Changes in annual extremes of daily temperature and precipitation in CMIP6 models

Chao Li, Francis W. Zwiers, Xuebin Zhang, Guilong Li, Ying Sun, & Michael Wehner

2021

Original citation:

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

Changes in Annual Extremes of Daily Temperature and Precipitation in CMIP6 Models

CHAO LI,[a,b,d] FRANCIS ZWIERS,[c,d] XUEBIN ZHANG,[e] GUILONG LI,[e] YING SUN,[f] AND MICHAEL WEHNER[g]

[a] *Key Laboratory of Geographic Information Science, Ministry of Education, East China Normal University, Shanghai, China*
[b] *School of Geographic Sciences, East China Normal University, Shanghai, China*
[c] *Pacific Climate Impacts Consortium, University of Victoria, Victoria, British Columbia, Canada*
[d] *Nanjing University of Information Science and Technology, Nanjing, China*
[e] *Climate Research Division, Environment and Climate Change Canada, Toronto, Ontario, Canada*
[f] *Laboratory for Climate Studies, National Climate Center, China Meteorological Administration, Beijing, China*
[g] *Computational Research Division, Lawrence Berkley National Laboratory, Berkley, California*

(Manuscript received 31 December 2019, in final form 19 November 2020)

ABSTRACT: This study presents an analysis of daily temperature and precipitation extremes with return periods ranging from 2 to 50 years in phase 6 of the Coupled Model Intercomparison Project (CMIP6) multimodel ensemble of simulations. Judged by similarity with reanalyses, the new-generation models simulate the present-day temperature and precipitation extremes reasonably well. In line with previous CMIP simulations, the new simulations continue to project a large-scale picture of more frequent and more intense hot temperature extremes and precipitation extremes and vanishing cold extremes under continued global warming. Changes in temperature extremes outpace changes in global annual mean surface air temperature (GSAT) over most landmasses, while changes in precipitation extremes follow changes in GSAT globally at roughly the Clausius–Clapeyron rate of ~7% $°C^{-1}$. Changes in temperature and precipitation extremes normalized with respect to GSAT do not depend strongly on the choice of forcing scenario or model climate sensitivity, and do not vary strongly over time, but with notable regional variations. Over the majority of land regions, the projected intensity increases and relative frequency increases tend to be larger for more extreme hot temperature and precipitation events than for weaker events. To obtain robust estimates of these changes at local scales, large initial-condition ensemble simulations are needed. Appropriate spatial pooling of data from neighboring grid cells within individual simulations can, to some extent, reduce the needed ensemble size.

KEYWORDS: Extreme events; Temperature; Precipitation; Climate change

1. Introduction

The Intergovernmental Panel on Climate Change (IPCC) estimated in its Special Report on Global Warming of 1.5°C (SR15; IPCC 2018) that global mean temperature has increased by about 1.0°C since the beginning of the Industrial Revolution. Simultaneously, changes in the frequency and intensity of various weather and climate extremes have been observed in many parts of the world (e.g., Alexander et al. 2006; Zwiers et al. 2011; Min et al. 2011; Zhang et al. 2013; Coumou and Rahmstorf 2012, Donat et al. 2013; Fischer and Knutti 2014; Kim et al. 2016; Li et al. 2017; Kharin et al. 2018; Li et al. 2018; Lorenz et al. 2019, Li et al. 2020), as has also been documented in the IPCC special report on Managing the Risks of Extreme Events to Advance Climate Change Adaptation (SREX; Seneviratne et al. 2012). Global warming is projected to continue into the future, including ongoing changes in weather and climate extremes (e.g., IPCC SR15).

Coordinated climate model intercomparison enabled by the Coupled Model Intercomparison Project (CMIP) has been central to international climate change assessments as well as many assessments conducted nationally or by individual groups. The first phase of CMIP was initiated 20 years ago under the auspices of World Climate Research Programme's (WCRP) Working Group on Coupled Modeling (WGCM). CMIP, which is now in its sixth phase, has helped to enable the development and evaluation of models that provide increasingly comprehensive representations of the climate system. It has also served as the catalyst for the development of an open, distributed, and heavily used archive of climate model simulations that follow prescribed experimental protocols. Phase 6 of CMIP (CMIP6; Eyring et al. 2016) continues this development, incorporating further improved Earth system models, a collection of affiliated projects focused on specific science questions, and a new suite of future forcing scenarios, that is, the shared socioeconomic pathway (SSP) scenarios. CMIP6 will be a central element informing the physical science basis for the upcoming 2021 IPCC Sixth Assessment Report (AR6). The objective of this paper is to evaluate the performance of this new-generation multimodel ensemble in simulating present-day multiyear to multidecade return period extremes of daily temperature and precipitation and to document their future projections.

Denotes content that is immediately available upon publication as open access.

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JCLI-D-19-1013.s1.

Corresponding author: Chao Li, cli@geo.ecnu.edu.cn

DOI: 10.1175/JCLI-D-19-1013.1

Among various indices representing different features of near surface temperature and precipitation extremes (such as those studied in Alexander et al. 2006; Tebaldi et al. 2006; Zhang et al. 2011; Sillmann et al. 2013a,b), the present study in particular focuses on extreme events with return periods of 2–50 years as estimated from annual maximum daily maximum temperature (TXx), annual minimum daily minimum temperature (TNn), annual maximum 1-day precipitation (Rx1day), and annual maximum 5-day precipitation (Rx5day). It follows on from previous studies, including Kharin et al. (2007), which evaluated CMIP3 simulations (Meehl et al. 2007), and Kharin et al. (2013, 2018), which evaluated CMIP5 simulations (Taylor et al. 2012). Such extremes are chosen for a number of reasons. First, these rare events often cause serious impacts on natural and societal systems (Easterling et al. 2000; Zwiers et al. 2011; IPCC SREX; IPCC SR15). Second, they are relevant to many aspects of infrastructure design. One of many examples is the National Building Code of Canada (National Research Council of Canada 2015), which includes specifications of building design criteria that require consideration of the 50-yr 1-day precipitation event in the location where the structure is to be built. Third, anthropogenic warming can cause disproportionally larger changes in the frequencies (e.g., Kharin et al. 2018) and intensities (e.g., Li et al. 2019b) of these types of extremes than for more moderate extremes that are expected to occur multiple times per year.

In the following, we start by introducing the datasets and methods used in this study (section 2). Turning to the results (section 3), we address two major questions: 1) How well does the CMIP6 multimodel ensemble reproduce present-day multiyear to multidecade return period extremes of daily temperature and precipitation? 2) How will the intensity and frequency of these extremes change in future under different emissions scenarios and at different global warming levels? We also discuss whether the rate of change of temperature and precipitation extremes per degree global warming depends on the choice of forcing scenario and model climate sensitivity, and demonstrate the role of large initial-condition ensemble simulations in robustly projecting future changes in such extremes at local scales. We summarize major findings in section 4. We focus on 50-yr events on a global scale in the main text, and provide results for relatively weaker events (e.g., 20-yr events) and for IPCC AR6 reference land regions (Iturbide et al. 2020; see Fig. S1 in the online supplemental material for their geographic boundaries) and the continents as online supporting information.

2. Datasets and methods

a. Datasets

We analyze TXx, TNn, Rx1day, Rx5day, and global annual mean surface air temperature (GSAT) for the period 1851–2100 calculated from daily output of the CMIP6 climate models driven by the updated historical forcings through 2014 (Hoesly et al. 2018; van Marle et al. 2017; Meinshausen et al. 2017) and extended respectively with the four tier-1 SSP future forcing scenarios (i.e., SSP1–2.6, SSP2–4.5, SSP3–7.0, and SSP5–8.5; O'Neill et al. 2016; Eyring et al. 2016; Meinshausen et al. 2020). Twenty models with simulations needed for computing TXx, TNn, Rx1day, and Rx5day for at least one of the five considered simulation experiments were available as of July 2020 (Fig. 1a). As only a few models have submitted multiple members, especially for future projections, we consider the first member for each model, except when illustrating the role of initial-condition ensemble simulations in robustly projecting changes of temperature and precipitation extremes at local scales, for which we use five ensemble members from four models (CanESM5, IPSL-CM6A-LR, MPI-ESM1–2-LR, and UKESM1–0-LL) that provide at least that many members. Table S1 lists the used simulation members from these models.

The new SSP-based scenarios for future forcing provide alternative radiative forcing pathways that can be attained through a collection of socioeconomic development pathways. In particular, the four tier-1 scenarios analyzed in the present study comprise four representative radiative forcing pathways with year 2100 radiative forcing levels of 2.6, 4.5, 7.0, and 8.5 Wm^{-2} respectively that are consistent with the emissions, concentrations, and land use of four socioeconomic development pathways as designated by SSP1, SSP2, SSP3, and SSP5. These four pathways represent respectively a pathway of sustainability-oriented growth and equality (SSP1), a ''middle of the road'' pathway where socioeconomic development trends roughly follow their historical patterns (SSP2), a relatively pessimistic pathway with resurgent nationalism and increasing inequalities (SSP3), and a pathway with unconstrained economic growth and energy use (SSP5). A more thorough introduction to the SSP scenarios can be found in O'Neill et al. (2016) and Riahi et al. (2017).

The new-generation CMIP6 models appear to have notably higher climate sensitivity than CMIP5 models (Fig. 1b), likely due to changes in the representation of cloud and aerosol processes (Wyser et al. 2019; Zelinka et al. 2020). Five of the twenty analyzed models show an equilibrium climate sensitivity above that of the highest model in CMIP5 (4.7°C; e.g., Andrews et al. 2012), producing greater projected warming. The warming of GSAT in 2071–2100 projected by the analyzed models ranges from 3.3° to 6.5°C relative to the 1851–1900 average under SSP5–8.5 and from 1.3° to 2.9°C under SSP1–2.6 (Fig. 1c). The overall higher sensitivity of the CMIP6 models points to the need to assess whether the rate of change of temperature and precipitation extremes per degree global warming depends on the sensitivities of climate models.

To evaluate the simulated present-day temperature and precipitation extremes, we rely on the latest European Centre for Medium-Range Weather Forecasts ERA5 reanalysis (Hersbach et al. 2018) as well as other reference datasets such as the ERA-Interim reanalysis (Dee et al. 2011), NCEP–Department of Energy (DOE) reanalysis 2 (NCEP2; Kanamitsu et al. 2002), and the recently updated Hadley Centre Global Climate Extremes Index 3 (HadEX3) dataset of land-based gridded observations (Dunn et al. 2020). Compared to the prior generation ERA-Interim, ERA5 features enhanced spatial resolution (31 km) and improved model physics and core dynamics and assimilates much more observational data, including precipitation information from ground-based radar observations, although only for 2009 onward. It is therefore expected to be able to produce more

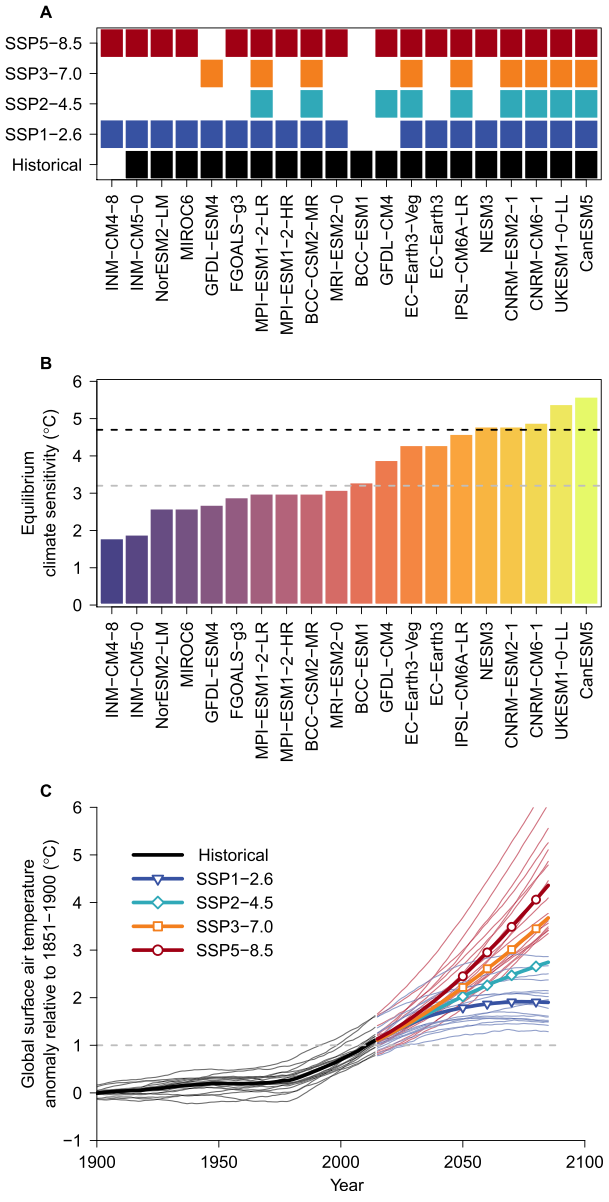

FIG. 1. Information describing the climate model simulations used in this study. (a) Diagram showing the CMIP6 climate model simulations. (b) Equilibrium climate sensitivities of the climate models, with the black and gray horizontal lines marking respectively the highest and mean value in the prior generation CMIP5 models. [These estimates are acquired from a Carbon Brief report "CMIP6: the next generation climate models explained" available from https://www.carbonbrief.org/cmip6-the-next-generation-of-climate-models-explained. Estimation is based on the method used in IPCC AR5 as described in Gregory et al. (2004)]. (c) Global annual mean surface air temperature anomalies relative to the 1851–1900 average simulated by the climate models. The thick lines (with symbols marked for future projections) show the multimodel median anomalies, and the thin lines show anomalies from individual models. Only anomalies under SSP1–2.6 and SSP5–8.5 are shown for individual models for clarity. The anomalies are smoothed by applying a 31-yr moving average. The horizontal line marks the 1.0°C anomaly line.

realistic simulations of temperature and precipitation extremes. We decided not to present evaluations against the HadEX3 observations in the main text as its spatial coverage for precipitation extremes is still limited (Fig. S6).

b. Methods

We define multiyear return period daily temperature and precipitation extremes as the upper or lower tail quantiles of the corresponding annual extremes modeled by a generalized extreme value (GEV) distribution. The use of the GEV distribution is motivated by the Fischer–Tippett theorem (Fisher and Tippett 1928; Leadbetter et al. 1983), which suggests that the distributions of block maxima, such as annual maxima, will converge to the GEV distribution as blocks become large if convergence occurs at all. Many previous studies have used the GEV distribution to describe the distribution of TXx, the negation of TNn, Rx1day, and Rx5day (e.g., Kharin et al. 2007, 2013, 2018; Zwiers et al. 2011; Sillmann et al. 2011; Xu et al. 2018; Li et al. 2019a), finding generally adequate fits for these indices calculated from climate model simulations as we consider here.

The cumulative distribution function for a GEV distributed random variable X is

$$F(x) = \begin{cases} \exp\left[-\left(1+\xi\dfrac{x-\mu}{\sigma}\right)^{-1/\xi}\right], & \xi \neq 0 \\ \exp\left[-\exp\left(-\dfrac{x-\mu}{\sigma}\right)\right], & \xi = 0 \end{cases},$$

where μ, $\sigma > 0$, and ξ are the location, scale, and shape parameters, respectively. These parameters are estimated by the method of probability weighted moments, which is found to be more robust than maximum likelihood estimation for relatively short samples (Hosking 1990). For TXx and TNn, we fit individual GEV distributions at each grid cell of each model's native grid using data for that grid cell only. In contrast, we consider data from a 3 × 3 grid cell region centered on each grid cell for Rx1day or Rx5day unless otherwise clarified. The weak spatial dependence of precipitation extremes means that pooling data from neighboring grid cells can help reduce uncertainty in the local estimates of GEV parameters caused by high-frequency internal variability in Rx1day and Rx5day (Li et al. 2019a,b). Spatial pooling is not used for TXx and TNn because the stronger spatial dependence in temperature extremes limit the uncertainty reductions obtainable from spatial pooling.

Unless otherwise noted, GEV estimation is implemented in different 30-yr time windows such as 1985–2014 for a recent past period and 2071–2100 for a long-term future period, thus assuming that temperature or precipitation extremes remain approximately stationary within 30-yr periods. This is a pragmatic choice that is made necessary by the fact that nonstationary models cannot be fitted reliably to short records (Li et al. 2019a). The resulting extreme quantile estimates can be somewhat biased in opposite directions at the ends of time windows when a secular trend is present (Kharin et al. 2013). These biases could be reduced by using a shorter, 20-yr time window, as used in the IPCC AR5 for assessing future changes

in climate variables, including temperature and precipitation extremes with return periods up to 20 years (e.g., Kharin et al. 2007). Estimating rarer extremes (e.g., 50-yr event) based on a short sample (e.g., 20 years) needs more aggressive extrapolation into the upper tail of the fitted GEV distribution, however, leading to more uncertain estimates for extreme quantiles and potential biases (Ben Alaya et al. 2020). We chose 30-yr time windows to balance estimation uncertainty from GEV tail extrapolation with that induced by secular trends.

Having fitted the GEV distribution, the temperature or precipitation extremes of interest can be defined, and their changes can be evaluated. The intensity (or return value) of an event with occurrence frequency p (or return period $T = 1/p$) can be estimated by the $1 - p$ quantile of the fitted GEV distribution, while the occurrence frequency of an event with intensity x can be estimated by the probability of exceeding x in the fitted GEV distribution. Changes in the intensities of extreme events with a given frequency or changes in the frequencies of extreme events with a given intensity can accordingly be estimated for different future periods or warming levels. Changes in future periods are expressed with respect to the recent past period 1985–2014, while changes at different warming levels are expressed relative to their intensity or frequency in a climate that is 1°C warmer above preindustrial. It is noted that most of the climate models considered in this paper simulate the present-day climate as being roughly 1°C warmer than their preindustrial climates (gray line in Fig. 1).

Although multiple methods have been used to define global warming levels (e.g., James et al. 2017 and references therein), we follow the definition used in the IPCC SR15 (IPCC 2018), by which a 1°C warmer world, for example, is taken as the first 31-yr period with average GSAT exceeding 1°C relative to the 1851–1900 average. To have as many lines of evidence as possible, we use projections under all of the SSP scenarios to obtain multiple (≤4) plausible climates for a given warming level from each climate model. Figure S2 presents the identified time windows for different global warming levels. The GEV fitting is conducted individually for each plausible time window for a given warming level.

To estimate the rate of change in the intensity of temperature and precipitation extremes with GSAT, we rely on future projections only, although including historical simulations does not qualitatively affect the estimates (not shown). For a given occurrence frequency, or equivalently, for a given return period, we estimate at each grid cell the corresponding intensity of the event of interest and the average GSAT in each of the six overlapping 30-yr periods (i.e., 2021–50, 2031–60, . . . , 2071–2100). We then compute changes in the estimated intensities from one period to a later period and the corresponding GSAT changes, which are then used to estimate the scaling rate for the intensity of extreme temperature or precipitation events that have the given return period by quantile regression for the median (Koenker 2005). We perform the estimation separately for each model and each SSP scenario so as to evaluate whether the estimated scaling rate depends on forcing scenarios and/or sensitivities of climate models.

To summarize multimodel ensemble statistics, we first regrid the estimated extremes and their changes to a common 2.5° × 2.5° grid using bilinear interpolation, and then take the multimodel median values for each 2.5° × 2.5° grid cell. Medians are used because they are more robust to unusually large or small values from individual ensemble members than the multimodel mean values. Grid cells are stippled if at least 80% of the available models or warming level windows agree on the sign of the corresponding ensemble median changes. Global, continental, and regional statistics are likewise obtained by computing the corresponding spatial medians.

3. Results

a. *The ability of CMIP6 models to simulate temperature and precipitation extremes*

Figure 2 compares the CMIP6 multimodel median estimates of 50-yr return values of TXx and TNn in 1985–2014 and the corresponding ERA5 estimates. Models reproduce quite well the spatial patterns of both hot and cold temperature extremes including the large-scale latitudinal gradients, land–sea contrast, as well as topography-induced regional patterns over the Tibetan Plateau, the Andes mountain range, and the Rockies (Figs. 2a–d), with an overall spatial pattern correlation being as high as 0.99 for both hot and cold extremes. For hot extremes over Eurasia, the 35°–45°C contours appear to be displaced a bit farther north in models, perhaps due to different land surface representations in ERA5 and the models or a manifestation of unforced multidecadal (>30 years) internal climate variability.

Models also reproduce well the magnitude of temperature extremes. For hot extremes, models tend to have a slight warm bias over most land areas (Fig. 2e), with a warm bias of 0.4°C on average over the globe. The largest biases of 5°–10°C are found mainly in central Asia, parts of South America, and inland Antarctica. In contrast, relative to ERA5, models tend to produce cold extremes that are too cold almost everywhere over land except north Asia, and over the ice-covered Arctic and Southern Oceans (Fig. 2f). The global average bias in cold extremes is −1.4°C, but cold biases below −10°C can be seen in the Arctic Ocean and at high elevations over the Tibetan Plateau and Rocky Mountains. Overall, models perform better for hot extremes than for cold extremes over land areas.

Figure 3 shows the corresponding comparison for the 50-yr return values of Rx1day and Rx5day. Models also do a reasonable job in capturing typical large-scale features of these precipitation extremes, such as intense precipitation extremes in the intertropical convergence zone (ITCZ) and weak precipitation extremes in dry areas over the eastern subtropical basins of the South Pacific and South Atlantic and in northern Africa (Figs. 3a–d). Nevertheless, a double-ITCZ bias over the equatorial central and eastern Pacific is evident in estimates of 50-yr Rx1day and Rx5day return values. A spurious double-ITCZ can appear as a result of different interacting factors such as biased sea surface temperatures, surface winds, and erroneous cloud microphysics over the equatorial Pacific (e.g., Dai 2006; Zhang et al. 2015; Woelfle et al. 2019). Overall, the spatial pattern agreement between models and ERA5 looks slightly better for Rx5day than for Rx1day, with pattern correlation being 0.91 for the latter and 0.88 for the former.

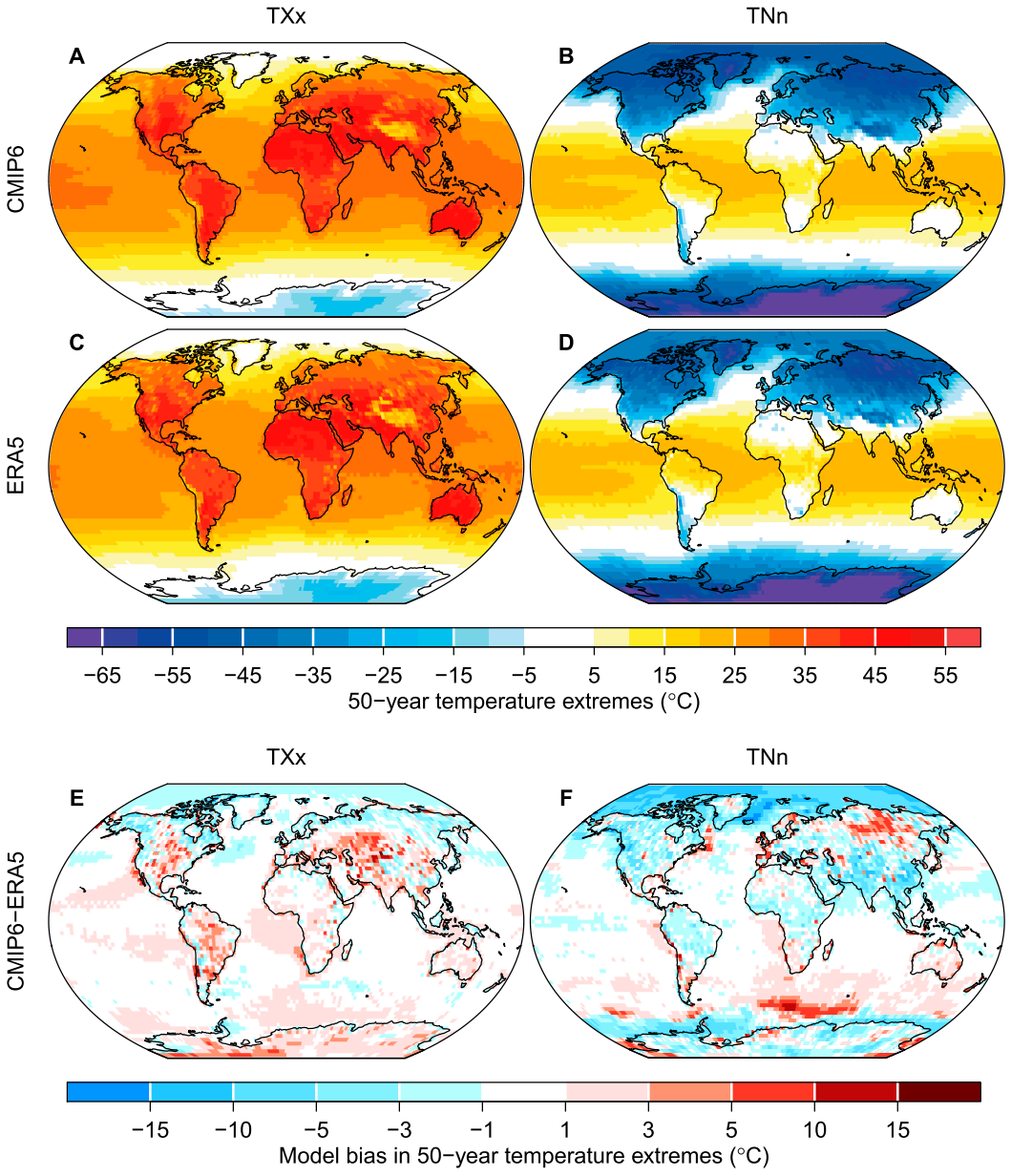

FIG. 2. Estimates of 50-yr return values of TXx and TNn (°C) in 1985–2014. (a)–(d) The CMIP6 multimodel median estimates of 50-yr return values of TXx and TNn in 1985–2014 and the corresponding estimates using ERA5 reanalysis. (e),(f) Differences between the CMIP6 multimodel median estimates of 50-yr return values of TXx and TNn in 1985–2014 and the corresponding ERA5 estimates.

Despite the reasonable reproduction of large-scale spatial features, there are biases in the magnitude of precipitation extremes relative to ERA5, particularly in the tropics and subtropics (Figs. 3e–3f) that were also seen in previous versions of CMIP, although evaluated with respect to different reanalyses (Kharin et al. 2007, 2013). These include precipitation extremes that are too weak in the intertropical convergence zone and too intense in the subtropical dry areas of northern Africa and the eastern subtropical basins of the South Pacific and South Atlantic. In the extratropics, models show better agreement with ERA5, consistent with previous phases of CMIP. The multimodel median estimates of the 50-yr Rx1day and Rx5day return values range from 80% to 170% of ERA5 values over the extratropics (north of 30°N and south of 30°S), with a regional mean value of about 115% for both event durations.

For both temperature and precipitation extremes, models show similar skill in simulating extreme events of different rarity levels on a global scale (bottom panels in Figs. S3 and S4). Nevertheless, notably better performance can be seen for

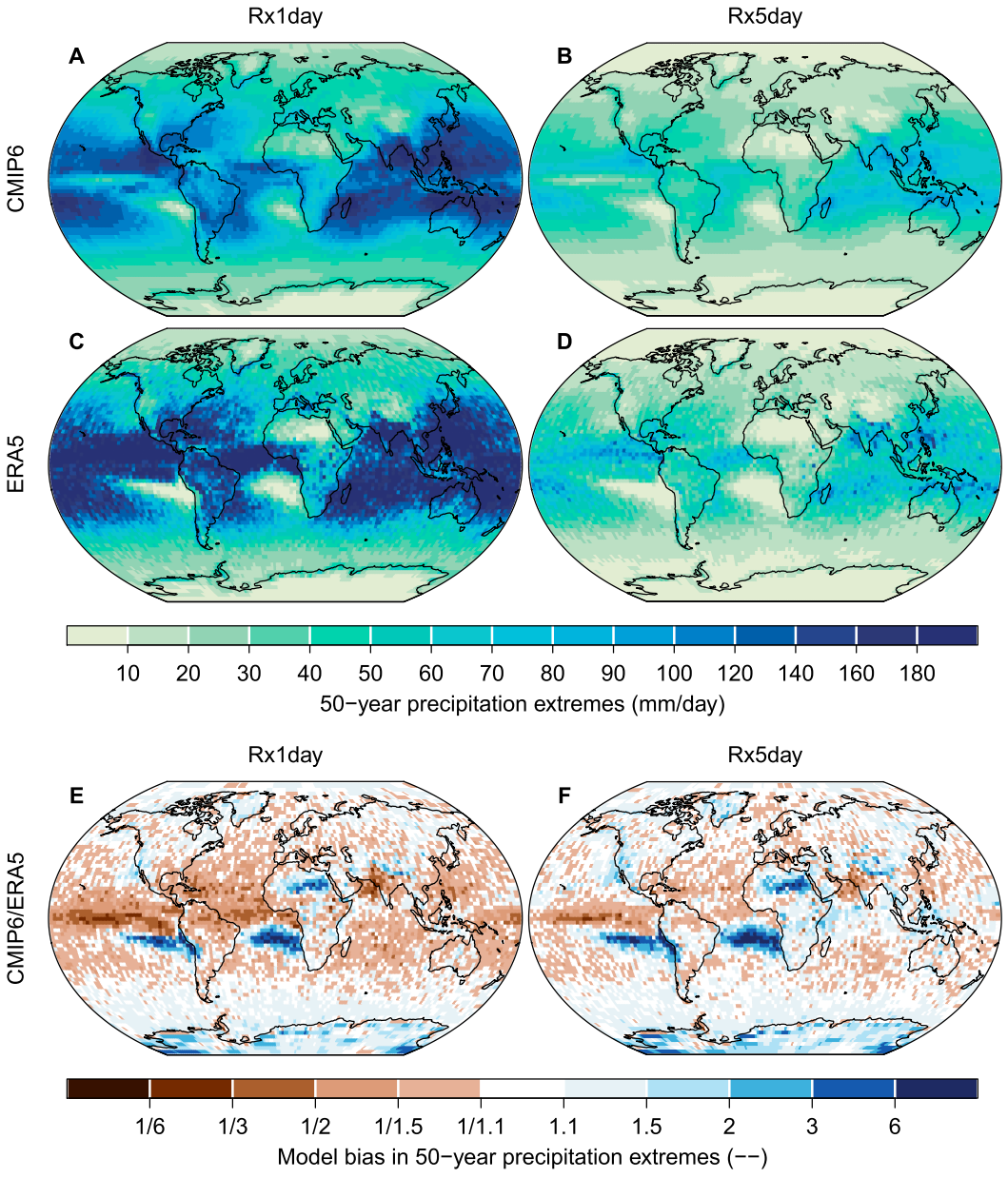

FIG. 3. As in Fig. 2, but for Rx1day and Rx5day [units in (a)–(d) are mm day^{-1} and in (e) and (f) are dimensionless]. Note that the differences between models and ERA5 are expressed as the ratios of the model estimates over the ERA5 estimates.

less extreme events in some regions (top two panels in Figs. S3 and S4). Comparing spatial patterns of the multimodel median biases (bottom panels in Figs. 2 and 3) and the standard deviations of return values estimated from individual models (which measure model agreement; Fig. S5), it is evident that regions with relatively large model biases generally coincide with regions with relatively low model agreement for all considered temperature and precipitation extremes. In line with previous studies, these findings are sensitive to the choice of reference dataset (Table 1), particularly over land and ice-covered regions for temperature extremes and in tropical and subtropical regions for precipitation extremes, as indicated by the different magnitudes and even signs of model biases (Fig. S6).

b. The projected changes in the intensity of temperature and precipitation extremes

Figures 4a–d present the CMIP6 multimodel median changes in 50-yr return values of TXx and TNn in 2071–2100 relative to 1985–2014 for the SSP1–2.6 and SSP5–8.5 scenarios (see Fig. S7 for changes in 20-yr return values). In line with evidence from CMIP3/5 models (Kharin et al. 2007, 2013), temperature extremes exhibit widespread warming across the entire globe

TABLE 1. Average differences between the CMIP6 multimodel median estimates of 50-yr return values of TXx, TNn, Rx1day, and Rx5day in 1985–2014 and the corresponding estimates based on different reference datasets over the globe and the global land area. For Rx1day and Rx5day, the differences are expressed as ratios of the model estimates over the reference estimates. The average absolute differences between models and observations for temperature extremes are also shown by numbers in parentheses. Rx1day and Rx5day units are dimensionless.

Reference dataset	Region	TXx (°C)	TNn (°C)	Rx1day	Rx5day
ERA5	Globe	0.4 (1.4)	−1.4 (2.7)	1.1	1.2
	Land	0.5 (2.2)	−1.0 (2.8)	1.2	1.3
ERA-Interim	Globe	0.3 (1.1)	−1.2 (2.5)	1.2	1.3
	Land	1.0 (1.6)	−1.8 (2.7)	1.2	1.4
NCEP2	Globe	−0.5 (1.4)	4.4 (5.4)	1.1	1.1
	Land	−1.1 (2.4)	7.8 (8.0)	1.1	1.1
HadEX3	Land	−0.3 (3.1)	−3.8 (5.8)	1.0	1.3

during the century regardless of the choice of forcing scenario, with several known large-scale features, such as the larger warming over land than over ice-free ocean areas, pronounced polar amplification, and more rapid warming in cold extremes than in hot extremes. Over land, the 50-yr return value of TXx is projected to increase 5.3°C on average (3.7°–7.3°C for the central 90% range of individual model estimates) by the end of this century under SSP5–8.5, while the corresponding warming for TNn is 5.0°C (4.0–8.2°C). Choosing a socioeconomic pathway with lower emissions such as SSP1–2.6 would reduce the warming in 50-yr return values of land TXx and TNn over the century to 1.3°C (0.8°–2.1°C) and 1.7°C (1.2°–3.2°C), respectively, indicating the dramatic impacts of societal decisions on future increases in temperature extremes. We note that over 80% of the models agree on the sign of the presented multimodel median warming almost everywhere across the globe for both hot and cold extremes under all considered emissions scenarios (stippling in Figs. 4a–d).

Although globally cold extremes warm more than hot extremes, there are regional exceptions. For example, models tend to project somewhat larger warming in hot extremes over the tropical belt land and sea area, the northern oceans except the Arctic, and the central Asia, and they tend to project comparable warming in hot and cold extremes over the Southern Oceans except along the edges of Antarctic ice shelves and sea ice, as also documented in previous studies (Kharin et al. 2007, 2013).

Precipitation extremes are projected to intensify over most of Earth's surface, with the largest percentage increases occurring in the tropics followed by the high latitudes (see Figs. 5a–d for the multimodel median relative changes in the 50-yr return values of Rx1day and Rx5day from the period of 1985–2014 to 2071–2100; also see Fig. S8 for relative changes in 20-yr return values). Decreases in extreme precipitation return values are confined to regions close to the descending branches of the Hadley circulation. It should be noted that these regions are also where climate models show the largest biases relative to ERA5 (Figs. 3e,f), indicating that the projected changes in these regions might be subject to high model uncertainty. On average over land, the projected intensification in the 50-yr Rx1day and Rx5day events by the end of this century relative to 1984–2015 are, respectively, 27.9% (17.6%–44.9% for the central 90% range of individual model estimates) and 23.1% (15.1%–39.6%) under SSP5–8.5, and 8.4% (4.7%–13.6%) and 7.7% (4.6%–12.6%) under SSP1–2.6, suggesting that societal decisions also play an important role in future intensification of precipitation extremes.

There is high model agreement under SSP5–8.5 on the direction of multimode median change in areas with increasing precipitation extremes (stippling in Figs. 5c,d). Areas where at least 80% of the climate models agree on the direction of multimodel median change in 2071–2100 cover 90% and 80% of Earth's surface for Rx1day and Rx5day, respectively. Under SSP1–2.6, however, the same level of model agreement occurs over less than half Earth's surface because the response to external forcing is smaller relative to unforced internal variability than in SSP5–8.5 (stippling in Figs. 5a,b). Projected near-term 2021–50 changes show similarly low model agreement even under the SSP5–8.5 scenario (not shown) as the signal of change is weak relative to unforced internal variability in near-term future. These results are consistent with the previous findings (e.g., Fischer et al. 2014; Zhang et al. 2017; Li et al. 2019a), indicating that it is hard to robustly estimate long-term changes of rare precipitation extremes at small spatial scales with short data records, especially when the external forcing is not strong.

The projected changes in temperature and precipitation extremes at different warming levels bear spatial patterns similar to those seen for changes in the 2071–2100 period under different forcing scenarios, as can be seen from the multimodel median changes in 50-yr return values of these extremes for global warming of 2° and 4°C above preindustrial levels (Figs. 4e–h and 5e–h). The available collection of CMIP6 models project that warming of 4°C above preindustrial levels would on average produce 50-yr TXx and TNn events over land that are, respectively, 4.1°C (3.5°–4.5°C for the central 90% range of estimates from the corresponding warming level windows) and 4.2°C (3.6°–4.9°C) hotter relative to the 1°C warming world, and 50-yr Rx1day and Rx5day events over land that are 22.8% (18.2%–30.3%) and 20.0% (16.3%–25.9%) more intense compared to the 1°C warming world. The projected corresponding changes would be reduced to 1.3°C (1.1°–1.6°C), 1.5°C (1.0°–1.7°C), 7.2% (5.7%–9.1%), and 6.3% (4.3%–8.4%) for 2°C global warming, and

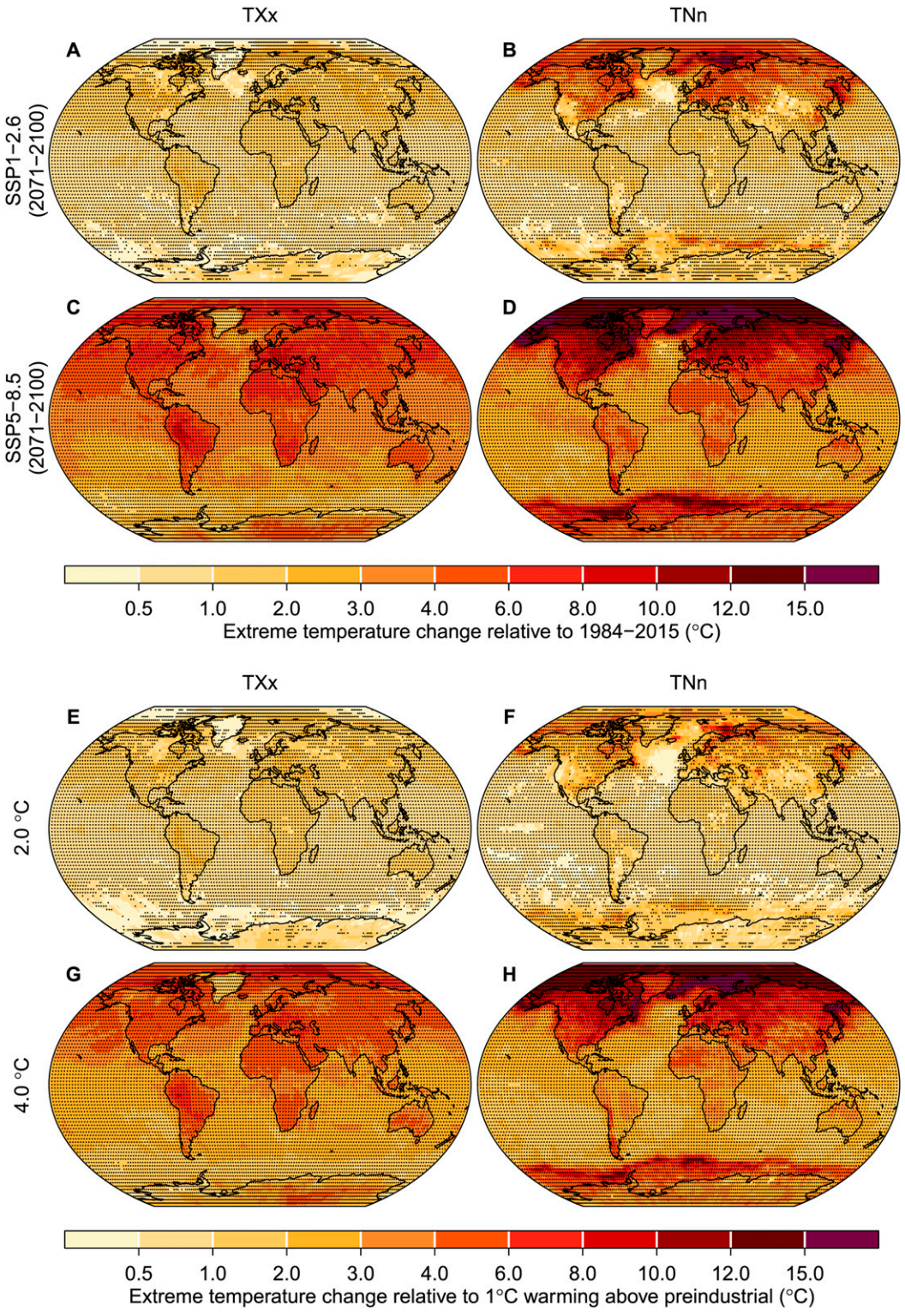

FIG. 4. Projected changes in 50-yr return values of TXx and TNn (°C). (a)–(d) The CMIP6 multimodel median changes in the 50-yr return values of TXx and TNn in 2071–2100 under the lower SSP1–2.6 and higher SSP5–8.5 scenarios relative to 1985–2014. (e),(f) The corresponding changes at 2.0° and 4.0°C global warming above pre-industrial relative to the 1.0°C global warming. Stippling marks grid cells where at least 80% of the available models or warming level windows agree on the sign of the corresponding ensemble median changes.

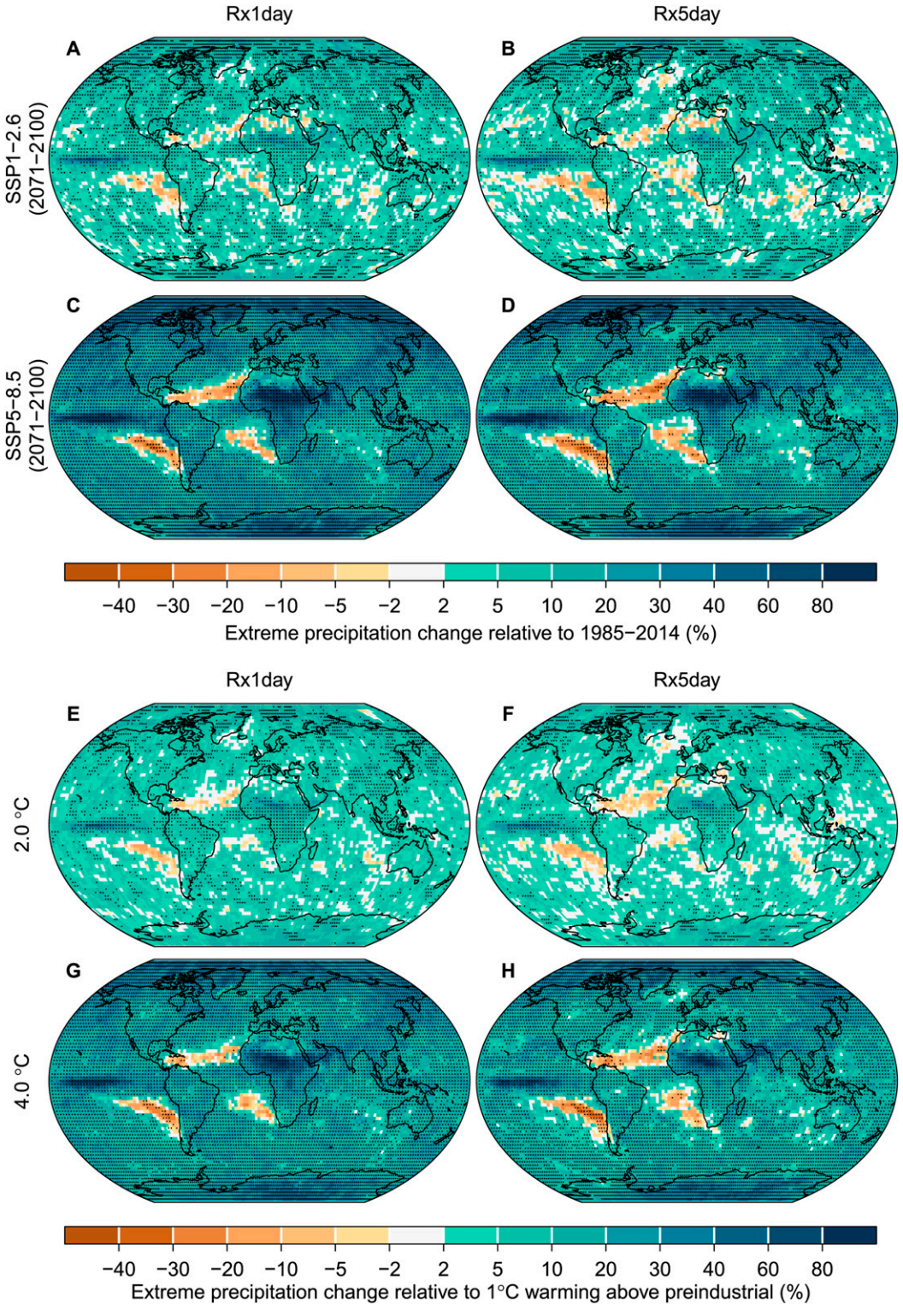

FIG. 5. As in Fig. 4, but for Rx1day and Rx5day (%).

further reduced to 0.6°C (0.4°–0.9°C), 0.7°C (04°–1.0°C), 3.7% (2.2%–4.3%), and 3.0% (2.0%–4.3%) for 1.5°C global warming. See Tables S2–S5 for the projected intensity changes at different global warming levels for the IPCC AR6 land regions, the continents, and the globe. A point worth of noting is that hot temperature and precipitation extremes are projected to exhibit an ''intense gets intenser'' tendency in most of the IPCC AR6 land regions (Fig. S9).

c. *The rates of change in temperature and precipitation extremes with warming*

We next consider whether the projected rates of change in the intensity of extreme temperature and precipitation events per 1°C of global warming depend on the choice of forcing scenario and the sensitivity of climate models. Figures 6a, 6b and 7a, 7b display the global land median changes in the 50-yr return values of TXx and TNn and percentage changes in the 50-yr return values of Rx1day and Rx5day as a function of GSAT change. Generally, increases in temperature extremes outpace warming in GSAT (Figs. 6a,b), while changes in precipitation extremes follow changes in GSAT at roughly the Clausius–Clapeyron rate of ~7% $°C^{-1}$ (Figs. 7a,b).

Globally, the rates of change of temperature and precipitation extremes do not appear to vary strongly across emissions scenarios (color bars in Figs. 6c, 6d and 7c, 7d). For TXx and Rx5day, the multimodel median rates of change under SSP1–2.6 appear to be markedly lower than those under other scenarios (blue bars in Figs. 6c and 7d). The range of the estimated rates of change among models, however, is particularly wide under this scenario (blue whiskers in Figs. 6c and 7d) due to the relatively weak signal of change compared to unforced internal climate variability. The high variation among models impedes a reliable evaluation of whether the rates of change in these two extreme variables with global warming under SSP1–2.6 really differ from those under other scenarios. Previous evaluations based on CMIP3/5 simulations showed overall independence of emissions scenarios of changes in temperature and precipitation extremes normalized by GSAT (Kharin et al. 2007, 2013; Pendergrass et al. 2015; Seneviratne et al. 2016).

The rates of change in temperature and precipitation extremes also tend not to vary strongly with the sensitivity of climate models (gray bars in Figs. 6c,d and 7c,d). In fact, CanESM5, which has the highest climate sensitivity among the 20 models considered (Fig. 1b), shows intermediate rates of change in temperature and precipitation extremes, and INM-CM4–8, which has the lowest climate sensitivity, shows intermediate rates of change for hot temperature and precipitation extremes and the second highest rate of change in cold extremes. We speculate that the differences in the response rates of temperature and precipitation extremes between climate models are more likely due to different representations of the relevant physical processes. Moreover, the global rate of change relative to GSAT change does not vary systematically over time for either temperature or precipitation extremes, as indicated by the linearly aligned scatter of changes in the corresponding extremes versus changes in GSAT (Figs. 6a,b and 7a,b).

Regional relations between changes in the intensity of temperature and precipitation extremes and changes in GSAT can differ substantially from those at the global scale, particularly for cold extremes and precipitation extremes. For example, cold extremes exhibit divergent rates of change with global warming across models and scenarios in northern high-latitude regions such as northwest North America and northern Europe (Figs. S10 and S11), and in dry regions such as central Asia (Fig. S12). Changes of precipitation extremes in the tropics usually do not scale well with changes in GSAT, particularly in some regions where the prevailing atmospheric circulations are subsiding such as the Sahara (Fig. S13), or where dynamical factors such as the expansion of the tropics may cause temporally nonuniform precipitation change such as Central America (Figs. S14; e.g., Pfahl et al. 2017).

d. *The projected changes in the frequency of temperature and precipitation extremes*

The top two panels in Figs. 8 and 9 present the CMIP6 multimodel median ratios of the frequencies of 50-yr TXx, TNn, Rx1day, and Rx5day events (defined in the reference period 1985–2014) in 2071–2100 compared to the reference frequency in the reference period (which is 1/50 = 0.02; also see Figs. S15, S16 for the frequency ratios for 20-yr events). As expected, the frequency ratios for hot extremes are projected to exceed 1 everywhere (Figs. 8a,c), indicating that extremely hot days become more frequent everywhere. Despite the pronounced polar amplification of intensities (Fig. 4), larger relative frequency increases are found in lower latitudes, with the largest increases over the tropical oceans and in the Arabian Peninsula, due to the lower interannual variability of temperatures in these regions (e.g., Mahlstein et al. 2011; Hawkins and Sutton 2012). The projected global land median frequency ratio is 6.9 (3.8–14.1 for the central 90% range of individual model estimates) by the end of the century under SSP1–2.6, indicating that days as hot as those expected with an annual probability of 1/50 in the reference period 1985–2014 may be about 7 times as likely on average over land (Fig. 8a). The frequency ratios saturate to a constant level of 50 over many parts of Earth's surface under SSP5–8.5 (Fig. 8c). This means that the annual hottest day at the end of the century is projected to be at least as hot under SSP5–8.5 as the 50-yr hot event in the 1985–2014 climate in these regions.

For cold extremes, the projected frequency ratios are less than 1 everywhere, indicating a reduction in the frequency of cold nights (Figs. 8b,d). Again, owing to the lower interannual variability, the largest relative frequency decreases are found over low-latitude oceans. Northern high-latitude decreases are also quite large because of the strong Arctic amplification effect on nighttime temperatures. The models project that there may not be a night at the end of this century (<1/1000 chance) that is as cold as the 50-yr TNn event in the 1985–2014 climate over about half of the global surface even under the SSP1–2.6 scenario (Fig. 8b) and over almost the entire globe under SSP5–8.5 (Fig. 8d).

Extreme precipitation events are projected to become more frequent over most areas, except in some subtropical regions where their intensities are projected to decline (Figs. 9a–d).

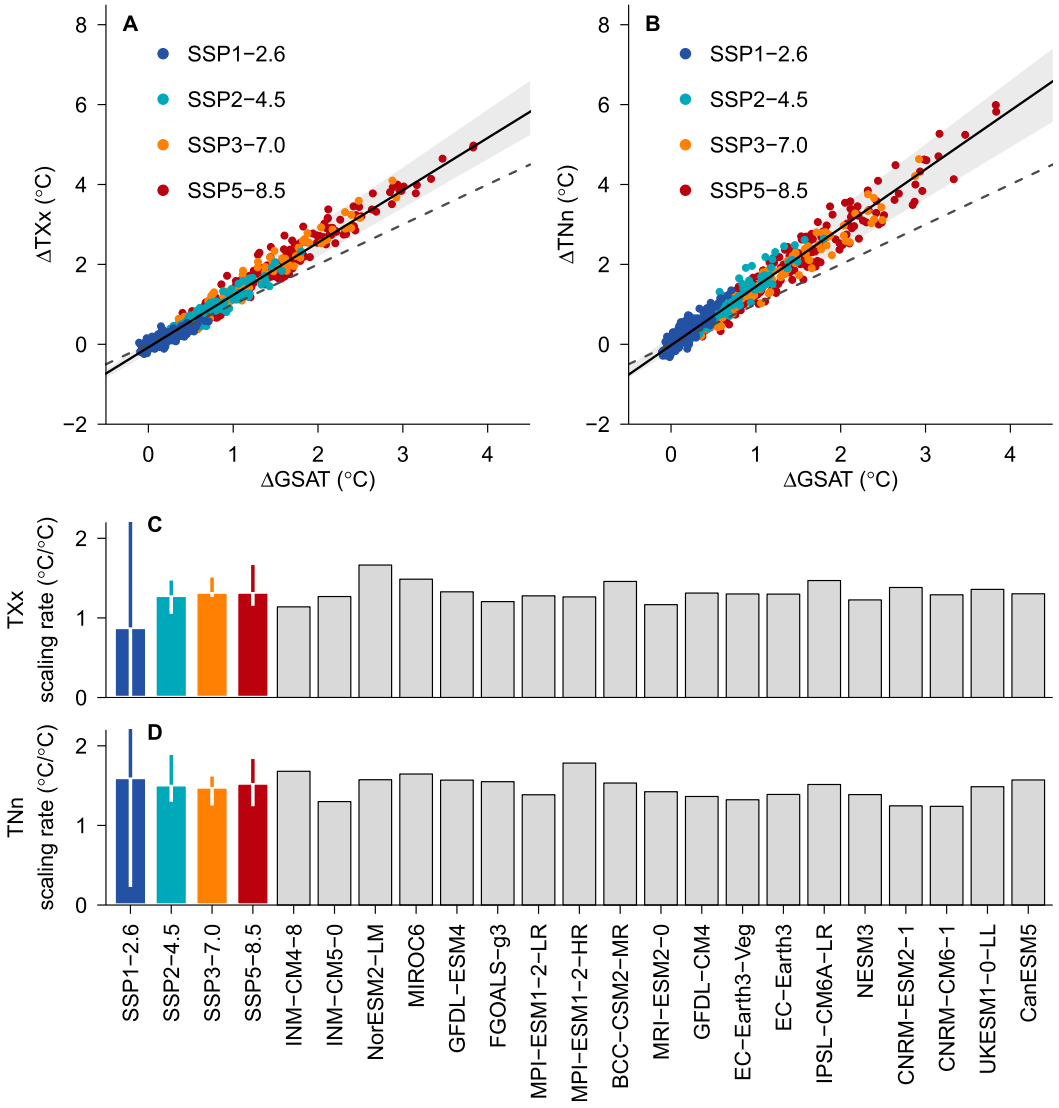

FIG. 6. The scaling of the global land median increases in the 50-yr return values of TXx and TNn with global warming. (a),(b) Scatterplots of the global land median changes in the 50-yr return values of TXx and TNn vs changes in GSAT (°C) in the CMIP6 multimodel ensemble simulations under different forcing scenarios. Each dot represents the median change in the intensity of 50-yr events over the global land area between two 30-yr periods in a CMIP6 simulation under a given SSP scenario (colors). The black solid lines show the quantile regression lines for the median through the scatter of points and the gray envelopes are bounded the corresponding quantile regression lines for the 5th and 95th quantiles of the scatter of points. The dashed lines show the 1-to-1 reference line. (c),(d) Estimates of the corresponding scaling rates (°C °C^{-1}) based on simulations under different forcing scenarios (color bars) and from different climate models (gray bars). Climate models from left to right exhibit increasing equilibrium climate sensitivity (see Fig. 1b). For scaling rate estimates under a given forcing scenario, the whiskers indicate the 5%–95% uncertainty range of the estimates from individual models, while for scaling rate estimates from a given model, simulations under different forcing scenarios were pooled together. Scenario uncertainty is not shown for estimates from individual models because most models have simulations only under the SSP1–2.6 and SSP5–8.5 scenarios (Fig. 1a).

The tropics and high latitudes are projected to see the strongest relative frequency increases. On average over land, daily precipitation events as extreme as the 1-in-50-yr event in 1985–2014 may be 3.9 (2.6–7.0 for the central 90% ensemble range of individual model estimates) times as likely by the end of the century under SSP5–8.5 (Fig. 9c), and 5-day precipitation events may be 3.4 (2.5–6.3) times as likely (Fig. 9d). That is, a 50-yr event in the reference climate may become a 13-yr event or so in the climate of 2071–2100. Due to the relatively weaker external forcing of SSP1–2.6, the projected relative frequency

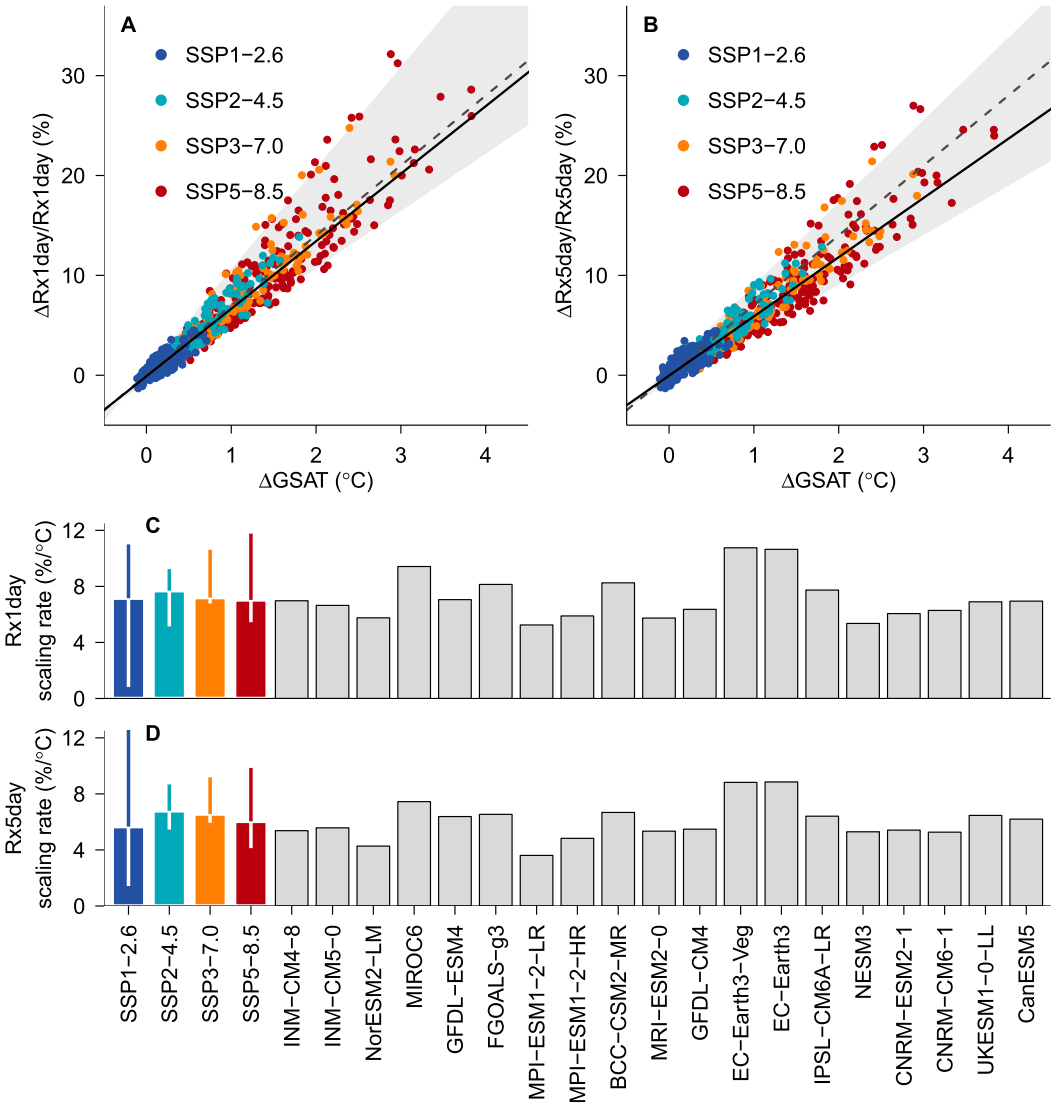

FIG. 7. As in Fig. 6, but for Rx1day and Rx5day [changes in precipitation extremes and GSAT in (a) and (b) are in % and °C, respectively; the scaling rate estimates for precipitation extremes in (c) and (d) are in % $°C^{-1}$]. The dashed line in (a) and (b) shows the 7% $°C^{-1}$ reference line.

changes are contaminated more by internal variability, particularly at small spatial scales, as indicated by the scattered patches with decreasing or increasing extreme precipitation frequency (Figs. 9a,b). Nevertheless, 1-in-50-yr Rx1day and Rx5day events in the reference climate are projected to become about 1.7 times as likely on average over land under SSP1–2.6 by 2071–2100.

The patterns of increasing and decreasing frequencies of extreme temperature and precipitation events at global warming levels are similar to those for changes in the 2071–2100 period, but with different magnitudes of frequency ratios (bottom two panels in Figs. 8 and 9; see also Tables S2-S5 for the projected changes at different global warming levels for the IPCC AR6 land regions and the continents). A 4°C warming may result in land median frequency ratios of 25.4 (22.1–30.3 for the central 90% range of estimates from the corresponding warming level windows), 1/221 (1/1150–1/38.7), 3.1 (2.4–4.5), and 2.7 (2.4–4.3), respectively, for the 50-yr TXx, TNn, Rx1day, and Rx5day events compared to a 1°C warmer world, while in a 2°C warmer world, the corresponding frequency ratios are 5.4 (3.8–7.1), 1/4.6 (1/8.3–1/2.9), 1.6 (1.3–1.7), and 1.4 (1.3–1.7), respectively (orange boxplots in Fig. 10). If further limiting global warming to 1.5°C, these frequency ratios would be 2.6 (1.9–3.5), 1/2 (1/3–1/1.5), 1.3 (1.2–1.4), and 1.2 (1.1, 1.3). Apparently, limiting global warming below 1.5°C would substantially reduce the occurrence of hot extremes and precipitation extremes, and substantially slowdown the disappearance of cold extremes, in line with findings on temperature and precipitation extremes from the SREX and SR15, the Benefits of Reduced Anthropogenic Climate Change projects analyzing

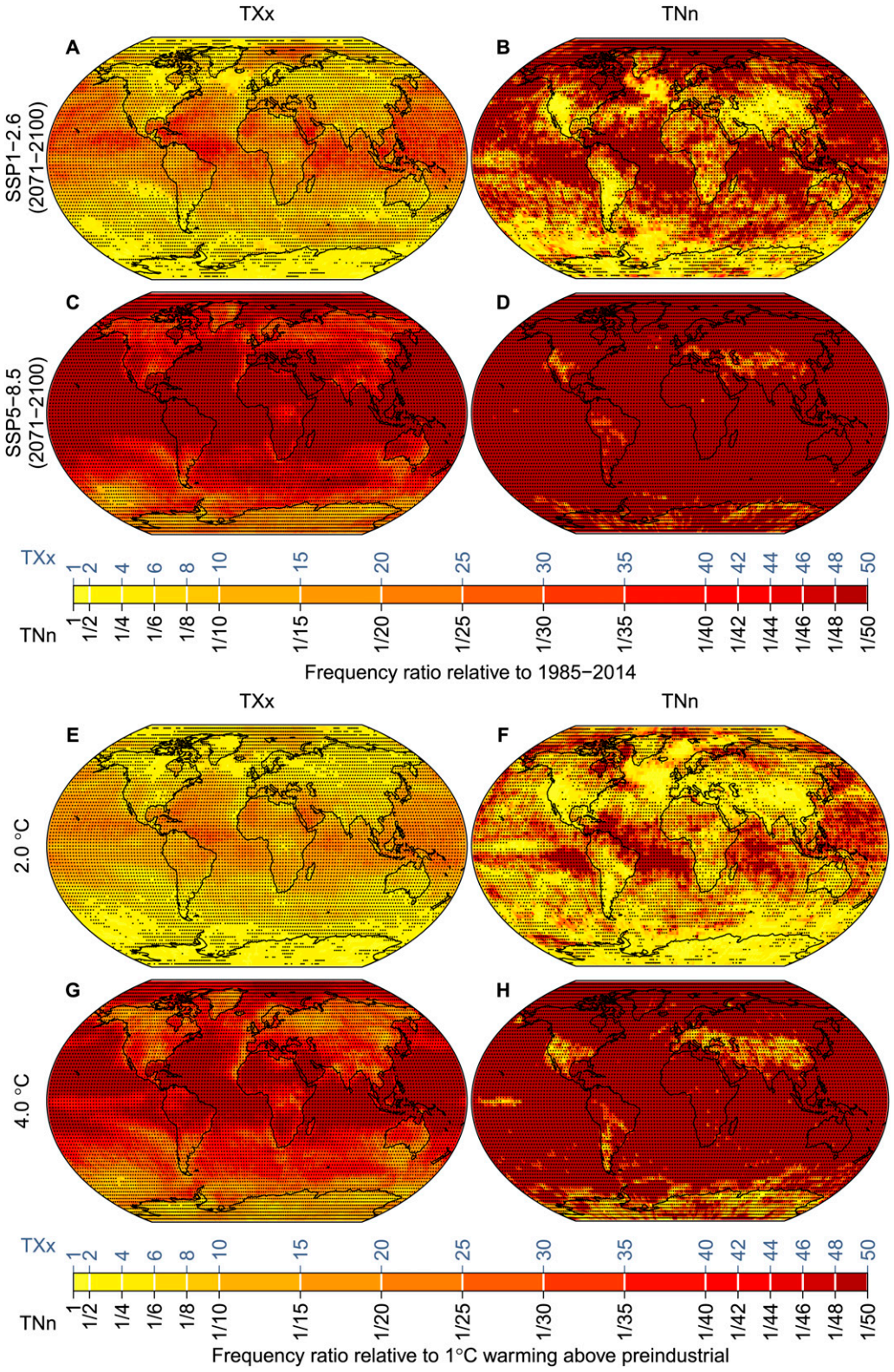

FIG. 8. Projected relative frequency changes of 50-yr TXx and TNn events. (a)–(d) The CMIP6 multimodel median ratios of the frequencies in 2071–2100 under the lower SSP1–2.6 and higher SSP5–8.5 scenarios for the 50-yr TXx and TNn events defined in the reference period 1985–2014 compared to the reference event frequency in 1985–2014 (which is 1/50 = 0.02). (e),(f) The corresponding ratios at 2.0° and 4.0°C global warming above pre-industrial compared to 1.0°C global warming. Stippling marks grid cells where at least 80% of the available models or warming level windows agree on the direction of the corresponding ensemble median relative frequency changes. Units are dimensionless.

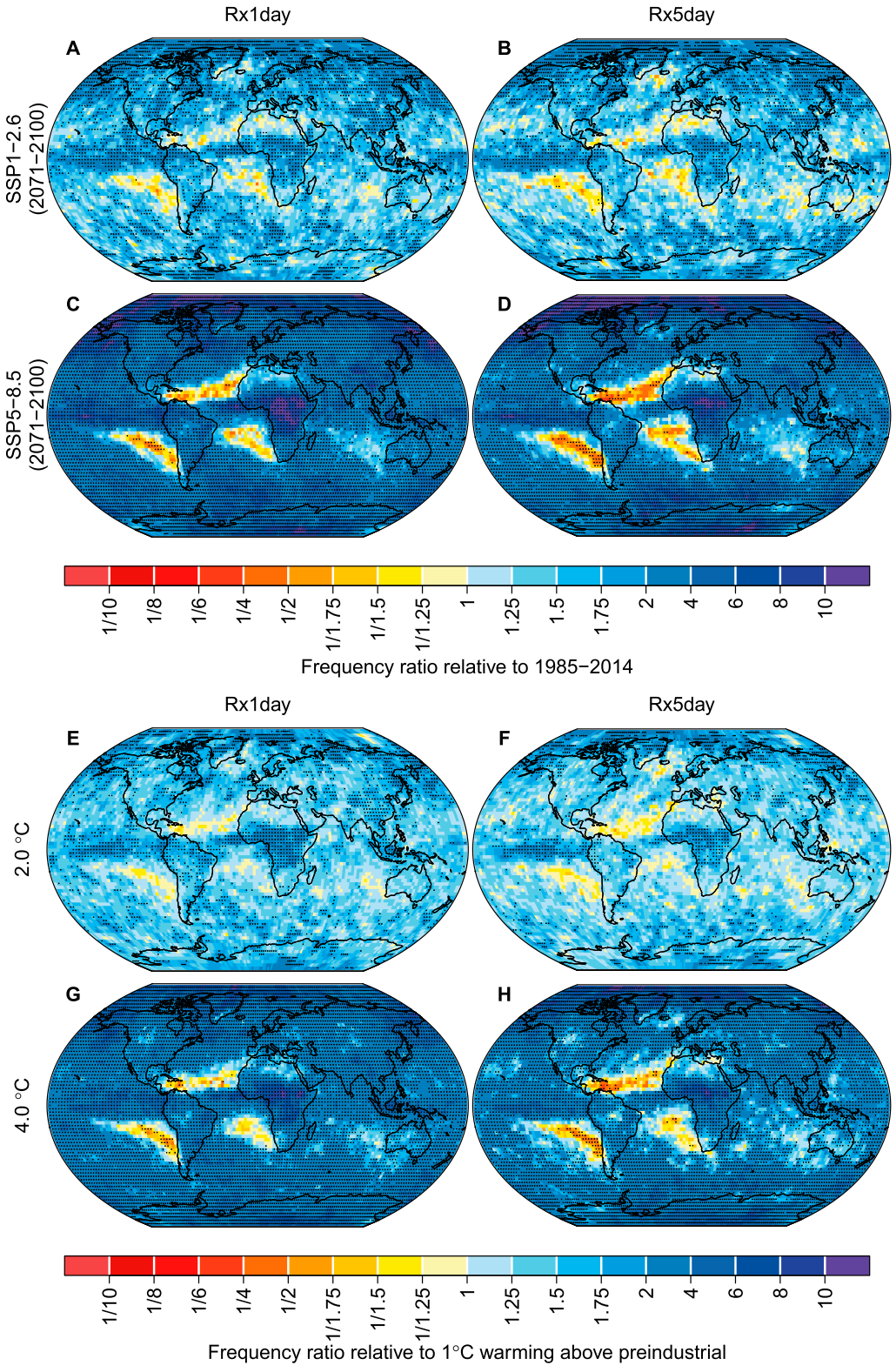

FIG. 9. As in Fig. 8, but for Rx1day and Rx5day. Units are dimensionless.

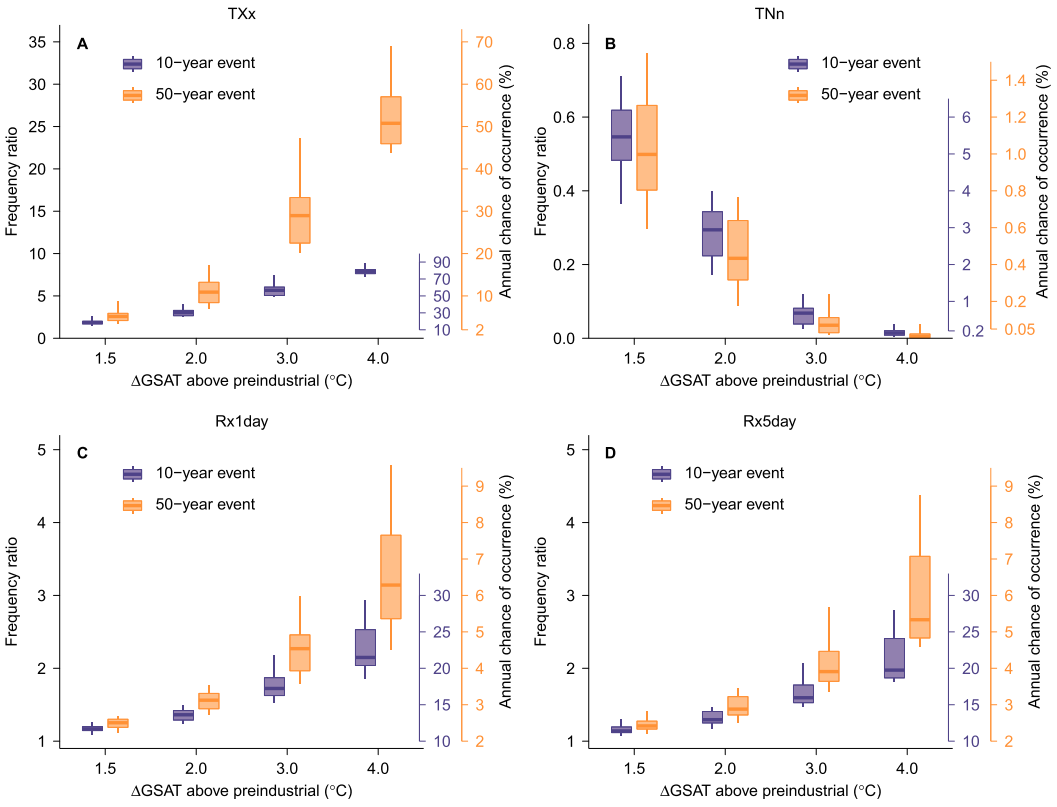

FIG. 10. Global land median changes in the frequency of temperature and precipitation extremes as function of global warming and event frequency. Panels show the global land median ratios of the frequencies at 1.5°, 2.0°, 3.0°, and 4.0°C global warming above preindustrial levels compared to their frequency in the 1.0°C global warming level climate above preindustrial levels for 10-yr and 50-yr (a) TXx, (b) TNn, (c) Rx1day, and (d) Rx5day events defined in the 1.0°C warming level climate. Annual chances of occurrence of the events corresponding to frequency ratios of the events are marked on the right side of the plots. The horizontal line and the box represent the median and central 66% uncertainty range of the frequency ratios estimated from plausible warming level windows, and the whisker extends to the full range of these frequency ratios.

CMIP5 simulations (e.g., Tebaldi and Wehner 2018) and simulations stabilizing at 1.5° and 2.0°C above preindustrial levels by the end of the century (Aerenson et al. 2018).

The projected frequency changes for 50-yr events are substantially larger than those for weaker events. For example, the land median frequency ratios for less intense 10-yr events in a 2°C warmer world are projected to be 3.0 (2.6–3.6), 1/3.4 (1/5.5–1/2.7), 1.4 (1.2–1.5), and 1.3 (1.2–1.4), respectively (purple boxplots in Fig. 10), which are remarkably smaller than the corresponding ratios for 50-yr events. The contrast in relative frequency changes between more extreme and weaker events is projected to become larger as climate warms (orange vs purple boxplots in Fig. 10). Furthermore, the larger relative frequency changes in more extreme hot temperature and precipitation events than weaker events are projected to occur in almost all IPCC AR6 land regions (Fig. S17).

The dependence of relative frequency changes in temperature and precipitation extremes on rarity has previously been reported by Kharin et al. (2018), who analyzed the same temperature and precipitation extremes in CMIP5 simulations but used a transient nonstationary GEV distribution with both location and scale parameters varying with GSAT to estimate the relative frequency changes of these extremes. The consistent results obtained here using a time-slice GEV estimation method that does not require assumptions on how the GEV parameters change with warming indicate that the larger relative frequency changes of more extreme events are unlikely to be due to the use of a particular analysis method. In fact, implementing an empirical analysis approach that does not require any distributional assumptions on simulations from models with five initial-condition ensemble members also produces qualitatively consistent results (Fig. S18).

e. The role of large ensemble initial-condition simulations in small-scale projections

Estimating long-term changes of climate extremes in the presence of internal climate variability is challenging, particularly when the extreme event of interest is truly rare and at small spatial scales, data records are short, and the response to external forcing is weak relative to internal variability (e.g., Li et al. 2019a), as is also confirmed by the reported results. Large ensemble initial-condition simulations enable

increased sampling of internal climate variability (magnitude and phase), thus leading to improved estimates of long-term changes in climate extremes with reduced sampling uncertainty. We here highlight the role that large ensemble initial-condition simulations can play in robustly projecting future changes in temperature and precipitation extremes at small spatial scales. As the influence of internal variability on precipitation extremes is substantially larger than for temperature extremes, we take the former as an illustrative example.

Due to the very high computational cost, truly large initial-condition ensemble simulations with, for example, 50 or more ensemble members, are very limited. We therefore implement spatial pooling of Rx1day data from 3×3 grid cell regions on 5-member initial-condition simulations from four CMIP6 models (Table S1). Doing so creates Rx1day samples that are effectively multiple times ($<5 \times 9 = 45$) the size of the samples obtained from a single simulation without spatial pooling, provided that the variations of Rx1day between neighboring grid cells are mainly due to high-frequency internal variability (Li et al. 2019a). This enlarged dataset is used to mimic the availability of larger initial-condition ensemble simulations.

The top two panels in Fig. 11 contrast the estimated changes in 50-yr Rx1day events in 2071–2100 relative to 1985–2014 under SSP1–2.6 and SSP5–8.5 based on the first CanESM5 simulation without spatial pooling and those obtained by pooling Rx1day data from 3×3 grid cells from all five simulations. The estimated changes from the first simulation show fragmented spatial patterns of increasing and decreasing precipitation extremes, with isolated patches showing extremely high or low changes, especially under the relatively weaker SSP1–2.6 scenario (Figs. 11a,c), whereas geographically organized spatial patterns that are potentially physically interpretable emerge when using the pooled simulations (Figs. 11b,d).

A simple measure of the uncertainty of an estimated change is the absolute value of the ratio between the estimate and its standard error (which is inferred using a spatiotemporal block bootstrap procedure with a spatial block that is the 3×3 grid cell region and a temporal block of consecutive 5 years; see a detailed treatment in Li et al. 2019b). An estimated change $\hat{\Delta}$ for which this ratio is greater than 5 indicates that the 95% confidence interval of this estimated change is *robustly* constrained by $\hat{\Delta} \pm 0.5\hat{\Delta}$.

If just one simulation is available and the spatial pooling is not implemented, the absolute value of this ratio for the estimated changes in the intensity of 50-yr Rx1day events is unlikely to exceed 5 almost everywhere across the globe (green lines within solid red bars in Fig. 11e). Implementing 3×3 spatial pooling can improve this result, but to a limited extent (white lines within solid red bars in Fig. 11e). This is because spatial pooling is unable to sample low-frequency internal variability at multidecadal time scales (Li et al. 2019a). The five initial-condition simulations together with 3×3 spatial pooling can produce such robust estimates over 60% of the global surface under SSP5–8.5 in all models with ensembles of this size except MPI-ESM1–2-LR (solid red bars in Fig. 11e). Overall, these results suggest that large ensemble initial-condition simulations are needed for robust projections of future precipitation extremes at local scales that are often impact relevant. This is particularly so when projecting very rare extremes (hatched bars for 2-yr events vs solid bars for 50-yr events in Fig. 11e) under relatively weaker forcing scenarios (blue bars for SSP1–2.6 vs red bars for SSP5–8.5). Spatial pooling can, to some extent, reduce the ensemble size needed for projecting climate variables with weak spatial dependence, such as precipitation extremes.

4. Conclusions

We have presented an evaluation of the new-generation CMIP6 multimodel ensemble in simulating present-day extremes of near surface daily temperature and precipitation that are expected once every 2–50 years on average, and their projected changes under the four CMIP6 tier-1 future forcing scenarios. The following summarizes the main conclusions, which are primarily focused on large spatial scales:

- Judged by their similarity to ERA5, the new-generation CMIP6 models simulate reasonably well large-scale spatial patterns of the present-day near surface temperature and precipitation extremes with a broad range of return periods from 2 to 50 years, with pattern correlations for these extremes being larger than 0.88.
- For temperature extremes, models perform better for hot extremes than for cold extremes and better over ice-free oceans than elsewhere. Models tend to underestimate precipitation extremes in the tropics and overestimate them in subtropical dry areas relative to ERA5. On average over land, the multimodel median return value estimate is about 0.5°C warmer for the 50-yr TXx events compared to ERA5, while it is 1.0°C colder for the 50-yr TNn events. The multimodel median estimates of the 50-yr land Rx1day and Rx5day return values are 120% and 130% of ERA5 values, respectively.
- Models consistently project increases in the frequency and intensity of hot extremes and decreases in cold extremes everywhere across the globe, with more rapid change in cold extremes than in hot extremes. The 50-yr TXx and TNn events over land are projected to warm 1.3°C (1.0°–1.6°C for the central 90% range of estimates from warming level windows as simulated in the considered climate models) and 1.5°C (1.0°–1.7°C), respectively, when global warming increases from 1° to 2°C above preindustrial levels. Consequently, days as hot as those expected with an annual probability of 1/50 in the reference 1°C warming world as measured by TXx would be about 5 times as likely on average over land in the 2°C warming world. While regional changes can differ substantially from these global-scale values, increases in warm extremes and decreases in cold extremes are projected in all regions.
- A large majority (>80%) of models agree on increases in the frequency and intensity of precipitation extremes over most of Earth's surface, except some subtropical regions with prevailing downwelling atmospheric circulations. Over land, 50-yr Rx1day and Rx5day events are projected to intensify 7.2% (5.7%–9.1% for the central 90% range of estimates from warming level windows as simulated in the considered

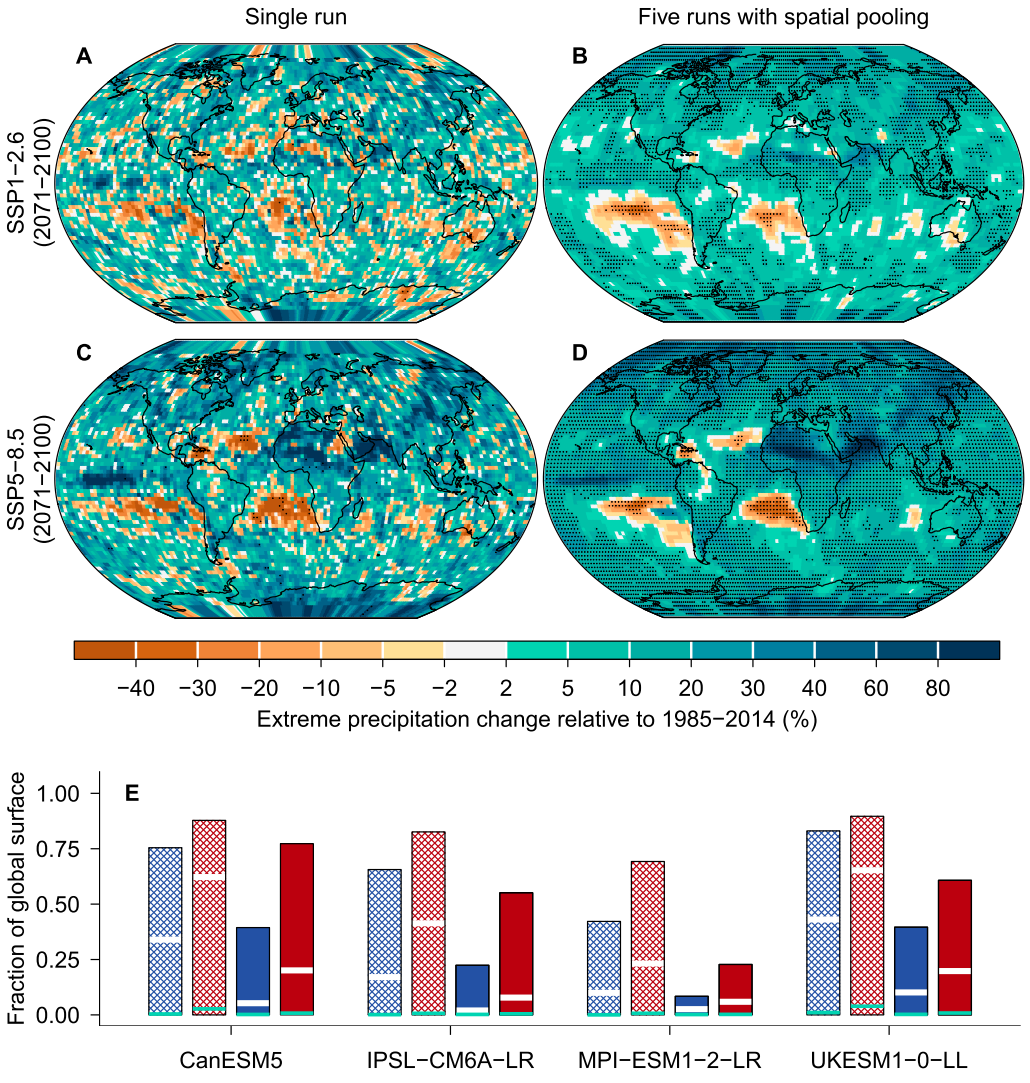

FIG. 11. (a)–(d) Estimated relative changes in the 50-yr return values of Rx1day in 2071–2100 under the lower SSP1–2.6 and higher SSP5–8.5 scenarios relative to 1985–2014 using a single simulation without spatial pooling and an ensemble of five simulations from CanESM5 with a 3×3 spatial pooling. Stippling marks grid cells with robustly constrained estimates, that is, for which the absolute ratio of an estimated change and its standard error is greater than 5. (e) The fractions of the global surface with robustly constrained estimates of relative changes in 50-yr (solid bars) and 2-yr (hatched bars) return values of Rx1day in 2071–2100 under the lower SSP1–2.6 (blue bars) and higher SSP5–8.5 (red bars) scenarios relative to 1985–2014 using an ensemble of five simulations with spatial pooling. The corresponding fractions obtained using a single simulation without and with spatial pooling are marked by green and white lines, respectively.

climate models), and 6.3% (4.3%–8.4%) when global warming increases from 1° to 2°C above preindustrial. Correspondingly, these reference climate events are projected to become approximately 1.5 times as likely. Again, markedly regional variations exist.

- Changes in temperature extremes outpace changes in global annual mean surface air temperature (GSAT) over the majority of landmasses, while, globally, changes in precipitation extremes follow changes in GSAT at roughly the Clausius–Clapeyron rate of ~7% $°C^{-1}$. Changes in temperature and precipitation extremes normalized with respect to GSAT do not depend sensitively on forcing scenarios or climate model sensitivity and do not vary strongly over time, but with notable regional variations.
- In most land regions, there exists an ''intense gets intenser'' tendency in hot temperature and precipitation extremes. In almost all land regions, the relative frequency changes are larger for more extreme such events than for weaker events.
- Limiting global warming to no more than 2°C above preindustrial levels can substantially suppress the intensification of hot extremes and precipitation extremes and slowdown the disappearance of cold extremes.

- To obtain robust projections of impact-relevant local temperature and precipitation extremes, large initial-condition ensemble simulations are needed. Appropriate spatial pooling data of neighboring grid cells can, to some extent, reduce the required ensemble size for climate extreme variable with weak spatial dependence, such as precipitation extremes.

Overall, these conclusions based on new CMIP6 simulations are in line with those from previous CMIP simulations, painting the same large-scale picture of more frequent and more intense hot temperature and precipitation extremes if climate warming is not limited. To obtain more robust estimates for the rare 50-yr extreme events, we chose to analyze 30-yr time windows rather than 20-yr windows as used in IPCC assessment reports. For reference, we provide projected changes in the intensity and frequency of temperature and precipitation extremes in 2081–2100 relative to 1995–2014 in Figs. S19 and S20, but caution that these results are somewhat more uncertain because they are more strongly affected by internal climate variability.

Acknowledgments. We thank the Program for Climate Model Diagnosis and Intercomparison and the World Climate Research Programme's Working Group on Coupled Modeling for making the WCRP CMIP multimodel dataset available (https://esgf-node.llnl.gov/search/cmip6/). We thank Jana Sillmann at the Center for International Climate Research, Norway, for retrieving TXx, TNn, Rx1day, and Rx5day indices from the CMIP6 simulations. We acquired ERA5 and ERA-Interim data from https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5, NCEP Reanalysis 2 data from https://www.esrl.noaa.gov/psd, and the HadEX3 observations from https://www.metoffice.gov.uk/hadobs/hadex3/index.html. Computer code for estimating extremes with GEV and empirical distributions are provided as online supporting materials. This study was supported by the National Key R&D Programs of China (2018YFC1507700). CL was also supported by National Natural Science Foundation of China (42075026). MW was supported by the Director, Office of Science, Office of Biological and Environmental Research of the U.S. Department of Energy through the Regional and Global Climate Modelling program (DE340AC02-05CH1123).

REFERENCES

Aerenson, T., C. Tebaldi, B. Sanderson, and J. Lamarque, 2018: Changes in a suite of indicators of extreme temperature and precipitation under 1.5 and 2.0 degrees warming. *Environ. Res. Lett.*, **13**, 035009, https://doi.org/10.1088/1748-9326/aaafd6.

Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, **111**, D05109, https://doi.org/10.1029/2005JD006290.

Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor, 2012: Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophys. Res. Lett.*, **39**, L09712, https://doi.org/10.1029/2012GL051607.

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

Coumou, D., and S. Rahmstorf, 2012: A decade of weather extremes. *Nat. Climate Change*, **2**, 491–496, https://doi.org/10.1038/nclimate1452.

Dai, A., 2006: Precipitation characteristics in eighteen coupled climate models. *J. Climate*, **19**, 4605–4630, https://doi.org/10.1175/JCLI3884.1.

Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, https://doi.org/10.1002/qj.828.

Donat, M. G., and Coauthors, 2013: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res. Atmos.*, **118**, 2098–2118, https://doi.org/10.1002/jgrd.50150.

Dunn, R. J. H., and Coauthors, 2020: Development of an updated global land in situ-based data set of temperature and precipitation extremes: HadEX3. *J. Geophys. Res. Atmos.*, **125**, e2019JD032263, https://doi.org/10.1029/2019JD032263.

Easterling, D. R., G. A. Meehl, C. Parmesan, S. A. Changnon, T. R. Karl, and L. O. Mearns, 2000: Climate extremes: Observations, modeling, and impacts. *Science*, **289**, 2068–2074, https://doi.org/10.1126/science.289.5487.2068.

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

Fischer, E. M., and R. Knutti, 2014: Detection of spatially aggregated changes in temperature and precipitation extremes. *Geophys. Res. Lett.*, **41**, 547–554, https://doi.org/10.1002/2013GL058499.

——, J. Sedláček, E. Hawkins, and R. Knutti, 2014: Models agree on forced response pattern of precipitation and temperature extremes. *Geophys. Res. Lett.*, **41**, 8554–8562, https://doi.org/10.1002/2014GL062018.

Fisher, R. A., and L. H. C. Tippett, 1928: Limiting forms of the frequency distribution of the largest or smallest members of a sample. *Proc. Cambridge Philos. Soc.*, **24**, 180–190, https://doi.org/10.1017/S0305004100015681.

Gregory, J. M., and Coauthors, 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, **31**, L03205, https://doi.org/10.1029/2003GL018747

Hawkins, E., and R. Sutton, 2012: Time of emergence of climate signals. *Geophys. Res. Lett.*, **39**, L01702, https://doi.org/10.1029/2011GL050087.

Hersbach, H., and Coauthors, 2018: Operational global reanalysis: Progress, future directions and synergies with NWP. ECMWF ERA Rep. Series 27, 65 pp., https://www.ecmwf.int/en/elibrary/18765-operational-global-reanalysis-progress-future-directions-and-synergies-nwp.

Hoesly, R. M., and Coauthors, 2018: Historical (1750-2014) anthropogenic emissions of reactive gases and aerosols from the Community Emissions Data System (CEDS). *Geosci. Model Dev.*, **11**, 369–408, https://doi.org/10.5194/gmd-11-369-2018.

Hosking, J. R. M., 1990: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc.*, **52B**, 105–124, https://doi.org/10.1111/j.2517-6161.1990.tb01775.x.

IPCC, 2018: *Global Warming of 1.5°C*. V. Masson-Delmotte et al., Eds., Cambridge University Press, 630 pp., https://www.ipcc.ch/site/assets/uploads/sites/2/2019/06/SR15_Full_Report_Low_Res.pdf.

Iturbide, M., and Coauthors, 2020: An update of IPCC climate reference regions for subcontinental analysis of climate model data: Definition and aggregated datasets. *Earth Syst. Sci. Data*, **12**, 2959–2970, https://doi.org/10.5194/essd-12-2959-2020.

James, R., R. Washington, C. Schleussner, J. Rogelj, and D. Conway, 2017: Characterizing half-a-degree difference: A review of methods for identifying regional climate responses to global warming targets. *Wiley Interdiscip. Rev.: Climate Change*, **8**, e457, https://doi.org/10.1002/wcc.457.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.*, **83**, 1631–1644, https://doi.org/10.1175/BAMS-83-11-1631.

Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl, 2007: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Climate*, **20**, 1419–1444, https://doi.org/10.1175/JCLI4066.1.

——, ——, ——, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, **119**, 345–357, https://doi.org/10.1007/s10584-013-0705-8.

——, G. M. Flato, X. Zhang, N. P. Gillett, F. W. Zwiers, and K. J. Anderson, 2018: Risks from climate extremes change differently from 1.5°C to 2.0°C depending on rarity. *Earth's Future*, **6**, 704–715, https://doi.org/10.1002/2018EF000813.

Kim, Y., S. Min, X. Zhang, F. W. Zwiers, L. V. Alexander, M. K. Donat, and Y. Tung, 2016: Attribution of extreme temperature changes during 1951-2010. *Climate Dyn.*, **46**, 1769–1782, https://doi.org/10.1007/s00382-015-2674-2.

Koenker, R., 2005: *Quantile Regression*. Cambridge University Press, 349 pp.

Leadbetter, M. R., G. Lindgren, and H. Rootzen, 1983: *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, 336 pp.

Li, C., X. Zhang, F. W. Zwiers, Y. Fang, and A. M. Michalak, 2017: Recent very hot summers in Northern Hemispheric land areas measured by wet bulb globe temperature will be the norm within 20 years. *Earth's Future*, **5**, 1203–1216, https://doi.org/10.1002/2017EF000639.

——, Y. Fang, K. Caldeira, X. Zhang, N. S. Diffenbaugh, and A. M. Michalak, 2018: Widespread persistent changes to temperature extremes occurred earlier than predicted. *Sci. Rep.*, **8**, 1007, https://doi.org/10.1038/s41598-018-19288-z.

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

——, and Coauthors, 2019b: Larger increases in more extreme local precipitation events as climate warms. *Geophys. Res. Lett.*, **46**, 6885–6891, https://doi.org/10.1029/2019GL082908.

——, Y. Sun, F. W. Zwiers, D. Wang, X. Zhang, G. Chen, and H. Wu, 2020: Rapid warming in summer wet bulb globe temperature in China with human-induced climate change. *J. Climate*, **33**, 5697–5711, https://doi.org/10.1175/JCLI-D-19-0492.1.

Lorenz, R., Z. Stalhandske, and E. M. Fischer, 2019: Detection of a climate change signal in extreme heat, heat stress, and cold in Europe from observations. *Geophys. Res. Lett.*, **46**, 8363–8374, https://doi.org/10.1029/2019GL082062.

Mahlstein, I., R. Knutti, S. Solomon, and R. W. Portmann, 2011: Early onset of significant local warming in low latitude countries. *Environ. Res. Lett.*, **6**, 034009, https://doi.org/10.1088/1748-9326/6/3/034009.

Meehl, G. A., C. Covey, T. Delworth, M. Latif, B. McAvaney, J. F. B. Mitchell, R. J. Stouffer, and K. E. Taylor, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383–1394, https://doi.org/10.1175/BAMS-88-9-1383.

Meinshausen, M., and Coauthors, 2017: Historical greenhouse gas concentrations for climate modelling (CMIP6). *Geosci. Model Dev.*, **10**, 2057–2116, https://doi.org/10.5194/gmd-10-2057-2017.

——, and Coauthors, 2020: The shared socio-economic pathway (SSP) greenhouse gas concentrations and their extensions to 2500. *Geosci. Model Dev.*, **13**, 3571–3605, https://doi.org/10.5194/gmd-13-3571-2020.

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

National Research Council of Canada, 2005: National Building Code of Canada: 2005. Canadian Commission on Building and Fire Codes, 1167 pp., https://doi.org/10.4224/40001245.

O'Neill, B. C., and Coauthors, 2016: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.*, **9**, 3461–3482, https://doi.org/10.5194/gmd-9-3461-2016.

Pendergrass, A. G., F. Lehner, B. M. Sanderson, and Y. Xu, 2015: Does extreme precipitation intensity depend on the emissions scenario? *Geophys. Res. Lett.*, **42**, 8767–8774, https://doi.org/10.1002/2015GL065854.

Pfahl, S., P. A. O'Gorman, and E. M. Fischer, 2017: Understanding the regional pattern of projected future changes in extreme precipitation. *Nat. Climate Change*, **7**, 423–427, https://doi.org/10.1038/nclimate3287.

Riahi, K., and Coauthors, 2017: The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environ. Change*, **42**, 153–168, https://doi.org/10.1016/j.gloenvcha.2016.05.009.

Seneviratne, S. I., and Coauthors, 2012: Changes in climate extremes and their impacts on the natural physical environment. *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, C. B. Field et al., Eds., Cambridge University Press, 109–230, https://www.ipcc.ch/site/assets/uploads/2018/03/SREX-Chap3_FINAL-1.pdf.

——, M. G. Donat, A. J. Pitman, R. Knutti, and R. L. Wilby, 2016: Allowable CO_2 emissions based on regional and impact-related climate targets. *Nature*, **529**, 477–483, https://doi.org/10.1038/nature16542.

Sillmann, J., M. Croci-Maspoli, M. Kallache, and R. W. Katz, 2011: Extreme cold winter temperatures in Europe under the influence of North Atlantic atmospheric blocking. *J. Climate*, **24**, 5899–5913, https://doi.org/10.1175/2011JCLI4075.1.

——, V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh, 2013a: Climate extremes indices in the CMIP5 multimodel ensemble: Part I. Model evaluation in the present climate. *J. Geophys. Res.*, **118**, 1716–1733, https://doi.org/10.1002/jgrd.50203.

——, ——, F. W. Zwiers, X. Zhang, and D. Bronaugh, 2013b: Climate extremes indices in the CMIP5 multimodel ensemble: Part II. Future climate projections. *J. Geophys. Res.*, **118**, 2473–2493, https://doi.org/10.1002/jgrd.50188.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, https://doi.org/10.1175/BAMS-D-11-00094.1.

Tebaldi, C., and M. F. Wehner, 2018: Benefits of mitigation for future heat extremes under RCP4.5 compared to RCP8.5. *Climatic Change*, **146**, 349–361, https://doi.org/10.1007/s10584-016-1605-5.

——, K. Hayhoe, J. Arblaster, and G. Meehl, 2006: Going to the extremes. An intercomparison of model-simulated historical and future changes in extreme events. *Climatic Change*, **79**, 185–211, https://doi.org/10.1007/s10584-006-9051-4.

van Marle, M. J. E., and Coauthors, 2017: Historic global biomass burning emissions for CMIP6 (BB4CMIP) based on merging satellite observations with proxies and fir models (1750-2015). *Geosci. Model Dev.*, **10**, 3329–3357, https://doi.org/10.5194/gmd-10-3329-2017.

Woelfle, M. D., C. S. Bretherton, C. Hannay, and R. Neale, 2019: Evolution of the double-ITCZ bias through CESM2 development. *J. Adv. Model. Earth Syst.*, **11**, 1873–1893, https://doi.org/10.1029/2019MS001647.

Wyser, K., T. van Noije, S. Yang, J. von Hardenberg, D. O'Donnell, and R. Döscher, 2019: On the increased climate sensitivity in the EC-Earth model from CMIP5 to CMIP6. *Geosci. Model Dev.*, **13**, 3465–3474, https:doi.org/10.5194/GMD-13-3465-2020.

Xu, Y., X. Gao, F. Giorgi, B. Zhou, Y. Shi, J. Wu, and Y. Zhang, 2018: Projected changes in temperature and precipitation extremes over China as measured by 50-yr return values and periods based on a CMIP5 ensemble. *Adv. Atmos. Sci.*, **35**, 376–388, https://doi.org/10.1007/s00376-017-6269-1.

Zelinka, M. D., T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, P., S. A. Klein, and K. E. Taylor, 2020: Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.*, **47**, e2019GL085782, https://doi.org/10.1029/2019GL085782.

Zhang, X., L. Alexander, G. C. Hegerl, P. Jones, A. Klein Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdiscip. Rev.: Climate Change*, **2**, 851–870, https://doi.org/10.1002/wcc.147.

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

——, F. W. Zwiers, G. Li, H. Wan, and A. J. Cannon, 2017: Complexity in estimating past and future extreme short-duration rainfall. *Nat. Geosci.*, **10**, 255–259, https://doi.org/10.1038/ngeo2911.

Zhang, X.-X., H. Liu, and M. Zhang, 2015: Double ITCZ in coupled ocean-atmosphere models: From CMIP3 to CMIP5. *Geophys. Res. Lett.*, **42**, 8651–8659, https://doi.org/10.1002/2015GL065973.

Zwiers, F. W., X. Zhang, and Y. Feng, 2011: Anthropogenic influence on long return period daily temperature extremes at regional scales. *J. Climate*, **24**, 881–892, https://doi.org/10.1175/2010JCLI3908.1.