Weather and Climate Extremes 29 (2020) 100269

Contents lists available at ScienceDirect

Weather and Climate Extremes

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

Evaluation of the CMIP6 multi-model ensemble for climate extreme indices

Yeon-Hee Kim [a], Seung-Ki Min [a,*], Xuebin Zhang [b], Jana Sillmann [c], Marit Sandstad [c]

[a] Division of Environmental Science and Engineering, Pohang University of Science and Technology (POSTECH), Pohang, South Korea
[b] Climate Research Division, Environment and Climate Change Canada, Toronto, ON, Canada
[c] Center for International Climate Research, Oslo, Norway

ARTICLE INFO

Keywords:
Climate extremes
ETCCDI indices
CMIP6
Model evaluation
GEV analysis

ABSTRACT

This study evaluates global climate models participating in the Coupled Model Intercomparison Project phase 6 (CMIP6) for their performance in simulating the climate extreme indices defined by the Expert Team on Climate Change Detection and Indices (ETCCDI). We compare global climatology patterns of the indices simulated by the CMIP6 models with those from HadEX3 and four reanalysis datasets and the CMIP5 multi-model ensemble using root-mean-square errors for the 1981–2000 period. Regional evaluations are conducted for 41 sub-regions, defined for the Intergovernmental Panel on Climate Change Sixth Assessment Report. In particular, regional mean biases are analyzed for the 20-year return values (20RV) of the warmest day and coldest night temperatures (TXx and TNn) and annual maximum of daily precipitation (RX1day) using a Generalized Extreme Value (GEV) analysis. Results show that the CMIP6 models generally capture the observed global and regional patterns of temperature extremes with limited improvements compared to the CMIP5 models. Systematic biases like a cold bias in cold extremes over high-latitude regions remain even in stronger amplitudes. The CMIP6 model skills for the precipitation intensity and frequency indices are also largely comparable to those of CMIP5 models, but precipitation intensity simulations are found to be improved with reduced dry biases. The GEV analysis results indicate that the regional biases in 20RV of temperature extremes are dominated by GEV location parameter (related to mean intensity) with relatively small contribution from GEV scale/shape parameters (related to interannual variability). CMIP6-simulated 20RV of RX1day is characterized by dry biases over the tropics and subtropical rain band areas, as in the CMIP5 models, for which biases in both GEV location and scale/shape parameters are important.

1. Introduction

Extreme weather and climate events exert huge impacts on human society and ecosystems. It is imperative to understand the causes of their observed changes and to produce creditable future projections for climate change adaptation planning. For temperature extremes, the observed changes are significant such as the overall increases in hot extremes and decreases in cold extremes, and there is ample evidence for anthropogenic influences (Seneviratne et al., 2012; Bindoff et al., 2013; Min et al., 2013; Kim et al., 2016; Lu et al., 2018; Yin and Sun, 2018; Ting et al., 2020). Recent studies also found significant anthropogenic contributions to the observed large-scale intensification of precipitation extremes (Min et al., 2011; Zhang et al., 2013; Donat et al., 2019; Paik et al., 2020). Climate models participating in the Coupled Model Intercomparison Project phase 5 (CMIP5) project a continuous warming of extreme temperatures globally, with increased hot extremes and decreased cold extremes in terms of frequency and severity (Collins et al., 2013). The Intergovernmental Panel on Climate Change (IPCC) Special Report on the impacts of global warming of 1.5 °C (SR15) concluded with high confidence that temperature extremes are expected to increase up to 3–4.5 °C (relative to a preindustrial condition) over northern mid-to-high latitudes in a 1.5 °C warmer world (Hoegh-Guldberg et al., 2018).

Global climate models (GCMs) have been used as a primary tool for examining the past and future changes in climate extremes, and the comprehensive evaluation of GCM performances is important for proper interpretation of the simulated results. Recently, a new generation of GCMs have been developed for the CMIP6 experiments (Eyring et al., 2016). CMIP6 models have an increased range of complexity from GCMs to Earth System Models with improvements in physical processes and higher spatial resolution. One of the scientific focuses of the CMIP6 experiment is to assess changes in climate extremes for the past and

* Corresponding author.
E-mail address: skmin@postech.ac.kr (S.-K. Min).

https://doi.org/10.1016/j.wace.2020.100269
Received 28 December 2019; Received in revised form 22 June 2020; Accepted 23 June 2020
Available online 26 June 2020

future periods and to understand associated physical processes (Eyring et al., 2016; Marotzke et al., 2017).

There have been two approaches applied to defining and analyzing climate extremes (Seneviratne et al., 2012). One approach is to use the climate extremes indices defined by the Expert Team on Climate Change Detection and Indices (ETCCDI) (Klein Tank et al., 2009; Zhang et al., 2011). The ETCCDI indices represent relatively frequent extreme meteorological events in a given year or month, so it is relatively easy to understand their changes (Kharin et al., 2013). The other approach is to analyze changes in more extreme climate statistics based on the extreme value theory. They usually conduct a Generalized Extreme Value (GEV) distribution analysis to fit annual maxima of daily temperature and precipitation and then examine changes in rare events such as 20-year return values, which provide more relevant information for long-term planning (e.g., Kharin et al., 2007, 2013).

Previous studies evaluated CMIP5 models' performance for climate extremes using either approach (Sillmann et al., 2013; Kharin et al., 2013). Examining the overall performance of CMIP5 models using the ETCCDI indices based on global mean root mean square error (RMSE) measures, Sillmann et al. (2013) found a reduced inter-model spread in the CMIP5 ensemble in extreme temperature indices and improved simulations of precipitation intensity compared to the CMIP3 ensemble. Kharin et al. (2013) evaluated the 20-year return values of the annual maxima of temperature and precipitation in an ensemble of CMIP5 models. They found systematic model bias including the cold biases over high latitude regions in the Northern Hemisphere and the dry biases over the tropical and subtropical regions with large uncertainties.

The aim of this study is to evaluate the performances of the CMIP6 multi-model ensemble for temperature and precipitation extremes using the two approaches. We first evaluate the overall performance of CMIP6 models in terms of the 27 ETCCDI climate extreme indices in comparison to those of CMIP5 models, conducting an updated analysis of Sillmann et al. (2013). In addition, CMIP6 models are evaluated for 20-year return values of the annual maxima daily temperatures and precipitation. Here we consider the 41 sub-regions, newly defined for the IPCC Sixth Assessment Report (AR6). Regional mean biases in 20-year return values for CMIP6 models are compared with those for CMIP5 models. Following previous studies (Min et al., 2009; Park and Min, 2019), we further investigate the relative contribution of GEV location and scale/shape parameters (related to mean intensity and interannual variability, respectively) to the total biases in 20-year return values.

2. Data and methods

2.1. Data

We used two reference datasets: HadEX3 observations and four reanalysis datasets. The HadEX3 dataset provides gridded extreme indices for which extreme indices are first calculated at each stations and then interpolated onto a global grid with 1.875° longitude × 1.25° latitude spatial resolution (Dunn et al., 2020). HadEX3 do not fully cover the global land due to station-data availability. To include the global land for model evaluation and also to make comparison with previous studies (Sillmann et al., 2013, hereinafter referred to as S13), we use four reanalysis datasets: ERA40 (Uppala et al., 2005), NCEP/NCAR Reanalysis 1 (NCEP1; Kalnay et al., 1996), NCEP-DOE Reanalysis 2 (NCEP 2; Kalnay et al., 2002), and ERA-5 (Hersbach and Dee, 2016). The available periods of ERA40 and NCEP1 are 1958–2001 and 1948–2018, respectively. NCEP2 and ERA-5 provide data from 1979 to 2018. These reanalyses are the same as those used in S13 except ERA-5 which is used instead of ERA-Interim (Dee et al., 2011). The newly released ERA-5 has higher spatial (30 km) and time resolution (hourly) than the other three reanalysis datasets, potentially making it more suitable to estimate and evaluate the climate extremes. This study used ERA-5 reanalysis (1° × 1° resolution version) as a main reference dataset to evaluate CMIP6 models in comparison with CMIP5 models. Pointing out large discrepancies among four reanalysis datasets for climate extreme indies, S13 suggested a thoughtful choice of reference datasets. In this regard, the inter-reanalysis spread in model skills was evaluated. ERA-5-based results were also compared with HadEX3-based results on global scale with applying same data availability to gain an insight into the influence of observational uncertainty.

We used daily minimum and maximum temperature (TN and TX) and daily precipitation amount (PR) from 32 CMIP6 models (Table S1, Eyring et al., 2016) and 35 CMIP5 models (Table S2, Taylor et al., 2012). The historical simulations from both model groups were used for model evaluation, which were performed under anthropogenic (greenhouse gases and anthropogenic aerosols) plus natural (solar and volcanic activities) forcings. Although some CMIP6 models have large ensemble members, up to 31 members, we use a single member for each model for fair comparisons, typically the first member (r1i1p1f1 for CMIP6 and r1i1p1 for CMIP5).

2.2. Climate extreme indices definition and calculation

Table 1 shows the 16 temperature and 11 precipitation extreme indices defined by the ETCCDI (Zhang et al., 2011). The capital "X" and "N" means the daily maximum and minimum temperature, respectively. The small letter "x" and "n" means the annual or monthly maximum and minimum value, respectively. The indices can be divided into four groups: (1) absolute indices like the hottest day and coldest night per year (TXx and TNn) or the annual daily and five day maximum PR (RX1day and RX5day); (2) threshold indices, which count the number of days exceeding a fixed threshold such as frost days (FD) and summer days (SU), (3) percentile-based threshold indices, which indicate the exceedance rates below 10th percentile (10p) or above 90th percentile (90p) derived from the 1961–1990 base period like TN10p, TX10p, TN90p, and TX90p; (4) duration indices, which represent the length of warm and cold spells (WSDI and CSDI, based on percentile thresholds) or dry and wet spell (CDD and CWD, based on absolute threshold). When calculating the percentile-based threshold indices, the same base period of 1961–1990 was applied to all climate models and two reanalyses (NCEP1 and ERA-40). However, different base periods were used for NCEP2 (1979–2008) and ERA-5 (1981–2010) due to the different starting years, which may induce some differences in those indices among reanalyses, thereby affecting model skills. For more details, see Klein Tank et al. (2009) and Zhang et al. (2011).

The 27 climate extreme indices were calculated using the R package 'climdex.pcic'. The R package and its documentation is available on the website (http://cran.r-project.org/web/packages/climdex.pcic/index.html). All indices of GCMs and three reanalyses (ERA-40, NCEP1 and NCEP2) were computed on original model/reanalysis grids and then interpolated into a common 1° × 1° grid using a bilinear remapping before taking area-average for global land and 41 sub-regions. For ERA-5, we first interpolated daily temperature and precipitation data at 30 km resolution into 1° × 1° grid and then calculated climate extreme indices in order to reduce the difference in spatial scales from CMIP6 models (1–2°). We conducted the regional analysis for 41 sub-regions (Fig. 1) following the domains of Iturbide et al. (2020) prepared for the IPCC AR6. HadEX3 indices were used in its original resolution.

2.3. Model performance metric

Following S13, this study employed metrics based on the RMSEs of the model climatology pattern for the period of 1981–2000. The equation for RMSE is as follows:

$$\mathrm{RMSE}_{XY} = \sqrt{\langle (X - Y)^2 \rangle} \quad (1)$$

where X and Y are denoted as the model and reanalysis climatology of an index, respectively. The angular brackets represent the spatial mean over global land. The relative model RMSE (RMSE'_{XY}) for each model is

Table 1
27 extreme indices by the ETCCDI.

Label	Index Name	Index Definition	Units
TN10p	Cold nights	Let TN_{ij} be the daily minimum temperature on day *i* in period *j* and let $TN_{in}10$ be the calendar day 10th percentile centred on a 5-day window for the base period 1961–1990. The percentage of time for the base period is determined where: $TN_{ij} < TN_{in}10$	%
TX10p	Cold days	Let TX_{ij} be the daily maximum temperature on day *i* in period *j* and let $TX_{in}10$ be the calendar day 10th percentile centred on a 5-day window for the base period 1961–1990. The percentage of time for the base period is determined where: $TX_{ij} < TX_{in}10$	%
TN90p	Warm nights	Let TN_{ij} be the daily minimum temperature on day *i* in period *j* and let $TN_{in}90$ be the calendar day 90th percentile centred on a 5-day window for the base period 1961–1990. The percentage of time for the base period is determined where: $TN_{ij} > TN_{in}90$	%
TX90p	Warm days	Let TX_{ij} be the daily maximum temperature on day *i* in period *j* and let $TX_{in}90$ be the calendar day 90th percentile centred on a 5-day window for the base period 1961–1990. The percentage of time for the base period is determined where: $TX_{ij} > TX_{in}90$	%
WSDI	Warm spell duration	Let TX_{ij} be the daily maximum temperature on day *i* in period *j* and let $TX_{in}90$ be the calendar day 90th percentile centred on a 5- day window for the base period 1961–1990. Then the number of days per period is summed where, in intervals of at least 6 consecutive days: $TX_{ij} > TX_{in}90$	days
CSDI	Cold spell duration	Let TN_{ij} be the daily minimum temperature on day *i* in period *j* and let $TN_{in}10$ be the calendar day 10th percentile centred on a 5- day window for the base period 1961–1990. Then the number of days per period is summed where, in intervals of at least 6 consecutive days: $TN_{ij} < TN_{in}10$	days
TXx	Max TX	Let TX_{kj} be the daily maximum temperatures in month *k*, period *j*. The maximum daily maximum temperature each month is then: $TXx_{kj} = \max(TX_{kj})$	°C
TXn	Min TX	Let TX_{kj} be the daily maximum temperatures in month *k*, period *j*. The minimum daily maximum temperature each month is then: $TXn_{kj} = \min(TX_{kj})$	°C
TNx	Max TN	Let TN_{kj} be the daily minimum temperatures in month *k*, period *j*. The maximum daily minimum temperature each month is then: $TNx_{kj} = \max(TN_{kj})$	°C
TNn	Min TN	Let TN_{kj} be the daily minimum temperatures in month *k*, period *j*. The minimum daily minimum temperature each month is then: $TNn_{kj} = \min(TN_{kj})$	°C
FD	Frost days	Let TN_{ij} be the daily minimum temperature on day *i* in period *j*. Count the number of days where $TN_{ij} < 0$ °C	days
ID	Ice days	Let TX_{ij} be the daily maximum temperature on day *i* in period *j*. Count the number of days where $TX_{ij} < 0$ °C	days
SU	Summer days	Let TX_{ij} be the daily maximum temperature on day *i* in period *j*. Count the number of days where $TX_{ij} > 25$ °C	days
TR	Tropical nights	Let TN_{ij} be the daily minimum temperature on day *i* in period *j*. Count the number of days where $TN_{ij} > 20$ °C	days
GSL	Growing season length	Let T_{ij} be the mean temperature ((TN + TX)/2) on day *i* in period *j*. Count the number of days between the first occurrence of at least 6 consecutive days with T > 5 °C and the first occurrence after 1st July (NH) or 1st January (SH) of at least 6 consecutive days with $T_{ij} < 5$ °C	days
DTR	Diurnal temperature range	Let TN_{ij} and TX_{ij} be the daily minimum and maximum temperature respectively on day *i* in period *j*. If *I* represents the number of days in *j*, then: $DTR_j = \sum_{n=1}^{I}(TX_{ij} - TN_{ij})/I$	°C
RX1day	Max 1 day precipitation	Let PR_{ij} be the daily precipitation amount on day *i* in period *j*. The maximum 1 day value for period *j* are: $RX1day_j = \max(PR_{ij})$	mm
RX5day	Max 5 day precipitation	Let PR_{kj} be the precipitation amount for the 5 day interval ending *k*, period *j*. Then maximum 5 day values for period *j* are: $RX5day_j = \max(PR_{kj})$	mm
SDII	Simple daily intensity	Let PR_{wj} be the daily precipitation amount on wet days, PR ≥ 1 mm in period *j*. If *W* represents number of wet days in *j*, then: $SDII_j = (\sum_{w=1}^{W} PR_{wj})/W$	mm
R1mm	Number of wet days	Let PR_{ij} be the daily precipitation amount on day *i* in period *j*. Count the number of days where $PR_{ij} > 1$ mm	days
R10mm	Heavy precipitation days	Let PR_{ij} be the daily precipitation amount on day *i* in period *j*. Count the number of days where $PR_{ij} > 10$ mm	days
R20mm	Very heavy precipitation days	Let PR_{ij} be the daily precipitation amount on day *i* in period *j*. Count the number of days where $PR_{ij} > 20$ mm	days
CDD	Consecutive dry days	Let PR_{ij} be the daily precipitation amount on day *i* in period *j*. Count the largest number of consecutive days where $PR_{ij} < 1$ mm	days
CWD	Consecutive wet days	Let PR_{ij} be the daily precipitation amount on day *i* in period *j*. Count the largest number of consecutive days where $PR_{ij} > 1$ mm	days
R95p	Very wet days	Let PR_{wj} be the daily precipitation amount on a wet day *w* (PR ≥ 1 mm) in period *i* and let $PR_{wn}95$ be the 95th percentile of precipitation on wet days in the 1961–1990 period. If *W* represents the number of wet days in the period, then: $R95p_j = \sum_{w=1}^{W} PR_{wj}$, where $PR_{wj} > PR_{wn}95$	mm
R99p	Extremely wet days	Let PR_{wj} be the daily precipitation amount on a wet day *w* (PR ≥ 1 mm) in period *i* and let $PR_{wn}99$ be the 99th percentile of precipitation on wet days in the 1961–1990 period. If *W* represents the number of wet days in the period, then: $R99p_j = \sum_{w=1}^{W} PR_{wj}$, where $PR_{wj} > PR_{wn}99$	mm
PRCPTOT	Total wet-day precipitation	Let PR_{ij} be the daily precipitation amount on day *i* in period *j*. If *I* represents the number of days in *j*, then: $PRCPTOT_j = \sum_{n=1}^{I} PR_{ij}$	mm

derived as

$$RMSE'_{XY} = \frac{RMSE_{XY} - RMSE_{median}}{RMSE_{median}} \quad (2)$$

where $RMSE_{median}$ represents the median of the RMSE for all models. The median of the RMSE is standardized by the spatial standard deviation (SD) of climatology in the reanalysis for each index as follows:

$$RMSE_{median,std} = RMSE_{median} \Big/ \sqrt{\langle (Y - \langle Y \rangle)^2 \rangle} \quad (3)$$

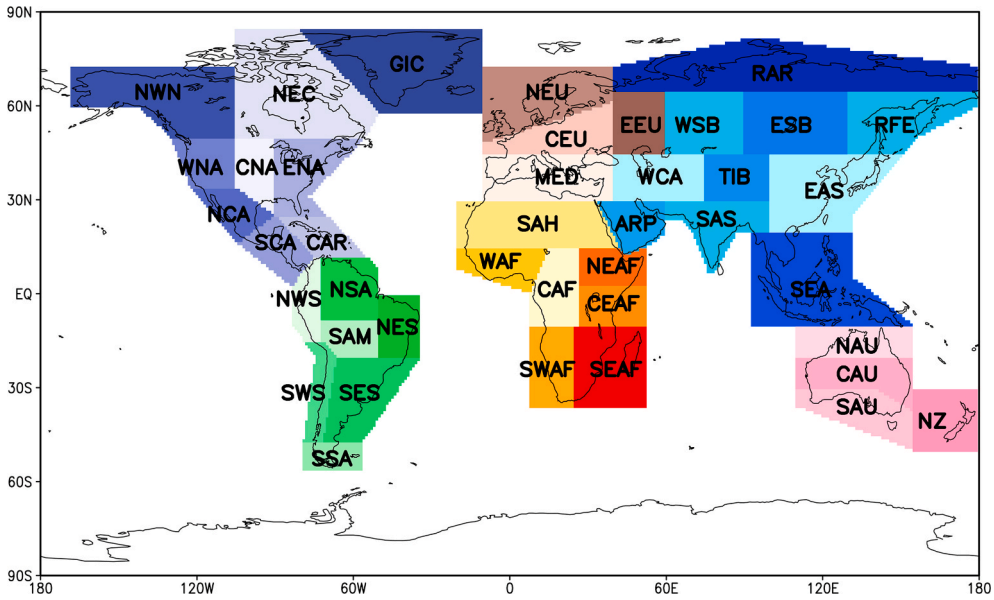

Fig. 1. 41 sub-regional domains over global land adopted from Iturbide et al. (2020). Shaded colors indicate different continents: North America (purple series), South America (green), Europe (brown), Africa (red), Asia (blue), and Oceania (pink series). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

This provides a measure of absolute magnitude of errors for the multimodel ensemble with respect to the given reanalysis. Note that for percentile-based threshold indices for extreme temperatures (TN10p, TX10p, TN90p, and TX90p), the median RMSE is divided by the expected exceedance rate 10% rather than the SD of reanalysis climatology, considering the design of the indices. We repeated RMSE calculations using four reanalyses and compared results by displaying them in the 'portrait' plot. For more details, refer to S13.

2.4. GEV analysis

Following previous studies (Kharin et al., 2007, 2013), we analyzed the return values of annual extremes for temperature (TXx and TNn) and precipitation (RX1day). Return values were calculated as the quantile functions of GEV distribution. To derive the return values, annual maxima of temperature and precipitation at every grid point during 1981–2000 were first fitted to the GEV distribution. The cumulative density function (CDF) of the GEV distribution for variable x is:

$$F(x;\mu,\sigma,\xi) = \begin{cases} \exp\left[-\exp\left\{-\frac{x-\mu}{\sigma}\right\}\right], \ \xi = 0 \\ \exp\left[-\left\{1+\xi\frac{x-\mu}{\sigma}\right\}^{-\xi^{-1}}\right], \ \xi \neq 0, 1+\xi\frac{x-\mu}{\sigma} > 0 \end{cases} \quad (4)$$

where μ, σ, and ξ represent the location, scale, and shape parameters, respectively. GEV parameters are estimated using the L-moments method (Hosking, 1990), which is suggested when the sample size is small (Kharin et al., 2007, 2013).

A return value is calculated as the exceedance of the annual extreme with probability p. For the CMIP5 and CMIP6 multi-model ensembles, the 20-year return values (20RV) of TXx, TNn and RX1day are calculated for the historical time period 1981–2000. The quantile function of GEV is derived by inverting a CDF for a given probability p as:

$$X_p = \begin{cases} \mu - \sigma\ln[-\ln(p)], \ \xi = 0 \\ \mu - \frac{\sigma}{\xi}\ln[1-(-\ln(p))^{-\xi}], \ \xi \neq 0. \end{cases} \quad (5)$$

The corresponding 20RV is obtained when $p = 0.95$ (or an annual extreme exceedance probability = 5%).

The quantile function of GEV (return values) includes two terms. The first term is the GEV location parameter itself and the second term describes the contribution of GEV scale and shape parameters (Min et al., 2009; Park and Min, 2019). By comparing the two terms of Eq. (5), we could evaluate which GEV parameters play a dominant role in the bias of 20RV. This information is useful to understand the origins of 20RV bias, i.e. whether the total biases are related to mean intensity bias, interannual variability bias, or both.

3. Results

3.1. Temperature indices

For absolute and threshold indices, the spatial climatology patterns are compared between HadEX3, ERA-5, the CMIP6 multimodel ensemble median (MEM) and the CMIP5 MEM. Fig. 2 shows the spatial distribution of the 1981–2000 climatology for TXx and TNn. The climatology patterns of the other reanalyses (ERA-40, ERA-Interim, NCEP1, and NCEP2) can be seen from SI Figs. 1 and 2 of S13. Overall, the CMIP6 models can reproduce the HadEX3 and reanalysis climatology pattern for TXx and TNn well (Fig. 2). ERA-5 shows a similar pattern to those of ERA-40 and ERA-Interim. Compared to ERA-5, CMIP6 models on average simulate lower TXx over northern high latitudes and higher TXx over eastern USA, west Asia, and South America. These systematic errors are consistent with those of the CMIP5 MEM, but the warm biases of CMIP6 are generally weaker than those of CMIP5, particularly over South America. TNn is colder in CMIP6 models across the global land except for northeastern Eurasia and southern midlatitudes. Overall bias pattern of TNn is similar to that of the CMIP5 but CMIP6 models tend to simulate lower TNn than CMIP5. These features of spatial patterns for absolute indices are also present in threshold indices such as FD and SU (SI Fig. 1). CMIP6 models simulate more FD and less SU than those from ERA-5 over northern high latitudes and the Tibetan Plateau due to the cold biases in TN and TX. These features were also observed in CMIP3 (Randall et al., 2007) and CMIP5 (SI Fig. 1, S13).

To evaluate regional performance of CMIP6 models, TXx and TNn climatology and biases for 41 sub-regions (cf. Fig. 1) are compared in Fig. 3. Climate extreme indices estimated from ERA-40 and NCEP1 are also shown for comparison with ERA-5, which exhibited the largest differences (S13). TXx and TNn estimated from ERA-40 are similar to those of ERA-5 while NCEP1 shows higher TXx and lower TNn than ERA-

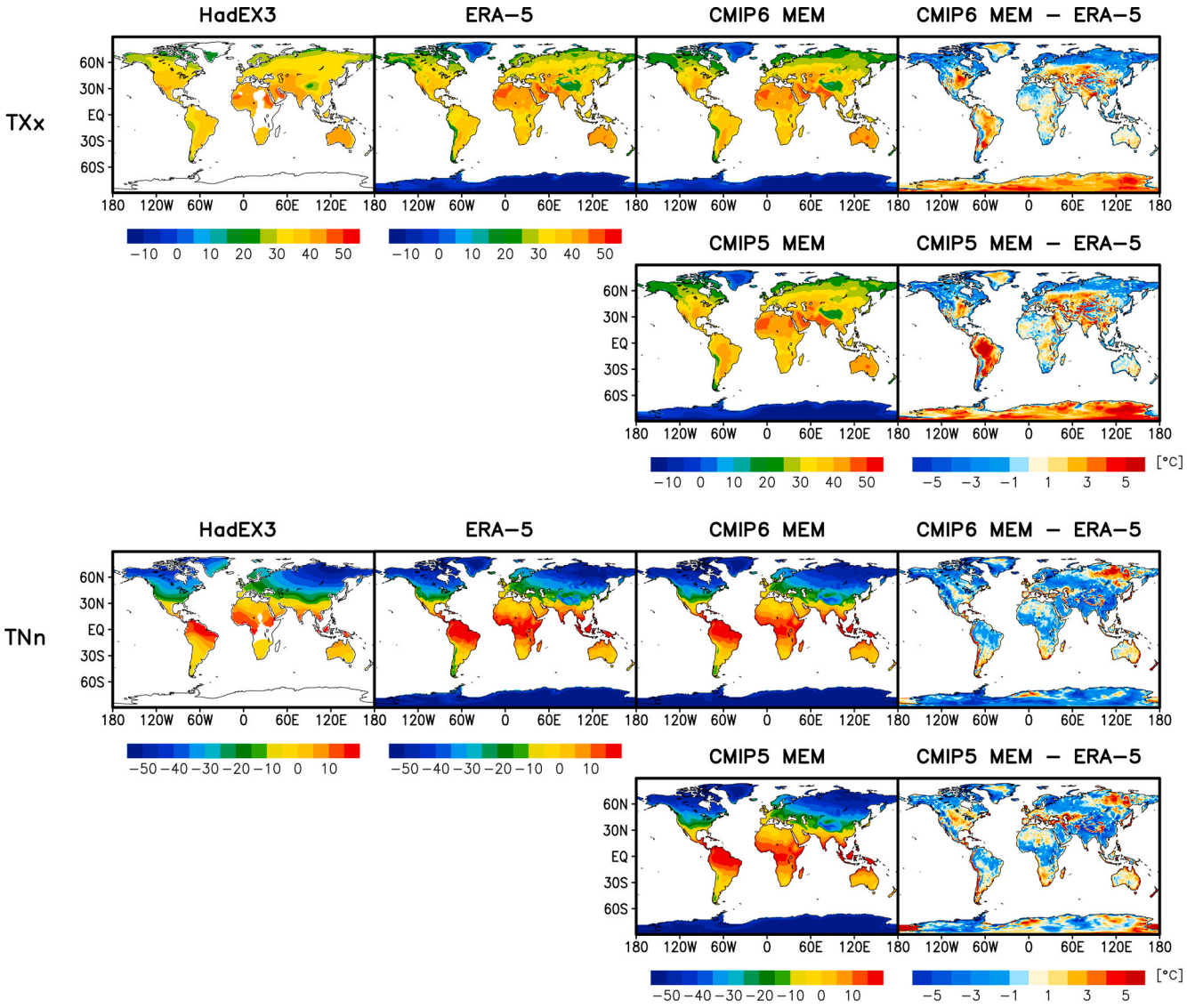

Fig. 2. Spatial distribution of climatology (1981–2000 mean) for TXx (upper panel) and TNn (lower panel) from HadEX3, ERA-5, CMIP6 multimodel ensemble median (MEM), CMIP5 MEM, and difference between CMIP6/CMIP5 MEM and ERA-5.

5. CMIP6 models can generally reproduce the ERA-5 regional mean TXx reasonably well. For some South American and mid-latitude Asian regions, CMIP6 models simulate higher TXx than ERA-5 while they simulate lower TXx over northern high latitude regions such as N.E. Canada (NEC), N.W. North-America (NWN), Russian-Arctic (RAR) and Russian-Far-East (RFE). CMIP6 models underestimate the TNn compared to ERA-5 in northern mid- and high latitude regions such as Greenland/Iceland (GIC), NEC, NWN, W. North-America (WNA), N. Europe (NEU), W.C. Asia (WCA), and Tibetan Plateau (TIB). The MEMs of CMIP6 are comparable with those of CMIP5 for most regions (Fig. 3a and b). Cold biases of TNn are slightly stronger in CMIP6 models than CMIP5 models. Due to the difference in the models between CMIP5 (35 models) and CMIP6 (32 models), it is difficult to directly compare the inter-model spread between them. To address this, we compare the spread of models for temperature extremes using the models from the same institutions participating in both CMIP5 (27 models) and CMIP6 (26 models) (SI Fig. 2). The inter-model spread of TXx (as measured by the interquartile range of the multimodel ensemble) of CMIP6 is found to be narrower than that of CMIP5 over many regions while the CMIP6 spread for TNn remains similar to that of CMIP5.

The regional evaluation results for the threshold indices (FD and SU) are similar to those for the absolute indices (SI Fig. 3). CMIP6 models simulate slightly larger FD than CMIP5 over northern high latitudes, consistent with the colder TNn. The median of SU calculated from CMIP6 simulations is similar to that of CMIP5. The model spread for FD and SU also resembles that of the absolute indices. The CMIP6 spread for SU is narrower than that of CMIP5 while FD spread is similar between CMIP6 and CMIP5 models when using same institution models (not shown).

3.2. *Precipitation indices*

Fig. 4 shows climatology patterns for very wet days (R95p), annual maximum 5 day precipitation (RX5day) and consecutive dry days (CDD) for HadEX3, ERA-5, the CMIP6 MEM, and the CMIP5 MEM. The CMIP6 climatology for RX5day and R95p compares well with HadEX3 and ERA-5 including the east-west pattern over North America and high rainfall over the Asian and South American tropical regions. There are wet biases over southern Africa, central South America, and northern Australia. Compared to extreme precipitation events (R95p and RX5day), CDD in HadEX3 has a broader data coverage because the dry climate has a larger spatial and temporal scale than extreme events (S13). CMIP6 models reproduce the CDD climatology pattern well with a large number of dry days in the Sahara. However, they overestimate

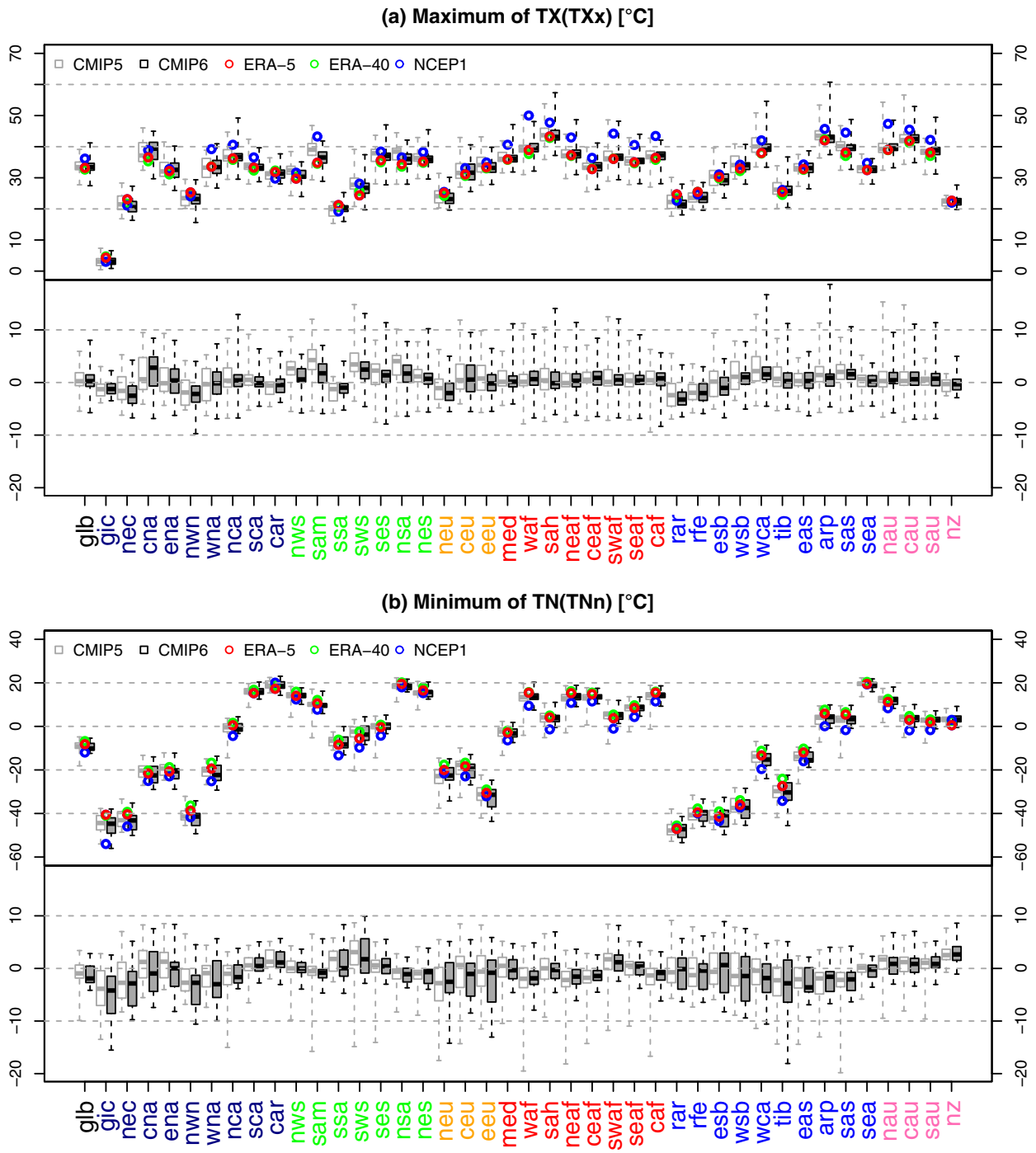

Fig. 3. Box-and-whisker plots for (a) TXx and (b) TNn calculated from 35 CMIP5 (gray) and 32 CMIP6 (black) models with raw indices (upper panel) and its biases (bottom panel). Biases are calculated with respect to ERA-5 values. The boxes indicate the interquartile model spread (range between the 25th and 75th percentiles), the black/gray solid marks within the boxes show the multimodel median and the whiskers indicate the total intermodel range. The reanalyses are indicated in different colors for ERA-5 (read), ERA40 (green) and NCEP1 (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

CDD over South America and underestimate CDD over Sahara region. The CMIP6 climatology patterns of precipitation extreme indices are overall similar to those of CMIP5. A noticeable difference is a reduced dry bias over central South America, South Asia and East Asia in CMIP6.

For regional analysis, box-and-whisker plots are presented in Fig. 5 for regional means and their biases of total wet-day precipitation (PRCPTOT), Simple Daily Intensity (SDII), and heavy precipitation days (R10mm). Three reanlayses show different estimates for precipitation extreme indices, particularly over wet sub-regions such as N.W. South-America (NWS) and S.E. Asia (SEA). The PRCPTOT and R10mm calculated from CMIP6 are similar to the reanalysis results, but there is an underestimation compared to ERA-5 over South America. The CMIP6 MEM of the PRCPTOT, SDII and R10mm are larger than CMIP5 across many regions. The more intense precipitation simulated by CMIP6 models are also seen in annual precipitation extremes, such as very wet days (R95p) and RX5day in most regions (SI Fig. 4), representing that CMIP6 models simulate more strong precipitation than CMIP5 models.

CDD estimated by CMIP6 is larger than that of CMIP5 over some

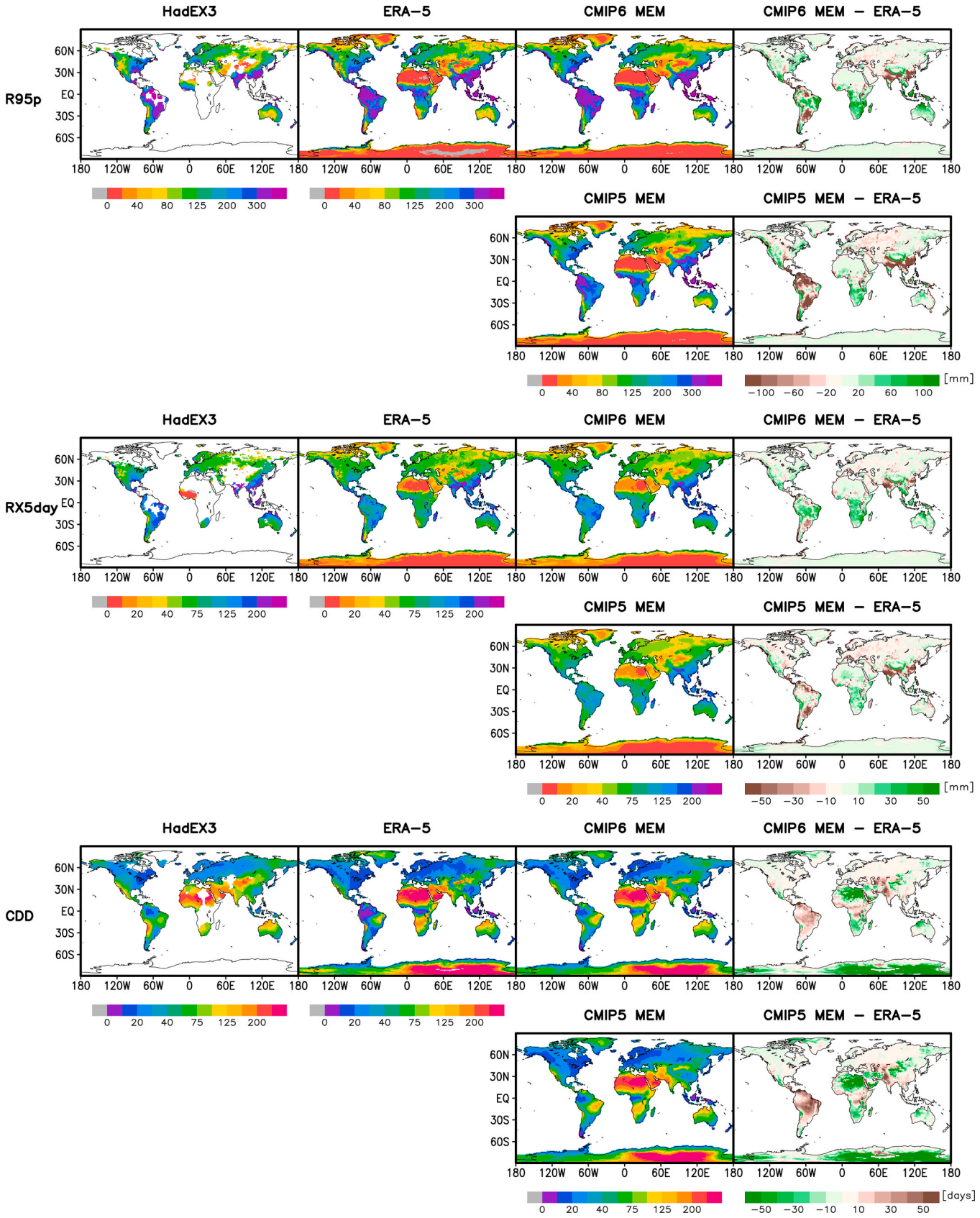

Fig. 4. Same as Fig. 2 but for (a) R95p, (b) RX5day, and (c) CDD.

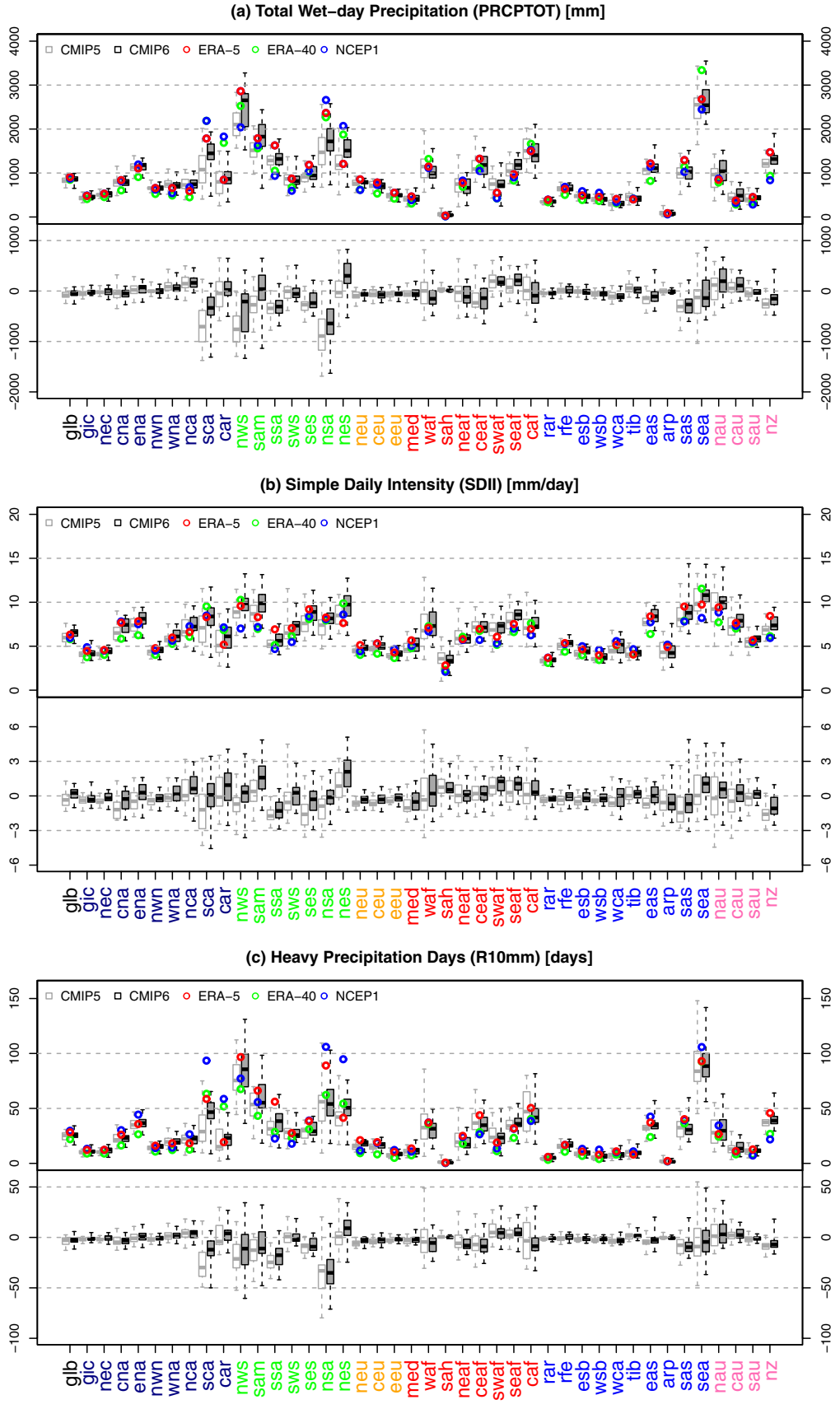

Fig. 5. Same as Fig. 3 but for (a) PRCPTOT, (b) SDII, and (c) R10mm.

African sub-regions (SI Fig. 5). As compared to ERA-5, CMIP6 models tend to overestimate the CDD over South American (SAM, NSA, and NES) and African (NEAF and CEAF) sub-regions but underestimate it over Sahara (SAH), SWAF, and TIB. The consecutive wet days (CWD) is simulated similarly between CMIP6 and CMIP5 models with some underestimation over SCA, NWS, NSA, SAS, and SEA (SI Fig. 5).

3.3. Metric analysis of model performance

Overall performance for individual models for climate extreme indices is summarized using a "portrait" diagram (Fig. 6). The portrait diagram consists of the relative magnitudes of global land mean RMSE for each index by rows and for each model by columns, and the average RMSE for all indices ($RMSE_{all}$) in top raw. In the relative magnitudes of the RMSEs, the blue series colors indicate that model performance is better than others and the red colors indicate models with relatively low skills, on average. Four triangles within each box indicates results from four reanalyses: ERA-5 (upper), NCEP1 (right), NCEP2 (bottom), and ERA40 (left). We also evaluate the performance of the multimodel mean and median of the CMIP6 models, which is displayed in the first two columns. To obtain RMSEs for the multimodel mean and median, we first obtain the multimodel mean and median of each extreme index and then calculate its relative RMSE. The performance of multimodel mean and median is largely better than individual models due to substantial reduction of the systematic errors in individual models (S13). Based on the RMSEall (top raw), NorESM2-MM shows the best performance, with negative relative RMSEs for all four reanalyses, followed by ACCESS-CM2, CNRM–ESM2-1, GFDL-ESM4, HadGEM3-GC31-LL, MPI-ESM1-2-HR, MRI-ESM2-0, and NorESM2-LM which exhibit relatively small errors for three reanalyses.

The right-hand side two columns with gray shading represent the median RMSE for CMIP6 and CMIP5 models, standardized by the spatial standard deviation (SD) in the reanalyses ($RMSE_{median,std}$). Values close to zero (white series colors) mean that absolute errors are lower than spatial variations over the global land for a given reanalysis. The $RMSE_{median,std}$ of precipitation indices is generally larger than temperature indices (except duration indices) in both CMIP6 and CMIP5 models. On the far right-hand, we plot the differences in $RMSE_{median,std}$ between CMIP6 and CMIP5 models. The performance of the CMIP6 models for modeling temperature extremes has overall been improved in

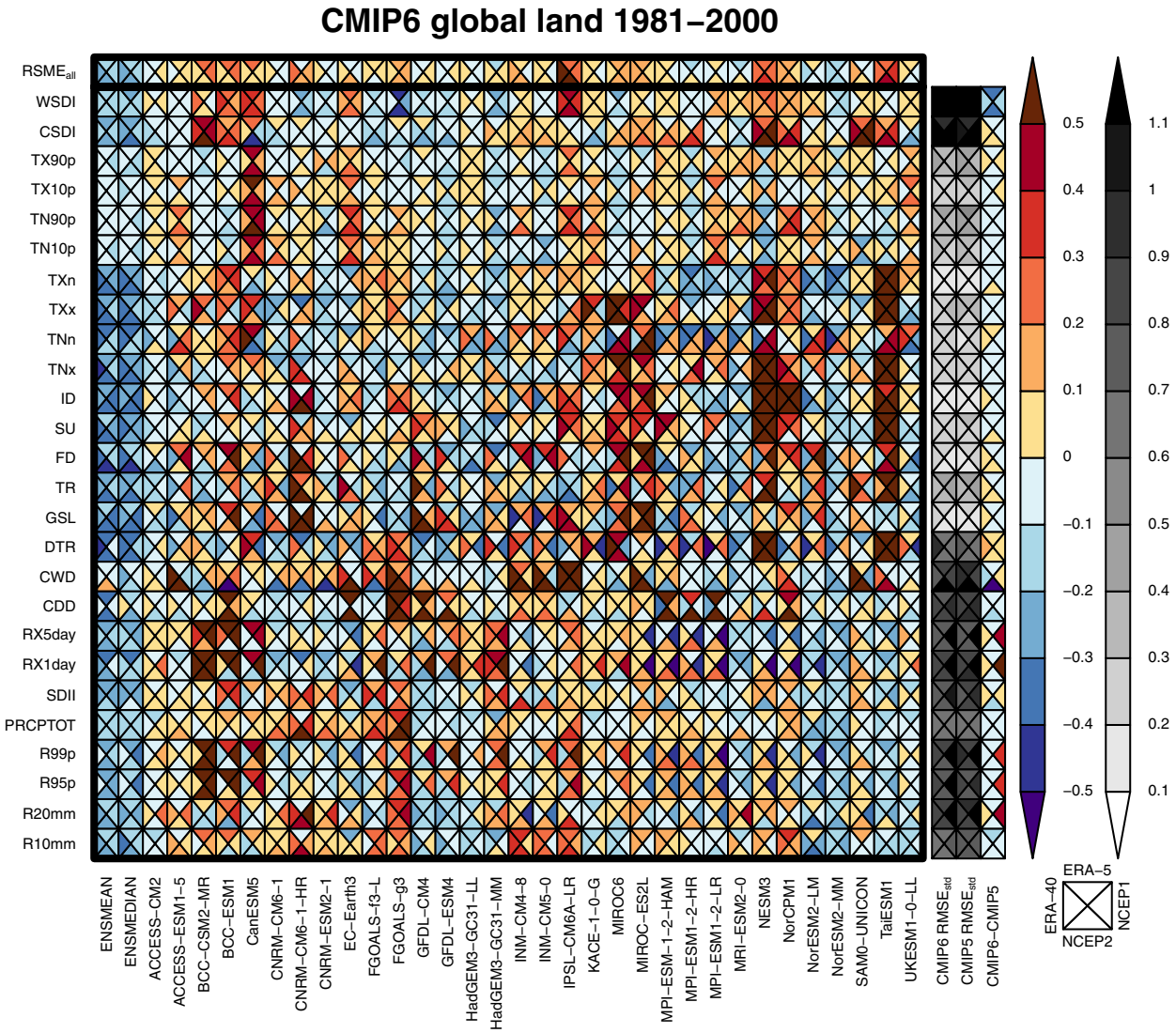

Fig. 6. The "portrait" diagram of relative RMSEs for global land grid points in the 1981–2000 climatologies of temperature and precipitation indices simulated by the CMIP6 models with respect to the four reanalyses: ERA40 (left triangle), ERA-5 (upper triangle), NCEP1 (right triangle), and NCEP2 (bottom triangle). The top row indicates the mean relative RMSE across all indices for a particular model. The gray-shaded columns and blue-red columns on the right side indicate the standardized median $RMSE_{median,std}$ for CMIP6 and CMIP5 and their differences. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

comparison with the CMIP5. Particularly in the WSDI, TN90p, TN10p, TXx, and TNx, CMIP6 models show reduced magnitude of absolute errors. Some improvements are also found in precipitation extremes indices, such as CDD, PRCPTOT, and R10mm.

As in S13, model performance varies substantially depending on reanalysis dataset used as a reference. When ranking the CMIP6 models based on $RMSE_{all}$, the orders of CMIP6 models are different among four reanalysis datasets. For example, NorESM2-MM shows the best performance when compared to NCEP2 and ERA-5 while MPI-ESM1-2-LR and NorESM2-LM have the smallest error against NCEP1 and ERA40, respectively. To quantify the associated uncertainty in model performance due to the different reanalyses, the inter-reanalysis spread (standard deviation) of relative RMSEs is calculated for each box of the portrait plot (SI Fig. 6). For temperature indices, TNn, FD, TR, GSL, and DTR exhibit a relatively large variation of relative RMSEs, indicating greater inter-reanalysis differences in nighttime temperature (TN) than daytime temperature (TX). For precipitation indices, duration indices (CDD and CWD), Rx1day, and percentile-based threshold indices (R99p and R95p) show larger differences in relative RMSEs among reanalyses. Overall, larger spread is observed for precipitation extreme indices than for temperature extreme indices. This seems to be partly due to stronger spatial inhomogeneity of precipitation (Donat et al., 2014) as well as no assimilation of precipitation observations into the reanalyses (Kalnay et al., 1996).

In order to further explore the possible influence of the observational uncertainty, we construct the same portrait plots using HadEX3 observations and compare the results with those based on ERA-5 reanalysis. Here, all extreme indices from ERA-5 and CMIP6 and CMIP5 models are interpolated onto the HadEX3 grid and masked with the HadEX3 data availability prior to analysis. In addition, to examine variable dependency of model performances, temperature and precipitation indices are illustrated separately with models being ranked based on the corresponding $RMSE_{all}$ values (SI Figs. 7 and 8). It can be seen that model ranks vary considerably between the two datasets, particularly for temperature indices. For example, HadGEM3-GC31-MM, MPI-ESM1-2-HR, and MRI-ESM2-0 correspond to a higher skill group when compared to EAR5, but they belong to a lower skill group from an evaluation against HadEX3 (SI Fig. 7). Model ranks for precipitation indices remain similar between the two datasets (SI Fig. 8). Another noticeable difference can be seen for the CMIP5-CMIP6 skill comparison. When using HadEX3 observations, the $RMSE_{std}$ values tend to increase and their CMIP6-CMIP5 difference also becomes weaker, which are found for both temperature (TN10p, TXn, ID, SU, FD, and DTR) and precipitation indices (PRCPTOT, R95p, and R99p). Overall, the high sensitivity of model performance to the use of different reanalyses and observations confirms the importance of reference data selection when evaluating climate models in terms of extreme indices (S13).

3.4. 20-year return level of annual maxima indices

Fig. 7a displays box-and-whisker plots for 20RV for TXx simulated by CMIP5 and CMIP6 models over global land and 41 sub-regions with the corresponding estimates from ERA-5, ERA-40, and NCEP1. The 20RV values of TXx estimated from ERA-5 and ERA-40 are almost same all across regions but those of NCEP1 are located far from the two ECMWF reanlayses, which can also be seen in 20-year mean TXx (Fig. 3a). For more detailed analysis, we plot the bias of 20RV based on ERA-5 (Fig. 7b). There are cold biases over northern high latitude regions (GIC, NEC, NWN, NEU, RAR, and RFE) and warm biases over South America except for SSA and mid-latitude Asia. To identify the relative contribution of GEV location parameter (related to mean intensity) and scale and shape parameters (related to interannual variability) to the biases in 20RV, we plot the bias in the GEV location parameter (first term of Eq. (5)) and the bias of 20RV associated with the GEV scale and shape parameters (second term of Eq. (5)) in Fig. 7c and d, respectively. It can be clearly seen that the biases in GEV location parameter are very similar in amplitude to those of 20RV while the bias related to GEV scale and shape parameters is much smaller. This indicates that mean intensity of TXx (i.e. near center location of GEV distribution) dominates the biases in 20RV of TXx with relatively small contribution from interannual variability of TXx (i.e. spread of GEV distribution). The CMIP6 patterns of 20RV and biases for global land and 41 sub-regions are generally similar to those of CMIP5. The CMIP6 models show reduced warm biases in 20RV of TXx over South American and Asian sub-regions compared to CMIP5 (Fig. 7b), which is well consistent with the results for TXx (Fig. 3a). This improved skill of CMIP6 models in 20RV of TXx is also found to be mainly related to the improved simulations of GEV location parameter (Fig. 7c).

The 20RV of TNn and their biases over global land and 41 sub-regions are displayed in Fig. 8. Compared to ERA-5, ERA-40 overestimates TNn and NCEP1 largely underestimates TNn (Fig. 8a), which is also seen in FD (SI Fig. 3). In CMIP6 models, the biases in 20RV of TNn are generally larger than those of TXx. CMIP6 models underestimate 20RV of TNn, as compared to ERA-5, especially over northern high-latitude areas such as GIC, NEC, and NWN and mid-latitude Asian areas like WCA, TIB, EAS, and SAS, with stronger amplitudes than CMIP5 results (Fig. 8b). The biases in GEV location parameter are very similar to those of 20RV while the biases related to GEV scale and shape parameters are overall very small (Fig. 8c and d). This indicates that for TNn, the mean intensity bias (near center location of GEV distribution) is the main contributor to the bias in 20RV as in the case of TXx.

Fig. 9 shows distributions of 20RV of RX1day and their biases for global land and 41 sub-regions. NCEP1 exhibits much lower values over all regions than ERA-5 and ERA-40, which are also seen in RX5day, R20mm, and R95p (SI Fig. 4), representing systematic dry biases. CMIP6 model skills for the extreme precipitation intensity are largely comparable to those of CMIP5 models over all across regions, reproducing ERA-5 values over many wet regions (Fig. 9a). However, dry biases remain over the tropics and the subtropical American (SCA, CAR, NWS, NSA, and NES) and Asian (EAS, SAS, and SEA) regions (Fig. 9b). These are consistent with CMIP5 results but biases are reduced over many regions. GEV analysis results show that both mean intensity (GEV location parameter) and interannual variability (GEV scale and shape parameters) contribute to the biases of 20RV of RX1day (Fig. 9c and d). In particular, stronger contribution of internanual variability than mean intensity is observed over the tropical and subtropical regions of the America and Asia (Fig. 9d). Improved simulation of extreme precipitation by CMIP6 models are also found to be related more to the improved simulation of the spread of GEV distribution (Fig. 9b and d). This suggests possible improvements of CMIP6 models in terms of simulating the interannual variability of summer monsoon (cf. Xin et al., 2020), details of which warrants further investigations.

4. Conclusions

This paper documents the performances of the GCMs participating in CMIP6 in terms of climate extremes. Historical simulations for 1981–2000 are analyzed and CMIP6 model skills are compared with those of CMIP5 models. The model-simulated values for the climate extreme indices defined by the ETCCDI are compared to HadEX3 observations and four reanalyses for the global and regional patterns of climatology. 20-year return values (20RV) of the warmest day and coldest night temperatures and the annual maximum daily precipitation are also evaluated for the 41 sub-regions using a GEV analysis.

We find that the CMIP6 models capture the observed climatology pattern of extreme temperature indices well, overall similar to CMIP5 models. For warm extreme indices, CMIP6 models tend to simulate reduced warm biases over the South America and mid-latitude Asia. For cold extreme indices, the cold biases over high latitude regions remain in the CMIP6 models with stronger amplitudes than those in the CMIP5 models. For the extreme precipitation indices, the overall performance of the CMIP6 models is comparable to that of the CMIP5 models, but the

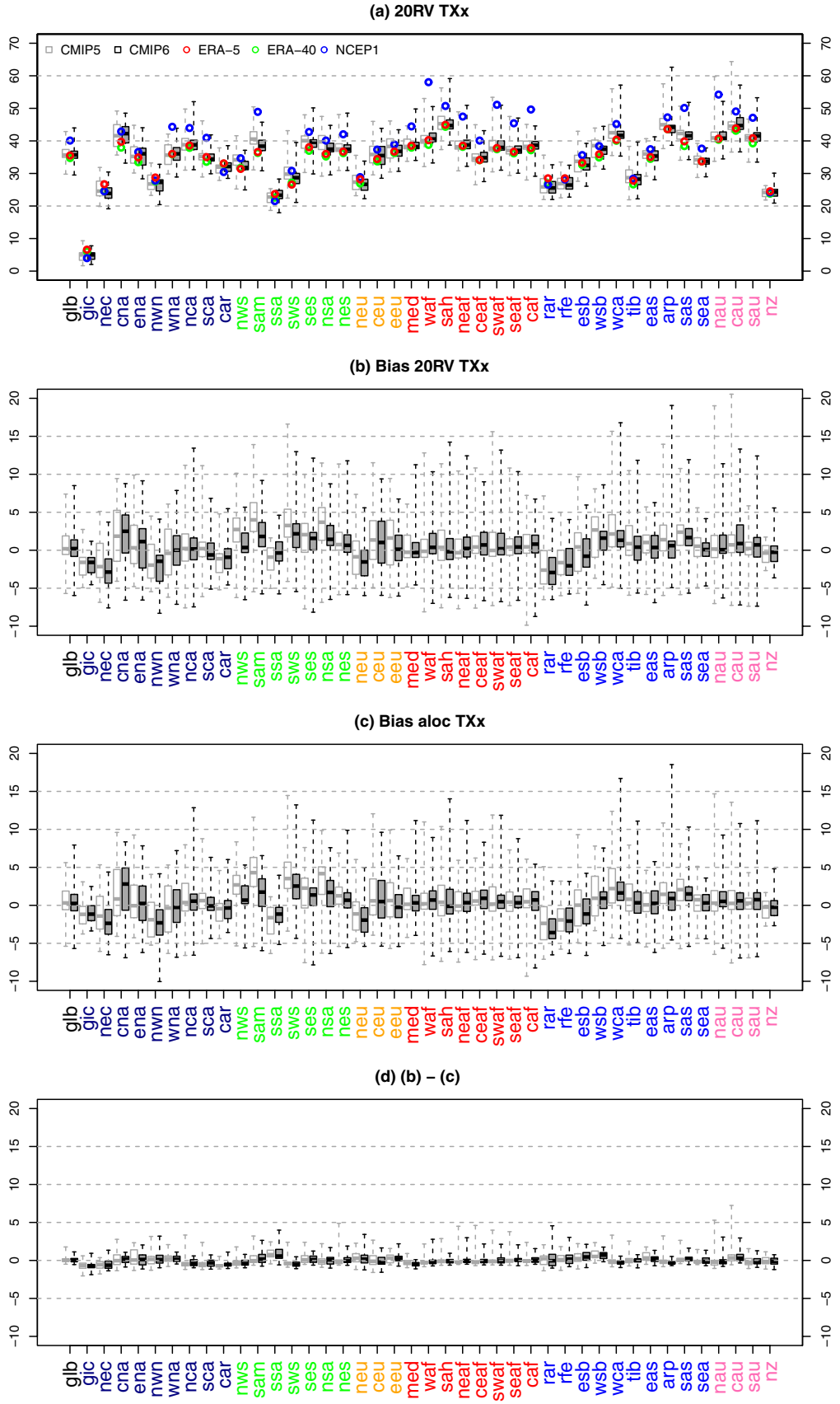

Fig. 7. Box-and-whisker plots for (a) 20-year return values of TXx (°C) from 35 CMIP5 (gray), 32 CMIP6 (black) models, and three reanalyses, (b) biases in 20RV of TXx, (c) biases of GEV location parameters of TXx, and (d) (b)–(c) representing the biases in 20RV associated with GEV scale and shape parameters. Biases are calculated with respect to ERA-5.

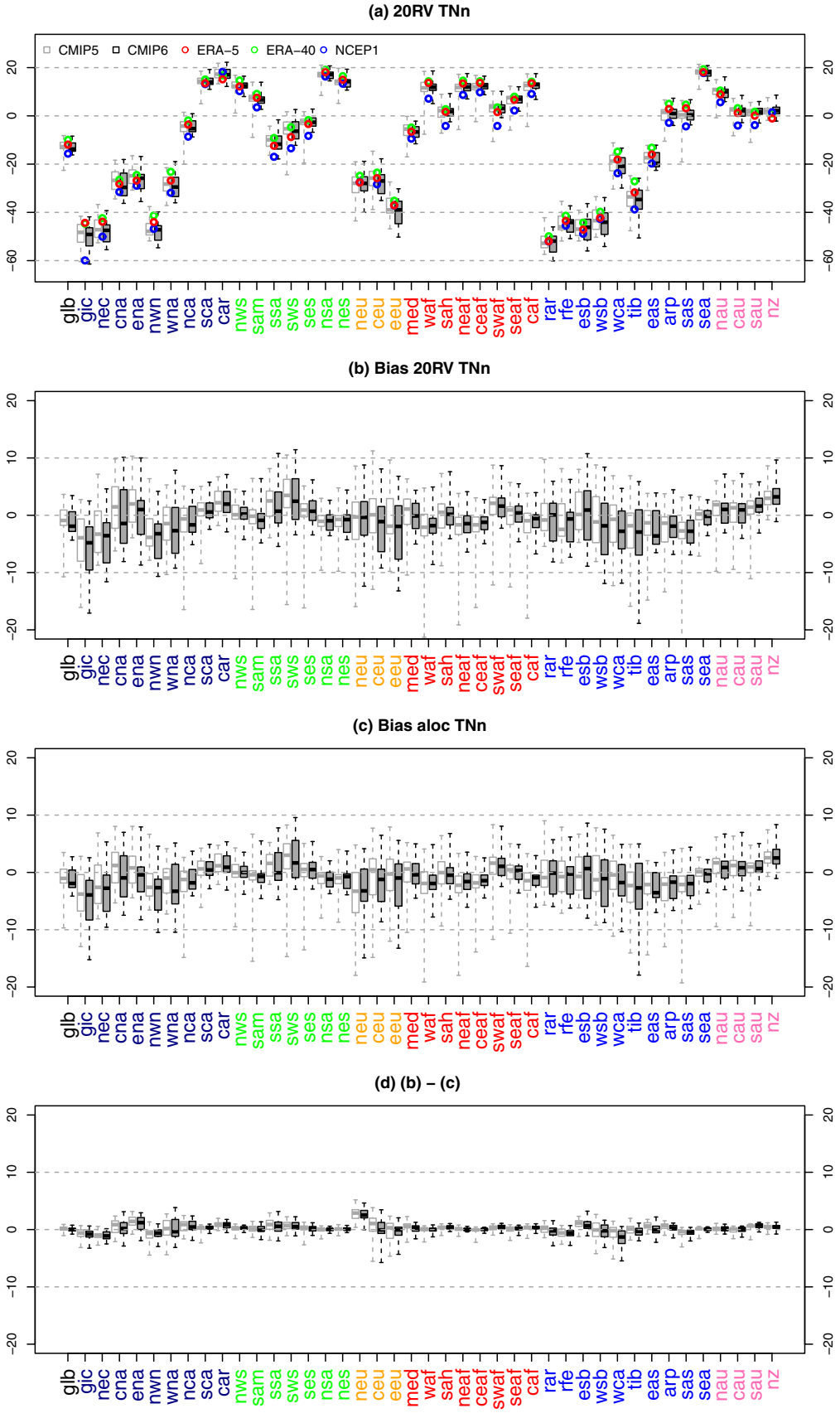

Fig. 8. Same as Fig. 7 but for TNn (°C).

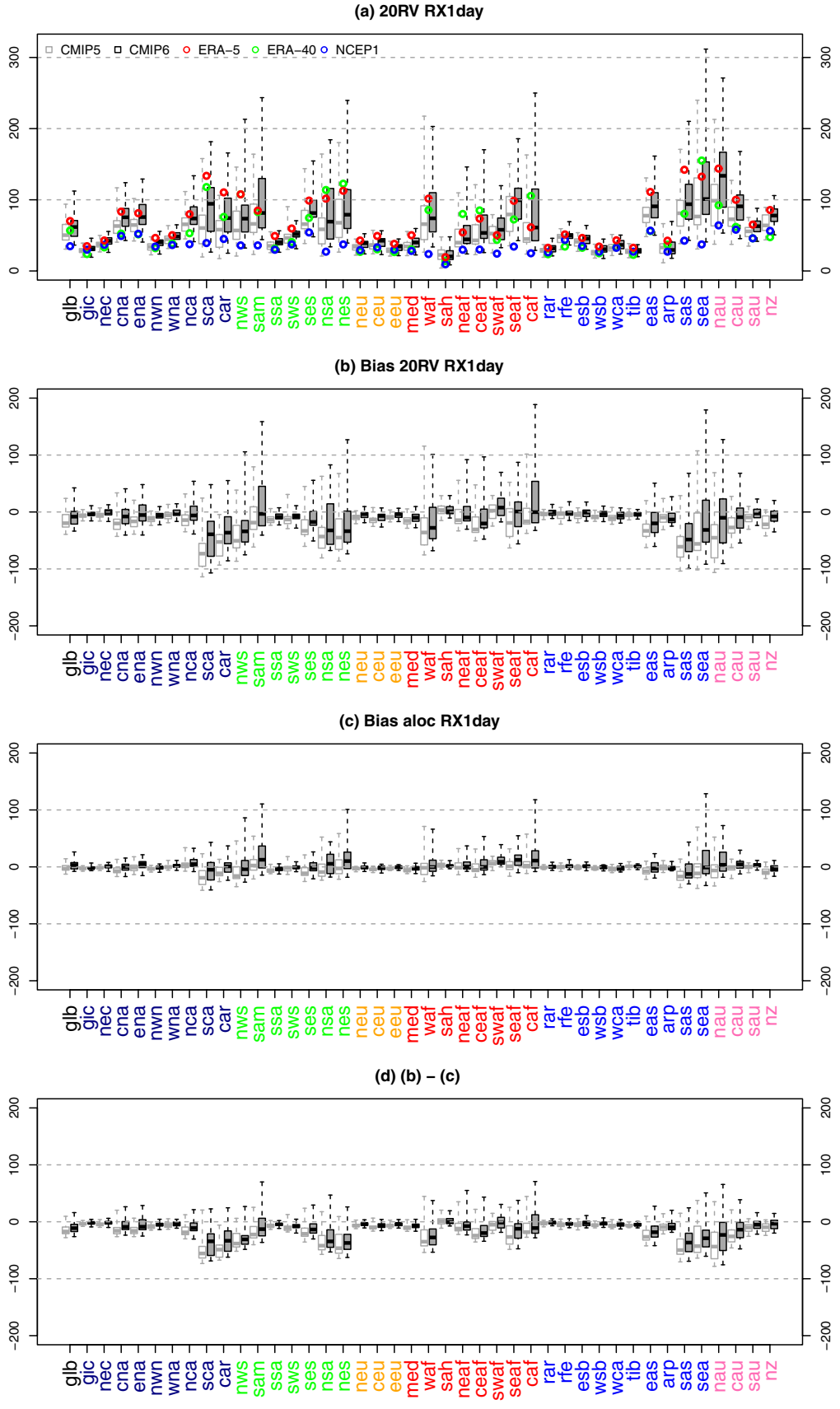

Fig. 9. Same as Fig. 7 but for RX1day (mm/day).

CMIP6 models are found to simulate extreme precipitation more strongly, with reduced dry biases over the tropical and subtropical rain band regions. Overall, there still exist large inter-model spreads.

Biases in 20RV of annual extreme temperature and precipitation are generally consistent with the biases in 20-year means, including the cold biases over high northern latitude and the dry biases in tropical and subtropical regions. GEV analysis results show that the biases in 20RV of temperature extremes are largely associated with biases in GEV location parameter (related to mean intensity). In contrast, the biases in 20RV of precipitation extreme are affected by both GEV location and GEV scale/shape parameters (related to interannual variability).

Overall, the skill of CMIP6 models is similar to those of CMIP5 models, indicating limited improvements in model skills for climate extremes. One notable exception is the improved reproduction of extreme precipitation intensity by CMIP6 models. Further investigation is needed to identify what modeling factors have contributed to the stronger extreme precipitation, including the influences of higher spatial resolution and the improved model physics (cf. Eyring et al., 2019; Paik et al., 2020). In addition, the sources of improved simulations of the interanual variability are needed to be examined, particularly for precipitation extremes over wet regions.

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

Yeon-Hee Kim: Data curation, Formal analysis, Writing - original draft. **Seung-Ki Min:** Conceptualization, Methodology, Writing - original draft. **Xuebin Zhang:** Writing - review & editing, Supervision. **Jana Sillmann:** Methodology, Writing - review & editing. **Marit Sandstad:** Validation, Writing - review & editing.

Acknowledgements

This study was supported by the Korea Meteorological Administration Research and Development Program under Grant KMI2018-01214 and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2018R1A5A1024958). We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled Modelling, coordinated and promoted CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP6 and ESGF.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wace.2020.100269.

References

Bindoff, N.L., Stott, P.A., AchutaRao, K.M., Allen, M.R., Gillett, N., Gutzler, D., Hansingo, K., Hegerl, G., Hu, Y., Jain, S., Mokhov, I.I., Overland, J., Perlwitz, J., Sebbari, R., Zhang, X., 2013. Detection and attribution of climate change: from global to regional. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W.J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A. J., Wehner, M., 2013. Long-term climate change: projections, commitments and irreversibility. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M. (Eds.), Climate Change 2013: the Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

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.M., 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., K€ohler, 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, 553–597. https://doi.org/10.1002/qj.

Donat, M.G., Sillmann, J., Wild, S., Alexander, L.V., Lippmann, T., Zwiers, F.W., 2014. Consistency of temperature and precipitation extremes across various global gridded in situ and reanalysis data sets. J. Clim. 27, 5019–5035.

Donat, M.G., Angélil, O., Ukkola, A.M., 2019. Intensification of precipitation extremes in the world's humid and water-limited regions. Environ. Res. Lett. 14 (6), 065003 https://doi.org/10.1088/1748-9326/ab1c8e.

Dunn, R.J.H., Alexander, L.V., Donat, M.G., Zhang, X., et al., 2020. Development of an updated global land in-situ-based dataset of temperature and precipitation extremes: HadEX3. J. Geophys. Res. Atmos. Accepted.

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

Eyring, V., Cox, P.M., Flato, G.M., Gleckler, P.J., Abramowitz, G., Caldwell, P., Collins, W.D., Gier, B.K., Hall, A.D., Hoffman, F.M., Hurtt, G.C., Jahn, A., Jones, C.D., Klein, S.A., Krasting, J.P., Kwiatkowski, L., Lorenz, R., Maloney, E., Meehl, G.A., Pendergrass, A.G., Pincus, R., Ruane, A.C., Russell, J.L., Sanderson, B.M., Santer, B. D., Sherwood, S.C., Simpson, I.R., Stouffer, R.J., Williamson, M.S., 2019. Taking climate model evaluation to the next level. Nat. Clim. Change 9, 102–110. https://doi.org/10.1038/s41558-018-0355-y.

Hersbach, H., Dee, D., 2016. ERA5 reanalysis is in production, ECMWF Newsletter 147. ECMWF, available at: https://www.ecmwf.int/en/newsletter/147/news/era5-reanalysis-production. last access: 24 March 2017.

Hoegh-Guldberg, O., Jacob, D., Taylor, M., Bindi, M., Brown, S., Camilloni, I., Diedhiou, A., Djalante, R., Ebi, K.L., Engelbrecht, F., Guiot, Hijioka, J., Mehrotra, Y. S., Payne, A., Seneviratne, S.I., Thomas, A., Warren, R., Zhou, G., 2018. Impacts of 1.5°C global warming on natural and human systems. In: Masson-Delmotte, V., Zhai, P., Portner, H.-O., Roberts, D., Skea, J., Shukla, P.R., Pirani, A., Moufouma-Okia, W., Pean, C., Pidcock, R., Connors, S., Matthews, J.B.R., Chen, Y., Zhou, X., Gomis, M.I., Lonnoy, E., Maycock, T., Tignor, M., Waterfield, T. (Eds.), Global Warming of 1.5°C. An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty. Press.

Hosking, J.R.M., 1990. L-moments: analysis and estimation of distributions using linear combinations of order statistics. J. Roy. Stat. Soc. 52, 105–124.

Iturbide, M., Gutiérrez, J.M., Alves, L.M., Bedia, J., Cimadevilla, E., Cofiño, A.S., Cerezo-Mota, R., Di Luca, A., Faria, S.H., Gorodetskaya, I., Hauser, M., Herrera, S., Hewitt, H.T., Hennessy, K.J., Jones, R.G., Krakovska, S., Manzanas, R., Marínez-Castro, D., Narisma, G.T., Nurhati, I.S., Pinto, I., Seneviratne, S.I., van den Hurk, B., Vera, C.S., 2020. An update of IPCC climate reference regions for subcontinental analysis of climate model data: definition and aggregated datasets. Earth Syst. Sci. Data Discuss. https://doi.org/10.5194/essd-2019-258 submitted for publication.

Kalnay, E., Kanamitsu, M., Kistler, R., et al., 1996. The NCEP/NCAR 40-year reanalysis project. Bull. Am. Meteorol. Soc. 77 (3), 437–472. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., Joseph, D., 2002. NCEP-DOE AMIP-II reanalysis (R-2). Bull. Am. Meteorol. Soc. 83, 1631–1643.

Kharin, V.V., Zwiers, F.W., Zhang, X., Hegerl, G.C., 2007. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. J. Clim. 20, 1419–1444.

Kharin, V.V., Zwiers, F.W., Zhang, X., Wehner, M., 2013. Changes in temperature and precipitation extremes in the CMIP5 ensemble. Climatic Change 119, 345–357.

Kim, Y.-H., Min, S.-K., Zhang, X., Zwiers, F.W., Alexander, L.V., Donat, M.G., Tung, Y.-S., 2016. Attribution of extreme temperature changes during 1951–2010. Clim. Dynam. 46, 1769–1782.

Klein Tank, A.M.G., Zwiers, F.W., Zhang, X., 2009. Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation. In: Climate Data and Monitoring WCDMP-No. 72, vol. 1500. WMO-TD No., p. 56

Lu, C., Sun, Y., Zhang, X., 2018. Multimodel detection and attribution of changes in warm and cold spell durations. Environ. Res. Lett. 13, 074013.

Marotzke, J., Jakob, C., Bony, S., Dirmeyer, P.A., O'Gorman, P.A., Hawkins, E., Perkins-Kirkpatrick, S., Le Quéré, C., Nowicki, S., Paulavets, K., Seneviratne, S.I., Stevens, B., Tuma, M., 2017. Climate research must sharpen its view. Nat. Clim. Change 7, 89–91.

Min, S.-K., Zhang, X., Zwiers, F.W., Friederichs, P., Hense, A., 2009. Signal detectability in extreme precipitation changes assessed from twentieth century climate simulations. Clim. Dynam. 32, 95–111.

Min, S.-K., Zhang, X., Zwiers, F.W., Hegerl, G.C., 2011. Human contribution to more-intense precipitation extremes. Nature 470, 378–381. https://doi.org/10.1038/nature09763.

Min, S.-K., Zhang, X., Zwiers, F.W., Shiogama, H., Tung, Y.-S., Wehner, M., 2013. Multimodel detection and attribution of extreme temperature changes. J. Clim. 26, 7430–7451.

Paik, S., Min, S.-K., Zhang, X., Donat, M.G., King, A.D., Sun, Q., 2020. Determining the anthropogenic greenhouse gas contribution to the observed intensification of extreme precipitation. Geophys. Res. Lett. 47 (12) https://doi.org/10.1029/2019GL086875 e2019GL086875.

Park, C., Min, S.-K., 2019. Multi-RCM near-term projections of summer climate extremes over East Asia. Clim. Dynam. 52, 4937–4952.

Randall, D.A., Wood, R.A., Bony, S., Colman, R., Fichefet, T., Fyfe, J., Kattsov, V., Pitman, A., Shukla, J., Srinivasan, J., Stouffer, R.J., Sumi, A., Taylor, K.E., 2007. Cilmate models and their evaluation. In: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M., Miller, H.L. (Eds.), Climate Change 2007: the Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Seneviratne, S.I., Nicholls, N., Easterling, D., Goodess, C.M., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., Zhang, X., 2012. Changes in climate extremes and their impacts on the natural physical environment. In: Field, C.B., Barros, V., Stocker, T.F., Qin, D., Dokken, D.J., Ebi, K.L., Mastrandrea, M.D., Mach, K.J., Plattner, G.-K., Allen, S.K., Tignor, M., Midgley, P.M. (Eds.), Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation, A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC). Cambridge University Press, Cambridge, UK, and New York, NY, USA, pp. 109–230.

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. Atmos. 118, 1716–1733. https://doi.org/10.1002/jgrd.50203.

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. https://doi.org/10.1175/BAMS-D-11-00094.1.

Ting, H., Sun, Y., Zhang, X., Min, S.-K., Kim, Y.-H., 2020. Human influence on frequency of temperature extremes. Environ. Res. Lett. 15, 064014 https://doi.org/10.1088/1748-9326/ab8497.

Uppala, S.M., KÅllberg, P.W., Simmons, A.J., Andrae, U., Bechtold, V.D.C., Fiorino, M., Gibson, J.K., Haseler, J., Hernandez, A., Kelly, G.A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R.P., Andersson, E., Arpe, K., Balmaseda, M.A., Beljaars, A.C.M., Berg, L.V.D., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B.J., Isaksen, L., Janssen, P.A.E.M., Jenne, R., Mcnally, A.P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N.A., Saunders, R.W., Simon, P., Sterl, A., Trenberth, K.E., Untch, A., Vasiljevic, D., Viterbo, P., Woollen, J., 2005. The ERA-40 Re-analysis. Quart. J. Roy. Meteor. Soc. 131, 2961–3012.

Xin, X., Wu, T., Zhang, J., Yao, J., Fang, Y., 2020. Comparison of CMIP6 and CMIP5 simulations of precipitation in China and the East Asian summer monsoon. Int. J. Climatol. https://doi.org/10.1002/joc.6590. In press.

Yin, H., Sun, Y., 2018. Detection of anthropogenic influence on fixed threshold indices of extreme temperature. J. Clim. 31, 6341–6352.

Zhang, X., Alexander, L., Hegerl, G.C., Jones, P., Klein Tank, A., Peterson, T.C., Trewin, B., Zwiers, F.W., 2011. Indices for monitoring changes in extremes based on daily temperature and precipitation data. WIREs Clim. Chang. 2, 851–870.

Zhang, X., Wan, H., Zwiers, F.W., Hegerl, G.C., Min, S.-K., 2013. Attributing intensification of precipitation extremes to human influence. Geophys. Res. Lett. 40 (19), 5252–5257. https://doi.org/10.1002/grl.51010.