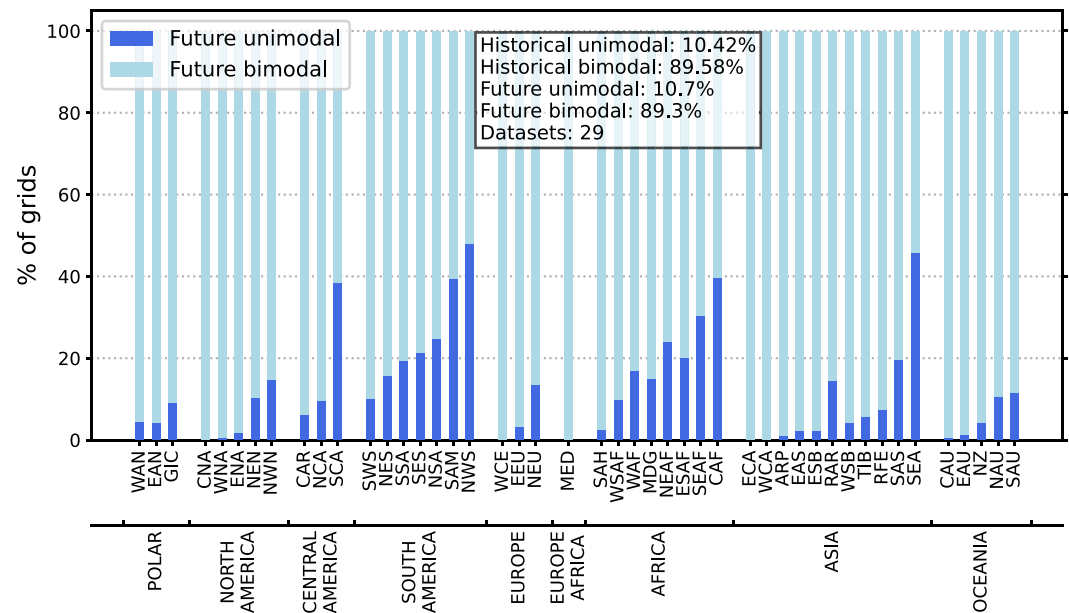
Figure 5. Same as Figure 3 but for future SSP5-8.5 scenario under global warming levels 3.0°C.

3. Results

First, we checked the change in the percentage of modalities from the present to the future time periods. For this, we analyzed the modality of the temperature data from each individual grid cell of an IPCC land region by counting the number of grid cells with each modality. We found that the percentages of grid cells with bimodal distributions stay almost the same under different warming levels. As some of the CMIP6 data sets do not exceed certain warming levels, the number of data sets are not identical for the historical and future period and therefore affect the change in percentages. We analyzed modalities of grid cells under different GWL for all SSP scenarios but we only present SSP5-8.5 results here, as the SSP5-8.5 scenario had data from 29 CMIP6 models and the GWL is scenario independent. Globally, almost 90% of all grid cells follow a bimodal distribution as shown in Figure 3 for the historical period, Figure 4 for the reanalysis data and Figure 5 for GWL 3.0°C for different regions grouped by continents (See Table S3 for other warming levels). Global averages and the number of data sets are shown in the white box in the upper center part of each figure. In the historical period, the grid cells in tropical and sub-tropical regions have slightly higher percentages of unimodal distributions compared to higher latitude regions. However, regions still mostly follow a bimodal distribution as shown in Figure 3. The multi-model mean percentage of unimodal distributions does not exceed 50% of grid cells in any of the regions, except in N.W.South-America (NWS) and South-American-Monsoon (SAM) regions where 51.94% and 50.33% of the grid cells follow a unimodal distribution, respectively, in the historical period. The higher percentage of unimodal distributions in lower latitudes is consistent with tropical climate features, where hot temperatures are observed all year round and the annual temperature range is small (Beck et al., 2018; Richter, 2016). This climate type is therefore expected to likely experience a temperature distribution close to a single Gaussian. All grid cells (99.9%) in CMIP6 models follow a bimodal distribution in the Mediterranean (MED) region in the historical period and under all future periods. In polar regions, more than 90% of the grid cells follow a bimodal distribution in the historical period. The percentage of grid cells with unimodal distributions in polar regions slightly increases under future GWL.

As previously mentioned in Section 2.2, large values of ΔT (see Equation 14) will cause the temperature distribution to change its modality for future GWL periods with respect to the historical base period of 1985–2014. We analyzed all regional grids for all CMIP6 models for the modality changes under GWL 1.5, 2, 3, and 4°C. Figure 6 shows the percentage of changes in grid cell distribution modalities under GWL3.0°C. Globally, the percentage of grids changing from a unimodal (bimodal) distribution in the historical period to a bimodal (unimodal) distribution in the future periods is between 2.79% (2.26%) and 6.02% (3.88%) for different scenarios and GWL as shown in Table 2. The change from unimodal to bimodal distribution in the future period is most prevalent in
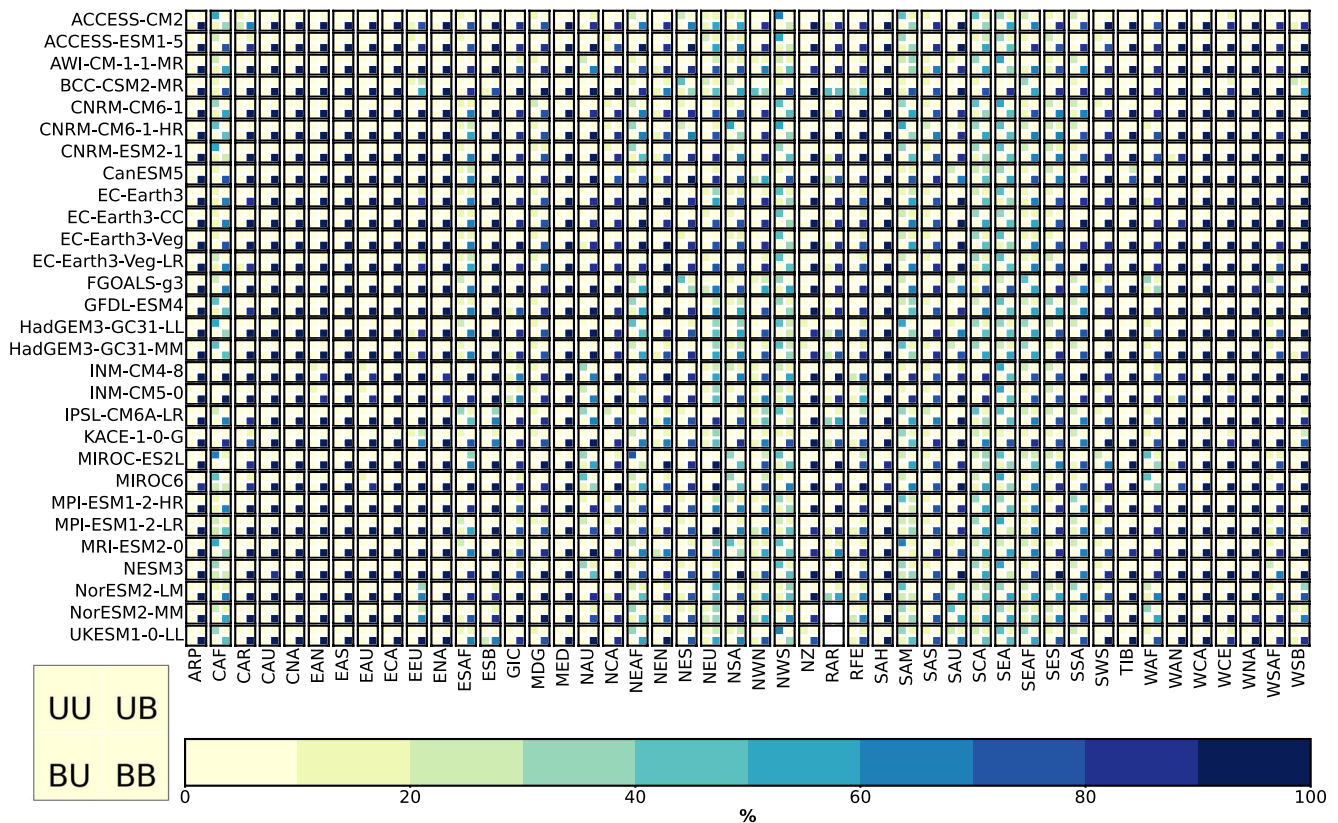
Figure 6. Percentage of changes in grid cell modalities relative to 1985–2014 distribution shape for SSP5-8.5 under GWL3.0°C. Each cell represents a region of a CMIP6 model and is divided into four quadrants. Each quadrant of squares, q_{ij}, uses index notation, where i represents the modality in the historical period and j represents the modality in the future period, 1 for a unimodal distribution and 2 for a bimodal distribution. The top-left quadrant, q_{11}, shows the percentage of grid cells with unimodal distribution both in the historical and the future periods, that is, unimodal to unimodal (UU). The top-right quadrant, q_{12}, shows the percentage of grid cells that change from unimodal distribution in the historical period to bimodal distribution in the future (UB). The bottom-left quadrant, q_{21}, shows bimodal to unimodal (BU). The bottom-right quadrant, q_{22}, shows bimodal to bimodal distribution (BB). The color of the quadrants shows the percentage of grid cells.

regions where the highest percentage of unimodality was observed in the historical period, as shown in Figure 3. This suggests that regions that were previously characterized by more consistent temperatures (as indicated by a unimodal temperature distribution) may experience more variability in temperature in the future. As our analysis uses the mean and standard deviation of the same component from the historical and future daily maximum temperature distributions, we only used the grid cells which have the same modality in the historical and future periods. We disregarded the grid cells with changing modalities, that is, unimodal to bimodal or vice versa, as this will affect the mean and standard deviation, and hence the return period analysis.

We also analyzed the movements of the Gaussian components relative to each other using the ΔT definition from Equation 14 in grid cells with a bimodal distribution. Figure 7 shows the ΔT results for all analyzed regions for SSP5-8.5 under 3.0°C warming (see Figures S7–S12 in Supporting Information S1 for other warming levels). Changes in distribution peaks are smaller for the lower warming levels. This is consistent with the fact that the time periods for exceeding warming levels are very close to the historical period as shown in Figure 1. For the future 3.0°C warming scenario, we observed that the mean temperatures are increasing in all regions. Temperature distributions for the European regions have negative ΔT values, $-0.42°$ on average. This will cause already bimodal peaks in the historical period to separate further from each other in the future, while the whole distribution moves toward higher temperatures. Divergence of peaks will result in more extreme hot temperatures in Europe, as the hot Gaussian moves faster. This result is in agreement with findings from the IPCC AR6 report, in which temperatures in Europe are reported to increase faster than the rest of the globe (IPCC, 2021). Polar regions, Northern America and parts of Northern Asia have positive ΔT values, that is, converging peaks in grid cells with bimodal distributions. The distribution shape shifts to warmer temperatures and approaches a unimodal distribution as the cold Gaussian part of the distribution moves toward the warmer temperatures faster

Table 2
Global Average Percentage of Grid Cells With Varying Distribution Modality Between the Historical and Future Periods

Experiment	GWL	Unimodal → unimodal	Unimodal → bimodal	Bimodal → unimodal	Bimodal → bimodal
SSP1-2.6	1.0°C	11.01%	2.79%	2.26%	83.94%
SSP1-2.6	2.0°C	10.31%	3.53%	2.45%	83.71%
SSP2-4.5	1.5°C	11.02%	2.78%	2.26%	83.95%
SSP2-4.5	2.0°C	10.24%	3.56%	2.70%	83.50%
SSP2-4.5	3.0°C	8.79%	4.71%	3.08%	83.42%
SSP2-4.5	4.0°C	7.21%	6.02%	3.42%	83.35%
SSP3-7.0	1.5°C	10.82%	2.95%	2.40%	83.83%
SSP3-7.0	2.0°C	10.15%	3.62%	2.89%	83.34%
SSP3-7.0	3.0°C	8.92%	4.68%	3.44%	82.96%
SSP3-7.0	4.0°C	7.80%	5.50%	3.81%	82.89%
SSP5-8.5	1.5°C	11.05%	2.85%	2.31%	83.78%
SSP5-8.5	2.0°C	10.32%	3.58%	3.04%	83.06%
SSP5-8.5	3.0°C	9.14%	4.76%	3.78%	82.32%
SSP5-8.5	4.0°C	8.21%	5.47%	3.88%	82.45%

than the hot Gaussian part. This convergence is also consistent with the slight increase in the percentage of unimodal distribution in polar regions as shown in Figure 5. This will cause polar regions to have more days with warmer temperatures also in the colder mode while also having an overall warmer climate. The convergence of peaks in three polar regions (EAN, WAN, GIC) and three northern regions (RAR, NEN and NWN) becomes clear when the regions are sorted by the mean temperature of cold Gaussian component as shown in Figure 8. High ΔT values in polar regions are also supported by previous studies reporting that Arctic regions are warming faster than the global average (Taylor et al., 2022). The lowest ΔT values are in MED and SAM regions, -0.90 and $-1.21°$ respectively, which will cause both bimodal peaks to diverge from each other while both are moving toward warmer temperatures. Regions in Oceania, Central- and parts of South-America have ΔT values close to zero, that is, the cold and hot Gaussian peaks shift toward the warmer temperatures at the same rate. This will cause these regions to have warmer cold and hot periods under future GWL compared to the historical period. When all regions are considered, we observe that the extreme temperature events will increase everywhere, as the mean temperatures increase in all regions compared to the historical distributions. The fact that the peaks are converging only in cold climate regions while diverging in other regions shows that shifts in the Gaussian components with respect to each other are essential for extreme temperature event analyses as these changes affect the overall distribution shape and extent. Also, these results are consistent with the change in skewness in temperature distribution as shown in previous studies (Skelton et al., 2020; Tamarin-Brodsky et al., 2020). Skelton et al. (2020) found an abrupt change in skewness in Europe. Tamarin-Brodsky et al. (2020) found that changes in skewness in winter and summer months will cause cold anomalies in Southern Europe, while warm anomalies intensify in Northeastern Europe. They emphasize the importance of analyzing the shape of temperature distributions.

After analyzing the distribution shapes and peak movements, we calculated the return periods -the average time between the occurrences of a certain event- of 1-year, 5-year, 10-year and 20-year events using only the grid cells
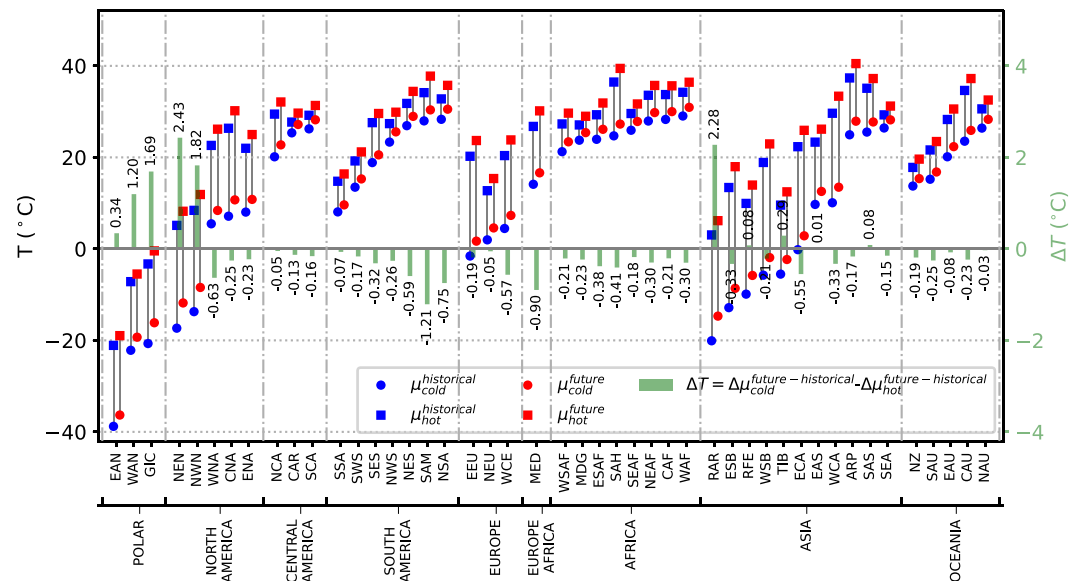

Figure 7. Multi-model peak mean change of region temperature distributions from bimodal grid cells for SSP5-8.5 under GWL3.0°C. Blue (red) dots and squares are the means for cold (hot) peaks of the historical (future) period, respectively. They are plotted on the left *y*-axis. Green bars describe ΔT, the change in the difference between the means of cold and hot Gaussian components, and are plotted on the right *y*-axis. The upward shift in markers represents the overall warming (see Figures S7–S9 in Supporting Information S1 for other warming levels).
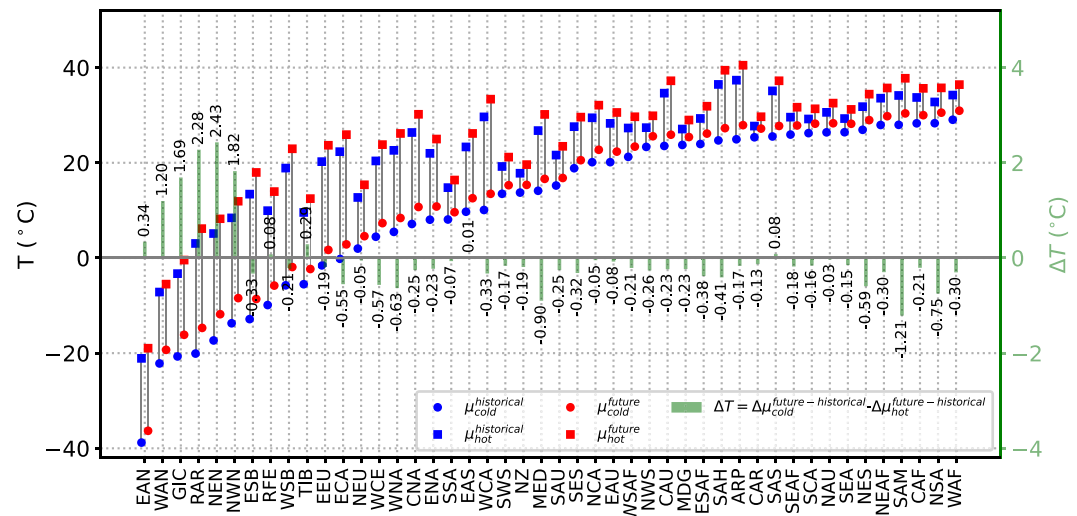
Figure 8. Multi-model peak mean change of region temperature distributions sorted by cold Gaussian mean temperatures (blue dots) for SSP5-8.5 under global warming levels 3.0°C. Blue (red) dots and squares are the means for cold and hot peaks of the historical (future) period, respectively. They are plotted on the left *y*-axis. Green bars describe ΔT, the change in the difference between the means of cold and hot Gaussian components, and are plotted on the right *y*-axis. The colder regions have positive ΔT values and their absolute values are higher than the other regions. The upward shift in blue dots shows that the temperature of cold days is getting warmer and this increase is faster in polar regions compared to the rest of the world (see Figures S10–S12 in Supporting Information S1 for other warming levels).

with constant modalities, that is, unimodal or bimodal both for the historical and future periods, as described in Equation 13. Instead of analyzing extreme temperatures within specific time blocks, our analysis focused on the extremes in the region's probability distribution of 30 (20)-years of daily maximum temperatures. Since we used the hottest component in the mixture of Gaussian components to define n-year events, we considered the number of data points falling under the Gaussian component to define *year*-length according to Equation 7. For example, globally a 10-*year* event was a temperature event once in every 1,880 days (10 years×188 $\frac{days}{year}$) (for bimodal distributions) in the historical period, but it will occur once in every 643, 355, 138, and 63 days under GWL 1.5, 2.0, 3.0, and 4.0°C scenarios, as shown in the plot showing global results in Figure 9 (also in Figure 10), respectively. In other words, historical 10-*year* events will be 3.42-*year*, 1.89-*year*, 0.73-*year* and 0.34-*year* events under the future GWL 1.5, 2.0, 3.0, and 4.0°C scenarios, respectively. After calculating the frequency of extreme events using the temperature distributions in each grid cell individually for an IPCC land region, we averaged the results for the whole region for a single model. The global map with box plots in Figure 9 shows multi-model 10-year event frequencies of each region for SSP5-8.5 scenario under different GWL, where the boxes from light to dark shades of red represent 1.5, 2.0, 3.0, and 4.0°C. Results for 1-year, 5-year, and 20-year events are left out for simplicity and presented in Figures S13–S27 of Supporting Information S1. The length of a "*year*" in each region that is used for return period calculations, that is, the number of days in 10 years, is shown on the top right corner of each sub-plot in Figure 9.

As shown in Figure 9, return periods of extreme temperature events are getting shorter for all regions under all GWL scenarios as the median of each box is smaller than the historical period. The frequency of extreme events is higher in lower latitudes compared to higher latitudes. For example, the return periods are getting prominently shorter in regions around the equator -where a higher percentage of unimodal grid cells was observed- compared to the other regions. Furthermore, CMIP6 models show narrower boxes and shorter whiskers in lower latitudes compared to wider boxes and longer whiskers in higher latitudes for all analyzed GWL. Among all analyzed regions, the Caribbean (CAR) region has the highest increase in the frequency of a 10-year event, from once in 1,910 days for the historical period to once in every 137.3, 35.32, 5.5, and 2.0 days under GWL 1.5, 2, 3, and 4°C, respectively. Regions around the equator (namely CAR, NSA, NWS, NES, SEA, SCA, SAM, MDG, WAF, and SEAF regions) are the top 10 regions with the highest increase in the frequency of extreme events under all GWL. The frequency of a temperature event equivalent to a 10-year event (historically once in every 1,610 days) in the Mediterranean (MED) region increases to once in 405.6, 215.7, 72.4, and 30.6 days in the future under GWL
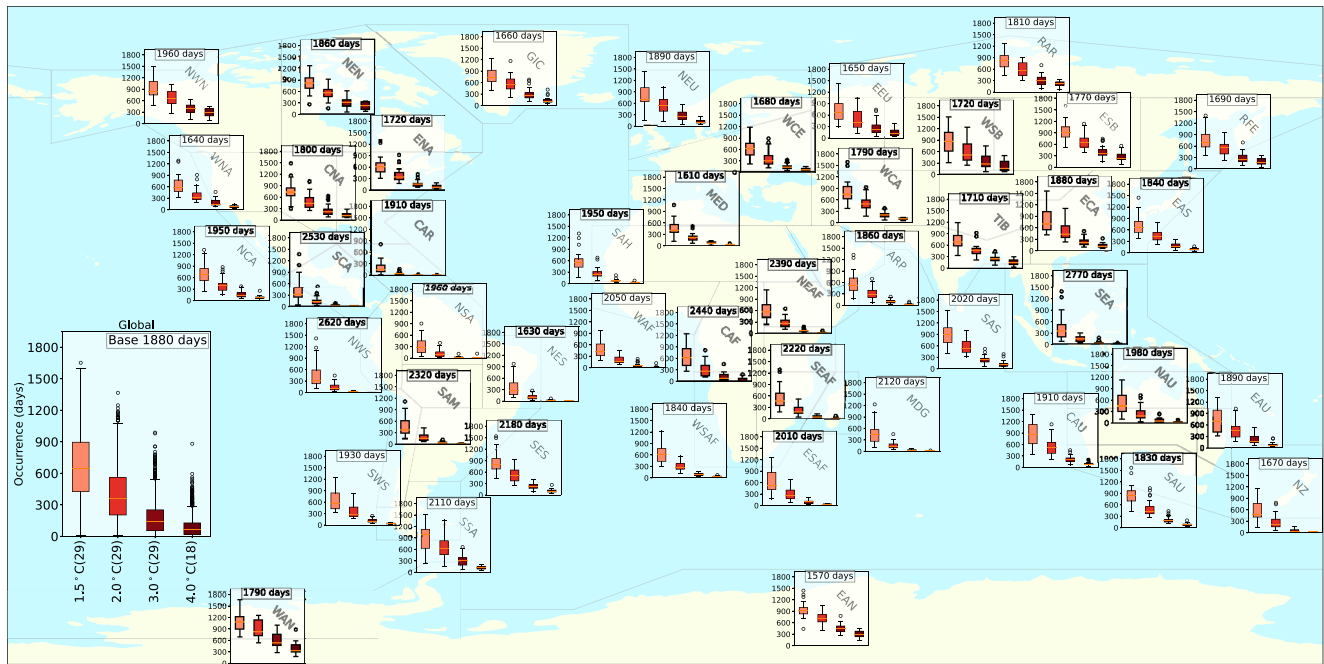
Figure 9. Multi-model median of event frequencies for 10-year hot temperature events compared to the 1985–2014 period under global warming levels 1.5, 2, 3 and 4°C relative to 1850–1900 baseline for SSP5-8.5 scenario. The boxes from light to dark shades red represent 1.5, 2.0, 3.0, and 4.0°C, respectively. The orange lines inside the boxes show the CMIP6 multi-model median, and the boxes extend between the first quartile (Q1) to the third quartile (Q3) of the data, that is, inter-quartile range (IQR). The vertical lines, that is, whiskers, stretch out 1.5 IQR from the box. The circles represent the models outside of the interquartile range, that is, outliers. The length of the hot period used for return period calculations, that is, number of days in 10 years, is shown in the top right corner of each plot. The global return periods are shown on the left. The more outlier points in the global box plot are because of the differences between regional return periods (See Figures S13–S27 in Supporting Information S1 for other return periods).

1.5, 2, 3, and 4°C, respectively. Within the European continent, the West&Central Europe (WCE) region has a higher increase in the frequency of extreme events compared to the Eastern Europe (EEU) and the North Eastern Europe (NEU) regions, where the latter two regions are among the regions with the least increase in extreme temperature event frequency. The smallest increase in the frequency of hot extremes is observed in the Western Antarctica (WAN) region, where the return periods of 10-year events will decrease from once in 1,790 days to once in 1,070.1, 827.6, 542.7 and 338.7 days under GWL 1.5, 2, 3, and 4°C, respectively. High latitude regions, such as WAN, NEU, EAN, NWN, ESB, GIC, RAR, SSA, TIB, and NEN regions are the 10 regions with the smallest decrease in return periods of extreme hot temperature events. Some of these regions are polar regions with positive ΔT values as shown in Figure 8. This will cause more days with warmer temperatures in the colder mode of these regions while having an increase in hot extremes.

4. Summary and Discussion

Detection of extreme events is important to mitigate their impact on natural and anthropogenic systems. Future projections suggest that the mean and standard deviations of maximum surface temperature will increase. This change in the shape of maximum surface temperature distributions increases the intensity and frequency of extreme events in the future. However, not only the shift to warmer temperatures but also the modality of temperature distribution affects the parameters of the entire distribution which is important to calculate the return periods as shown in this study.

GMM are a promising method for calculating the return periods of extreme events, and additionally determining the shape of the entire distribution for daily maximum temperature data. GMM can provide information on different climate features in different regions such as cold and hot periods, and their changes. We showed that bimodality is a prominent characteristic observed in multi-year daily near-surface maximum temperature data. To understand the underlying factors of this bimodal pattern, we analyzed temperature distributions from grid cells with distinct bimodality across different months, seasons and 6-month running windows. We observed
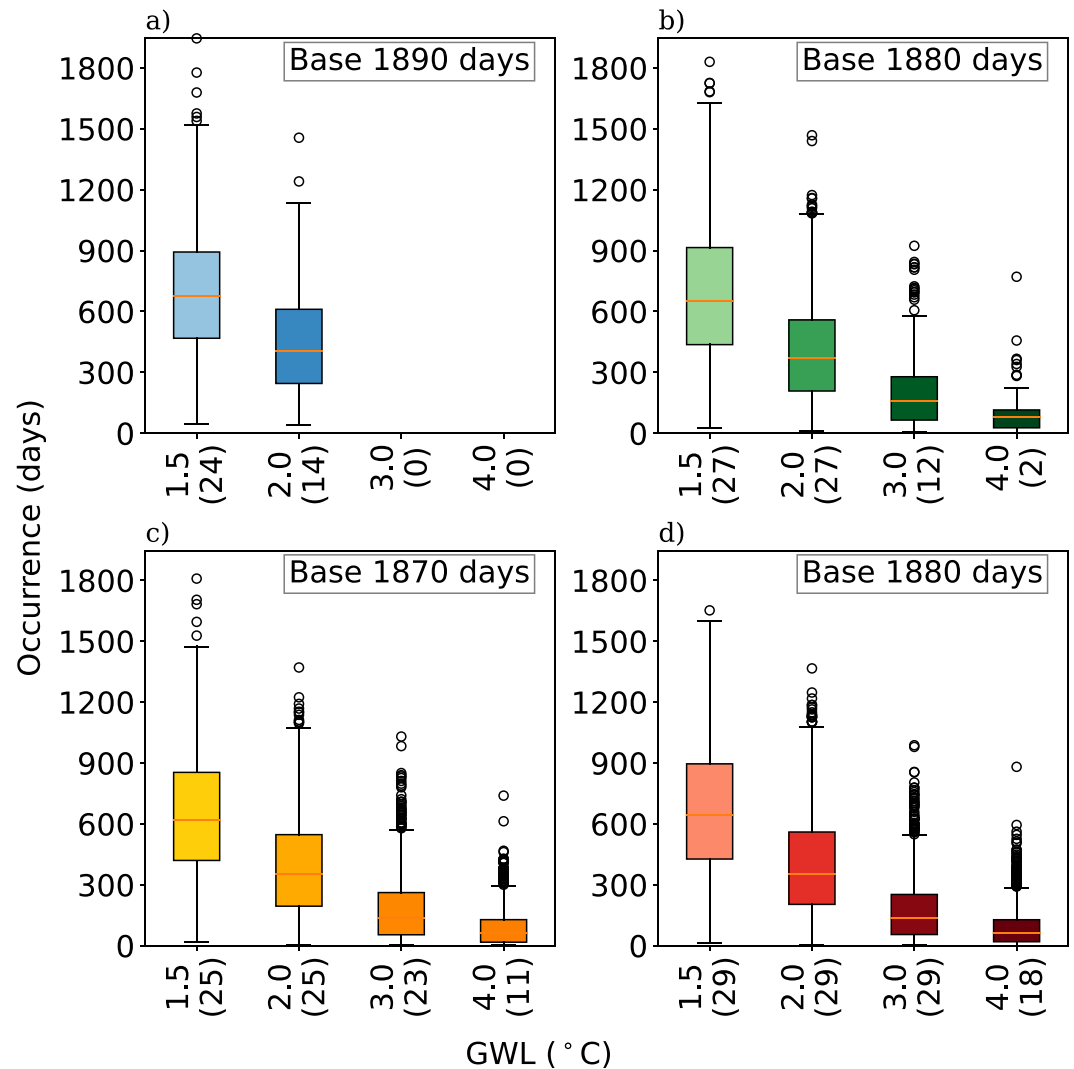
Figure 10. Global multi-model median of event frequencies for 10-year temperature events under 1.5, 2, 3 and 4°C warming levels for (a) SSP1-2.6, (b) SSP2-4.5, (c) SSP3-7.0 and (d) SSP5-8.5 scenarios. The orange lines inside the boxes show the CMIP6 multi-model median, and the boxes extend between the first quartile (Q1) to the third quartile (Q3) of the data, that is, inter-quartile range (IQR). The vertical lines, that is, whiskers, stretch out 1.5 IQR from the box. The circles represent the models outside of the interquartile range, that is, outliers. The length of the hot period used for return period calculations, that is, number of days in 10 years, is shown in the top right corner of each plot. The number of data sets is given in parenthesizes. All plots show similar results for different Shared Socioeconomic Pathway scenarios as the global warming levels are scenario-independent.

that the winter and summer seasons emerged as the primary contributors to the peaks observed in the bimodal distribution. In grid cells of different regions with distinct bimodal distributions, the transition from winter to summer occurs swiftly, leading to a more distinct separation of the temperature modes. Consequently, the distributions during transitional seasons, such as spring and autumn, appeared to be wider (covering a broader value range) compared to the more distinct distributions observed during winter and summer (covering a very small value range). Furthermore, analyses of 6-month running windows also showed an agglomeration of similar temperatures around winter (summer) months from November (May) to April (October) that creates the peaks in the bimodal distribution (See Figure S2 in Supporting Information S1 for the distributions of seasons and months.). Here, the advantage of GMM becomes evident. For analyses to uncover the origins of bimodality, we had to select certain seasons or months. Seasonal periods are commonly used in previous studies to analyze extreme events (Prodhomme et al., 2022; Qian & Zhang, 2015, 2019; Walt & Fitchett, 2021). For example, Qian and Zhang (2015) found that the seasonality is weakening in the northern high-latitude regions and East Asia

while strengthening in the Mediterranean. This can also be seen in Figure 8, as the northern regions and east/central Asia regions have converging peaks which means that these regions will have a distribution closer to a unimodal distribution. Meanwhile, diverging peaks in MED will introduce more distinct cold and warm periods. However, onsets and length of seasons are predicted to change with climate change (Wang et al., 2021). Therefore, the definition of current seasonal periods or months will not necessarily be valid for future climates. One can utilize GMM to determine the hot Gaussian component of a region to define the length of the analysis period instead of using fixed seasonal definitions. Moreover, the bimodality analysis also shows how peaks are changing in the future, effectively changing the expected climate of the area.

ETCCDI indices are commonly used in extreme event analysis as they offer a simple and concise way to define extremes (Vogel, Hauser, & Seneviratne, 2020; X. Zhang et al., 2011; Zhao et al., 2021). ETCCDI indices use block maxima methods such as TXx (Monthly maximum value of daily max temperature), TNx (Monthly maximum value of daily min temperature) or percentile-over-threshold (POT) methods such as TX90p (Percentage of time when daily max temperature >90th percentile), TN90p (Percentage of time when daily min temperature <90th percentile) (X. Zhang et al., 2011). These exceedances can be modeled with GEV distributions or generalized Pareto distribution. However, GEV distributions are a better fit for longer block sizes than for shorter blocks like daily data. If the available data set is short, the longer block sizes will produce fewer data which can increase the variability in parameter estimation (Huang et al., 2016; Wang et al., 2016). For example, if there is more than one extremely hot day in the block (month, season or year), for example, several consecutive days, block maxima methods consider only the hottest, and hence only 1 day in a block, while GMM considers all days hotter than the threshold. Assuming that a heat wave lasts usually days to a few weeks, a substantial number of hot days might not be seen by block maxima methods as long as they fall into the same block. Percentile-over-threshold methods together with count-day indices such as WSDI (Warm spell duration indicator) are useful for analyzing the durations of events. However, the derivation of percentiles is strongly affected by the choice of the base period, a right shift in the distribution will result in a higher threshold and erroneously reduce the frequency of extreme events (Yosef et al., 2021). Seasonality in temperature extremes adds complexity to the process of selecting percentiles to define extreme temperatures (Huang et al., 2016). The advantage of GMM is that the model analyses the distribution of temperatures without any previous assumption and learns the hot periods from the data. Also, GMM uses all available data in contrast to block maxima methods, which makes it useful if the available data is short or bimodality exists (Ben Alaya et al., 2020; Knoben et al., 2019; Sardeshmukh et al., 2015; Wang et al., 2016).

However, since the Gaussian components of GMM are not bounded, it is important to only calculate the return periods of extreme events equal to or less than the study period when applying GMM. Additionally, we only used grid cells which have the same number of Gaussian components in their temperature distribution, that is, unimodal or bimodal distribution, both for the historical and future periods. Grid cells with changing distribution shapes, for example, transforming from a bimodal distribution in the historical period to a unimodal distribution in the future or vice versa, were found in less than 10% of the grid cell for each GWL as shown in Table 2, and were disregarded in the analysis as calculating the temperature thresholds becomes problematic with the abrupt change in means and standard deviations.

For the first time, the IPCC AR6 Report includes a new dedicated chapter on weather and climate extreme events (IPCC, 2021). This emphasizes the importance of robust methods of extreme event detection to be able to mitigate the impact of such events. IPCC AR6 reports that the return periods of 10-year events will increase around the world, with the highest changes projected to happen in some mid-latitude and semi-arid regions. Our findings are in agreement with these results. Furthermore, IPCC AR6 projects the warming rate in mid-latitudes to be higher than the average global warming rate. GMM might explain why these regions are projected to have higher warming, as we observed that grid cells in these regions predominantly follow a bimodal distribution in the historical (future) period as shown in Figure 3 (5). Furthermore, these regions have diverging peaks as shown in Figure 8, that is, mode for warm temperatures moving toward warmer temperatures faster than the mode for colder temperatures. These diverging bimodal peaks will create distinct Gaussian components in the entire multi-year daily maximum temperature, which in turn results in a higher increase in extremes in these regions. For example, almost all grid cells in the Mediterranean region follow a bimodal distribution, and the peaks of bimodal distribution will diverge in the future as shown in Figure 7. Mediterranean region is identified as one of the most responsive regions to climate change and a hot spot of climate extremes (Feng et al., 2022; IPCC, 2021). Similarly, Arctic regions are projected to have the highest increase in temperature of the coldest days (IPCC, 2021; C. Li et al., 2021). Our results are also consistent with these increases as shown in Figure 7,

where diverging bimodal peaks in mid-latitude regions will shift the mode for warm temperatures, that is, hot Gaussian, to the higher temperature ranges. This shift in the Gaussian components of temperature distribution will cause those land regions to have warmer temperature extremes and can explain the higher average warming rate than the global average. Likewise, converging peaks in polar regions as shown in Figure 7 will move the cold Gaussian part toward warmer temperatures, thereby introducing higher warming on the coldest days.

According to our analyses, 10-year events will increase almost 3-fold under GWL 1.5°C compared to the historical period for all SSP scenarios as shown in Figure 10 when looking at the whole globe. This means a temperature event that occurs once in every 10 years (1,880 days) will be expected to occur 2.9 times in every 10 years under GWL 1.5°C. 10-year extreme temperature events will become even more frequent globally under GWL 2, 3, and 4°C; 5.3, 13.6, and 29.5 times every 10 years, respectively. In other words, current 10-year events will be 3.42-, 1.89-, 0.73-, and 0.34-*year* events in the future under GWL 1.5, 2, 3, and 4°C, respectively. Our results show a higher increase compared to the IPCC AR6 report, where the frequency of 10-year events is projected to increase approximately 3, 4, 5.5 and 9-fold under GWL 1.5, 2, 3, and 4°C, respectively (IPCC, 2021), using a block maxima method for determining the extreme events. The higher increase in our method compared to IPCC AR6 can most likely be explained by the fact that we used GMM to model the distribution of temperatures and GMM considers all days hotter than the threshold, while the block maxima method only uses the maximum of a block. Another important point deduced from the analyses of different regions for several CMIP6 models is that the ensemble of analyzed CMIP6 models shows coherent results for regions as shown in the regional box plots in Figure 9. Most of the individual model results fall within the first and third quartile, and only a few models fall outside this range. The higher number of outlier points in the global box plot in Figure 9, and also shown for different SSP scenarios in Figure 10, are caused by the differences between regional return periods. All SSP scenarios show similar results with each other as the return periods are calculated for GWL which have the same forcing on climate.

Return periods of extreme events become shorter in every region, which means that the frequency of extreme temperature events increases. This will become larger with increasing GWL. Some climate models have already exceeded GWL 1.5°C with respect to the 1850–1900 period as shown in Figure 1. This fact further emphasizes the importance of robust methods to detect extreme events. Even though there is a delay in taking the necessary precautions to reduce the speed of the warming of the climate, as time goes by, tomorrow's projections become today's reality.

Data Availability Statement

The recipes to extract regional data from CMIP6 models using ESMValTool, python scripts to analyse extreme events and to produce all figures of this manuscript are accessible in the following GitHub repository: https://github.com/EyringMLClimateGroup/pacal23jgr_GaussianMixtureModels_Extremes. The regional output files amount to hundreds of GB. The latest release of ESMValTool is publicly at https://github.com/ESMValGroup/ESMValTool (Andela et al., 2022).

References

Alexander, L. V., Zhang, X., Peterson, T. C., Caesar, J., Gleason, B., Klein Tank, A. M. G., et al. (2006). Global observed changes in daily climate extremes of temperature and precipitation. *Journal of Geophysical Research*, *111*(D5), D05109. https://doi.org/10.1029/2005JD006290

Alizadeh, M. R., Adamowski, J., Nikoo, M. R., AghaKouchak, A., Dennison, P., & Sadegh, M. (2020). A century of observations reveals increasing likelihood of continental-scale compound dry-hot extremes. *Science Advances*, *6*(39), eaaz4571. https://doi.org/10.1126/sciadv.aaz4571

Andela, B., Broetz, B., de Mora, L., Drost, N., Eyring, V., Koldunov, N., et al. (2022). ESMValTool [Software]. Zenodo. https://doi.org/10.5281/zenodo.3401363

Anderson, A., & Kostinski, A. (2010). Reversible record breaking and variability: Temperature distributions across the globe. *Journal of Applied Meteorology and Climatology*, *49*(8), 1681–1691. https://doi.org/10.1175/2010JAMC2407.1

Baker, H. S., Millar, R. J., Karoly, D. J., Beyerle, U., Guillod, B. P., Mitchell, D., et al. (2018). Higher CO_2 concentrations increase extreme event risk in a 1.5°C world. *Nature Climate Change*, *8*(7), 604–608. https://doi.org/10.1038/s41558-018-0190-1

Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., & García-Herrera, R. (2011). The hot summer of 2010: Redrawing the temperature record map of Europe. *Science*, *332*(6026), 220–224. https://doi.org/10.1126/science.1201224

Beck, H. E., Zimmermann, N. E., McVicar, T. R., Vergopolan, N., Berg, A., & Wood, E. F. (2018). Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data*, *5*(1), 180214. https://doi.org/10.1038/sdata.2018.214

Ben Alaya, M. A., Zwiers, F., & Zhang, X. (2020). An evaluation of block-maximum-based estimation of very long return period precipitation extremes with a large ensemble climate simulation. *Journal of Climate*, *33*(16), 6957–6970. https://doi.org/10.1175/JCLI-D-19-0011.1

Benestad, R. (2004). Record-values, nonstationarity tests and extreme value distributions. *Global and Planetary Change*, *44*(1–4), 11–26. https://doi.org/10.1016/j.gloplacha.2004.06.002

Bentsen, M., Oliviè, D. J. L., Seland, Y., Toniazzo, T., Gjermundsen, A., Graff, L. S., et al. (2019). NCC NorESM2-MM model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.8040

Bertossa, C., Hitchcock, P., DeGaetano, A., & Plougonven, R. (2021). Bimodality in ensemble forecasts of 2 m temperature: Identification. *Weather and Climate Dynamics*, *2*(4), 1209–1224. https://doi.org/10.5194/wcd-2-1209-2021

Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., et al. (2020). Quantifying progress across different CMIP phases with the ESMValTool. *Journal of Geophysical Research: Atmospheres*, *125*(21), e2019JD032321. https://doi.org/10.1029/2019JD032321

Boucher, O., Denvil, S., Levavasseur, G., Cozic, A., Caubel, A., Foujols, M.-A., et al. (2018). IPSL IPSL-CM6A-LR model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.5195

Byun, Y.-H., Lim, Y.-J., Sung, H. M., Kim, J., Sun, M., & Kim, B.-H. (2019). NIMS-KMA KACE1.0-G model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.8378

Cai, W., Borlace, S., Lengaigne, M., van Rensch, P., Collins, M., Vecchi, G., et al. (2014). Increasing frequency of extreme El Niño events due to greenhouse warming. *Nature Climate Change*, *4*(22), 111–116. https://doi.org/10.1038/nclimate2100

Cao, J., & Wang, B. (2019). NUIST NESMv3 model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.8769

Cho, H., & Jeong, S. (2016). Estimation and comparative analysis on the distribution functions of air and water temperatures in Korean coastal seas. *Journal of Korean Society of Coastal and Ocean Engineers*, *28*(3), 171–176. https://doi.org/10.9765/KSCOE.2016.28.3.171

Christidis, N., Jones, G. S., & Stott, P. A. (2015). Dramatically increasing chance of extremely hot summers since the 2003 European heatwave. *Nature Climate Change*, *5*(11), 46–50. https://doi.org/10.1038/nclimate2468

Clarkson, D., Eastoe, E., & Leeson, A. (2022). Melt probabilities and surface temperature trends on the Greenland ice sheet using a Gaussian mixture model. *The Cryosphere*, *16*(5), 1597–1607. https://doi.org/10.5194/tc-16-1597-2022

Copernicus Climate Change Service (C3S). (2023). *ERA5 hourly data on single levels from 1940 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.adbb2d47

Demiroglu, O. C., Saygili-Araci, F. S., Pacal, A., Hall, C. M., & Kurnaz, M. L. (2020). Future holiday climate index (HCI) performance of urban and beach destinations in the Mediterranean. *Atmosphere*, *11*(9), 911. https://doi.org/10.3390/atmos11090911

Dix, M., Bi, D., Dobrohotoff, P., Fiedler, R., Harman, I., Law, R., et al. (2019). CSIRO-ARCCSS ACCESS-CM2 model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.4271

Donat, M. G., & Alexander, L. V. (2012). The shifting probability distribution of global daytime and night-time temperatures. *Geophysical Research Letters*, *39*(14), L14707. https://doi.org/10.1029/2012GL052459

EC-Earth Consortium (EC-Earth). (2019a). EC-Earth-Consortium EC-Earth3 model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.4700

EC-Earth Consortium (EC-Earth). (2019b). EC-Earth-Consortium EC-Earth3-Veg model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.4706

EC-Earth Consortium (EC-Earth). (2020). EC-Earth-Consortium EC-Earth3-Veg-LR model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.4707

EC-Earth Consortium (EC-Earth). (2021). EC-Earth-Consortium EC-Earth-3-CC model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.4702

Eyring, V., Bock, L., Lauer, A., Righi, M., Schlund, M., Andela, B., et al. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0 – An extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of Earth system models in CMIP. *Geoscientific Model Development*, *13*(7), 3383–3438. https://doi.org/10.5194/gmd-13-3383-2020

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Inter-comparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Eyring, V., Gillett, N. P., Achuta Rao, K. M., Barimalala, R., Barreiro Parrillo, M., Bellouin, N., et al. (2021). Human influence on the climate system. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Pean, S. Berger, et al. (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (Chapter 3)*. Cambridge University Press.

Feng, X., Qian, C., & Materia, S. (2022). Amplification of the temperature seasonality in the Mediterranean region under anthropogenic climate change. *Geophysical Research Letters*, *49*(20), e2022GL099658. https://doi.org/10.1029/2022GL099658

Fischer, E. M., & Schär, C. (2010). Consistent geographical patterns of changes in high-impact European heatwaves. *Nature Geoscience*, *3*(6), 398–403. https://doi.org/10.1038/ngeo866

García-León, D., Casanueva, A., Standardi, G., Burgstall, A., Flouris, A. D., & Nybo, L. (2021). Current and projected regional economic impacts of heatwaves in Europe. *Nature Communications*, *12*(1), 5807. https://doi.org/10.1038/s41467-021-26050-z

Grace, W. (1995). A binormal model of frequency distributions of daily maximum temperature. *Australian Meteorological Magazine*, *42*, 151–161.

Guerreiro, S. B., Dawson, R. J., Kilsby, C., Lewis, E., & Ford, A. (2018). Future heat-waves, droughts and floods in 571 European cities. *Environmental Research Letters*, *13*(3), 034009. https://doi.org/10.1088/1748-9326/aaaad3

Hajat, S., Gampe, D., Sarsour, A., & Abuzerr, S. (2022). Climate change and diarrhoeal disease burdens in the gaza strip, Palestine: Health impacts of 1.5°C and 2°C global warming scenarios. *International Journal of Environmental Research and Public Health*, *19*(8), 4898. https://doi.org/10.3390/ijerph19084898

Hajima, T., Abe, M., Arakawa, O., Suzuki, T., Komuro, Y., Ogura, T., et al. (2019). MIROC MIROC-ES2L model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.5602

Hao, Z., Hao, F., Singh, V. P., & Zhang, X. (2018). Changes in the severity of compound drought and hot extremes over global land areas. *Environmental Research Letters*, *13*(12), 124022. https://doi.org/10.1088/1748-9326/aaee96

Hauser, M., Engelbrecht, F., & Fischer, E. M. (2022). Transient global warming levels for CMIP5 and CMIP6 [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7390473

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., et al. (2018). *ERA5 hourly data on single levels from 1940 to present*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.adbb2d47

Huang, W. K., Stein, M. L., McInerney, D. J., Sun, S., & Moyer, E. J. (2016). Estimating changes in temperature extremes from millennial-scale climate simulations using generalized extreme value (GEV) distributions. *Advances in Statistical Climatology, Meteorology and Oceanography*, *2*(1), 79–103. https://doi.org/10.5194/ascmo-2-79-2016

IPCC. (2021). In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Pean, S. Berger, et al. (Eds.), *Climate Change 2021: The Physical Science Basis*. Cambridge University Press. (In Press).

Iturbide, M., Gutiérrez, J. M., Alves, L. M., Bedia, J., Cerezo-Mota, R., Cimadevilla, E., et al. (2020). An update of IPCC climate reference regions for subcontinental analysis of climate model data: Definition and aggregated datasets. *Earth System Science Data*, *12*(4), 2959–2970. https://doi.org/10.5194/essd-12-2959-2020

Iyakaremye, V., Zeng, G., Ullah, I., Gahigi, A., Mumo, R., & Ayugi, B. (2022). Recent observed changes in extreme high-temperature events and associated meteorological conditions over africa. *International Journal of Climatology*, *42*(9), 4522–4537. https://doi.org/10.1002/joc.7485

Iyakaremye, V., Zeng, G., & Zhang, G. (2021). Changes in extreme temperature events over africa under 1.5 and 2.0c global warming scenarios. *International Journal of Climatology*, *41*(2), 1506–1524. https://doi.org/10.1002/joc.6868

Jungclaus, J., Bittner, M., Wieners, K.-H., Wachsmann, F., Schupfner, M., Legutke, S., et al. (2019). MPI-M MPI-ESM1.2-HR model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.6594

Kharin, V. V., Zwiers, F. W., Zhang, X., & Wehner, M. (2013). Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, *119*(2), 345–357. https://doi.org/10.1007/s10584-013-0705-8

Knoben, W. J. M., Woods, R. A., & Freer, J. E. (2019). Global bimodal precipitation seasonality: A systematic overview. *International Journal of Climatology*, *39*(1), 558–567. https://doi.org/10.1002/joc.5786

Kodra, E., & Ganguly, A. R. (2014). Asymmetry of projected increases in extreme temperature distributions. *Scientific Reports*, *4*(1), 5884. https://doi.org/10.1038/srep05884

Kollu, R., Rayapudi, S. R., Narasimham, S., & Pakkurthi, K. M. (2012). Mixture probability distribution functions to model wind speed distributions. *International Journal of Energy and Environmental Engineering*, *3*(1), 27. https://doi.org/10.1186/2251-6832-3-27

Krasting, J. P., John, J. G., Blanton, C., McHugh, C., Nikonov, S., Radhakrishnan, A., et al. (2018). NOAA-GFDL GFDL-ESM4 model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.8597

Lauer, A., Eyring, V., Bellprat, O., Bock, L., Gier, B. K., Hunter, A., et al. (2020). Earth System Model Evaluation Tool (ESMValTool) v2.0 – Diagnostics for emergent constraints and future projections from Earth system models in CMIP. *Geoscientific Model Development*, *13*(9), 4205–4228. https://doi.org/10.5194/gmd-13-4205-2020

Lelieveld, J, Proestos, Y., Hadjinicolaou, P., Tanarhte, M., Tyrlis, E., & Zittis, G. (2016). Strongly increasing heat extremes in the Middle East and North Africa (MENA) in the 21st century. *Climatic Change*, *137*(1–2), 245–260. https://doi.org/10.1007/s10584-016-1665-6

Lewis, S. C., King, A. D., & Perkins-Kirkpatrick, S. E. (2017). Defining a new normal for extremes in a warming world. *Bulletin of the American Meteorological Society*, *98*(6), 1139–1151. https://doi.org/10.1175/BAMS-D-16-0183.1

Li, C., Zwiers, F., Zhang, X., & Li, G. (2019). How much information is required to well constrain local estimates of future precipitation extremes? *Earth's Future*, *7*(1), 11–24. https://doi.org/10.1029/2018EF001001

Li, C., Zwiers, F., Zhang, X., Li, G., Sun, Y., & Wehner, M. (2021). Changes in annual extremes of daily temperature and precipitation in CMIP6 models. *Journal of Climate*, *34*(9), 3441–3460. https://doi.org/10.1175/JCLI-D-19-1013.1

Li, L. (2019). CAS FGOALS-g3 model output prepared for CMIP6 CMIP historical [dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.3356

Linz, M., Chen, G., & Hu, Z. (2018). Large-scale atmospheric control on non-Gaussian tails of midlatitude temperature distributions. *Geophysical Research Letters*, *45*(17), 9141–9149. https://doi.org/10.1029/2018GL079324

Loikith, P. C., & Neelin, J. D. (2019). Non-Gaussian cold-side temperature distribution tails and associated synoptic meteorology. *Journal of Climate*, *32*(23), 8399–8414. https://doi.org/10.1175/JCLI-D-19-0344.1

Mallick, J., Islam, A. R. M. T., Ghose, B., Islam, H. M. T., Rana, Y., Hu, Z., et al. (2022). Spatiotemporal trends of temperature extremes in Bangladesh under changing climate using multi-statistical techniques. *Theoretical and Applied Climatology*, *147*(1–2), 307–324. https://doi.org/10.1007/s00704-021-03828-1

Meehl, G. A., & Tebaldi, C. (2004). More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, *305*(5686), 994–997. https://doi.org/10.1126/science.1098704

Mueller, B., Zhang, X., & Zwiers, F. W. (2016). Historically hottest summers projected to be the norm for more than half of the world's population within 20 years. *Environmental Research Letters*, *11*(4), 044011. https://doi.org/10.1088/1748-9326/11/4/044011

Nikulin, G., Kjellström, E., Hansson, U., Strandberg, G., & Ullerstig, A. (2011). Evaluation and future projections of temperature, precipitation and wind extremes over Europe in an ensemble of regional climate simulations. *Tellus A*, *63*(1), 41–55. https://doi.org/10.1111/j.1600-0870.2010.00466.x

O'Neill, B. C., Kriegler, E., Riahi, K., Ebi, K. L., Hallegatte, S., Carter, T. R., et al. (2014). A new scenario framework for climate change research: The concept of shared socioeconomic pathways. *Climatic Change*, *122*(3), 387–400. https://doi.org/10.1007/s10584-013-0905-2

O'Neill, B. C., Tebaldi, C., van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., et al. (2016). The scenario model Intercomparison Project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, *9*(9), 3461–3482. https://doi.org/10.5194/gmd-9-3461-2016

Ozturk, T., Saygili-Araci, F. S., & Kurnaz, M. L. (2021). Projected changes in extreme temperature and precipitation indices over CORDEX-MENA domain. *Atmosphere*, *12*(55), 622. https://doi.org/10.3390/atmos12050622

Paciorek, C. J., Stone, D. A., & Wehner, M. F. (2018). Quantifying statistical uncertainty in the attribution of human influence on severe weather. *Weather and Climate Extremes*, *20*, 69–80. https://doi.org/10.1016/j.wace.2018.01.002

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Perera, A. T. D., Nik, V. M., Chen, D., Scartezzini, J.-L., & Hong, T. (2020). Quantifying the impacts of climate change and extreme climate events on energy systems. *Nature Energy*, *5*(2), 150–159. https://doi.org/10.1038/s41560-020-0558-0

Perkins-Kirkpatrick, S. E., & Lewis, S. C. (2020). Increasing trends in regional heatwaves. *Nature Communications*, *11*(1), 3357. https://doi.org/10.1038/s41467-020-16970-7

Perron, M., & Sura, P. (2013). Climatology of non-Gaussian atmospheric statistics. *Journal of Climate*, *26*(3), 1063–1083. https://doi.org/10.1175/JCLI-D-11-00504.1

Pfleiderer, P., Schleussner, C.-F., Kornhuber, K., & Coumou, D. (2019). Summer weather becomes more persistent in a 2°C world. *Nature Climate Change*, *9*(9), 666–671. https://doi.org/10.1038/s41558-019-0555-0

Prodhomme, C., Materia, S., Ardilouze, C., White, R. H., Batté, L., Guemas, V., et al. (2022). Seasonal prediction of European summer heatwaves. *Climate Dynamics*, *58*(7–8), 2149–2166. https://doi.org/10.1007/s00382-021-05828-3

Qian, C., & Zhang, X. (2015). Human influences on changes in the temperature seasonality in mid- to high-latitude land areas. *Journal of Climate*, *28*(15), 5908–5921. https://doi.org/10.1175/JCLI-D-14-00821.1

Qian, C., & Zhang, X. (2019). Changes in temperature seasonality in China: Human influences and internal variability. *Journal of Climate*, *32*(19), 6237–6249. https://doi.org/10.1175/JCLI-D-19-0081.1

Rahmstorf, S., & Coumou, D. (2011). Increase of extreme events in a warming world. *Proceedings of the National Academy of Sciences*, *108*(44), 17905–17909. https://doi.org/10.1073/pnas.1101766108

Raymond, C., Matthews, T., & Horton, R. M. (2020). The emergence of heat and humidity too severe for human tolerance. *Science Advances*, *6*(19), eaaw1838. https://doi.org/10.1126/sciadv.aaw1838

Ribeiro, A. F. S., Brando, P. M., Santos, L., Rattis, L., Hirschi, M., Hauser, M., et al. (2022). A compound event-oriented framework to tropical fire risk assessment in a changing climate. *Environmental Research Letters*, *17*(6), 065015. https://doi.org/10.1088/1748-9326/ac7342

Richter, M. (2016). Temperatures in the tropics. In L. Pancel & M. Köhl (Eds.), *Tropical forestry handbook* (pp. 343–361). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-54601-3_38

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., & Andrews, T. (2019a). MOHC HadGEM3-GC31-LL model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.6109

Ridley, J., Menary, M., Kuhlbrodt, T., Andrews, M., & Andrews, T. (2019b). MOHC HadGEM3-GC31-MM model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.6112

Righi, M., Andela, B., Eyring, V., Lauer, A., Predoi, V., Schlund, M., et al. (2020). Earth system model evaluation Tool (ESMValTool) v2.0 – Technical overview. *Geoscientific Model Development*, *13*(3), 1179–1199. https://doi.org/10.5194/gmd-13-1179-2020

Robine, J.-M., Cheung, S. L. K., Le Roy, S., Van Oyen, H., Griffiths, C., Michel, J.-P., & Herrmann, F. R. (2008). Death toll exceeded 70,000 in Europe during the summer of 2003. *Comptes Rendus Biologies*, *331*(2), 171–178. https://doi.org/10.1016/j.crvi.2007.12.001

Robinson, A., Lehmann, J., Barriopedro, D., Rahmstorf, S., & Coumou, D. (2021). Increasing heat and rainfall extremes now far outside the historical climate. *npj Climate and Atmospheric Science*, *4*(11), 1–4. https://doi.org/10.1038/s41612-021-00202-w

Sardeshmukh, P. D., Compo, G. P., & Penland, C. (2015). Need for caution in interpreting extreme weather statistics. *Journal of Climate*, *28*(23), 9166–9187. https://doi.org/10.1175/JCLI-D-15-0020.1

Sardeshmukh, P. D., & Sura, P. (2009). Reconciling non-Gaussian climate statistics with linear dynamics. *Journal of Climate*, *22*(5), 1193–1207. https://doi.org/10.1175/2008JCLI2358.1

Schär, C., Vidale, P. L., Lüthi, D., Frei, C., Häberli, C., Liniger, M. A., & Appenzeller, C. (2004). The role of increasing temperature variability in European summer heatwaves. *Nature*, *427*(6972), 332–336. https://doi.org/10.1038/nature02300

Seferian, R. (2018). CNRM-CERFACS CNRM-ESM2-1 model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.4068

Seland, Ø., Bentsen, M., Oliviè, D. J. L., Toniazzo, T., Gjermundsen, A., Graff, L. S., et al. (2019). NCC NorESM2-LM model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.8036

Semmler, T., Danilov, S., Rackow, T., Sidorenko, D., Barbi, D., Hegewald, J., et al. (2018). AWI AWI-CM1.1MR model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2686

Seneviratne, S., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., et al. (2012). Changes in climate extremes and their impacts on the natural physical environment. In C. B. Field, V. Barros, T. F. Stocker, D. Qin, D. J. Dokken, K. L. Ebi, et al. (Eds.), *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (pp. 109–230). Cambridge University Press.

Seneviratne, S., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., et al. (2021). Weather and climate extreme events in a changing climate. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Pean, S. Berger, et al. (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (chapter 11)*. Cambridge University Press. https://doi.org/10.1017/9781009157896.013

Shin, J.-Y., Kang, M., & Kim, K. R. (2022). Outdoor thermal stress changes in South Korea: Increasing inter-annual variability induced by different trends of heat and cold stresses. *Science of the Total Environment*, *805*, 150132. https://doi.org/10.1016/j.scitotenv.2021.150132

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres*, *118*(4), 1716–1733. https://doi.org/10.1002/jgrd.50203

Sillmann, J., Kharin, V. V., Zwiers, F. W., Zhang, X., & Bronaugh, D. (2013). Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *Journal of Geophysical Research: Atmospheres*, *118*(6), 2473–2493. https://doi.org/10.1002/jgrd.50188

Skelton, A., Kirchner, N., & Kockum, I. (2020). Skewness of temperature data implies an abrupt change in the climate system between 1985 and 1991. *Geophysical Research Letters*, *47*(23), e2020GL089794. https://doi.org/10.1029/2020GL089794

Stone, D. A. (2019). A hierarchical collection of political/economic regions for analysis of climate extremes. *Climatic Change*, *155*(4), 639–656. https://doi.org/10.1007/s10584-019-02479-6

Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). CCCma CanESM5 model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.3610

Tamarin-Brodsky, T., Hodges, K., Hoskins, B. J., & Shepherd, T. G. (2019). A dynamical perspective on atmospheric temperature variability and its response to climate change. *Journal of Climate*, *32*(6), 1707–1724. https://doi.org/10.1175/JCLI-D-18-0462.1

Tamarin-Brodsky, T., Hodges, K., Hoskins, B. J., & Shepherd, T. G. (2020). Changes in northern hemisphere temperature variability shaped by regional warming patterns. *Nature Geoscience*, *13*(66), 414–421. https://doi.org/10.1038/s41561-020-0576-3

Tamarin-Brodsky, T., Hodges, K., Hoskins, B. J., & Shepherd, T. G. (2022). A simple model for interpreting temperature variability and its higher-order changes. *Journal of Climate*, *35*(1), 387–403. https://doi.org/10.1175/JCLI-D-21-0310.1

Tang, Y., Rumbold, S., Ellis, R., Kelley, D., Mulcahy, J., Sellar, A., et al. (2019). MOHC UKESM1.0-LL model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.6113

Tatebe, H., & Watanabe, M. (2018). MIROC MIROC6 model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.5603

Taylor, P., Boeke, R., Boisvert, L., Feldl, N., Henry, M., Huang, Y., et al. (2022). Process drivers, inter-model spread, and the path forward: A review of amplified arctic warming. *Frontiers in Earth Science*, *9*. https://doi.org/10.3389/feart.2021.758361

Thompson, V., Kennedy-Asser, A. T., Vosper, E., Lo, Y. T. E., Huntingford, C., Andrews, O., et al. (2022). The 2021 western North America heat wave among the most extreme events ever recorded globally. *Science Advances*, *8*(18), eabm6860. https://doi.org/10.1126/sciadv.abm6860

Twardosz, R., & Kossowska-Cezak, U. (2013). Exceptionally hot summers in central and Eastern Europe (1951–2010). *Theoretical and Applied Climatology*, *112*(3), 617–628. https://doi.org/10.1007/s00704-012-0757-0

Vogel, M. M., Hauser, M., & Seneviratne, S. I. (2020). Projected changes in hot, dry and wet extreme events' clusters in CMIP6 multi-model ensemble. *Environmental Research Letters*, *15*(9), 094021. https://doi.org/10.1088/1748-9326/ab90a7

Vogel, M. M., Zscheischler, J., Fischer, E. M., & Seneviratne, S. I. (2020). Development of future heatwaves for different hazard thresholds. *Journal of Geophysical Research: Atmospheres*, *125*(9), e2019JD032070. https://doi.org/10.1029/2019JD032070

Voldoire, A. (2018). CMIP6 simulations of the CNRM-CERFACS based on CNRM-CM6-1 model for CMIP experiment historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.4066

Voldoire, A. (2019). CNRM-CERFACS CNRM-CM6-1-HR model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.4067

Volodin, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., et al. (2019). INM INM-CM5-0 model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.5070

Volodin, E. M., & Yurova, A. Y. (2012). Summer temperature standard deviation, skewness and strong positive temperature anomalies in the present day climate and under global warming conditions. *Climate Dynamics*, *40*(5–6), 1387–1398. https://doi.org/10.1007/s00382-012-1447-4

von, E., Mortikov, E., Gritsun, A., Lykossov, V., Galin, V., Diansky, N., et al. (2019). INM INM-CM4-8 model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.5069

von Hippel, P. T. (2005). Mean, median, and skew: Correcting a textbook rule. *Journal of Statistics Education*, *13*(2). https://doi.org/10.1080/10691898.2005.11910556

Walt, A. J., & Fitchett, J. M. (2021). Trend analysis of cold extremes in South Africa: 1960–2016. *International Journal of Climatology*, *41*(3), 2060–2081. https://doi.org/10.1002/joc.6947

Wang, J., Guan, Y., Wu, L., Guan, X., Cai, W., Huang, J., et al. (2021). Changing lengths of the four seasons by global warming. *Geophysical Research Letters*, *48*(6), e2020GL091753. https://doi.org/10.1029/2020GL091753

Wang, J., Han, Y., Stein, M. L., Kotamarthi, V. R., & Huang, W. K. (2016). Evaluation of dynamically downscaled extreme temperature using a spatially-aggregated generalized extreme value (GEV) model. *Climate Dynamics*, *47*(9–10), 2833–2849. https://doi.org/10.1007/s00382-016-3000-3

Wehner, M., Stone, D., Shiogama, H., Wolski, P., Ciavarella, A., Christidis, N., & Krishnan, H. (2018). Early 21st century anthropogenic changes in extremely hot days as simulated by the C20C+ detection and attribution multi-model ensemble. *Weather and Climate Extremes*, *20*, 1–8. https://doi.org/10.1016/j.wace.2018.03.001

Weigel, K., Bock, L., Gier, B. K., Lauer, A., Righi, M., Schlund, M., et al. (2021). Earth System Model Evaluation Tool (ESMValTool) v2.0 – Diagnostics for extreme events, regional and impact evaluation, and analysis of Earth system models in CMIP. *Geoscientific Model Development*, *14*(6), 3159–3184. https://doi.org/10.5194/gmd-14-3159-2021

Wieners, K.-H., Giorgetta, M., Jungclaus, J., Reick, C., Esch, M., Bittner, M., et al. (2019). MPI-M MPI-ESM1.2-LR model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.6595

Wilks, D. S. (2002). Smoothing forecast ensembles with fitted probability distributions. *Quarterly Journal of the Royal Meteorological Society*, *128*(586), 2821–2836. https://doi.org/10.1256/qj.01.215

Wu, T., Chu, M., Dong, M., Fang, Y., Jie, W., Li, J., et al. (2018). BCC BCC-CSM2MR model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.2948

Yosef, Y., Aguilar, E., & Alpert, P. (2021). Is it possible to fit extreme climate change indices together seamlessly in the era of accelerated warming? *International Journal of Climatology*, *41*(S1), E952–E963. https://doi.org/10.1002/joc.6740

Yukimoto, S., Koshiro, T., Kawai, H., Oshima, N., Yoshida, K., Urakawa, S., et al. (2019). MRI MRI-ESM2.0 model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.6842

Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., et al. (2011). Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Climate Change*, *2*(6), 851–870. https://doi.org/10.1002/wcc.147

Zhang, Y., Li, Q., Ge, Y., Du, X., & Wang, H. (2022). Growing prevalence of heat over cold extremes with overall milder extremes and multiple successive events. *Communications Earth and Environment*, *3*(1), 73. https://doi.org/10.1038/s43247-022-00404-x

Zhao, Y., Qian, C., Zhang, W., He, D., & Qi, Y. (2021). Extreme temperature indices in Eurasia in a CMIP6 multi-model ensemble: Evaluation and projection. *International Journal of Climatology*, *41*(11), 5368–5385. https://doi.org/10.1002/joc.7134

Ziehn, T., Chamberlain, M., Lenton, A., Law, R., Bodman, R., Dix, M., et al. (2019). CSIRO ACCESS-ESM1.5 model output prepared for CMIP6 CMIP historical [Dataset]. Earth System Grid Federation. https://doi.org/10.22033/ESGF/CMIP6.4272

Zittis, G., Hadjinicolaou, P., Almazroui, M., Bucchignani, E., Driouech, F., El Rhaz, K., et al. (2021). Business-as-usual will lead to super and ultra-extreme heatwaves in the Middle East and North Africa. *npj Climate and Atmospheric Science*, *4*(1), 20. https://doi.org/10.1038/s41612-021-00178-7