Weather and Climate Extremes 28 (2020) 100255

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: http://www.elsevier.com/locate/wace

Contribution of mean climate to hot temperature extremes for present and future climates

Alejandro Di Luca [a,*], Ramón de Elía [b], Margot Bador [a], Daniel Argüeso [c]

[a] Climate Change Research Centre, ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, Australia
[b] Servicio Meteorológico Nacional, Buenos Aires, Argentina
[c] Department of Physics, University of Balearic Islands, Palma de Mallorca, Spain

ARTICLE INFO

Keywords:
Future changes
Climate extremes
Model evaluation
Temperature decomposition
Extreme anomaly
Stationarity

ABSTRACT

The occurrence of very high temperatures (hot extremes) is often linked with negative impacts in human health, natural ecosystems and the economy (e.g., energy, water supply and agriculture). Studies have invariably shown that the intensity and frequency of hot extremes will increase in the future thus increasing their associated risks. While much progress has been made in quantifying and understanding hot temperature extremes and their future changes, there are still open questions. This paper focusses on the sources of hot extremes and their changes by applying a simple and unambiguous methodology that describes daily hot extremes as the superposition of four well known physical terms that include information on the annual mean temperature, the amplitude of the annual cycle, the diurnal temperature range and the local temperature anomaly on the day of the extreme. The methodology was applied to 30-year daily temperature records from 6 observation-based datasets and 31 atmosphere-ocean global climate models from the Coupled Model Intercomparison Project Phase 5 (CMIP5). The comparison between observed and simulated hot extremes shows a remarkably consistent picture where most CMIP5 models overestimate the term describing the local temperature extreme anomaly over most regions of the globe regardless of the observed dataset considered. Simultaneously, CMIP5 models show a systematic cold bias in the annual mean temperature and in the diurnal temperature range terms leading to substantial error compensation over some regions. This prompted us to define a new error estimator as the sum of errors in individual terms that appears to be much more effective at characterising model's performance compared to the traditional bias estimator. The assessment of future changes in hot extremes shows that changes are dominated by changes in annual mean temperatures with varying contributions from the other terms that strongly depend on the specific region being considered. Western Europe appears as a hot spot for extreme temperature changes (increases of ~8 °C by the end of the century) due to significant contributions from all decomposition terms including the summer mean anomaly, the diurnal temperature range and the daily extreme anomaly. Tropical South America also appears as a hot spot for extreme temperature changes (increases of ~7 °C) largely due to an increase in the daily extreme anomaly term (explaining about 30% of the full change) making this region one of the most sensitive regions in the world in terms of hot extremes. The analysis reveals that the separation of future changes according to terms describing mean, variability and tails is very sensitive to the specific way the mean component is defined including assumptions about stationarity.

1. Introduction

Record breaking temperatures and their underlying warm spells and heatwaves often have substantial social and environmental impacts including large economic losses, excessive mortality and severe environmental effects, as observed for example over Europe in summer 2003 (Garcia-Herrera et al., 2010).

For this reason, a great deal of attention has been devoted to quantifying and understanding current hot extremes and their future changes. Studies include the analysis of changes in heat waves (Schär et al., 2004; Imada et al., 2014; Perkins et al., 2014), high percentiles of daily mean temperature (Kharin et al., 2013; Sillmann et al., 2013) and record breaking temperatures (Meehl et al., 2009; Coumou and Rahmstorf, 2012; Bador et al., 2016, 2017). Regardless of the region and the

* Corresponding author.
E-mail address: a.diluca@unsw.edu.au (A. Di Luca).

https://doi.org/10.1016/j.wace.2020.100255
Received 27 July 2019; Received in revised form 16 December 2019; Accepted 10 March 2020
Available online 16 March 2020

season considered, these studies have shown that the intensity and frequency of hot extremes will increase by the end of the 21st century largely controlled by the emission scenario considered (e.g., Field et al., 2012; Collins et al., 2013, and references therein).

Generally, daily extremes result from the superposition of different physical processes operating and possibly interacting with each other at various temporal scales including subdaily, daily, seasonal and decadal time scales (Sillmann et al., 2017). In mid and high latitudes, daily hot extremes require persistent synoptic patterns (e.g., atmospheric blocking patterns) that favour positive anomalies of warm advection and shortwave incoming radiation over multi-day periods leading to warm spells and heatwaves (e.g., Miralles et al., 2014). At seasonal time scales, daily hot extremes co-vary with seasonal mean temperatures which are largely controlled by the incoming solar radiation, the elevation and the proximity to the ocean and can be influenced by low-frequency modes of variability such as El Niño Southern Oscillation and the North Atlantic Oscillation (e.g., Lewis and King, 2015). At decadal time scales, daily hot extremes co-vary with long-term mean temperatures which are affected by changes in greenhouse gas concentrations, changes in the general circulation and a number of feedback processes (Orlowsky and Seneviratne, 2012; Rhein et al., 2013; Seneviratne et al., 2016). Surface conditions such as the amount of soil moisture have been shown to affect temperatures at daily, seasonal and decadal time scales (Fischer et al., 2012; Seneviratne et al., 2010; Hirschi et al., 2011; Vogel et al., 2017; Donat et al., 2017, 2018). Similarly, the presence/absence of snow and ice at the surface in high latitudes and high elevation areas has been shown to play a key role in determining near surface temperatures and their future changes (e.g., Chen et al., 2017; Pithan and Mauritsen, 2014).

Previous studies have shown that future changes in hot extremes show a strong spatial dependence (e.g., Kharin et al., 2013; Sillmann et al., 2013; Seneviratne et al., 2010; Hirschi et al., 2011; Bador et al., 2016; Vogel et al., 2017; Donat et al., 2017, 2018) and a quasi-linear relation with global mean temperature changes (Seneviratne et al., 2016). The largest changes in hot extremes have been projected to occur in some mid-latitude regions with changes attaining about 9 C° in western Europe by 2081–2100 according to the RCP8.5 scenario of concentrations (Collins et al., 2013). With the aim of quantifying the way future changes might occur, some studies focused on looking into whether future changes only arise due to a shift in the daily temperature distribution (change in the mean) and changes in other moments of the distribution (Kodra and Ganguly, 2014; Fischer and Schär, 2009; Fischer et al., 2012; Della-Marta et al., 2007; Lewis and King, 2017; Wehner et al., 2018; Seneviratne et al., 2016). Della-Marta et al. (2007) analysed historical trends of in situ observations in western Europe and found that the increases in the variance of daily summer maximum temperature could explain up to 40% of the changes in hot days over the period 1880–2005. Fischer and Schär (2009) analysed an ensemble of RCM simulation over Europe and found a consistent increase in the standard deviation of daily mean temperatures over south-central Europe in agreement with other studies (Schär et al., 2004; Schoetter et al., 2015; Donat et al., 2017).

Argüeso et al. (2016) and Lau and Nath (2012, 2014) separated future changes in extremes according to the mean and the variability/tails by comparing extremes from a future simulation with those obtained from a hypothetical scenario derived by artificially ‘‘shifting’’ the historical distribution using the long-term future mean change. Lau and Nath (2012, 2014) found that projected increases in heatwaves duration and frequency can be accounted for by the shift in the climatological mean in both North America and Europe. At the global scale, Argüeso et al. (2016) showed that in most mid- and high-latitude areas, mean temperature changes (i.e., a shift in the distribution) largely dominate changes in future heat waves with only a few regions showing some consistent contribution from an increased variability. In a recent study, Donat et al. (2017) calculated the ratio between the future change in extremes and its corresponding mean change thus identifying regions where extremes are projected to warm faster/slower than the long-term mean. They found hot spots for temperature changes in Europe, North America, South America, and Southeast China.

Some of the apparent disagreements in the role of mean, variability, and tails in the above-mentioned studies are caused by the way the ‘‘mean’’ term is defined. A number of studies assume that mean changes are given by changes in annual mean temperatures (e.g., Donat et al., 2017; Seneviratne et al., 2016) while others take into account the seasonality of these changes (e.g., Lau and Nath, 2012, 2014; Argüeso et al., 2016; Wehner et al., 2018). Most studies assume that mean changes are given by changes in the daily mean temperature thus effectively discarding possible effects of changes in the diurnal temperature range that have been shown to be substantial in some regions (e.g., Lindvall and Svensson, 2015). Similarly, the use of the long-term mean to characterise a non-stationary period can be misleading as, if changes are monotonous through the period (as is the case for mean temperatures), the mean temperature differs at the beginning and the end of the period.

In this article we present a framework that allows to quantify in a consistent manner the effect that processes occurring at different temporal scales have on hot extremes. The framework is based on a simple decomposition that includes information from long-term local means, the seasonality, the diurnal temperature range and the local temperature anomaly of the day of the extreme. In addition, we assess and discuss some artifacts that arise when decomposing extremes due to historical and future climate periods being non-stationary. The decomposition is applied to historical and future daily temperature time series derived from four observed gridded datasets, two reanalyses and 31 climate models. Present climate results allow to disentangle climate model biases and to identify those terms that are at the origin of major errors. Quantifying the effect of different terms in future changes of temperature extremes helps to identify whether the changes result simply from a general trend in global mean temperatures or from more complex changes associated with other local processes. We focus on hot extremes as defined by the absolute 30-year daily maximum temperatures, but in order to ensure the robustness of the analysis we also consider less extreme quantities such as the 95th and 99th percentile of daily maximum temperatures.

The paper is organised as follows. The next section presents the data (Section 2.1) and discusses in detail the decomposition method (Section 2.2) and the comparison approach (Section 2.3). Section 3.1 shows results for the observed extremes while Sections 3.2 and 3.3 assess biases and errors in CMIP5 models including a discussion about the compensation of errors in models. Future changes in hot extremes are presented Section 3.4 and Section 3.5 discusses the effects of dealing with non-stationary time series. A broad discussion of the results is presented in Section 4 while conclusions are given in Section 5.

2. Data and methods

2.1. Observed and simulated temperature data

Four land-only daily minimum and maximum temperature gridded datasets are used in this study. The BEST dataset is a product from the Berkeley Earth group (Rohde et al., 2013a, b) and consists on a homogenised gridded 1° latitude by 1° longitude daily temperature for land-only grid points. It was constructed by combining the interpolated daily temperature anomalies from individual stations with a monthly gridded dataset that contains substantially more records (nearly 37000 records during the period 1880–2011). The CPC dataset has been developed by the Climate Prediction Center (CPC) from the National Oceanic and Atmospheric Administration (NOAA) and provides daily minimum and maximum temperatures over a 0.5° latitude-longitude grid mesh for the period from 1979 to present. The HadGHCND dataset (Caesar et al., 2006; Donat et al., 2013a) provides daily minimum and maximum temperatures over a 2.5° latitude by 3.75° longitude for the period from 1950 to 2014. HadGHCND contains around 3000

quality-controlled stations with good coverage in North America, Europe, Japan, China and Australia although quite poor in most of the Southern Hemisphere including Africa and South America. Finally, monthly minimum and maximum temperature from version 3.2.1 of the Climate Research Unit (CRU-TS321) dataset (Harris et al., 2014) has been used as an alternative dataset to calculate those terms only involving long-term mean quantities.

In addition, daily minimum and maximum temperatures from the two latest reanalyses developed by the European Centre for Medium-Range Weather Forecasts are also included in this study. The ERA-Interim (ERAI) reanalysis (Dee et al., 2011) has been used in a number of studies looking at temperature extremes (e.g., Kharin et al., 2013; Sillmann et al., 2013). ERAI uses a T255 horizontal resolution, which corresponds to approximately 79-km spacing on a reduced Gaussian grid and is available from 1 January 1979 onward. The ERA5 reanalysis (Hersbach et al., 2018) has an enhanced horizontal resolution of 31 km and is also available over the whole globe from January 1979 onwards.

Daily minimum and maximum temperatures from 31 Atmosphere-Ocean Global Climate Models from the Coupled Model Intercomparison Project Phase 5 (CMIP5, Taylor et al. (2012) are used in this study (Table S1) in Supplementary Information). Results presented here are based on a single realisation from the historical (1971–2000, HIST) and the representative concentration pathway (RCP) 8.5 scenario (2071–2100, RCP8.5) experiments for each model. RCP8.5 represents a high emission scenario of greenhouse gases.

2.2. 30-Year absolute daily maximum temperatures decomposition

For any given day d, the daily maximum temperature TX can be expressed as

$$TX(d) = T_m(d) + \frac{DTR(d)}{2}, \tag{1}$$

where the first term in the right-hand side corresponds to the daily mean temperature and the second term is the diurnal temperature range (DTR), defined as $TX - TN$ with TN the daily minimum temperature. Eq. (1) is an identity only if the term $T_m(d)$ is defined as $(TN + TX)/2$. It should be noted however that the actual mean temperature of a given day might differ from $(TN+TX)/2$ because the diurnal temperature variation can be asymmetric. Neglecting the effect of the diurnal temperature asymmetry is common in climate studies including observations for which only TN and TX are available.

From the time series in Eq. (1) we can calculate hot temperature extremes. For the sake of simplicity, in what follows we will consider TX_x to be the absolute maximum of $TX(d)$ over the whole 30-year period:

$$TX_x = T_m(d_x) + \frac{DTR(d_x)}{2} \tag{2}$$

where d_x denotes the day when the extreme occurs. Given that the temperature in a given day can be decomposed in the long-term mean $\overline{()}$ and a daily anomaly $()'$ we have:

$$\begin{aligned} TX_x &= T_m(d_x) + \frac{DTR(d_x)}{2}, \\ &= \overline{T_m} + \frac{\overline{DTR}}{2} + T_m'(d_x) + \frac{DTR'(d_x)}{2}, \end{aligned} \tag{3}$$

The first two terms of the decomposition contain long-term mean values and includes the daily mean temperature and the mean DTR. The second two terms, denoted as primed quantities, include the specific temperature anomalies of the day d_x when the absolute maximum occurs (i.e., denoted as daily extreme anomaly terms).

In most places, with the exception of some tropical regions, absolute maximum temperatures occur during the summer season. This implies that the actual distribution of daily absolute maximum temperatures is better defined by the distribution of daily maximum temperatures over the summer season (see also Wehner et al., 2018). In order to highlight the impact of using summer months instead of the entire year, the summer-mean temperature can be trivially separated into the annual mean and a departure as:

$$\begin{aligned} \overline{T_m} \equiv \overline{T_m^s} &= \overline{T_m^a} + \left(\overline{T_m^s} - \overline{T_m^a}\right), \\ &= \overline{T_m^a} + \overline{T_m^s}'. \end{aligned} \tag{4}$$

where the prime stands for the departure of the summer mean with respect to the annual mean.

The full decomposition can then be obtained by replacing Eq. (4) into Eq. (3):

$$TX_x = \overline{T_m^a} + \overline{T_m^s}' + \frac{\overline{DTR}}{2} + T_m'(d_x) + \frac{DTR'(d_x)}{2}. \tag{5}$$

The first term in the right-hand side denotes the local annual-mean temperature while the second term informs about local anomalies arising from the consideration of summer mean instead of annual mean values (i.e., seasonality effect). The third term describes the effect of the local summer mean DTR. The last two terms denote the daily temperature anomaly on the day of the extreme as compared with the local summer mean and includes a component coming from the anomalous DTR of that day ($\frac{DTR'(d_x)}{2}$). The distinction between the anomaly compared with the mean DTR or with the mean daily temperature is of little interest at this stage and thus we will generally present the combined effect of these last two terms (i.e., $TX'(d_x) = TX'_x = T'_m(d_x) + \frac{DTR'(d_x)}{2}$):

$$TX_x = \overline{T_m^a} + \overline{T_m^s}' + \frac{\overline{DTR}}{2} + TX'_x. \tag{6}$$

For illustration purposes, Fig. 1 shows daily maximum temperature in two grid points located near Paris in France (left panels) and Montréal in Canada (right panels) according to HadGHCND data. Plots also show the annual, seasonal and DTR long-term mean values together with the resulting extreme anomaly term. A yearly and weekly window centred in the absolute maximum of the 30-years period (TX_x; large red dot) is shown in middle and bottom panels respectively.

Future changes in the absolute maximum of the 30-year maximum temperature can then be expressed as

$$\Delta TX_x = TX_x^{FUT} - TX_x^{HIST} = \Delta\overline{T_m^a} + \Delta\overline{T_m^s}' + \Delta\frac{\overline{DTR}}{2} + \Delta TX'_x \tag{7}$$

where $HIST$ and FUT denote temperature terms evaluated using some historical and future periods respectively.

2.3. Regridding extremes

As discussed by Chen and Knutson (2008), the interpretation of model output as areal mean values (instead of grid-point estimations) implies that observed and simulated datasets with different horizontal resolutions provide information at different spatial scales. Consequently, in the evaluation process, the direct comparison between extremes from datasets with different resolutions is ambiguous because it is generally not possible to determine whether differences are related to differences in the spatial scale or in the quality of the dataset. This issue is commonly approached by postprocessing (remapping) observed and simulated output so all data provide information at (approximately) the same spatial scale.

In this paper, before applying the temperature decomposition, we have regridded all observed and simulated daily minimum and maximum temperatures onto the coarsest resolution grid mesh (i.e., the 2.5° x 3.75° HadGHCND grid) using a conservative mapping technique. For simulated and reanalysis data, only grid points covered by land in at least 90% of the total surface are considered in the regridding process. This remapping ensures that information at spatial scales finer than HadGHCND's grid spacing is filtered out. It should be noted, however,

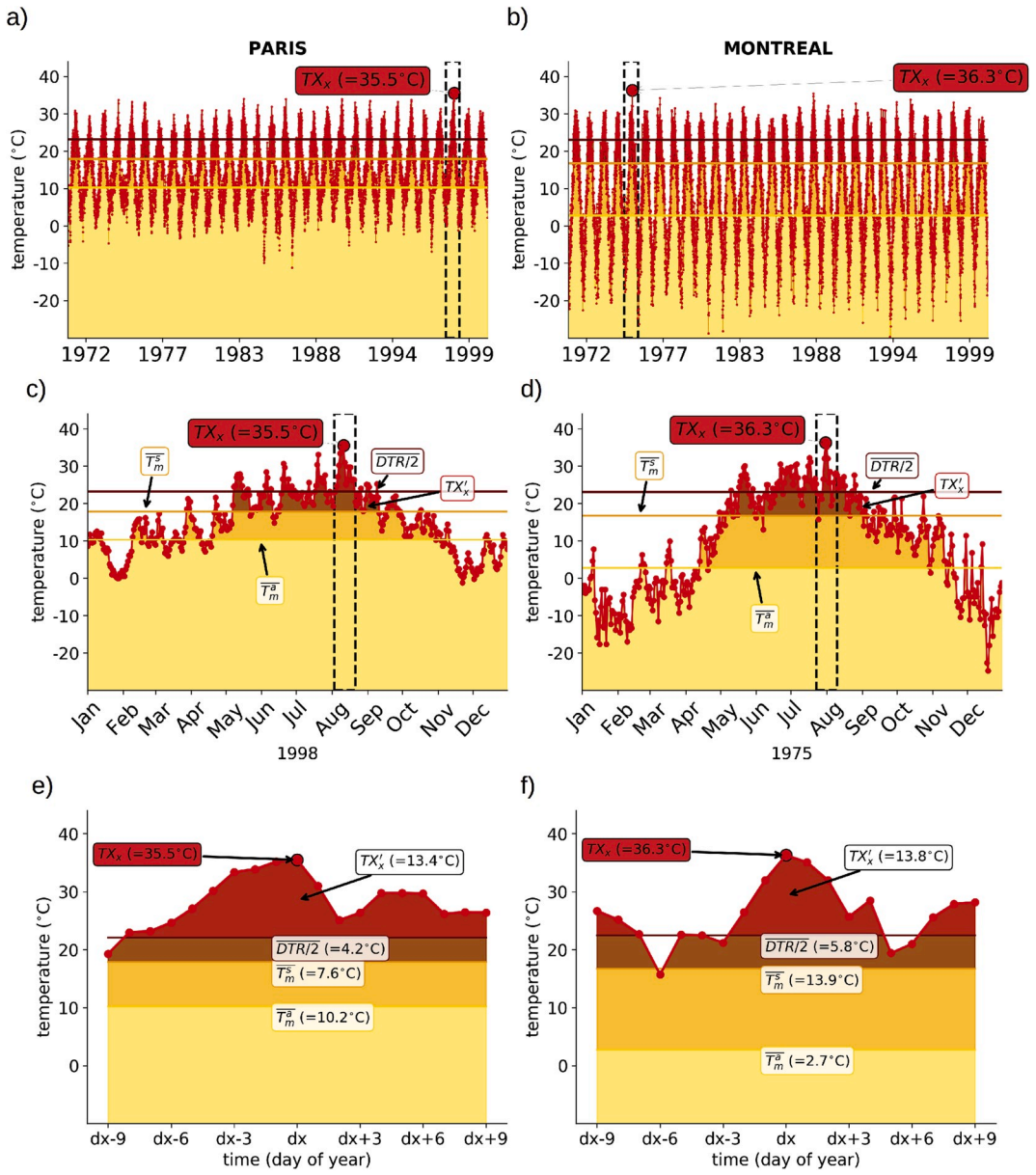

Fig. 1. Daily time series of maximum temperature (red) from the HadGHCND dataset in two individual grid points: one near Paris, France (left panels) and another near Montréal, Canada (right panels). Top panels show the full period 1971–2000, middle panels show the year when the absolute maximum occurred and bottom panels show a 19-day window centred on the day when the absolute maximum occurred. The absolute maximum (TX_x) is denoted with a large red dot. The annual mean temperature ($\overline{T_m^a}$), seasonal mean temperature ($\overline{T_m^s}$) and mean diurnal temperature range ($\overline{\frac{DTR}{2}}$) are also indicated in the plots. Bottom panels also include the contribution from each term to the absolute maximum temperature.

that such a methodology does not alter the influence that fine scales could have had on scales larger than the HadGHCND grid. Once the remapping has been applied, the temperature decomposition as described by Eq. (6) was applied to observed and simulated regridded datasets. Moreover, a mask that accounts for observed missing data is applied to all observed and simulated datasets to ensure that spatial averages across datasets represent the same areas. The time-invariant mask was created by discarding grid points with 20% or more missing daily values over the full period when considering the HadGHCND dataset.

In order to use data from the whole globe at a higher resolution, future changes are presented using a common, regular 2° by 2° grid mesh. In this case the remapping was performed using a simple nearest neighbour approach thus ensuring that each dataset retains the original information. Only grid points with at least 90% of land and with at least 16 models are considered in the analysis.

Regardless of the output grid used, global mean values were calculated using land grid points by weighting values by the area of grid boxes. Also, absolute maximum temperatures were calculated assuming that they occur during the summer season thus considering JJA and DJF months for the northern and southern hemispheres respectively. It should be noted that in some tropical or monsoonal regions, the calendar summer season might not correspond to the warmest season of the year. This issue is discussed in detail in Section 3.5.

3. Results

3.1. Observed hot extremes

Fig. 2 shows the absolute maximum temperature (Fig. 2a) and

decomposition terms (Fig. 2b–e) as estimated using the HadGHCND data. Figs. S1, S2, S3, S4 and S5 in the Supplementary Information show decomposition terms for BEST, CPC, ERAI, ERA5 and CRU-TS321 (only for long-term mean terms) respectively.

Absolute temperature extremes (Fig. 2a) vary between about 13 ˚C in some grid points in Greenland to nearly 48 ˚C in the hottest areas such as Australia, northern Africa and Middle East with a land-only global mean of 35.1 ˚C. According to the HadGHCND dataset, the local annual mean temperature (Fig. 2b) varies between about 28 ˚C in low latitudes and −19 ˚C in high latitudes with a land-only globally averaged annual mean temperature of 8.7 ˚C (note missing data in large tropical areas). Differences between summer and annual mean temperatures (Fig. 2c) are largest at high latitudes over the Northern Hemisphere where they can attain 25 ˚C and smallest in the subtropics where they are close to 0 ˚C. The globally averaged difference between summer and annual mean values is 11.1 ˚C suggesting a large contribution from the seasonal cycle to the annual hot extremes. Similarly, temperature variations through the day (Fig. 2d) contribute to temperature hot extremes by between 2 and 10 ˚C with the largest values occurring over dry regions such as southwest North America, Patagonia and southern Australia. Globally averaged the diurnal temperature range term ($\frac{DTR}{2}$) contributes by 5.9 ˚C to hot extremes.

Finally, the local daily extreme temperature anomaly (Fig. 2e) varies between about 2 ˚C in some grid points in the tropics and about 17 ˚C in some continental grid points at high latitudes showing a general increase towards higher latitudes. According to the HadGHCND dataset, globally averaged the local daily extreme temperature anomaly is 9.4 ˚C, thus of similar magnitude to the seasonal cycle term.

Fig. 3 shows the observational uncertainty range in each grid point for daily extreme temperatures (Fig. 3a) and the four decomposition terms (Fig. 3b–e). The observational uncertainty range is lower than 3 ˚C in several midlatitude regions including North America, Europe, Central Asia and Australia while the largest values (about 20 ˚C) occur over poorly sampled regions such as Greenland, northern Africa, Himalayas and central Andes. Analysing the observational uncertainty according to the four decomposition terms shows that the largest contributions are given by the daily extreme anomaly followed by the annual mean and the diurnal temperature range terms. Uncertainties in the annual mean term are largest in areas with important topographic forcings such as the Himalayas, Greenland and central Andes while the daily extreme anomaly term shows a more spatially homogeneous pattern with less topographic dependence and values generally increasing towards higher latitudes.

Fig. 4a shows the globally averaged observational range for different subsets of observations created by removing, one at a time, individual observed datasets. Fig. 4a shows that not all datasets contribute similarly to the observational range and that this contribution also depends on the specific term being considered. For example, the observational uncertainty of the 30-year absolute maximum and the annual mean terms is largest when including the HadGHCND dataset that explains nearly 25 and 35% of the full range for those two terms respectively (see normalised observational range in Fig. 4b). For the other terms, including the summer mean anomaly, the mean DTR and the mean extreme anomaly terms, the observational uncertainty is largest when

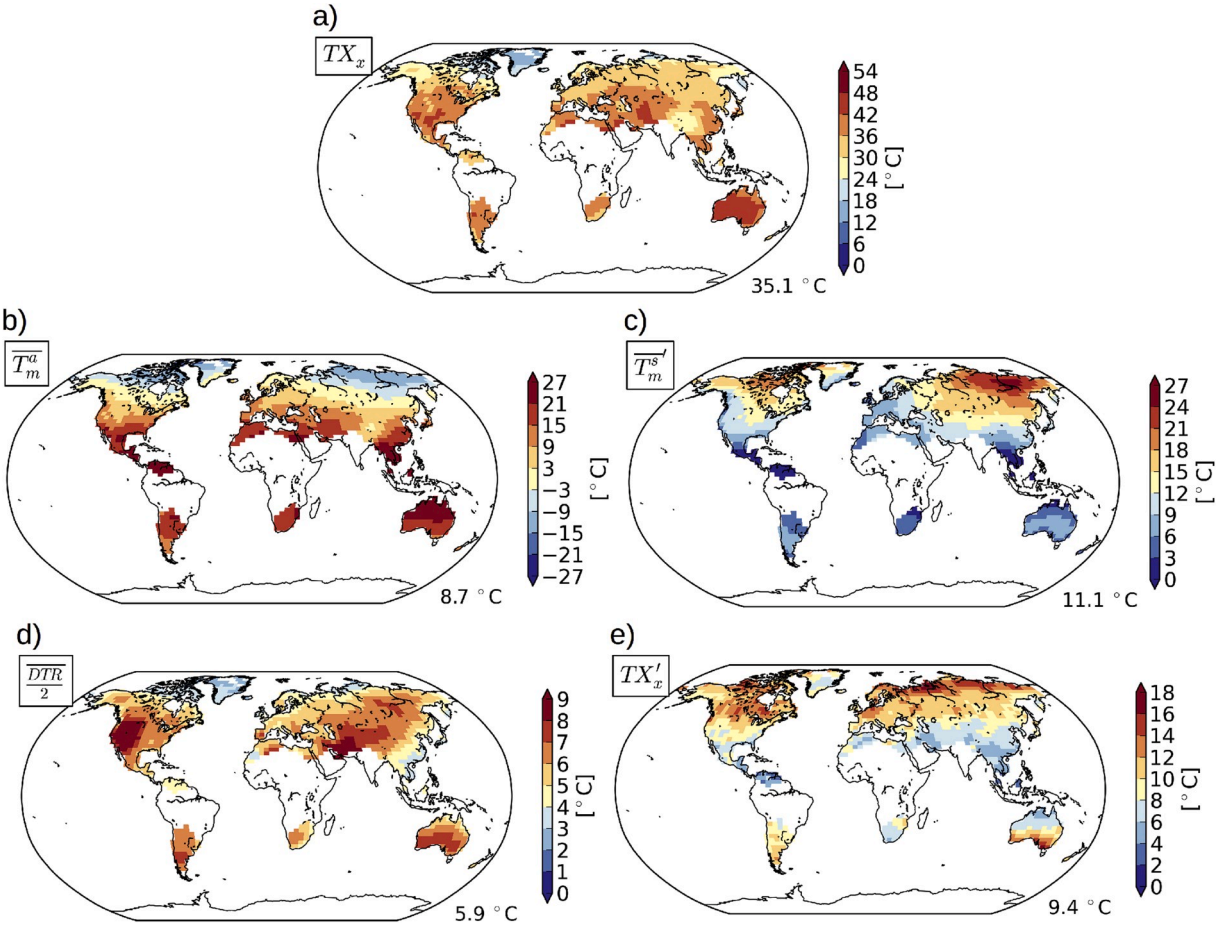

Fig. 2. (a) Absolute maximum temperature (TX_x), (b) local annual mean temperature ($\overline{T_m^a}$), (c) summer mean anomaly ($\overline{T_m^s}$), (d) summer diurnal temperature range ($\overline{\frac{DTR}{2}}$) and (e) daily extreme anomaly (TX'_x) from the HadGHCND dataset. See Eq. (6) for details on the decomposition. For each plot, the (area-weighted) global average is shown in the bottom-right corner. Note that the range of values plotted is different for different terms.

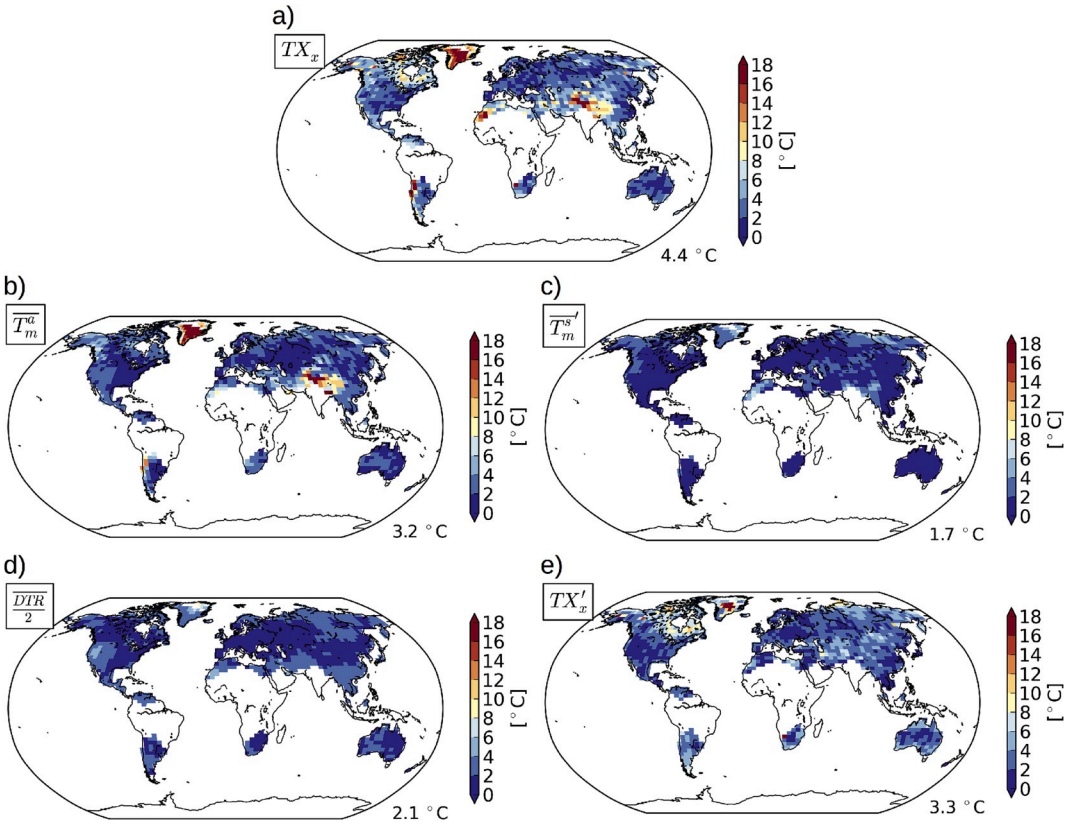

Fig. 3. Observational uncertainty range for the absolute maximum temperature (a) and the four decomposition terms (b–e) as presented in Eq. (6). The observational uncertainty range was calculated in each grid point as the largest difference between any two observation-based datasets, including the reanalyses. For each plot, the (area-weighted) global average is shown in the bottom-right corner.

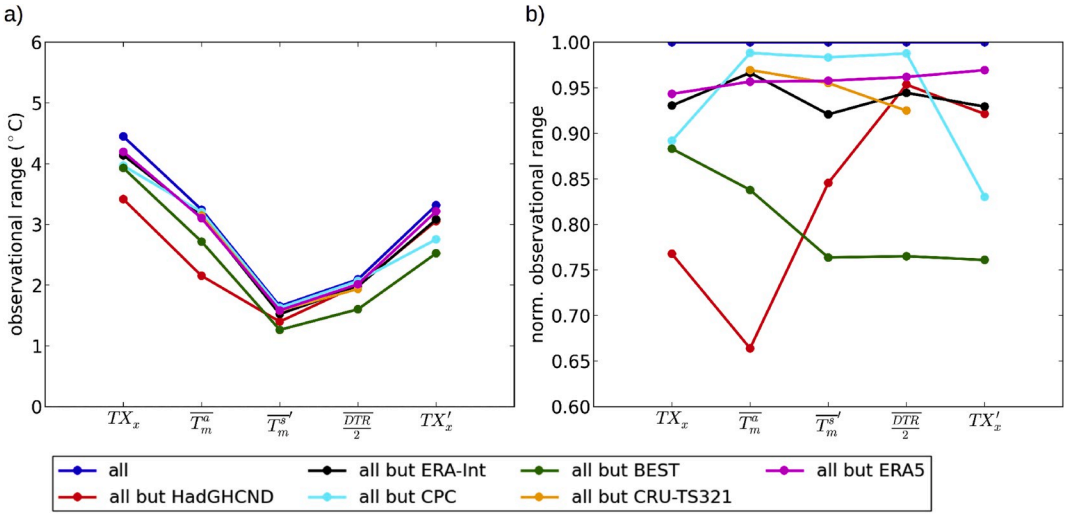

Fig. 4. a) Globally averaged observational uncertainty range for various subsets of observation-based datasets. b) Same as a) but normalised by the full range. The "all" subset provides the range when using all six observed datasets. The other subsets provide the range when excluding, one at a time, each of the six observed datasets.

including the BEST dataset that increases the full range by about 25%. At least for the mean DTR term, this seems to arise from some problem in the BEST dataset that shows negligible spatial variability compared to other observed datasets (see Fig. S1g).

It should be noted that hot extremes can occur on different days in the various observation-based datasets. Fig. S6 a in the Supplementary Information shows the percentage of events that occur at the same time for each pair of observation-based datasets. Extreme events are assumed to occur at the same time in two given datasets when the day of the extreme in one dataset occurs within (less or equal than) 3 days in the other dataset. Results show that, globally averaged, the percentage of events that are the same in two given datasets vary between 26 and 55% with the largest matching rates found between ERAI and ERA5 reanalyses and lowest between BEST and ERAI datasets. Fig. S6b presents the number of observation pairs that were matched in each grid point (the maximum is 10) showing that the largest agreement across datasets are

found over midlatitude regions while low latitude regions usually show low agreement. Further investigation is needed to better understand these values and their spatial distribution.

3.2. CMIP5 biases

Fig. 5 shows CMIP5 ensemble mean biases compared with the HadGHCND dataset for absolute maximum temperatures (Fig. 5a) and individual decomposition terms (Fig. 5b–e). The stippling denotes those grid points where at least 80% of the models agree on the sign of the bias. A measure of the global model's agreement on the sign of the bias is given by the land area agreement (LAA) parameters. LAA provides the proportion of land area where models agree in a positive or a negative bias respectively thus indicating whether models share similar systematic errors.

Multi-model mean biases in absolute maximum temperatures vary between −19 (in Greenland) and 17 °C (in northern Africa) with a globally averaged absolute maximum temperature bias of 1.4 °C when comparing with HadGHCND. As shown in Fig. 3, the observational uncertainty range in those two regions is particularly large (about 20 °C) suggesting caution when interpreting these biases. The first column in Table 1 shows global mean values of absolute maximum temperatures for the multi-model mean and observed datasets. The resulting multi-model mean biases compared to BEST, CPC, ERAI and ERA5 are 2.6, 0.7, 1.9 and 2.4 °C respectively showing that, despite observational uncertainties being large over specific regions, all observed datasets agree on a warm bias of hot extremes at the global scale.

The CMIP5 multi-model mean bias of 1.9 °C estimated here using ERAI data is much larger than the 0.5 °C bias reported by Kharin et al. (2013). Besides the use of a different base period (1981–2000 Vs. 1986–2005), there are a few reasons for this discrepancy. The main one is that in this paper the mean bias is calculated using only land grid points while Kharin et al. (2013) also included ocean grid points. Biases over the ocean appear to be much smaller thus substantially lowering the globally averaged bias. Second, we further mask areas where the HadGHCND data is not available thus excluding large regions over South America, Africa and India. When including these areas, the global mean bias compared with ERAI decreases by about 0.5 °C (from 1.9 to 1.4 °C). Third, a difference of 0.4 °C is explained by the regridding method used here to address the scale mismatch between models and observed products. Specifically, the regridding of ERAI from its original resolution to the HadGHCND grid using the conservative mapping technique results in hot temperature extremes being 0.4 °C lower when averaged over land areas.

Table 1
Globally averaged (weighted by area) values of the 30-year maximum absolute temperature (TX_x) and the four decomposition terms (see Eq. (6)) for the CMIP5 multi-model mean (MMM) and observation-based datasets. The global mean is calculated using only land grid points and a mask that excludes grid points with substantial missing observations. Bolded values denote those terms for which the MMM is outside the observational uncertainty range.

	TX_x	$\overline{T_m^a}$	$\overline{T_m^s}'$	$\overline{\frac{DTR}{2}}$	TX_x'
MMM	**36.5**	**7.1**	**11.9**	5.3	**12.2**
HadGHCND	35.1	8.7	11.1	5.9	9.4
BEST	33.9	8.0	11.6	5.7	8.5
CPC	35.8	8.1	11.2	5.5	11.0
ERAI	34.6	7.9	11.1	5.0	10.5
ERA5	34.1	7.5	11.3	5.0	10.3
CRU-TS321	n/a	7.6	11.3	5.8	n/a

Fig. 5b–e shows CMIP5 biases compared with HadGHCND for the

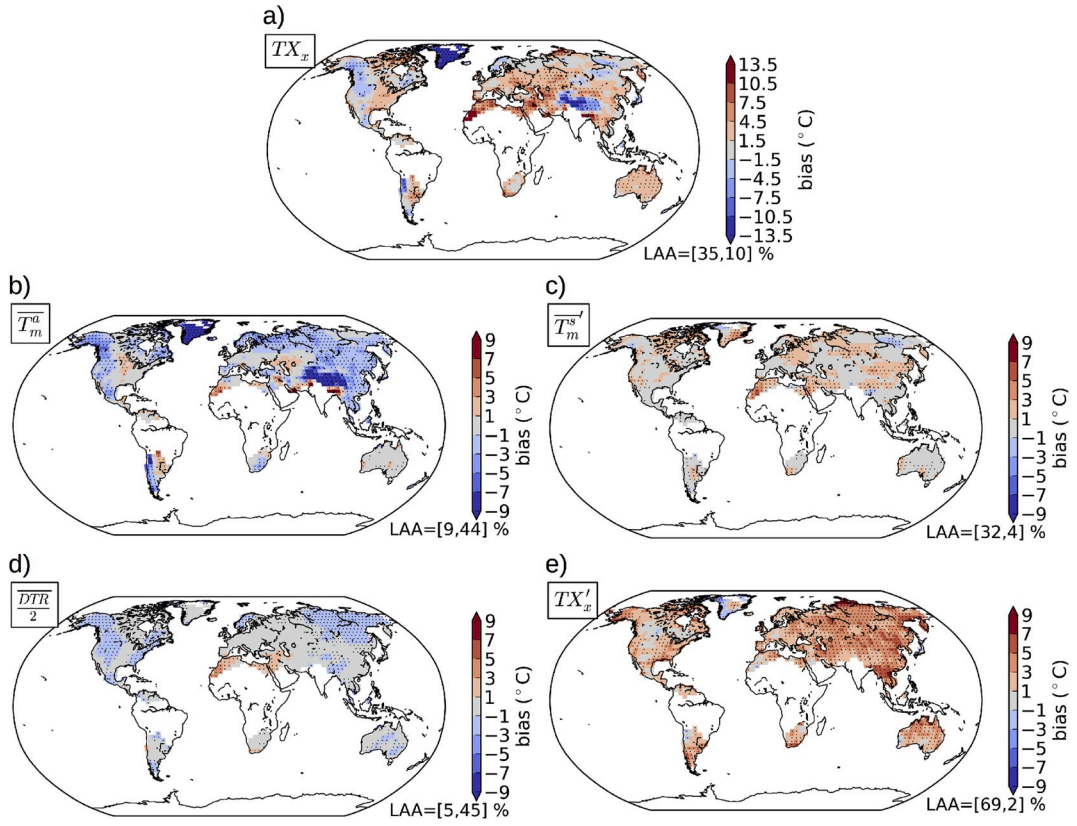

Fig. 5. CMIP5 ensemble mean biases for the absolute maximum temperature (a) and the four decomposition terms as presented in Eq. (6). Robust biases (stippled with small black dots) were identified when at least 80% of the models agree on the sign of the bias. For each plot, the proportion of land area showing positive and negative biases for at least 80% of the models is denoted by the LAA values in the bottom-right corner. Note that the range of values plotted in panel a) is different from panels b–e).

four decomposition terms (see Eq. (6)). The consideration of the various decomposition terms shows that the overall positive bias in absolute maximum temperatures is the net result of both positive and negative biases at different temporal scales. Specifically (see Table 1), the CMIP5 ensemble shows a negative bias in the globally averaged annual mean temperature compared to all observation-based datasets (−1.7, −0.8, −0.9, −0.8, −0.4 and −0.5 ˚C for the HadGHCND, BEST, CPC, ERAI, ERA5 and CRU-TS321 datasets respectively) and in the globally averaged mean DTR compared to all but the two reanalyses (−0.6, −0.4, −0.2, 0.3, 0.3 and −0.5 ˚C for HadGHCND, BEST, CPC, ERAI, ERA5 and CRU-TS321). While overall negative biases are observed between the CMIP5 multi-model mean and all observed datasets, the magnitude of these biases varies significantly in specific regions. For example, the well-known CMIP5 underestimation of annual mean temperatures over the Tibetan Plateau (Su et al., 2013; Chen et al., 2017) appears to be much stronger when compared with HadGHCND (−6.8 ˚C, areal mean over 22˚-40˚N and 68˚-106˚E) than with CRU-TS321 (−2.9 ˚C) or ERAI (−1.6 ˚C). The higher near-surface temperatures in CRU-TS321 compared with ERAI over the Tibetan Plateau are in agreement with values reported by Chen et al. (2017). Similarly, the globally averaged underestimation of DTR by CMIP5 models has been reported and discussed in more detail by Lindvall and Svensson (2015).

The CMIP5 ensemble mean shows warm biases compared to all observed datasets when considering the globally averaged summer-mean term (0.8, 0.3, 0.7, 0.8, 0.6 and 0.6 for HadGHCND, BEST, CPC, ERAI, ERA5 and CRU-TS321; see Table 1) and the globally averaged local daily extreme anomalies (2.8, 3.7, 1.2, 1.7 and 1.9 ˚C for HadGHCND, BEST, CPC, ERAI and ERA5). In the case of the local daily extreme anomaly, CMIP5 warm biases are not only evident at the global scale (see last column in Table 1) but are spatially widespread and robust across the ensemble. For example, the proportion of land area where models agree on a positive bias is 69, 87, 36, 40 and 48% for HadGHCND, BEST, CPC, ERAI and ERA5 datasets respectively. While the overall positive bias in the extreme anomaly term seems to be robust across regions and models, there are still some significant differences among observations. BEST and HadGHCND show a global mean value of the extreme temperature anomaly about 2.0 ˚C lower than the CPC dataset (see Table 1). Gross et al. (2018) reported that daily fields in the HadGHCND dataset could be over-smoothed due to the use of a relatively long radius in the interpolation process thus suggesting that HadGHCND might be underestimating the extreme temperature anomaly although it is unclear to what extent. The warm bias in the extreme anomaly term is even more robust and widespread when considering less extreme quantities than the 30-year absolute maximum temperature as shown in Fig. S7 for the 99th (left panels) and 95th (right panels) percentiles presumable due to a lesser influence of natural variability.

Fig. 6 shows mean differences between HadGHCND and observed and simulated datasets for each decomposition term over four distinct regions. In three of the four regions, the multi-model mean shows agreement regarding positive biases in daily temperature extreme anomalies with at least 25 out of 31 models showing a positive bias, which leads to a robust bias in the absolute temperature extremes in southeast South America and western Europe. In central Asia and eastern US, biases in the absolute maximum temperature are usually the result of a compensation between a negative bias in the annual mean term or the DTR and a large positive bias in the extreme anomaly term. Such compensations are present in other regions and will be discussed in more detail in the following section.

3.3. CMIP5 absolute and additive errors

The use of the mean bias to describe the performance of individual CMIP5 models to simulate temperature extremes can be misleading due to the compensation of errors of different sign arising from the decomposition terms. The decomposition can then be used to define a new estimator of the error, denoted here as the *additive error*, that explicitly accounts for individual errors as follows:

$$ae_{sum} = ae\left(\overline{T_m^a}\right) + ae\left(\overline{T_m^s}'\right) + ae\left(\frac{\overline{DTR}}{2}\right) + ae(TX'_x), \quad (8)$$

where *ae* denotes the absolute error and ae_{sum} the sum of the individual absolute errors. Defined in this way, $ae_{sum} >= ae(TX_x)$ and $ae_{sum} = ae(TX_x)$ if and only if at least three out of the four individual errors are zero. According to this error definition, a model is said to be right for the right reasons whenever the additive error ae_{sum} is small, implying that all individual errors are small.

Fig. 7 shows the additive error ae_{sum} as a function of the absolute error $ae(TX_x)$ for CMIP5 models (blue dots) and observed datasets (red dots) as compared to the HadGHCND product. Results are shown for the same four regions as in Fig. 5. The horizontal and vertical dashed lines represent the largest ae_{sum} and $ae(TX_x)$ errors between HadGHCND and any other observed dataset thus providing information on the observational uncertainty. The four quadrants delimited by the horizontal and vertical dashed lines can then be used to distinguish model's errors as follow:

1. Bottom-left quadrant: model errors are within the observational uncertainty
2. Top-left quadrant: model individual error ($ae(TX_x)$) is within the observational uncertainty
3. Bottom-right quadrant: model additive error (ae_{sum}) is within the observational uncertainty
4. Top-right quadrant: model errors are outside the observational uncertainty

Fig. 7 shows that in all regions there are some models that lie close to the one-to-one line meaning that, for those models, the additive error is largely driven by the error in a single term (usually the error in the extreme anomaly). Fig. 7 shows that the observational uncertainty is relatively small in eastern US, western Europe and southeast South America with the additive error between HadGHCND and other observed datasets lower than 4 ˚C. In contrast, the additive error attains 6 ˚C in central Asia. The number of models within the observational uncertainty region varies between 6 out of 31 in western Europe to 22 out of 31 in central Asia when considering the absolute error of the extremes ($ae(TX_x)$). However, when considering the additive error (ae_{sum}), there are at most 9 out of 31 models that lie outside the observational uncertainty proving that the additive error is a more sensitive estimator of the total error compared to the absolute error in temperature extremes. Moreover, in most regions the observational uncertainty range is given by the BEST dataset in part due to large differences in the DTR field that shows virtually no spatial variability compared with other observed datasets (see Fig. S1g). Fig. S8 shows the same as Fig. 7 but discarding the BEST dataset from the observational range. In each of the four regions, there are at most 2 models within the observational uncertainty of the additive error, showing that the additive error becomes much more efficient to separate models and observations.

It should be noted that the observational uncertainty as quantified here provides limited information. Small values in the observational uncertainty can be interpreted in two different ways. They can arise because independent datasets provide a similar estimation of the observed temperatures or because dependent datasets (i.e., based on the same input data) provide a similar estimation of the observed temperatures. The fact that we cannot ruled out the possibility of the underlying data being the same across different datasets means that we cannot disregard the second interpretation.

3.4. CMIP5 future changes

Fig. 8 shows the ensemble mean future changes (2071–2100 minus 1971–2000) of absolute 30-year maximum temperature (Fig. 8a) and the

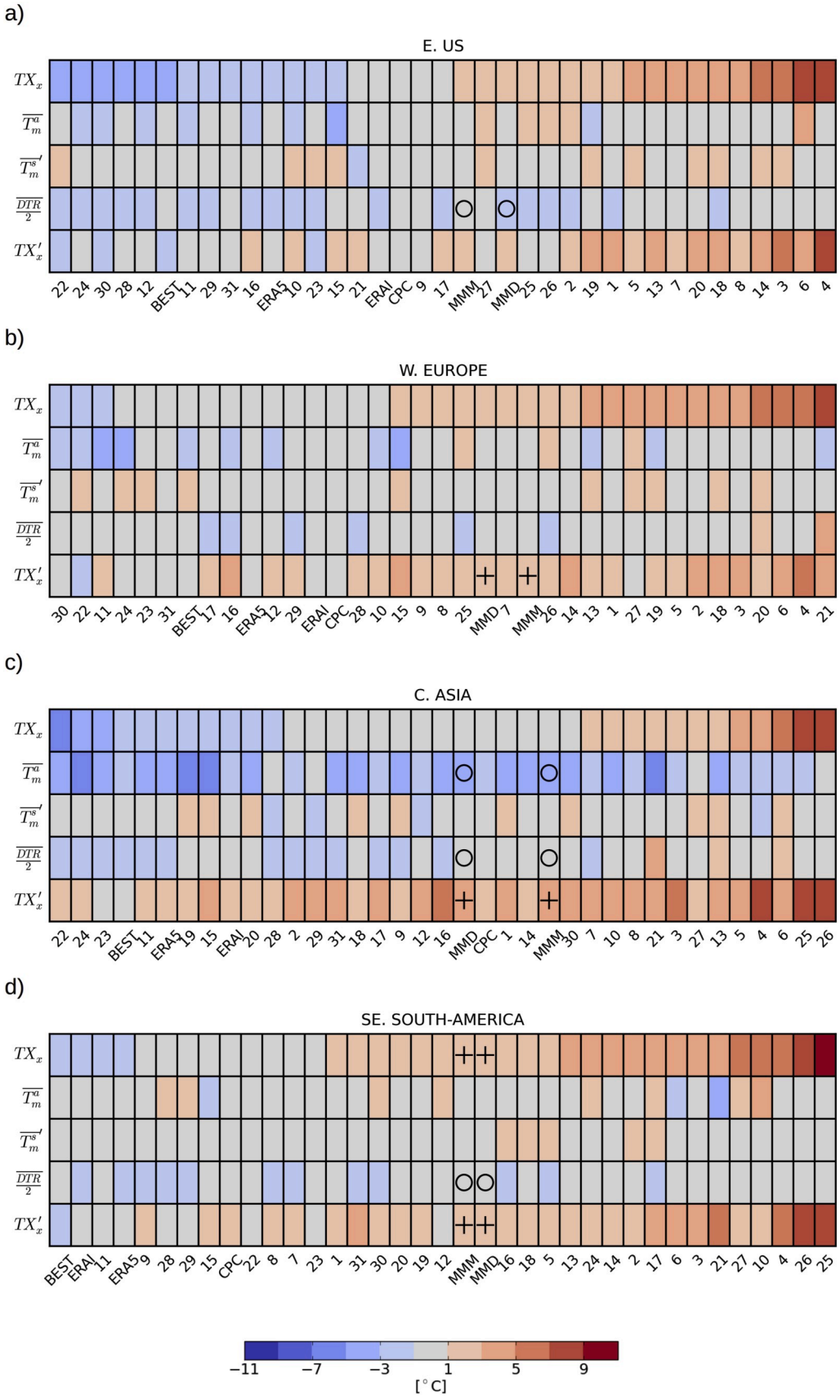

Fig. 6. Mean biases for CMIP5 models and observed datasets as compared to the HadGHCND dataset in four regions (see Fig. S11 for regions details). For each region, the various decomposition terms are shown in rows and individual model/observations are shown in columns including the CMIP5 multi-model mean (MMM) and multi-model median (MMD). Datasets are ranked from low to high according to the bias in absolute maximum temperatures (TX_x). Plus (circle) signs denote those terms for which at least 80% of the models show a positive (negative) bias. See Table S1 in the Supplementary Information to find out which model corresponds to each number.

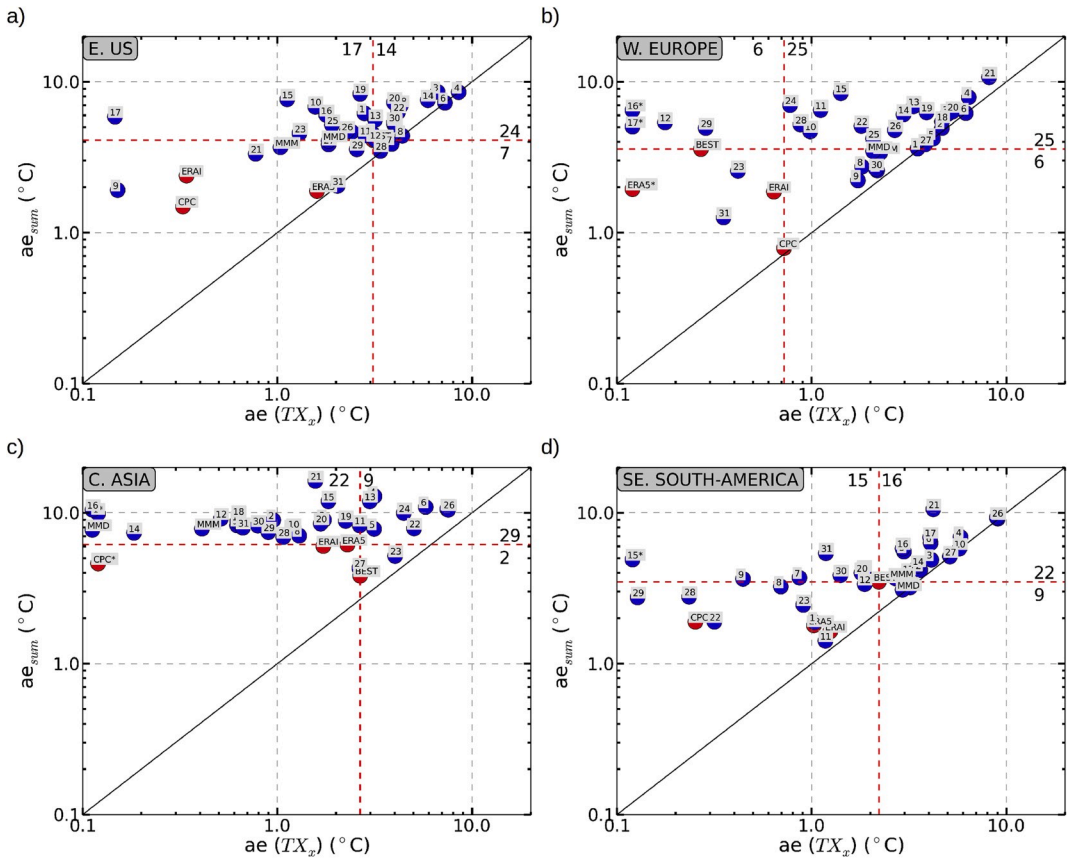

Fig. 7. Scatter plot of the sum of absolute errors (ae_{sum}; see Eq. (8)) and the absolute error of the absolute maximum temperature ($ae(TX_x)$) for individual CMIP5 models (blue circles) and observations (red circles) compared to the HadGHCND dataset. MMM and MMD denote the multi-model mean and multi-model median respectively. Red dashed lines show the largest difference between HadGHCND and other observations and thus indicate the observational uncertainty range. See Table S1 in the Supplementary Information to find out which model corresponds to each number.

four decomposition terms (see Eq. (7)). The stippling denotes grid points where at least 80% of the models agree on the direction (sign) of the change. In this case, the proportion of land area where models agree in the direction of the change is denoted by LAA with the first value for positive changes and the second for negative. It should be reminded that changes are calculated on a regular 2° by 2° grid as opposed to the HadGHCND grid used in the assessment of biases.

Averaged across CMIP5 models, future absolute maximum temperatures increase by between 0.6 and 8.7 °C, depending on the specific region considered, with a global mean increase of 5.6 °C (see Table 2) and at least 80% of the models showing a positive change in all grid points (i.e., LAA is 100%). The largest warming rates usually occur over midlatitude regions with changes consistently larger than 7 °C in central and east North America and Europe. Warming rates over 7 °C are also projected over tropical South America. Over land areas the lowest warming rates are projected over Greenland, Antarctica, Patagonia and most of Australia where changes are usually below 4.5 °C.

The largest overall contribution to the increases in absolute maximum temperatures is given by the annual-mean term that shows a globally averaged future change of 5.2 °C for the multi-model mean (Fig. 8b). This value is somewhat larger than the 4.8 °C value reported by Collins et al. (2013) using the same scenario (RCP8.5) presumably because they used a different number of models (39 instead of 32) and different present and future periods (1986–2005 and 1981–2100). Future changes in annual mean temperatures show a strong regional variation with values attaining 10 °C over the Arctic region and only about 2.5 °C in eastern Patagonia.

While changes in globally averaged annual mean temperatures are nearly as large as the globally averaged changes in absolute maximum temperatures, in order to explain regional differences, it is necessary to include the contribution from other terms. Fig. 8c shows that only in three land areas future changes in summer season will be larger than the annual changes: a large region around the Mediterranean Sea, a belt that extends through central US and a region that includes most of Argentina and central Chile. On the contrary, summer temperatures in high latitudes of the Northern Hemisphere show smaller warming rates than the annual mean value due to the dominant role played by polar amplification during the cold season (e.g. Collins et al., 2013).

Widespread and robust changes in DTR are only observed in a few regions. Over western Europe, changes in DTR contribute to increases in daily maximum temperatures. Over southern Sahara, southern Arabian Peninsula, Greenland and Antarctica, DTR changes contribute to decreases in daily maximum temperatures. In central US some positive contributions are also observed although the agreement between models is relatively low.

Future changes in the local daily extreme anomaly are often positive with a global mean value of 0.6 °C and local values attaining nearly 4 °C over the South American Monsoon (SAM) region and −3 °C over Greenland and Antarctica. At the global scale only 18% of grid points show positive changes that are robust across models and most of these grid points are located in land tropical areas. As found for the bias fields, future changes obtained using the 99th or the 95th percentiles (Fig. S9) are qualitatively very similar to those found when using the 30-year absolute maximum temperatures. There is, however, a clear increase in the robustness of the changes as less extreme metrics are considered. For example, the fraction of the area with robust positive changes increase from 18% for the 30-year absolute maximum temperature to 26 and 31% for the 99th and 95th percentiles respectively.

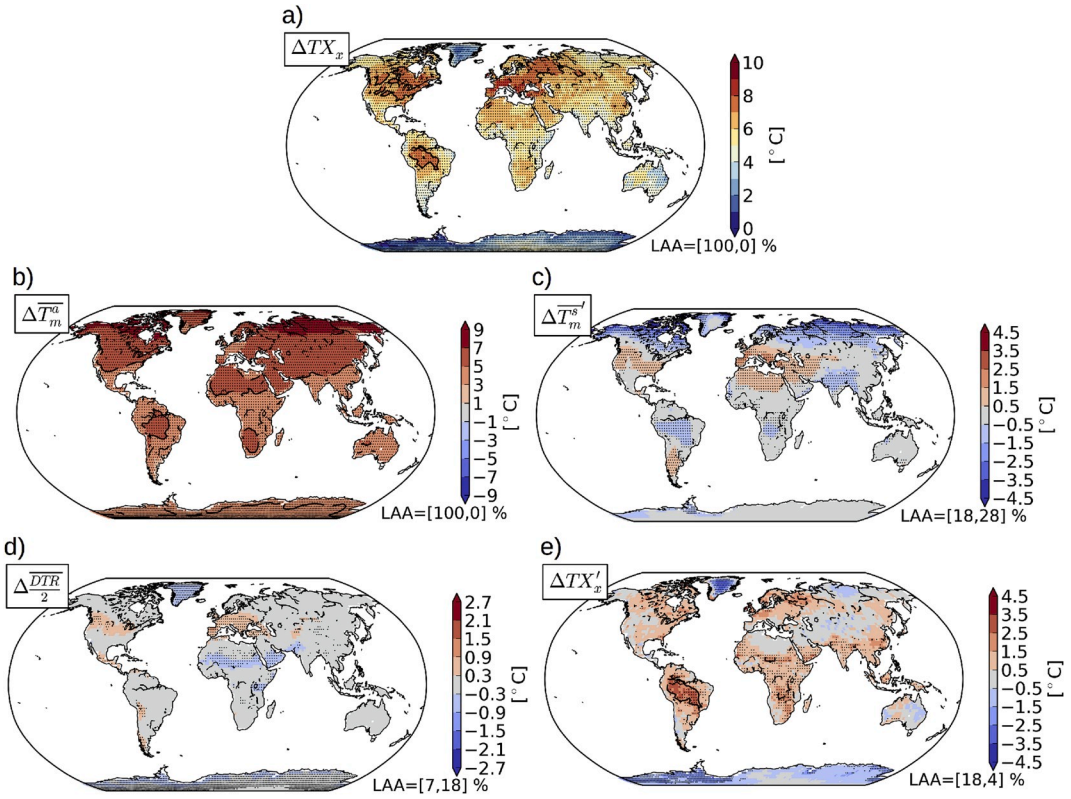

Fig. 8. CMIP5 ensemble mean future changes for absolute maximum temperature (a) and the four different decomposition terms (b–e) as presented in Eq. (7). Robust future changes (stippled with small black dots) were identified when at least 80% of the models agree on the sign of the change. Black contours denote values larger than 7 C° in a) and larger than 2 C° in b)-e). The proportion of land area showing a positive or a negative change with overall agreement among models are denoted by the LAA values in the bottom-right corner. Note that the range of values plotted is different for different terms.

Table 2
Globally averaged (weighted by area) values of the 30-year maximum absolute temperature (TX_x) and the four decomposition terms (see Eq. (6)) for the CMIP5 multi-model mean (MMM) in the historical period and the future changes. The global mean is calculated using only land grid points.

	TX_x	$\overline{T_m^{a,l}}$	$\overline{T_m^{s,*}}$	$\frac{\overline{DTR}}{2}$	TX_x'
historical	34.0	8.2	9.3	5.0	11.5
change (C°)	5.6	5.2	−0.2	0.0	0.6
LAA pos. (%)	100	100	18	7	18
LAA neg. (%)	0	0	28	18	4

Fig. 9 shows regional mean future changes for each decomposition term over four distinct regions: eastern US, western Europe, Amazon and eastern Australia (see Fig. S10 details on regions). In western Europe, all decomposition terms show positive and robust changes showing that the total change in absolute maximum temperatures result from the superposition of positive changes at various temporal scales. A similar behaviour is observed over eastern US. Over the South American Monsoon region, the large increases in hot extremes come from a very large contribution from the extreme anomaly term. Over eastern Australia, future changes are mostly explained by changes in the annual mean with negligible contributions from other terms.

3.5. Stationarity assumption and artifacts of the decomposition

This analysis would not be complete without a discussion of some fundamental hypothesis that underpins this development. In the derivation of Eq. (5), we have made two important assumptions that relate to the long-term stationarity of the climate. First, Eq. (5) assumes that long-term mean temperatures associated with daily extreme temperatures can be approximated using long-term mean summer (DJF or JJA) values. However, in areas where hot extremes do not occur systematically during summer (e.g., tropical and monsoonal regions) or areas with a large annual cycle (e.g., high-latitude regions), the summer mean might not be a good estimation of the mean temperature at the moment when daily extremes occur. In turn, since the sum in the right-hand side of Eq. (5) has to be equal to the extreme on the left-hand side, the incorrect estimation of the mean component will directly translate in an incorrect estimation of the extreme anomaly term. To quantify the effect of using the actual long-term mean temperature of the calendar day of the extreme, we have first calculated the mean temperature for each day using the 30-year period (i.e., 30 values) and a 10-day window centred upon the day of interest (Zhang et al., 2011). The resulting temperature extreme anomaly estimated using the calendar-day mean is smaller than the one estimated using the seasonal mean nearly everywhere with a global averaged difference of 1.2 °C (see left panels in Fig. S10 in the Supplementary Information). The largest differences between both estimations occur over India (monsoonal region), Antarctica and Siberia (large seasonality regions) where differences attain 3.0 °C. Fig. 10a) shows the resulting future change of the extreme anomaly term when using the calendar-day mean temperature. It shows a very similar spatial pattern compared with the change when using the seasonal mean (Fig. 8e) but the global average decreases by 0.04 °C and the percentage of land area where models agree in a significant increase decreases from 18 to 10%.

The second assumption about the stationarity is related to using a single value to characterise the full 30-year period without taking into account variations within the period. In the context of global warming, the use of the 30-year mean temperature to characterise present and future periods effectively means that mean temperatures are overestimated over the first half of the period and underestimated over the

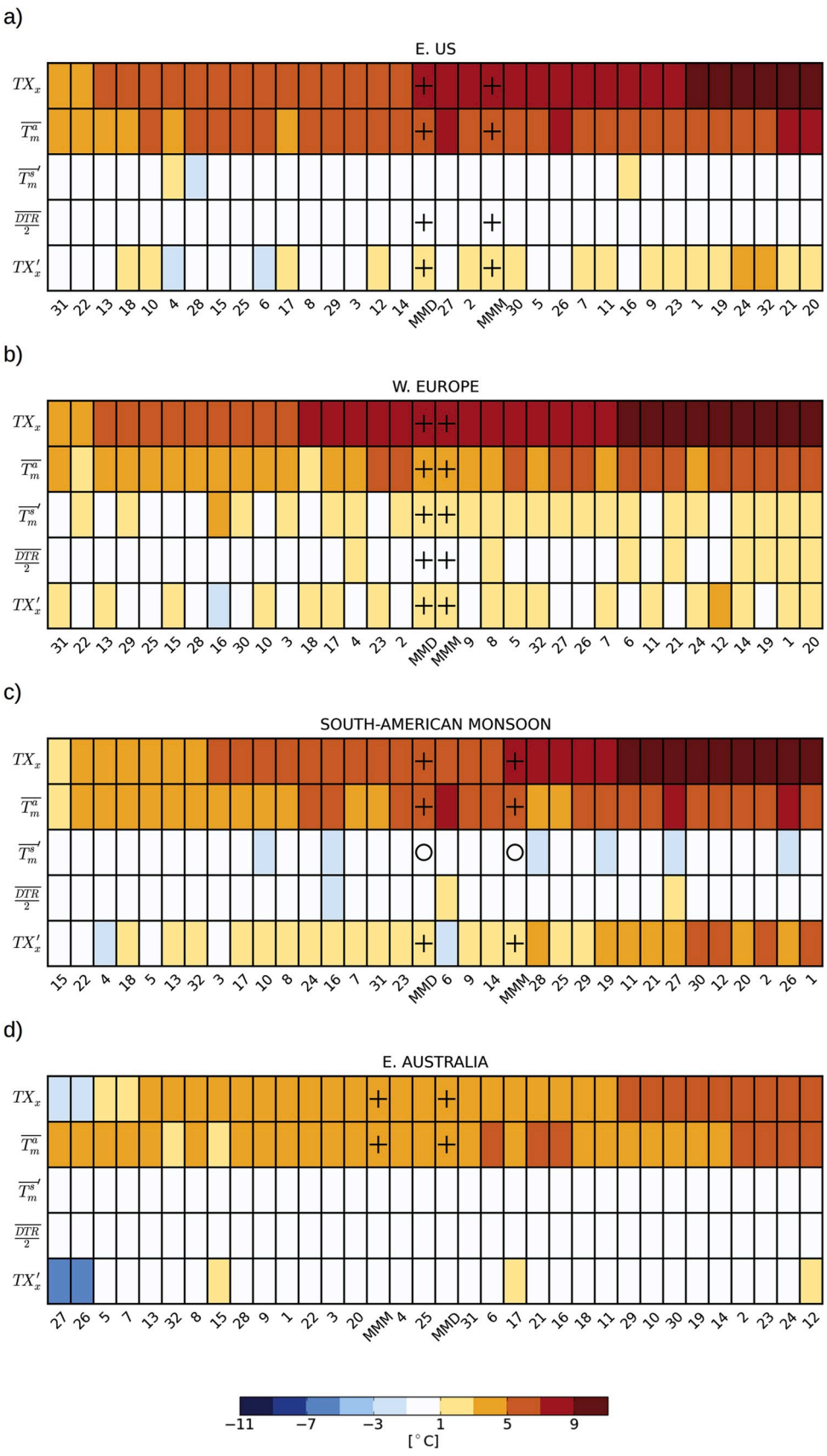

Fig. 9. Mean future changes for CMIP5 models in four regions (see Fig. S11 for regions details). For each region, the various decomposition terms are shown in rows and individual model are shown in columns including the multi-model mean (MMM) and median (MMD). Datasets are ranked from low to high according to the future change in absolute maximum temperatures (TX_x). Plus (circle) signs denote those terms for which at least 80% of the models show a positive (negative) change. See Table S1 in the Supplementary Information to find out which model corresponds to each number.

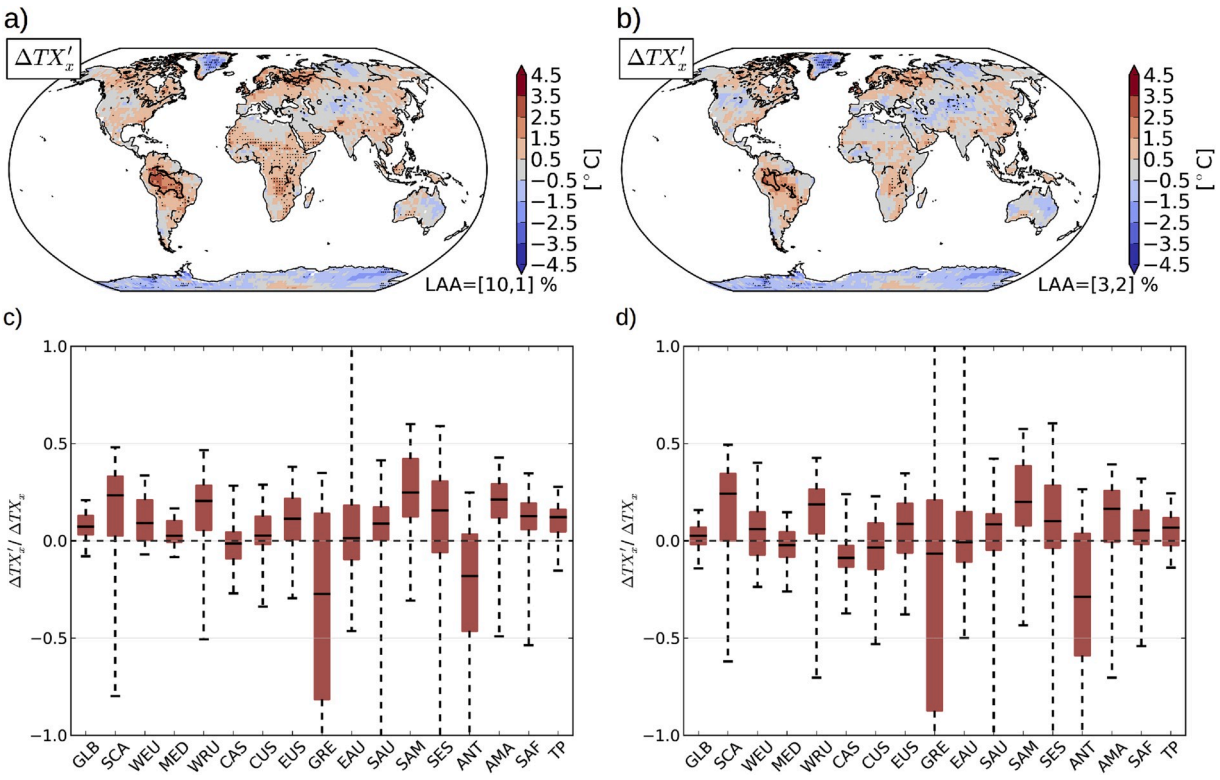

Fig. 10. Top panels show future changes in the daily extreme anomaly term as obtained using (a) the calendar-day mean and (b) the calendar-day mean plus the summer linear trend estimations. Bottom panels show multi-model boxplots of the fraction of the future change in the absolute maximum temperature that is explained by the extreme anomaly term in several regions (see Fig. S11 for regions details).

second half (see also Fischer and Schär, 2009; Della-Marta et al., 2007). This translates into an underestimation of the extreme anomaly term for the first half of the period and an overestimation in the second half. In addition, since warming rates (i.e. trends) in the RCP8.5 scenario are usually stronger at the end of the 21st century compared to the end of the 20th century, this trend effect leads to differences in future changes.

To quantify the trend effect, we have calculated summer mean linear trends at each grid point and for each model and period. We have then modified the calendar-day mean value by adding or subtracting a correction based on the value of the trend and the year when the extreme occurred. As found for the calendar correction, the temperature extreme anomaly estimated using the new correction is smaller than the one estimated using the seasonal mean nearly everywhere with a global averaged difference of 1.3 ˚C in the present climate and of 1.6 ˚C in the future climate (see right panels in Fig. S10 in the Supplementary Information). Results for the corrected extreme anomaly terms are shown in the right panels of Fig. 10. The future change in the extreme anomaly term when adjusted for the summer linear trend is still positive when averaged over the whole globe (0.2 ˚C) but much smaller than the non-adjusted value of 0.5 ˚C. In addition, the percentage of land area where models agree in a positive change in the extreme anomaly decreases from 10 to 3% when including the linear trend correction on top of the calendar-day correction. It should be noted that the trend correction applied here is only an approximation of the actual forced trend because of the influence of internal variability and the assumption of linearity. These effects should be tested more thoroughly in future works.

Bottom panels in Fig. 10 show the fractional contribution from the extreme anomaly term to the total future change in the absolute maximum temperature for calendar-day (Fig. 10c) and calendar-day plus linear trend (Fig. 10d) calculations. Box plots show the inter-model median, interquartile range and full range for the different regions shown in Fig. S11. The fraction of the future change in hot extremes explained by the extreme anomaly term varies between about 0.25 in the Scandinavian and South American Monsoon regions to −0.25 in Greenland and Antarctica. In the case of the full corrected estimation, only 2 regions (South American Monsoon and western Russia), out of the 17 regions considered, show a positive fraction in at least 80% of the CMIP5 ensemble.

4. Discussion

This paper addresses the long-standing issue of separating climate extreme temperatures according to long-term mean quantities (i.e., slow processes) and day-to-day weather variability (i.e., fast processes) (Gross et al., 2018; Donat et al., 2016, 2018). With this aim, we introduce a new methodology that describes temperature extremes as the superposition of four well known and physically coherent terms that include the annual mean temperature, the amplitude of the annual cycle, the mean diurnal temperature range and daily temperature extreme anomalies. The former three terms are calculated using long-term mean quantities and thus belong to the “mean component’’ while the daily temperature extreme anomaly term provides information on the specific synoptic and subsynoptic conditions surrounding the occurrence of the extreme and thus belong to the “variability/tail component’’.

The methodology is applied to address two key aspects of CMIP5 simulated hot temperature extremes: performance and projected changes. The evaluation is performed by taking into account the spatial-scale mismatch between the observed and simulated datasets involved and the observational uncertainty. This constitutes a novelty regarding earlier studies (e.g., Donat et al., 2016; Sillmann et al., 2013; Kharin et al., 2013) that performed comparisons between models and observations without explicitly addressing the spatial-scale mismatch. For example, the HadEX2 dataset (Donat et al., 2013b) has been widely used to evaluate climate models (e.g., Donat et al., 2016; Alexander and Arblaster, 2017; Sillmann et al., 2013) but its adequacy to identify biases in simulated extremes is questionable since HadEX2 data represent

grid-point (i.e., station) instead of areal-mean estimations. This issue has been acknowledged and discussed in many papers (e.g., Alexander and Arblaster, 2017; Sillmann et al., 2013).

The comparison of hot extremes from different observation-based datasets shows that the magnitude of the observational uncertainty strongly depends on the region being considered. In areas with relatively good station coverage such as Europe, North America and Australia, the observational uncertainty range is usually lower than 3 ˚C. However, in areas with more scarce coverage and/or characterised by complex topography the observational uncertainty range can be very large, attaining 20 ˚C in Greenland, northern Africa, Himalayas and central Andes. Decomposition terms show that the largest differences across observations arise from the representation of the extreme anomaly term and the annual mean term particularly in mountainous regions where a high density of stations is needed to well characterise the climatology.

While observational uncertainties can be substantial, the analysis of biases of different decomposition terms shows a remarkably consistent picture where CMIP5 models exhibit a widespread warm bias in the simulation of daily extreme anomalies that is often (regionally dependent) compensated by a cold bias in long-term mean terms. This finding is valid over most land areas of the globe and appears to be independent of observations. The superposition of a warm bias in the extreme anomaly term and (generally) a cold bias in the mean terms results in absolute hot extreme temperatures being overestimated although exhibiting important regional variations. While this general overestimation pattern has been reported before (Donat et al., 2016; Sillmann et al., 2013; Kharin et al., 2013), its partition according to different terms provides new insights on the performance of the models.

We define a new error estimator, called the *additive error*, as the sum of absolute errors in individual decomposition terms thus avoiding, to some extent, the error's compensation issue (see Palmer, 2016). The additive error provides a more sensitive characterisation of the total error in temperature extremes compared with the commonly used bias or absolute error. The additive error appears to be much more efficient at separating observation-based from simulated datasets proving to give a better characterisation of the overall performance of models. For example, at the global scale, the additive error of 30 out of 31 CMIP5 models is larger than the observational uncertainty while only 15 out of 31 models lie outside the observational uncertainty when the traditional absolute error is used. Palmer (2016) introduced the idea of "climatic Turing test'' to assess whether the origin of a given dataset, that is whether the data was produced using models or was obtained from observations, could be determined only by questioning the data without any a priori knowledge of its origin. In this sense, it seems that the additive error is a much more difficult "climatic Turing test'' for models to pass than the more classic absolute error suggesting that the additive error could be used more widely to characterise model's performance.

The application of the methodology to decompose future changes in hot temperature extremes provides insights on the sources of the changes and allows to identify regions that might undergo similar changes but due to different mechanisms. Western Europe belongs to a situation where very large increases in hot extremes are projected due to the superposition of positive changes in all decomposition terms: the annual mean, summer anomaly, the diurnal temperature range and the daily extreme anomaly term. On the other hand, tropical South America emerges as a future hot spot for extreme temperature increases due to changes in the annual mean term but mostly due to changes in the daily extreme anomaly term which shows the most robust and some of the largest values in the globe. The high contribution from the extreme anomaly term (about 20% of the change in the full extreme) makes tropical South America one of the most sensitive regions in the world in terms of hot extremes. There are also a few regions such as central Asia, Greenland and Antarctica where the extreme anomaly term is expected to decrease.

Our results show that the separation of future changes in extremes according to a mean component and variability/tails is very sensitive to the specific way the *mean component* is defined including assumptions about stationarity. We found that when properly estimating the various decomposition terms, most regions of the globe show no robust changes in the extreme anomaly term suggesting little influence of variability/tail changes in future extremes.

The metric used here to characterise hot extremes is the most extreme possible using a 30-year period (at least at daily time scales): the absolute maximum temperature. The consideration of more moderate extremes (i.e., 99th and 95th percentiles) shows qualitatively similar results both for the evaluation part and the future changes. There are, however, some quantitative differences. First, the uncertainties on the local daily extreme anomaly are reduced when considering more moderate extremes due to a substantial reduction in the influence of natural variability. This leads to a larger proportion of land area showing model agreement in the changes. For example, while only 2 (1) % of land area show a positive (negative) change in agreement across models for the absolute maximum temperature, the proportion increase to 8 (5) % and 12 (7) % when considering the 99th and 95th percentiles.

5. Conclusions

A simple and physically meaningful decomposition was used to separate hot extremes and their changes as the superposition of four well known physical terms that include information on the annual mean temperature, the amplitude of the annual cycle, the diurnal temperature range and the local temperature anomaly on the day of the extreme. Our results clearly show that the general warm bias in hot extremes as simulated by CMIP5 models results from a robust and widespread bias in the extreme anomaly term, even at expenses of a (general) cold bias in the long-term mean component. Furthermore, individual errors of decomposition terms were used to define a new error metric that allows to fix to some extent the issue of error compensation and whether simulated hot extremes were "right for the right reasons''.

The decomposition was also used to show that future changes in hot temperature extremes as projected using CMIP5 models are dominated by changes in long-term mean terms. This is valid over most of the globe including hot spots of extreme temperature changes such as western Europe and eastern US where changes in the variability/tail of daily temperature distribution contribute by at most 10% to the total change in hot extremes. The largest and most robust future changes in the variability/tail of the daily temperature distribution are projected to occur in tropical South America (South American Monsoon and Amazon regions) where median contributions attain nearly 20% of the total change. Other regions where changes in the variability/tail are important are Scandinavia and western Russia where changes tend to increase long-term mean changes and Greenland and Antarctica where changes tend to decrease long-term mean changes. It might appear paradoxical but a key implication of the finding above is that, in most regions, to understand future changes in hot extremes we need to focus on the seasonality and regional signature of long-term mean changes.

We believe that the decomposition presented here can help to better characterise model's overall performance and future projections when extending the analysis to other extremes (e.g., cold extremes) and other seasons. Ultimately, however, our confidence on the projected future changes relies on the understanding of the physical mechanisms behind the projected changes. With this aim, perhaps the decomposition could be applied to some of the key variables driving temperature changes (e.g., sensible and latent heat, temperature advection) in order to shed light on the physical sources behind the changes.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Alejandro Di Luca: Data curation, Conceptualization, Methodology, Software, Writing - original draft, Supervision, Writing - review & editing, Investigation, Visualization. **Ramón de Elía:** Conceptualization, Methodology, Writing - review & editing. **Margot Bador:** Conceptualization, Writing - review & editing. **Daniel Argüeso:** Validation, Conceptualization, Writing - review & editing.

Acknowledgments

We thank Olivier Geoffroy for helpful discussions and Andy Pitman for his useful comments of an earlier draft of the paper. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank each of the climate modelling groups for producing and making available their model output. For CMIP the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. CMIP5 data were obtained from the Australian node of the Earth System Grid Federation (ESGF), hosted at the National Computing Infrastructure (NCI). Furthermore, we acknowledge the efforts of the ECWMF for providing reanalysis data. CPC Global Temperature data was provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their Web site at https://www.esrl.noaa.gov/psd/. BEST data was provided by the Berkeley Earth Group and can be downloaded from http://berkeleyearth.org/data/. The HadGHCND dataset can be downloaded from https://www.metoffice.gov.uk/hadobs/hadghcnd/download.html. A. Di Luca was supported by the Australian Research Council (ARC) grants DE170101191 and CE170100023. D. Argüeso has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 743547. M. Bador was supported by the ARC Centre of Excellence for Climate Extremes (CE170100022).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wace.2020.100255.

References

Alexander, L.V., Arblaster, J.M., 2017. Historical and projected trends in temperature and precipitation extremes in Australia in observations and CMIP5. Weather Clim. Extremes 15, 34–56. https://doi.org/10.1016/j.wace.2017.02.001. ISSN 2212-0947. http://www.sciencedirect.com/science/article/pii/S2212094716300780.

Argüeso, D., Di Luca, A., Perkins-Kirkpatrick, S.E., Evans, J.P., 2016. Seasonal mean temperature changes control future heat waves. Geophys. Res. Lett. 43 (14), 7653–7660. https://doi.org/10.1002/2016gl069408. ISSN 00948276. https://search.crossref.org/?q=Seasonal+mean+temperature+changes+control+future+heat+waves.

Bador, M., Terray, L., Boé, J., 2016. Detection of anthropogenic influence on the evolution of record-breaking temperatures over Europe. Clim. Dynam. 46 (9–10), 2717–2735. https://doi.org/10.1007/s00382-015-2725-8. ISSN 0930-7575, 1432-0894. http://link.springer.com/10.1007/s00382-015-2725-8.

Bador, M., Terray, L., Boé, J., Somot, S., Alias, A., Gibelin, A.-L., Dubuisson, B., 2017. Future summer mega-heatwave and record-breaking temperatures in a warmer France climate. Environ. Res. Lett. 12 (7) https://doi.org/10.1088/1748-9326/aa751c}, 074025, ISSN 1748-9326. http://stacks.iop.org/1748-9326/12/i=7/a=074025?key=crossref.e9a5cefc0eb40086114731375b05ee1c.

Caesar, J., Alexander, L., Vose, R., 2006. Large-scale changes in observed daily maximum and minimum temperatures: creation and analysis of a new gridded data set. J. Geophys. Res.: Atmosphere 111 (D5), D05101. https://doi.org/10.1029/2005JD006280. ISSN 2156-2202. http://onlinelibrary.wiley.com/doi/10.1029/2005JD006280/abstract.

Chen, C.-T., Knutson, T., 2008. On the verification and comparison of extreme rainfall indices from climate models. J. Clim. 21 (7), 1605–1621. https://doi.org/10.1175/2007JCLI1494.1.

Chen, X., Liu, Y., Wu, G., 2017. Understanding the surface temperature cold bias in CMIP5 AGCMs over the Tibetan Plateau. Adv. Atmos. Sci. 34 (12), 1447–1460. https://doi.org/10.1007/s00376-017-6326-9. ISSN 1861-9533.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A., Wehner, M., 2013. Long-term Climate Change: Projections, Commitments and Irreversibility, Book Section 12. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1029–1136. https://doi.org/10.1017/CBO9781107415324.024. ISBN ISBN 978-1-107-66182-0. www.climatechange2013.org.

Coumou, D., Rahmstorf, S., 2012. A decade of weather extremes. Nat. Clim. Change 1758–678X, 1758–6798. https://doi.org/10.1038/nclimate1452.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., Mcnally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137 (656), 553–597. https://doi.org/10.1002/qj.828. ISSN 00359009.

Della-Marta, P.M., Haylock, M.R., Luterbacher, J., Wanner, H., 2007. Doubled length of western European summer heat waves since 1880. J. Geophys. Res. 112 (D15), D15103–D15111.

Donat, M., Alexander, L., Yang, H., Durre, I., Vose, R., Caesar, J., 2013a. Global land-based datasets for monitoring climatic extremes. Bull. Am. Meteorol. Soc. 94 (7), 997–1006. https://doi.org/10.1175/BAMS-D-12-00109.1.

Donat, M.G., Alexander, L.V., Yang, H., Durre, I., Vose, R., Dunn, R.J.H., Willett, K.M., Aguilar, E., Brunet, M., Caesar, J., Hewitson, B., Jack, C., Klein Tank, A.M.G., Kruger, A.C., Marengo, J., Peterson, T.C., Renom, M., Oria Rojas, C., Rusticucci, M., Salinger, J., Elrayah, A.S., Sekele, S.S., Srivastava, A.K., Trewin, B., Villarroel, C., Vincent, L.A., Zhai, P., Zhang, X., Kitching, S., 2013b. Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset. J. Geophys. Res. Atmos. 118 (5), 2098–2118. https://doi.org/10.1002/jgrd.50150. ISSN 2169-8996.

Donat, M.G., Alexander, L.V., Herold, N., Dittus, A.J., 2016. Temperature and precipitation extremes in century-long gridded observations, reanalyses, and atmospheric model simulations. J. Geophys. Res. 121 (19), 11174–11189. https://doi.org/10.1002/2016JD025480. ISSN 21562202.

Donat, M.G., Pitman, A.J., Seneviratne, S.I., 2017. Regional warming of hot extremes accelerated by surface energy fluxes: accelerated Warming of Hot Extremes. Geophys. Res. Lett. 44 (13), 7011–7019. https://doi.org/10.1002/2017GL073733. ISSN 00948276. https://search.crossref.org/?q=Regional+warming+of+hot+extremes+accelerated+by+surface+energy+fluxes%3A+accelerated+Warming+of+Hot+Extremes.

Donat, M.G., Pitman, A.J., Angélil, O., 2018. Understanding and reducing future uncertainty in midlatitude daily heat extremes via land surface feedback constraints. Geophys. Res. Lett. 45 (19), 10,627–10,636. https://doi.org/10.1029/2018GL079128. ISSN 19448007.

Field, C., Barros, V., Stocker, T., Qin, D., Dokken, D., Ebi, K., Mastrandrea, M., Mach, K., Plattner, G.-K., Allen, S., Tignor, M., Midgley, P., 2012. Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaption: Special Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, New York, NY. ISBN 978-1-107-02506-6 978-1-107-60780-4.

Fischer, E.M., Schär, C., 2009. Future changes in daily summer temperature variability: driving processes and role for temperature extremes. Clim. Dynam. 33 (7–8), 917–935. https://doi.org/10.1007/s00382-008-0473-8. ISSN 0930-7575, 1432-0894. http://link.springer.com/10.1007/s00382-008-0473-8.

Fischer, E.M., Rajczak, J., Schär, C., 2012. Changes in European summer temperature variability revisited. Geophys. Res. Lett. 39 (19) https://doi.org/10.1029/2012GL052730 n/a–n/a, ISSN 00948276, doi:\bibinfo{doi}{10.1029/2012GL052730}.

Garcia-Herrera, R., Diaz, J., Trigo, R.M., Luterbacher, J., Fischer, E.M., 2010. A review of the European summer heat wave of 2003. Crit. Rev. Environ. Sci. Technol. 40, 267–306. https://doi.org/10.1080/10643380802238137.

Gross, M.H., Donat, M.G., Alexander, L.V., Sisson, S.A., 2018. The sensitivity of daily temperature variability and extremes to dataset choice, 0 (0). J. Clim.. https://doi.org/10.1175/JCLI-D-17-0243.1. null.

Harris, I., Jones, P., Osborn, T., Lister, D., 2014. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. Int. J. Climatol. 34 (3), 623–642. https://doi.org/10.1002/joc.3711. ISSN 1097-0088.

Hersbach, H., de Rosnay, P., Bell, B., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Alonso-Balmaseda, M., Balsamo, G., Bechtold, P., Berrisford, P., Bidlot, J.-R., de Boisséson, E., Bonavita, M., Browne, P., Buizza, R., Dahlgren, P., Dee, D., Dragani, R., Diamantakis, M., Flemming, J., Forbes, R., Geer, A.J., Haiden, T., Hólm, E., Haimberger, L., Hogan, R., Horányi, A., Janiskova, M., Laloyaux, P., Lopez, P., Munoz-Sabater, J., Peubey, C., Radu, R., Richardson, D., Thépaut, J.-N., Vitart, F., Yang, X., Zsótér, E., Zuo, H., 2018. Operational global reanalysis: progress, future directions and synergies with NWP. https://www.ecmwf.int/node/18765.

Hirschi, M., Seneviratne, S.I., Alexandrov, V., Boberg, F., Boroneant, C., Christensen, O. B., Formayer, H., Orlowsky, B., Stepanek, P., 2011. Observational evidence for soil-moisture impact on hot extremes in southeastern Europe. Nat. Geosci. 4, 17–21. https://doi.org/10.1038/ngeo1032.

Imada, Y., Shiogama, H., Watanabe, M., Mori, M., Kimoto, M., Ishii, M., 2014. The contribution of anthropogenic forcing to the Japanese heat waves of 2013. Bull. Am. Meteorol. Soc. 95 (9), s52–s54.

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. ISSN 0165-0009, 1573-1480. http://link.springer.com/10.1007/s10584-013-0705-8.
Kodra, E., Ganguly, A.R., 2014. Asymmetry of projected increases in extreme temperature distributions. Sci. Rep. 4, 5884.
Lau, N.-C., Nath, M.J., 2012. A model study of heat waves over North America: meteorological aspects and projections for the twenty-first century. J. Clim. 25 (14), 4761–4784.
Lau, N.-C., Nath, M.J., 2014. Model simulation and projection of European heat waves in present-day and future climates. J. Clim. 27 (10), 3713–3730.
Lewis, S.C., King, A.D., 2015. Dramatically increased rate of observed hot record breaking in recent Australian temperatures. Geophys. Res. Lett. 42 (18), 7776–7784. https://doi.org/10.1002/2015GL065793. ISSN 19448007.
Lewis, S.C., King, A.D., 2017. Evolution of mean, variance and extremes in 21st century temperatures. Weather Clim. Extremes 1.
Lindvall, J., Svensson, G., 2015. The diurnal temperature range in the CMIP5 models. Clim. Dynam. 44 (1), 405–421. https://doi.org/10.1007/s00382-014-2144-2. ISSN 1432-0894.
Meehl, G.A., Tebaldi, C., Walton, G., Easterling, D., McDaniel, L., 2009. Relative increase of record high maximum temperatures compared to record low minimum temperatures in the U.S. Geophys. Res. Lett. 36 (23) https://doi.org/10.1029/2009GL040736. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009GL040736.
Miralles, D.G., Teuling, A.J., van Heerwaarden, C.C., Vilà-Guerau de Arellano, J., 2014. Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. Nat. Geosci. 7 (5), 345–349.
Orlowsky, B., Seneviratne, S.I., 2012. Global changes in extreme events: regional and seasonal dimension. Climatic Change 110 (3), 669–696. https://doi.org/10.1007/s10584-011-0122-9. ISSN 1573-1480.
Palmer, T.N., 2016. A personal perspective on modelling the climate system. Proc. Math. Phys. Eng. Sci. 472 https://doi.org/10.1098/rspa.2015.0772.
Perkins, S.E., Moise, A., Whetton, P., Katzfey, J., 2014. Regional changes of climate extremes over Australia - a comparison of regional dynamical downscaling and global climate model simulations. Int. J. Climatol. 34 (12), 3456–3478.
Pithan, F., Mauritsen, T., 2014. Arctic amplification dominated by temperature feedbacks in contemporary climate models. Nat. Geosci. 7 (3), 181–184. https://doi.org/10.1038/ngeo2071. ISSN 1752-0894, 1752-0908. http://www.nature.com/doifinder/10.1038/ngeo2071.
Rhein, M., Rintoul, S., Aoki, S., Campos, E., Chambers, D., Feely, R., Gulev, S., Johnson, G., Josey, S., Kostianoy, A., Mauritzen, C., Roemmich, D., Talley, L., Wang, F., 2013. Observations: Ocean, Book Section 3. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 255–316. https://doi.org/10.1017/CBO9781107415324.010. ISBN ISBN 978-1-107-66182-0. www.climatechange2013.org.
Rohde, R., Muller, R., Jacobsen, R., Muller, E., Perlmutter, S., 2013a. A new estimate of the average Earth surface land temperature spanning 1753 to 2011. Geoinf. Geostat. Overview 1 (1).
Rohde, R., Muller, R., Jacobsen, R., Perlmutter, S., Rosenfeld, A., 2013b. Berkeley Earth temperature averaging process. Geoinf. Geostat. Overview 1 (2).
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.
Schoetter, R., Cattiaux, J., Douville, H., 2015. Changes of western European heat wave characteristics projected by the CMIP5 ensemble. Clim. Dynam. 1–16.
Seneviratne, S., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., Teuling, A.J., Seneviratne, S.I., Corti, T., Davin, E.L., Hirschi, M., Jaeger, E.B., Lehner, I., Orlowsky, B., Teuling, A.J., 2010. Investigating soil moisture-climate interactions in a changing climate: a review. Earth Sci. Rev. 99 (3–4), 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004. ISSN 00128252.
Seneviratne, S.I., Donat, M.G., Pitman, A.J., Knutti, R., Wilby, R.L., 2016. Allowable CO 2 emissions based on regional and impact-related climate targets. Nature 529 (7587), 477–483 doi:\bibinfo{doi}{10.1038/nature16542}.
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. J. Geophys. Res.: Atmosphere 118 (4), 1716–1733. https://doi.org/10.1002/jgrd.50203. ISSN 2169-8996. http://onlinelibrary.wiley.com/doi/10.1002/jgrd.50203/abstract.
Sillmann, J., Thorarinsdottir, T., Keenlyside, N., Schaller, N., Alexander, L.V., Hegerl, G., Seneviratne, S.I., Vautard, R., Zhang, X., Zwiers, F.W., 2017. Understanding, modeling and predicting weather and climate extremes: challenges and opportunities. Weather Clim. Extremes 18 (August), 65–74. https://doi.org/10.1016/j.wace.2017.10.003. ISSN 22120947.
Su, F., Duan, X., Chen, D., Hao, Z., Cuo, L., 2013. Evaluation of the global climate models in the CMIP5 over the Tibetan plateau. J. Clim. 26 (10), 3187–3208. https://doi.org/10.1175/JCLI-D-12-00321.1.
Taylor, K.E., Stouffer, R.J., Meehl, G.A., 2012. An overview of CMIP5 and the experiment design. Bull. Am. Meteorol. Soc. 93 (4), 485–498.
Vogel, M.M., Orth, R., Cheruy, F., Hagemann, S., Lorenz, R., van den Hurk, B.J., Seneviratne, S.I., 2017. Regional amplification of projected changes in extreme temperatures strongly controlled by soil moisture-temperature feedbacks. Geophys. Res. Lett. 44 (3), 1511–1519. ISSN 19448007, doi:\bibinfo{doi}{10.1002/2016GL071235}.
Wehner, M., Stone, D., Mitchell, D., Shiogama, H., Fischer, E., Graff, L.S., Kharin, V.V., Lierhammer, L., Sanderson, B., Krishnan, H., 2018. Changes in extremely hot days under stabilized 1.5 and 2.0 °C global warming scenarios as simulated by the HAPPI multi-model ensemble. Earth Syst. Dynam. 9 (1), 299–311. https://doi.org/10.5194/esd-9-299-2018.
Zhang, X., Alexander, L., Hegerl, G.C., Jones, P., Tank, A.K., Peterson, T.C., Trewin, B., Zwiers, F.W., 2011. Indices for monitoring changes in extremes based on daily temperature and precipitation data. Wiley Interdiscipl. Rev.: Clim. Change 2 (6), 851–870. https://doi.org/10.1002/wcc.147. https://onlinelibrary.wiley.com/doi/abs/10.1002/wcc.147.