Changes in Temperature and Precipitation Extremes in the IPCC Ensemble of Global Coupled Model Simulations

VIATCHESLAV V. KHARIN AND FRANCIS W. ZWIERS

Canadian Centre for Climate Modelling and Analysis, Environment Canada, Victoria, British Columbia, Canada

XUEBIN ZHANG

Climate Data and Analysis Section, Environment Canada, Toronto, Ontario, Canada

GABRIELE C. HEGERL

Nicholas School for the Environment and Earth Science, Duke University, Durham, North Carolina

(Manuscript received 15 August 2005, in final form 5 September 2006)

ABSTRACT

Temperature and precipitation extremes and their potential future changes are evaluated in an ensemble of global coupled climate models participating in the Intergovernmental Panel on Climate Change (IPCC) diagnostic exercise for the Fourth Assessment Report (AR4). Climate extremes are expressed in terms of 20-yr return values of annual extremes of near-surface temperature and 24-h precipitation amounts. The simulated changes in extremes are documented for years 2046–65 and 2081–2100 relative to 1981–2000 in experiments with the Special Report on Emissions Scenarios (SRES) B1, A1B, and A2 emission scenarios.

Overall, the climate models simulate present-day warm extremes reasonably well on the global scale, as compared to estimates from reanalyses. The model discrepancies in simulating cold extremes are generally larger than those for warm extremes, especially in sea ice–covered areas. Simulated present-day precipitation extremes are plausible in the extratropics, but uncertainties in extreme precipitation in the Tropics are very large, both in the models and the available observationally based datasets.

Changes in warm extremes generally follow changes in the mean summertime temperature. Cold extremes warm faster than warm extremes by about 30%–40%, globally averaged. The excessive warming of cold extremes is generally confined to regions where snow and sea ice retreat with global warming. With the exception of northern polar latitudes, relative changes in the intensity of precipitation extremes generally exceed relative changes in annual mean precipitation, particularly in tropical and subtropical regions. Consistent with the increased intensity of precipitation extremes, waiting times for late-twentieth-century extreme precipitation events are reduced almost everywhere, with the exception of a few subtropical regions. The multimodel multiscenario consensus on the projected change in the globally averaged 20-yr return values of annual extremes of 24-h precipitation amounts is that there will be an increase of about 6% with each kelvin of global warming, with the bulk of models simulating values in the range of 4%–10% K^{-1}. The very large intermodel disagreements in the Tropics suggest that some physical processes associated with extreme precipitation are not well represented in models. This reduces confidence in the projected changes in extreme precipitation.

1. Introduction

Human activities and the environment are greatly affected by climate and weather extremes. A growing interest in extreme climate events is motivated by the vulnerability of our society to the impacts of such events. There is growing evidence suggesting that the anthropogenic forcing is affecting the present climate (International Ad Hoc Detection and Attribution Group 2005) and will continue to do so in the future (Cubasch et al. 2001). The impacts of the changing climate will likely be felt most strongly through changes in intensity and frequency of climate extremes. It is therefore important to document future changes that might be caused by anthropogenic activities.

Corresponding author address: Viatcheslav Kharin, Canadian Centre for Climate Modelling and Analysis, University of Victoria, P.O. Box 1700, STN CSC, Victoria, BC, Canada.
E-mail: slava.kharin@ec.gc.ca

DOI: 10.1175/JCLI4066.1

Simulations with global coupled ocean–atmosphere general circulation models (CGCMs) forced with projected greenhouse gas and aerosol emissions are the primary tools for studying possible future changes in climate mean, variability, and extremes. Changes in rainfall distributions have attracted much attention because of the particular vulnerability of human activities to hydrological extreme events such as flood-producing rains and droughts. The intensity of extreme precipitation is projected to increase under global warming in many parts of the world, even in the regions where mean precipitation decreases (e.g., Kharin and Zwiers 2000, 2005; Semenov and Bengtsson 2002; Voss et al. 2002; Wilby and Wigley 2002; Wehner 2004). Future increases in heavy precipitation are accompanied by reduction in the probability of wet days, implying a more extreme future climate with higher probabilities of droughts and heavy precipitation events.

Changes in temperature extremes tend to follow mean temperature changes in many parts of the world. However, Kharin and Zwiers (2000, 2005) reported that cold temperature extremes warm faster than warm extremes in mid- and high latitudes, mainly as a result of snow and sea ice melting in winter under global warming. Increased temperature variability has been reported in some studies over land in summer (Gregory and Mitchell 1995; Kharin and Zwiers 2005), implying potentially larger relative increases in warm extremes than in mean summertime temperature.

The ability of the recent generation of atmospheric general circulation models to simulate temperature and precipitation extremes was recently documented by Kharin et al. (2005) for models participating in the second phase of the Atmospheric Model Intercomparison Project (AMIP2). The purpose of the present study is to document the performance of the current generation of CGCMs in simulating present-day extremes of temperature and precipitation and their potential changes under different projections for the evolution of the anthropogenic forcing, using model output submitted to the Program for Climate Model Diagnosis and Intercomparison (PCMDI; http://www-pcmdi.llnl.gov) in support of the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report (AR4).

The paper is organized as follows. Datasets are described in the next section. Extreme value methodology is summarized in section 3. The ability of the models to simulate present-day precipitation and temperature extremes is documented in section 4. Their changes under several emission scenarios are examined and discussed in section 5. The paper is concluded by a summary in section 6.

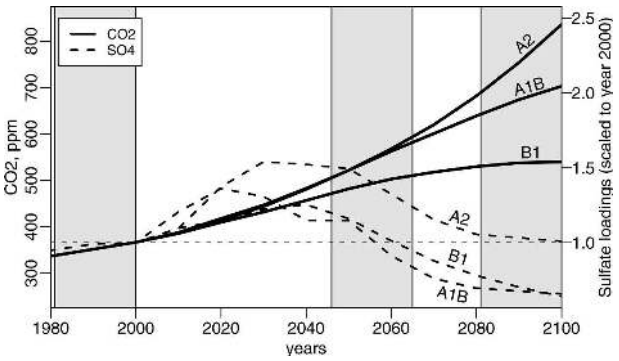

FIG. 1. The time evolution of the CO_2 concentrations (solid lines, *y* axis on the left-hand side) and globally averaged sulfate aerosol loadings scaled to year 2000 (dashed lines, *y* axis on the right-hand side) as prescribed in the IPCC SRES B1, A1B, and A2 experiments. The gray shaded areas indicate the time periods analyzed in the present study.

2. Datasets

The Working Group on Coupled Modeling (WGCM) of the World Climate Research Program (WCRP) requested that modeling groups submit daily model output for a number of 20-yr time periods to PCMDI in support of the IPCC AR4. In the present study we analyze annual extremes of daily maximum and minimum surface air temperature and of 24-h precipitation amounts for the time period 1981–2000 from simulations of the twentieth-century climate (20C3M), and for two 20-yr time periods 2046–65 and 2081–2100 from the Special Report on Emissions Scenarios (SRES) B1, A1B, and A2 experiments. Figure 1 illustrates the time evolution of carbon dioxide concentrations and sulfate aerosol loadings in these three emission scenarios. The gray shaded areas indicate the 20-yr time periods for which daily temperature and precipitation output was available for most of the models.

The B1 emission scenario, also known as the 550-ppm stabilization experiment, envisions the slowest growth of anthropogenic greenhouse gas concentrations, followed by the A1B experiment, or the 720-ppm stabilization experiment, with somewhat more rapid forcing. Many groups continued these simulations up to year 2300 with the concentrations held at the year-2100 level, but these stabilizations phases are not considered in the present study. The fastest growing greenhouse gas concentrations are specified in the A2 experiment with roughly 1% per year of CO_2 increase in the second half of the twenty-first century. The CO_2 concentrations are similar in the A1B and A2 emission scenarios up to the middle of the twenty-first century, but the A2 scenario also specifies somewhat greater sulfate aerosol concentrations, which are thought to have a cooling

TABLE 1. List of IPCC global coupled climate models analyzed in the present study and their horizontal and vertical resolutions. Model resolution is characterized by the size of a horizontal grid on which model output was available, and by the number of vertical levels. Spectral models are also characterized by their spectral truncations. Equilibrium climate sensitivity is provided where available.

Model label and climate sensitivity	Resolution	Institution and reference
CGCM3.1(T47) 3.6 K	96 × 48 L32 T47	Canadian Centre for Climate Modelling and Analysis (http://www.cccma.ec.gc.ca/models/cgcm3.shtml)
CGCM3.1(T63) 3.4 K	128 × 64 L32 T63	Canadian Centre for Climate Modelling and Analysis (http://www.cccma.ec.gc.ca/models/cgcm3.shtml)
CNRM-CM3 n/a	128 × 64 L45 T63	Centre National de Recherche Météorologique, France (Salas-Mélia et al. 2006, manuscript submitted to *Climate Dyn.*)
ECHAM5/MPI-OM 3.4 K	192 × 96 L31 T63	Max-Planck-Institut für Meteorologie, Germany (Jungclaus et al. 2006)
ECHO-G 3.2 K	96 × 48 L19 T30	Meteorological Institute of the University of Bonn, Germany, Meteorological Research Institute, South Korea (Min et al. 2005)
GFDL-CM2.0 2.9 K	144 × 90 L24	Geophysical Fluid Dynamics Laboratory (Delworth et al. 2006; Gnanadesikan et al. 2006)
GFDL-CM2.1 3.4 K	144 × 90 L24	Geophysical Fluid Dynamics Laboratory (Delworth et al. 2006; Gnanadesikan et al. 2006)
GISS-AOM n/a	90 × 60 L12	Goddard Institute for Space Studies Laboratory (Russell et al. 1995; http://aom.giss.nasa.gov)
GISS-ER 2.7 K	72 × 46 L20	Goddard Institute for Space Studies Laboratory (Schmidt et al. 2006; Russell et al. 2000)
INM-CM3.0 2.1 K	72 × 45 L21	Institute of Numerical Mathematics, Russia (Diansky and Volodin 2002)
IPSL-CM4.0 4.4 K	96 × 72 L19	Institut Pierre-Simon Laplace, France (http://dods.ipsl.jussieu.fr/omamce/IPSLCM4/DocIPSLCM4)
MIROC3.2(hires) 4.3 K	320 × l60 L56 T106	Center for Climate System Research, Japan (Hasumi and Emori 2004)
MIROC3.2(medres) 4.0 K	128 × 64 L20 T42	Center for Climate System Research, Japan (Hasumi and Emori 2004)
MRI-CGCM2.3.2 3.2 K	128 × 64 L30 T42	Meteorological Research Institute, Japan (Yukimoto et al. 2001, 2006)
NCAR-CCSM3 2.7 K	256 × l28 L26 T85	National Center for Atmospheric Research (Collins et al. 2006)
NCAR-PCM 2.1 K	128 × 64 L26 T42	National Center for Atmospheric Research (Washington et al. 2000; Meehl et al. 2006)

effect on surface temperature (e.g., Ramanathan et al. 2001).

The CGCMs that we analyzed are listed in Table 1 together with their horizontal grid resolutions and the number of vertical levels in the corresponding atmospheric components. Spectral atmospheric models are also characterized by the spectral type and truncation. Model output was available on a variety of grids with resolution ranging from 72 × 45 to 320 × 160, with the median resolution being 128 × 64. The vertical resolution varies from 12 levels to 56 levels with the median of 26 levels. Table 1 also lists estimates of equilibrium climate sensitivities compiled from the PCMDI IPCC model documentation Web site (http://www-pcmdi.llnl.gov/ipcc/model_documentation/ipcc_model_documentation.php and references therein). Equilibrium climate sensitivity is defined as the global surface air temperature change under CO_2 doubling in slab ocean experiments and ranges from 2.1 K in the Institute of Numerical Mathematics Coupled Model version 3.0 (INM-CM3.0) and National Center for Atmospheric Research (NCAR) Parallel Climate Model (PCM) to 4.3 K and larger in the L'Institut Pierre-Simon Laplace Coupled Model version 4 (IPSL-CM4) and Model for Interdisciplinary Research on Climate 3.2, high-resolution version [MIROC3.2(hires)]. A number of modeling groups submitted daily output from several ensemble members per scenario. These models will be identified when the results of the extreme value analysis are presented in the next sections.

Daily precipitation and daily temperature output for all three scenarios was not available for all models listed in Table 1. Daily model output from the A2 experiment was not available for two models: the Goddard Institute for Space Studies (GISS) Atmosphere–Ocean Model (AOM) and the MIROC3.2(hires). Daily temperature extremes were not available for the NCAR Community Climate System Model version 3 (CCSM3). Daily temperature output from the NCAR-PCM model was excluded from the analysis because daily maximum and minimum temperature extremes appear to be (erroneously) identical in 1981–2000. In total, daily model output for years 1981–2000 was available from 14 models for temperature and from 16 models for precipitation.

To ensure consistency of the results for all three sce-

narios and to minimize possible effects of different multimodel ensembles on the multimodel mean response, the analysis of changes in climate extremes is performed only for models for which daily model output was available for all three emission scenarios. As a result, analysis of changes in precipitation extremes was performed for 14 models [all models in Table 1 except for GISS AOM and MIROC3.2(hires)]. Changes in temperature extremes are analyzed for 12 models (excluding also NCAR-CCSM3 and NCAR-PCM). For completeness, the analysis was repeated for all available models, but the conclusions of the study remained essentially unaffected.

Several diagnostics describing simulated 1981–2000 climate extremes are compared to those derived from four reanalyses. The two older reanalyses are the National Centers for Environmental Prediction (NCEP)–NCAR reanalysis (Kalnay et al. 1996) denoted hereafter as NCEP1, and the 15-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-15: Gibson et al. 1997). The two more recent ones are the NCEP–Department of Energy (DOE) AMIP-II reanalysis (Kanamitsu et al. 2002), denoted as NCEP2, and 40-yr ECMWF Re-Analysis (ERA-40; Simmons and Gibson 2000). We also performed an analysis of annual extremes of nonoverlapping 5-day mean precipitation rates (pentads), and used for verification the Climate Prediction Center (CPC) Merged Analysis of Precipitation (CMAP) pentad dataset that is a blend of gauge observations, satellite observations, and precipitation fields from the NCEP–NCAR reanalysis (Xie et al. 2003). These are essentially the same validation sources that are used in the recent atmospheric model intercomparison study by Kharin et al. (2005) but updated for the 1981–2000 period whenever possible.

3. Methodology

Climate extremes are multifaceted meteorological phenomena and can be characterized in terms of intensity, frequency, or duration of one or more climatological parameters. To address the multitude of possible extreme value statistics, the WCRP/WGCM also requested that modeling groups submit a number of extremes indices, as described in Frich et al. (2002). These indices are not analyzed here but are the subject of several other diagnostic subprojects (http://www-pcmdi.llnl.gov/ipcc/diagnostic_subprojects.php; e.g., Tebaldi et al. 2006).

Here we follow the approach of Zwiers and Kharin (1998), Kharin and Zwiers (2000), and Kharin et al. (2005) and analyze extremes of surface air temperature and precipitation in terms of return values, or return levels, of their annual extremes. Note that there seems to be no universally agreed definition of return values. A conventional but somewhat loose definition of a $\mathcal{T}$-year return level as the level that is exceeded on average every $\mathcal{T}$ years is problematic in a nonstationary environment. We more precisely define a $\mathcal{T}$-year return value as the threshold that is exceeded by an annual extreme in any given year with the probability $p = 1/\mathcal{T}$, where $\mathcal{T}$ is expressed in years. In particular, a 20-yr return value is the level that an annual extreme exceeds with probability $p = 5\%$. The quantity $\mathcal{T} = 1/p$ indicates the "rarity" of an extreme event and is usually referred to as the return period, or the waiting time for an extreme event.

Return values defined as above are essentially the quantiles of a distribution of annual extremes and are estimated from a generalized extreme value (GEV) distribution fitted at every grid point to samples of annual temperature and precipitation extremes. The "three type" GEV distribution comprises the three classical asymptotic extreme value models, Gumbel, Frèchet, and Weibull (Jenkinson 1955). Its three parameters, location, scale, and shape, are estimated by the robust method of "L-moments" (Hosking 1990, 1992), also known as the method of probability-weighted moments, with the minor modification of Dupuis and Tsao (1998) to ensure the feasibility of the parameter estimates (i.e., to ensure that all observed or simulated annual extremes are in fact permitted by the estimated GEV distribution). This method of return value estimation is well documented in the aforementioned studies and is therefore not presented here.

We note that the GEV distribution theory is valid only asymptotically, that is, when extremes are drawn from increasingly larger samples. In the present study, annual extremes are drawn from samples of size 365 (or 366 for leap years). However, serial correlation and the presence of an annual cycle may substantially reduce the effective sample size. Therefore, it is imperative to evaluate whether the asymptotic GEV distribution provides a reasonable description of the behavior of a sample of observed annual extremes by performing goodness-of-fit tests. We routinely conduct standard Kolmogorov–Smirnov goodness-of-fit tests (Stephens 1970) that measure the overall difference between the empirical and fitted cumulative distributions for all available samples. These tests indicate that a GEV distribution is generally a reasonable approximation for a distribution of annual extremes of the considered variables in most models. The goodness-of-fit is diminished for annual precipitation extremes in extremely dry regions in some models, most notably in IPSL-CM4.0. The GEV fit is also somewhat problematic for annual

precipitation extremes in the Tropics in both GFDL models. Tropical annual precipitation extremes in these two models exhibit a somewhat intermittent behavior when more moderate annual extremes in some years are alternated with very large values in other years. As an additional check, we routinely estimate empirical quantiles of annual extremes for moderate return periods and compare them to the corresponding L-moment return value estimates. In most cases regionally averaged empirical and parametric return value estimates compare reasonably well and are not overly too different even in situations where a GEV fit appears to be problematic.

The choice of the L-moment method over the frequently used method of maximum likelihood for estimating the parameters of a GEV distribution is primarily dictated by relatively short 20-yr samples as are available for analysis. The standard maximum likelihood estimator is less efficient than the L-moment estimator in short samples for typical values of the shape parameter (Hosking et al. 1985). Coles and Dixon (1999) argue that this is mainly due to unreliable estimates of the shape parameter that translates to poor performance for return values. There have been efforts to improve the efficiency of the maximum likelihood estimator. For example, Martins and Stedinger (2000) propose a generalized maximum likelihood analysis by specifying a geophysical prior distribution to restrict the shape parameter to a physically plausible interval within a Bayesian framework. Coles and Dixon (1999) modify the likelihood function by introducing a penalty term to restrict the shape parameter values to the range for which the GEV distribution has finite mean. Both approaches require user decisions about the specification of the prior distribution or the weight and form of the penalty term. The benefits of these, more general and potentially more powerful but also somewhat more complex techniques, do not override, in our opinion, the simplicity of the L-moment method in the present setting.

A potential drawback of the L-moment method in a transient climate change setting is that it assumes the stationarity of annual extremes. Kharin and Zwiers (2005) demonstrated that the violation of this assumption may introduce bias in return value estimates that is comparable to sampling variance. Their finding was based on three-member ensemble simulations with a single CGCM, but its significance is diminished for the present multimodel study. First, as will be demonstrated further on, sampling errors in local return value estimates for moderate return periods are generally smaller than discrepancies between individual models and therefore do not represent the main source of uncertainty. Second, the bias is minor as compared to sampling variance when return values are estimated from short 20-yr samples from a single realization that are available for the majority of models in the present study. Third, the short sample size prohibits the use of more complex statistical models with time-varying GEV distribution parameters, as was done by Kharin and Zwiers (2005). Such models can be fitted with the maximum likelihood method but are less competitive than models with constant parameters in short samples. Any benefits that might be gained in reducing the bias by employing a more complex statistical model are likely to be offset by increased sampling variance. Overall, the L-moment method appears to be an appropriate and viable technique for the task in the present setting.

Alternatives to the annual extremes approach include peak-over-threshold techniques based on a generalized Pareto distribution, and r-largest extremes analysis with a GEV distribution (e.g., Palutikof et al. 1999; Zhang et al. 2004). Successful implementation of these methods generally requires more decisions from the user (e.g., declustering of extremes, specification of a sufficiently large threshold, dealing with the annual cycle, etc.). Thus applying these techniques in an automated manner in a multimodel ensemble setting across a variety of very different climatological zones is a rather difficult task. The main argument for using one of these alternative techniques is that they may use the available information more efficiently, which could potentially result in more accurate return value estimates. However, as will be demonstrated further on, sampling errors are not the main source of uncertainty in the multimodel/multiscenario setting. We therefore do not consider the use of other methods in the present study.

Most of the analysis that follows is performed for the return period of 20 yr, or equivalently, for the exceedance probability by annual extremes of 5%. Longer return periods, such as 50 yr (exceedance probability of 2%), or even 100 yr (exceedance probability of 1%), are less advisable given the relatively short 20-yr samples and considering the fact that only one climate simulation was available for each emission scenario for most models. Estimating return levels for very long return periods is prone to larger sampling errors and potentially larger biases due to inexact knowledge of the shape of the tails of a distribution of annual extremes.

The GEV distribution methodology also allows us to examine changes in the exceedance probability of events of a certain size. In particular, we examine projected changes in the exceedance probability p of late-twentieth-century 20-yr return levels and express these changes in terms of changes in waiting times $\mathcal{T} = 1/p$.

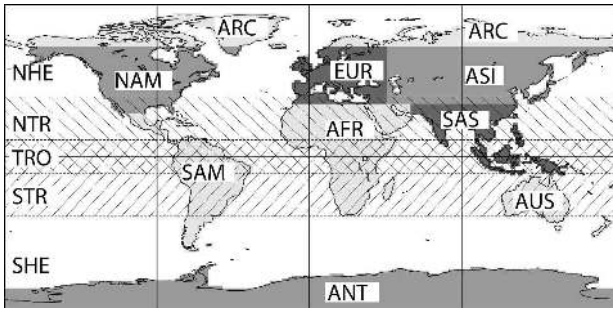

FIG. 2. Continent-wide regions and zonal bands considered in the present study. The coordinates of the regions are given in Table 2.

TABLE 2. Coordinates of continental-scale regions, as described in Fig. 2.

Region	Label	Latitudes	Longitudes
Global scale			
Globe	GLB	180° to 180°	90°S–90°N
Land	LND	180° to 180°	90°S–90°N
Zonal bands			
NH extratropics	NHE	180° to 180°	35°–90°N
SH extratropics	SHE	180° to 180°	90°–35°S
Tropics	TRO	180° to 180°	10°S–10°N
NH subtropics	NTR	180° to 180°	10°–35°N
SH subtropics	STR	180° to 180°	35°–10°S
Subcontinents			
Africa	AFR	20°W–60°E	40°S–30°N
Central Asia	ASI	45°E–180°	30°–65°N
Australia	AUS	105°E–180°	45°–10°S
Europe	EUR	20°W–45°E	30°–65°N
North America	NAM	165°–30°W	25°–65°N
South America	SAM	115°–30°W	55°S–25°N
South Asia	SAS	60°–160°E	10°S–30°N
Arctic	ARC	180° to 180°	65°–90°N
Antarctic	ANT	180° to 180°	90°–65°S

For example, we anticipate that late-twentieth-century warm extremes will generally be exceeded more frequently in a warmer climate, and therefore their waiting times will decrease, while waiting times for occurrences of late-twentieth-century cold extremes will increase.

Return values of cold and warm annual temperature extremes, and of annual 24-h precipitation extremes, are estimated for each model on its native grid. The resulting statistics are then interpolated onto a common 256×128 Gaussian grid for averaging and intercomparison purposes. Regionally averaged extreme value statistics are evaluated for a number of extratropical, subtropical, and tropical zonal bands, and also the continental regions displayed in Fig. 2 and defined in Table 2. The purpose of spatial averaging in the present study is twofold: 1) to reduce sampling errors and perhaps reduce some uncertainties associated with modeling errors at local scales and 2) to provide a condensed summary of typical and regionally representative amplitudes of extreme events, their uncertainties, and possible future changes.

There are no known analytical expressions for calculating standard errors and confidence intervals of the L-moment estimates, similar to those available for the maximum-likelihood estimates. We therefore use the nonparametric bootstrap (Efron and Tibshirani 1993), a resampling technique that allows us to estimate the uncertainties in return values that result from in-sample variability. For each model or observational dataset, 1000 bootstrap samples are generated by randomly sampling with replacement global fields of annual extremes from the original dataset. Global fields are resampled to preserve possible spatial dependencies. Local return values and their regional averages are calculated for each bootstrap sample. The resulting collection of 1000 resampled statistics is used to derive bootstrap confidence intervals.

The L-moment return value estimates based on short samples are slightly biased, even when the stationarity assumption is satisfied (Hosking et al. 1985). The bias is generally negligible when compared to standard errors of local estimates but becomes noticeable when local estimates are averaged over large regions so that sampling variance is greatly reduced. As a result, the bootstrap distribution of regionally averaged return values is not centered at the return value estimate obtained for the original sample. In the following we corrected regionally averaged return values for the bootstrap estimate of bias, defined as the mean of resampled statistics minus the statistic for the original sample, when presenting regionally averaged statistics.

4. Simulated late-twentieth-century climate extremes

We start the analysis of temperature and precipitation extremes by documenting their present-day climatologies. For space reasons, we are not able to display individual maps of extremes for each of the models analyzed in this study. Instead we limit the presentation by showing zonally and regionally averaged statistics simulated by individual models, together with maps of the multimodel ensemble mean and a measure of the discrepancy between models expressed in terms of the standard deviation about the multimodel mean.

a. Temperature extremes

Zonally averaged 20-yr return values of 1981–2000 annual warm and cold extremes simulated by 14 IPCC

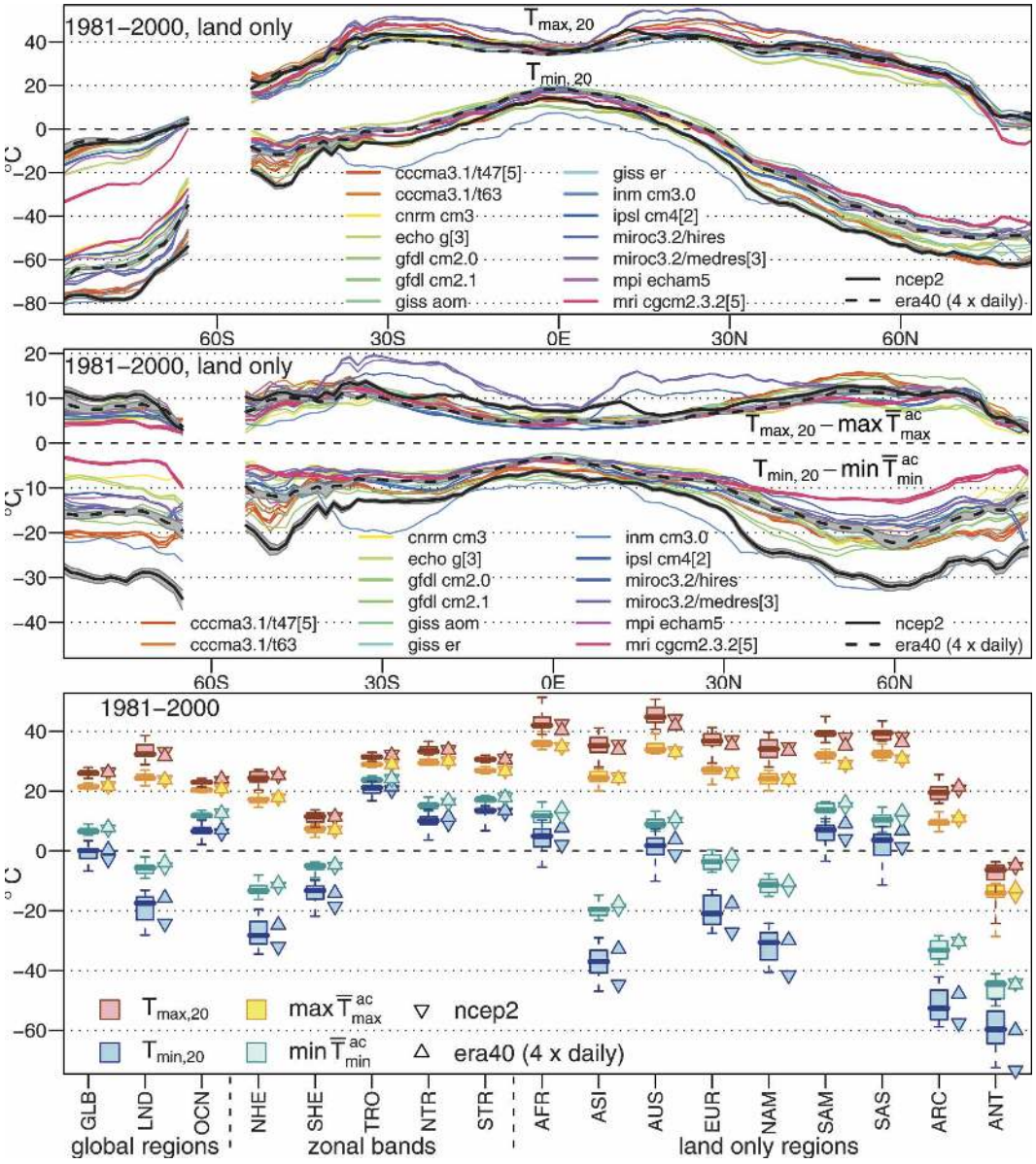

FIG. 3. (top) Zonally averaged 1981–2000 $T_{max,20}$ and $T_{min,20}$ as simulated over land by 14 IPCC AR4 models. Several models are represented by several climate simulations, one curve for each ensemble member. The ensemble size is indicated in brackets after the model labels. The NCEP2 and ERA-40 estimates are displayed in black together with the 95% bootstrap confidence intervals in gray. (middle) The difference between zonally averaged 1981–2000 temperature extremes $T_{max,20}$ and $T_{min,20}$ and the corresponding extremes of the annual cycle max $\overline{T}^{ac}_{max}$ and min $\overline{T}^{ac}_{min}$. (bottom) Boxplots of regionally averaged 1981–2000 $T_{max,20}$, $T_{min,20}$, max $\overline{T}^{ac}_{max}$, and min $\overline{T}^{ac}_{min}$. Boxplots indicate the central 50% intermodel range, the median, and the lower and upper bounds. The downward- and upward-pointing triangles represent the regionally averaged statistics estimated from NCEP2 and ERA-40, respectively.

models over land and the corresponding estimates from the NCEP2 and ERA-40 reanalyses are displayed in Fig. 3 (top). The model results are represented by colored curves, one curve for each ensemble member if there is more than one ensemble realization. The ensemble size is indicated in brackets after the model name in the legends. In principle, all ensemble members could have been concatenated together into one longer sample from which more accurate estimates of return values could have been obtained. Here, we plotted the return values for each ensemble member separately to get an idea of the uncertainty that arises due to

the interannual variability of annual extremes in 20-yr samples, as compared to model-to-model differences. The NCEP2 and ERA-40 reanalyses are represented by the black solid line and dashed line curves, respectively. We also display in gray the 95% bootstrap confidence intervals for the zonally averaged estimates of 20-yr return values for the two reanalyses.

Sampling errors do not appear to play a significant role in the uncertainty of zonally averaged estimates of 20-yr return values. The corresponding bootstrap confidence intervals are very narrow in comparison to the differences between individual models or reanalyses (the width of the confidence intervals in Fig. 3 is only marginally larger than the thickness of the curves). This is also supported by the fact that the curves obtained for individual ensemble members lay nearly on top of each other for models with more than one realization. The latter indicates that sampling errors are generally small and that possible natural variability on decadal and longer time scales has only a small effect on the amplitude of return values, at least for the zonally averaged return value statistics in the models considered.

The two reanalyses agree fairly well on the magnitude of zonally averaged warm extremes. The NCEP2 warm extremes tend to be only slightly warmer than the corresponding ERA-40 extremes over landmasses. However, NCEP2 cold extremes are much colder than their ERA-40 counterparts in many regions, by as much as 15°C and more, and are colder than those simulated by the majority of the models. Note that the ERA-40 temperature extremes are derived from data that are sampled every 6 h (4 times daily). Kharin et al. (2005) found that this coarser temporal resolution does not seriously compromise the accuracy of return values of annual temperature extremes. In particular, zonally averaged 20-yr return values of NCEP2 annual temperature extremes calculated from 6-hourly sampled data (not shown here) nearly coincide with the estimates based on the original diurnal temperature extremes. The discrepancies between the reanalyses are therefore unlikely to be due to the difference in temporal resolution of two datasets.

Similar to the AMIP2 study (Kharin et al. 2005), cold extremes are generally less reliably simulated by the models than warm extremes. The discrepancies between the models (and between the reanalyses) are generally larger for cold extremes than for warm extremes. However, there are some exceptions. In particular, there is a substantial warm bias in the warm extremes in subtropical regions in the Model for Interdisciplinary Research on Climate 3.2, medium-resolution version [MIROC3.2(medres)], and to a lesser degree in MIROC3.2(hires) and Canadian Centre for Climate Modelling and Analysis (CCCMA) CGCM3.1. There is also a large cold bias over Antarctica in the Meteorological Research Institute (MRI) CGCM2.3.2 model.

Some, but not all, of the biases can be attributed to differences in the model climatologies. Since warm and cold extremes tend to occur during the time of year when mean temperatures are the warmest or coldest, respectively, we examine the differences between warm and cold extremes relative to the respective warm or cold climatological mean temperatures. The middle panel of Fig. 3 displays the difference between the zonally averaged warm and cold extremes displayed in the upper panel, and the corresponding maximum and minimum of the climatological annual cycle, denoted as $\max \overline{T}^{\mathrm{ac}}_{\mathrm{max}}$ and $\min \overline{T}^{\mathrm{ac}}_{\mathrm{min}}$., respectively. The annual cycle is defined as the 1981–2000 average of monthly T_{max} or T_{min} for each calendar month. The magnitude of deviations from zero indicates the extent to which temperature extremes deviate from the mean temperature conditions in individual models and reanalyses.

There is better agreement between models and reanalyses with respect to such deviations for warm extremes than for cold extremes, except for MIROC3.2 in subtropical regions. Differences among models and reanalyses are largest over snow and sea ice–covered regions. Most notably, temperature occasionally deviates farther below the climatological mean temperature in NCEP2 as compared to ERA-40 or most models. The NCEP2 reanalysis (and the older NCEP1 reanalysis; not shown here) is somewhat exceptional in this regard (Kharin et al. 2005).

A boxplot summary of regionally averaged cold and warm extremes is displayed in the bottom panel of Fig. 3. The regions are defined in Table 2 and displayed in Fig. 2. Boxplots indicate the central 50% intermodel range (25th–75th percentiles), the median, and the lower and upper bounds in the multimodel ensemble. Generally speaking, simulated warm extremes compare well with the reanalyses on the considered regional scales, although the models have a tendency for a warm bias in tropical and subtropical regions in the models. Consistent with the findings above, regional differences are larger for cold extremes, both among the models and among the two reanalyses.

Figure 4 displays the multimodel ensemble mean of warm and cold extremes, and the differences between the ensemble mean extremes and the corresponding temperature extremes in NCEP2 and ERA-40. Warm extremes tend to be slightly colder in the models, on average, than in the reanalyses over oceans in the Northern Hemisphere but slightly warmer over South and Central America, North Africa, the Middle East,

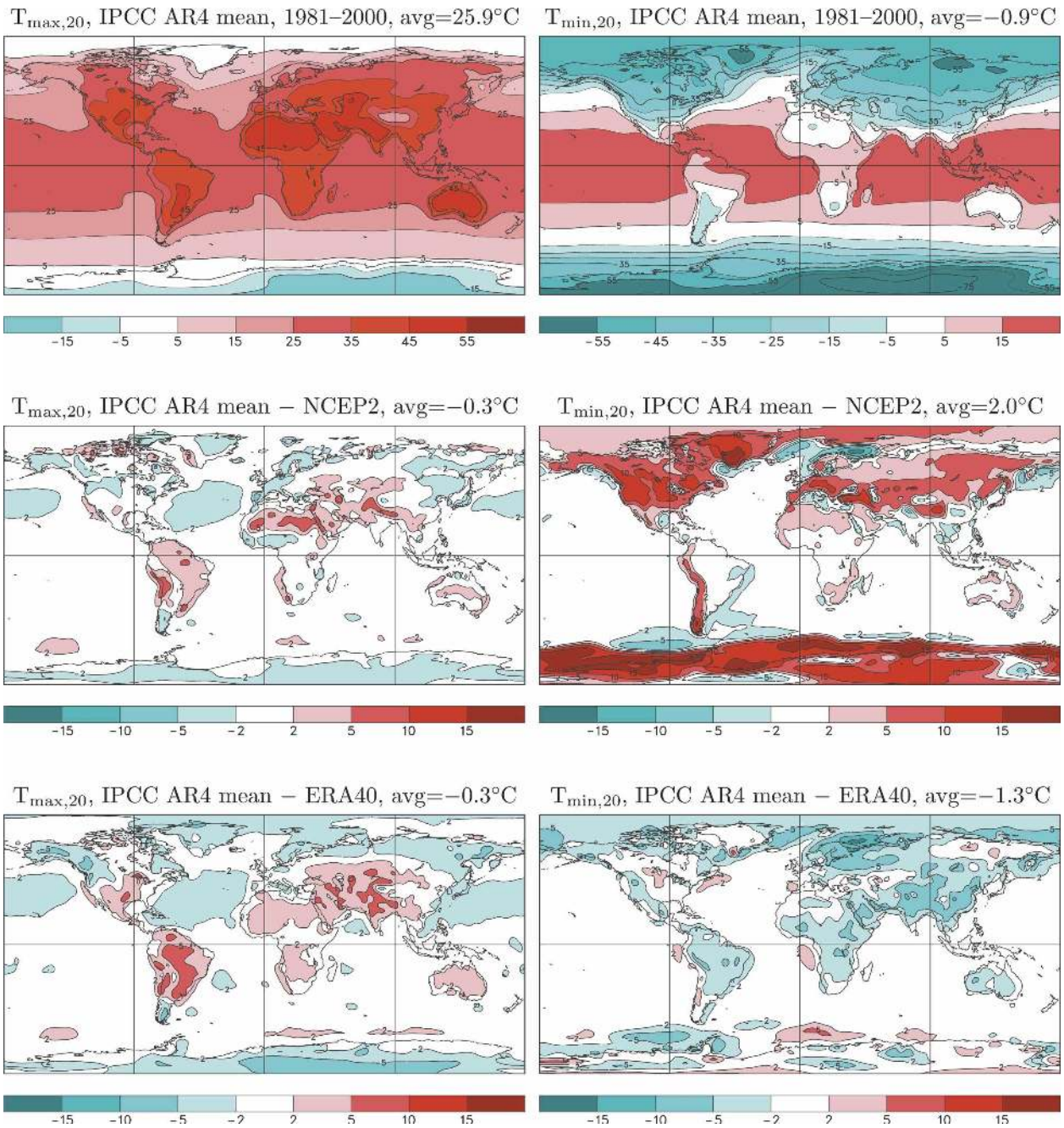

FIG. 4. (top) The multimodel ensemble mean average of 20-yr return values of 1981–2000 (left) annual maximum temperature ($T_{max,20}$) and (right) annual minimum temperature ($T_{min,20}$) as simulated by 14 IPCC AR4 models. (middle) The difference between the multimodel ensemble means of $T_{max,20}$ and $T_{min,20}$ and the corresponding temperature extremes estimated from NCEP2. (bottom) The difference between the multimodel ensemble mean of $T_{max,20}$ and $T_{min,20}$ and the corresponding extremes estimated from ERA-40. Units are °C. Global averages are indicated in the titles.

and Central Asia. Cold extremes are well simulated over ice free oceans, as compared to the reanalyses but, as noted previously, NCEP2 cold extremes are more severe over land and sea-ice-covered regions than in the models and in ERA-40.

Globally and land-only averaged statistics are summarized in Table 3. The multimodel mean of globally averaged simulated 1981–2000 warm extremes corresponds closely to that in NCEP2 and ERA-40. The multimodel mean of globally averaged cold extremes is

TABLE 3. The multimodel ensemble mean average of 20-yr return values of annual maximum and minimum temperature ($T_{max,20}$ and $T_{min,20}$) and the corresponding maximum and minimum of the annual cycle, max $\overline{T}^{ac}_{max}$ and min $\overline{T}^{ac}_{min}$, averaged over the globe and land only as simulated by 14 IPCC AR4 models in 1981–2000 in the twentieth-century experiment (20C3M) and in the NCEP2 and ERA-40 reanalyses. The central 50% intermodel range is displayed to the right of the ensemble mean value.

	$T_{max,20}$ (°C)		$T_{min,20}$ (°C)		max$\overline{T}^{ac}_{max}$ (°C)		max$\overline{T}^{ac}_{min}$ (°C)	
	Globe	Land	Globe	Land	Globe	Land	Globe	Land
20C3M	$26.0^{26.8}_{25.3}$	$33.2^{35.2}_{31.7}$	$-0.9^{+0.4}_{-2.5}$	$-18.7^{-15.6}_{-22.8}$	$21.5^{21.8}_{21.0}$	$24.5^{25.3}_{23.7}$	$6.6^{7.4}_{5.8}$	$-6.0^{-4.7}_{-7.5}$
NCEP2	26.3	33.1	−2.9	−24.3	21.7	23.7	7.3	−5.5
ERA-40	26.2	31.6	0.4	−15.8	21.6	23.8	7.8	−3.9

warmer than in NCEP2 but colder than in ERA-40. The discrepancies amongst the models and reanalyses are substantially smaller for warm and cold mean temperatures than for extreme temperatures.

Figure 5 summarizes intermodel differences of local 20-yr return value estimates and the estimated return value sampling standard errors. The upper two panels display the intermodel standard deviation of simulated warm and cold extremes. Model differences are larger over land and sea ice than over ice free oceans and are generally larger for cold extremes than for warm extremes, particularly over snow-covered regions. The bottom two panels display the estimate of sampling standard errors of local 20-yr return values obtained as the multimodel average of standard deviations of 1000 bootstrap resamples obtained for each model. Sampling errors are generally larger for cold extremes than for warm extremes due to generally larger interannual vari-

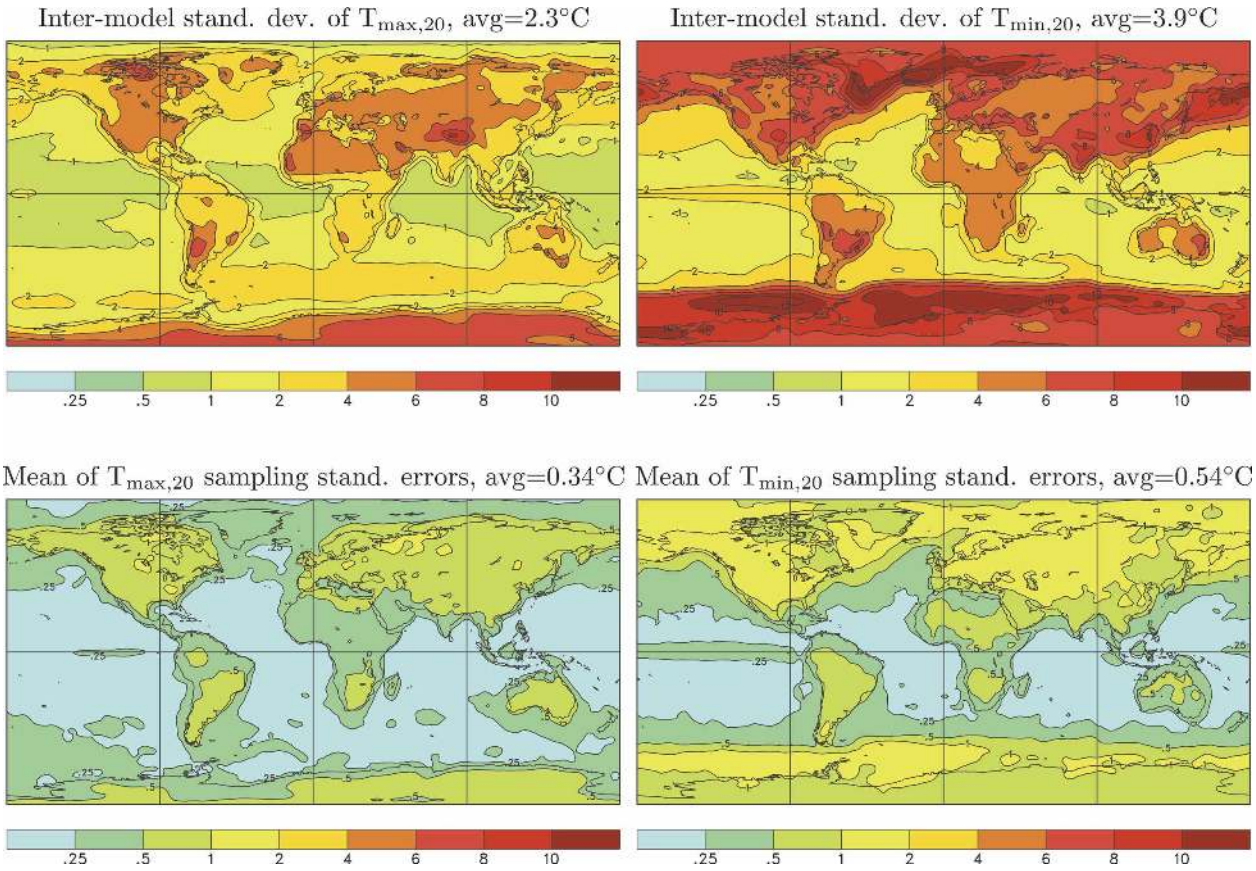

FIG. 5. (top) The intermodel standard deviation of 1981–2000 (left) $T_{max,20}$ and (right) $T_{min,20}$ as simulated by 14 IPCC AR4 models. (bottom) The multimodel ensemble mean of the bootstrap sampling standard errors of local (left) $T_{max,20}$ and (right) $T_{min,20}$. Units are °C. Global averages are indicated in the titles.

ability of cold temperatures. However, sampling errors constitute only a small fraction of the total uncertainty in local estimates of temperature extremes.

b. Precipitation extremes

The upper two panels of Fig. 6 display zonally averaged 20-yr return values of 1981–2000 annual extremes of 24-h precipitation amounts (P_{20}) and of nonoverlapping 5-day mean precipitation rates ($\overline{P}^5_{20}$) as simulated by 16 IPCC AR4 models and estimated from reanalyses and CMAP. The precipitation extremes are fairly consistently simulated in the moderate and high latitudes but much less so in the Tropics and subtropical regions. With the exception of the older NCEP1 reanalysis, the amplitude of precipitation extremes in other reanalyses in the Tropics is larger, zonally averaged, than in any of the models. Kharin et al. (2005) speculated that the weak tropical precipitation extremes in NCEP1 are perhaps not very trustworthy due to a known "spinup" deficiency for convection in the forecast model in that reanalysis.

The differences between simulated 5-day extremes are somewhat smaller than those between daily extremes but are still very large in tropical regions. The CMAP 5-day extremes are more moderate than those in the more recent reanalyses. Coincidently, the multimodel ensemble mean of 5-day precipitation extremes is closer to the CMAP extremes than to those estimated from any of the reanalyses. The 20-yr return values of annual 5-day precipitation extremes appear somewhat problematic in the CMAP dataset over Antarctica. The very large return value estimates are mainly caused by exceptionally large annual extremes in a single year, 1987, that are well in excess of 80 mm day^{-1} in some Antarctic regions, which perhaps points toward some extrapolation or other postprocessing problems associated with Antarctica's sparse observational network.

A number of models are represented by several curves in Fig. 6, one for each ensemble member. The ensemble size is indicated in brackets after the model label in the legends. We also show the 95% bootstrap confidence intervals derived from the observational datasets. It is evident that the differences in zonally averaged extremes between individual model realizations performed with the same model are generally much smaller than the differences between different models. The bootstrap confidence intervals that characterize sampling variability of zonally averaged 20-yr return values are also relatively narrow, as compared to the intermodel differences. The very wide confidence interval for the CMAP extremes over Antarctica is due to the aforementioned peculiarity in this dataset.

The bottom panel of Fig. 6 displays a boxplot summary of regionally averaged 24-h (in red) and 5-day (in blue) precipitation extremes plotted on a log scale. Symbols to the right of the boxplots indicate the corresponding regional statistics estimated from observationally based datasets. The height of the symbols corresponds to the 95% bootstrap confidence intervals of the regionally averaged observational estimates. In most cases, sampling errors of regionally averaged extremes are small compared to the corresponding intermodel differences. Not unexpectedly, model-to-model discrepancies (as indicated by the central 50% intermodel range) are generally smaller for 5-day precipitation extremes than for daily precipitation extremes. The intermodel uncertainties are relatively small for regions located well outside of the Tropics, such as Europe (EUR), North America (NAM), or North and Central Asia (ASI) but are much larger for tropical regions.

Multimodel mean P_{20} is displayed in the upper-left panel of Fig. 7, and the ratio of the multimodel mean extreme precipitation over that in ERA-40 is shown the upper-right panel. The ensemble mean amplitude of 20-yr return values of annual precipitation extremes is comparable to that in ERA-40 in the extratropics where departures from ERA-40 are generally within the $\pm 20\%$ range. The models simulate, on average, more intense precipitation extremes in the generally very dry regions of northern Africa and off the subtropical west coasts of Africa and North and South America. However, they simulate much weaker extremes in the narrow band along the equator. Similar features are present when the ensemble mean extreme precipitation is compared to NCEP2 (not shown), except that the maximum of tropical extreme precipitation in NCEP2 is broader than in ERA-40. Some regional statistics are summarized in Table 4.

The lower two panels of Fig. 7 display the magnitude of intermodel differences and the typical amplitude of sampling errors of P_{20} estimates derived from 20-yr samples. The estimated intermodel standard deviation of P_{20} displayed in the lower-left panel is normalized by the ensemble mean P_{20} There is better agreement between models in midlatitudes where intermodel standard deviations are about 20% of the ensemble mean amplitude. Differences amongst the simulated precipitation extremes are much larger in the Tropics and subtropical regions where they become comparable to the ensemble mean in some regions. The bootstrap sampling standard errors obtained for individual models for 1981–2000 are also normalized by the respective estimates of P_{20}. The lower-right panel of Fig. 7 shows the ensemble mean of such normalized standard errors, which are typically smaller than 10% in midlatitudes

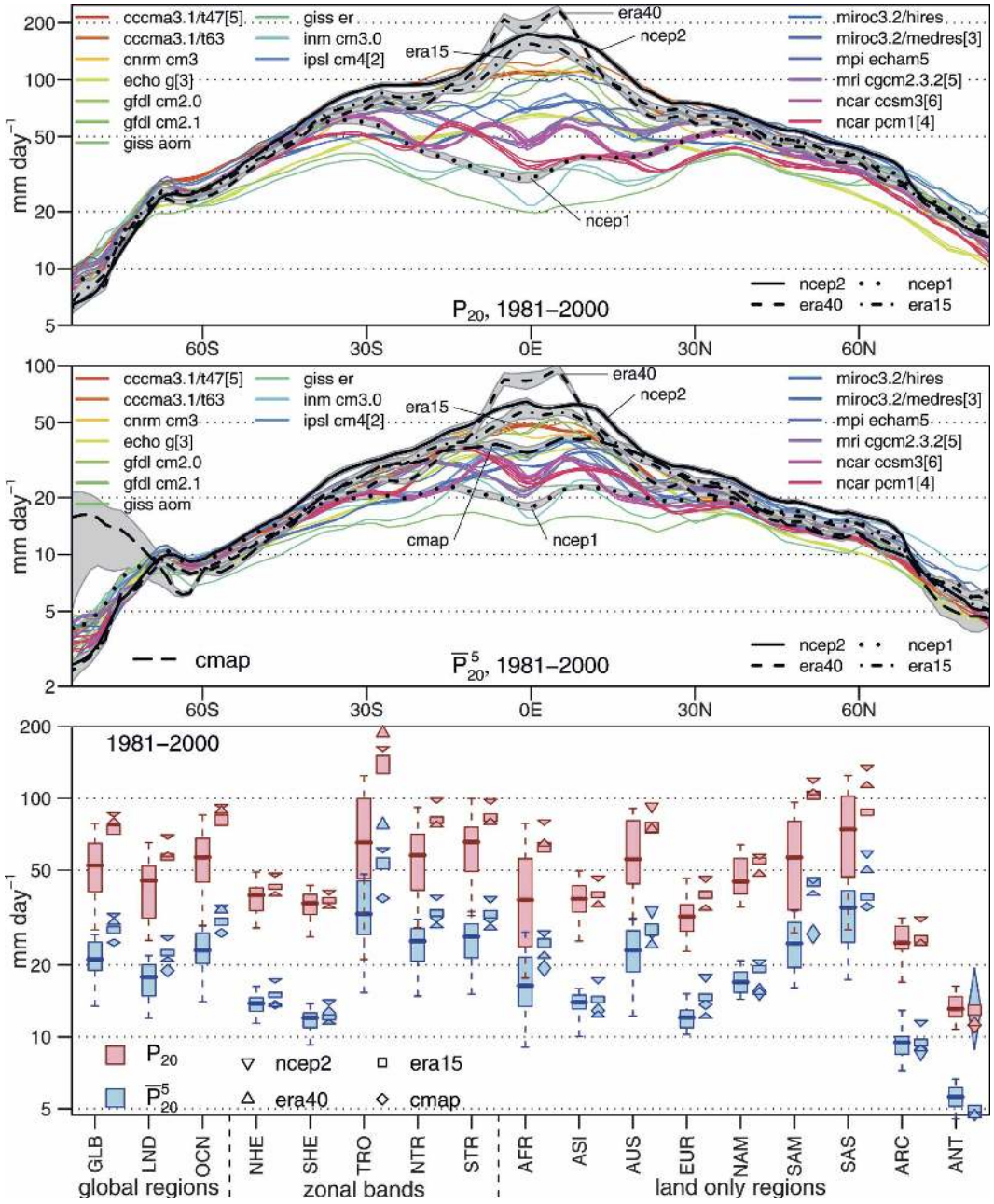

FIG. 6. Zonally averaged 20-yr return values of 1981–2000 annual extremes of (top) 24-h precipitation rates (P_{20}) and (middle) nonoverlapping 5-day mean precipitation rates ($\overline{P}^5_{20}$) as simulated by 16 IPCC AR4 models plotted on a log scale. Units are mm day^{-1}. Some models are represented by several ensemble members, one curve for each ensemble member. The ensemble size is indicated in brackets after the model labels. Precipitation extremes estimated from the reanalyses and CMAP pentad dataset are displayed in black together with the 95% bootstrap confidence intervals in gray. (bottom) Boxplots of simulated regionally averaged 1981–2000 P_{20} and $\overline{P}^5_{20}$. Symbols to the right of the boxplots indicate the corresponding statistics estimated from the reanalyses and CMAP pentad dataset. The height of the symbols corresponds to the 95% bootstrap confidence interval of the corresponding regional means.

and are slightly larger in tropical and polar regions. Overall, sampling variance is generally only a small fraction of total intermodel variability.

As in Kharin et al. (2005), the dependence of the magnitude of precipitation extremes on the spatial resolution in different models is found to be weak. In the extratropics, there is some evidence of somewhat stronger precipitation extremes in models with higher

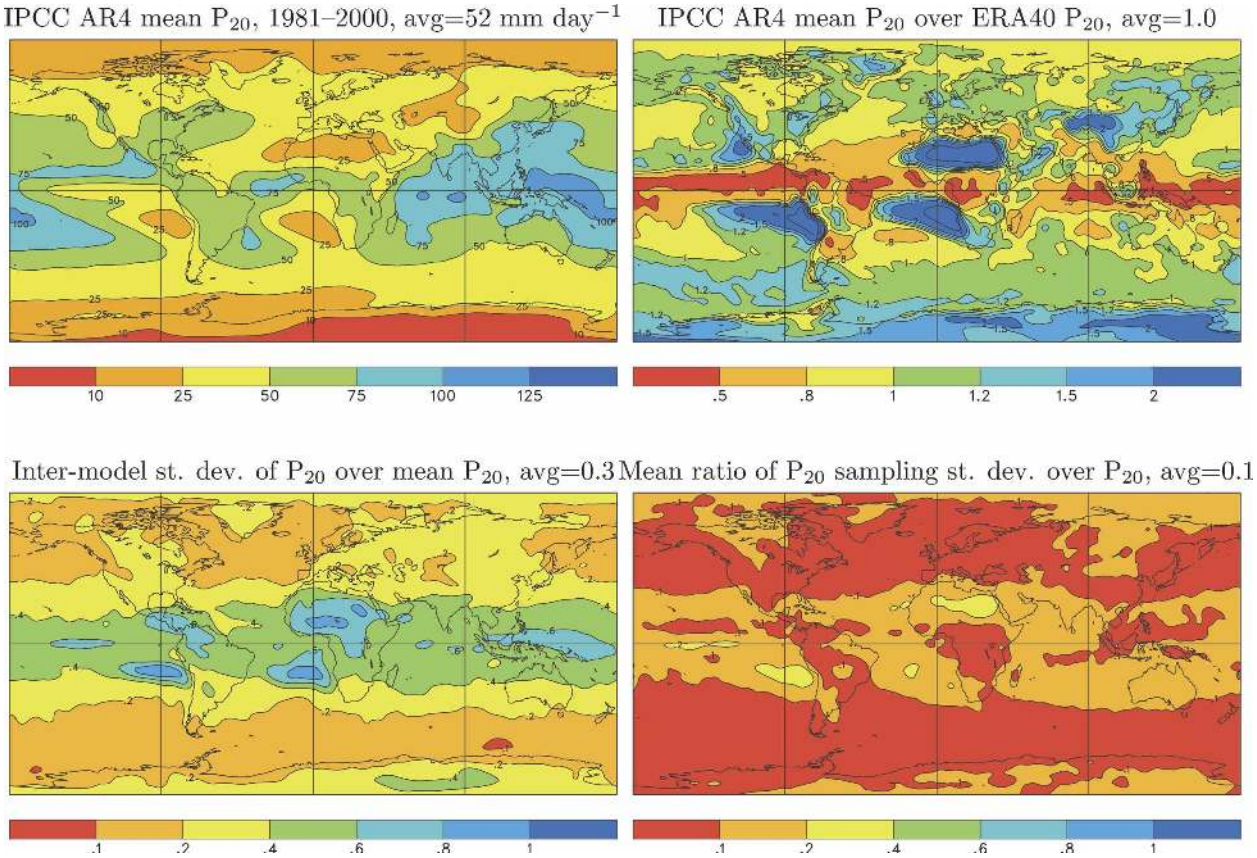

FIG. 7. (top left) The multimodel ensemble mean of 20-yr return values of 1981–2000 annual extremes of daily precipitation (mm day^{-1}) as simulated by 16 IPCC AR4 models. (top right) The ratio of the multimodel ensemble mean of P_{20} estimates over P_{20} estimated from ERA-40. (bottom left) The intermodel standard deviation of P_{20} estimates divided by their multimodel ensemble mean. (bottom right) The multimodel ensemble mean of the ratios of the bootstrap sampling standard deviations over the corresponding 1981–2000 P_{20} estimates for individual models. Global averages are indicated in the titles.

horizontal resolution. For example, the weakest extratropical extremes are simulated in the lower-resolution models, GISS-ER and GISS-AOM, while the strongest extremes are simulated in MIROC3.2(hires), which has the highest spatial resolution. The amplitude of extremes increases with resolution in simulations performed with models from the same modeling group. For example, the amplitude of precipitation extremes is about 15% larger, on average, in the simulation performed with the higher-resolution model CGCM3.1(T63), as compared to the lower-resolution version CGCM3.1(T47). A more dramatic increase in spatial resolution from T42 L20 in MIROC3.2(medres) to T106 L56 in MIROC3.2(hires) is accompanied by a

TABLE 4. The multimodel ensemble mean and the central 50% intermodel range of 20-yr return values of annual extremes of daily precipitation P_{20} (mm day^{-1}) and 5-day precipitation $\overline{P}^5_{20}$ (mm day^{-1}) averaged over the globe, land, the extratropical Northern Hemisphere (NHE; 35°–90°N), and the Tropics (TRO; 10°S–10°N) as simulated by 16 IPCC AR4 models in 1981–2000 in the twentieth-century experiment and the corresponding estimates from the NCEP2 and ERA-40 reanalyses and CMAP dataset.

	P_{20} (mm day^{-1})				$\overline{P}^5_{20}$ (mm day^{-1})			
	Globe	Land	NHE	TRO	Globe	Land	NHE	TRO
20C3M	$51.9^{64.7}_{40.6}$	$43.1^{52.1}_{31.5}$	$38.3^{42.2}_{33.9}$	$70.5^{100.1}_{46.1}$	$21.4^{25.0}_{19.0}$	$17.3^{20.0}_{14.8}$	$13.8^{14.6}_{12.8}$	$33.8^{45.0}_{26.8}$
NCEP2	82.9	66.8	46.5	155.8	31.3	25.1	16.7	58.6
ERA-40	77.8	56.4	38.7	184.6	29.6	20.5	13.3	75.8
CMAP	—	—	—	—	24.1	18.5	13.3	37.0

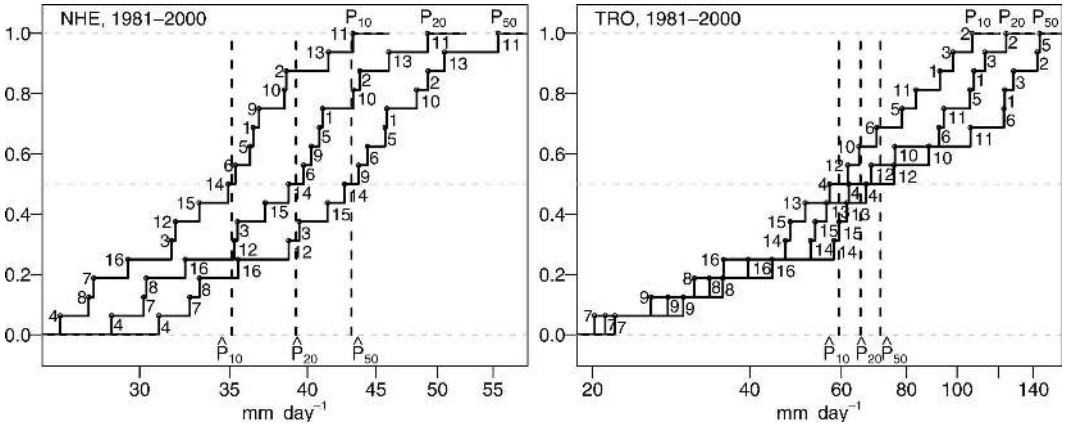

FIG. 8. Empirical cumulative distribution functions of regional estimates of 10-, 20-, and 50-yr return values of annual precipitation extremes averaged over (left) the northern extratropics (35°–90°N) and (right) the Tropics (10°S–10°N) as simulated by 16 IPCC AR4 models in 1981–2000 in the twentieth-century experiments. The *x* axis is on a log scale. The vertical dashed lines indicate the multimodel ensemble median values. Models are indicated by numbers as 1: CCCMA CGCM3.1/T47, 2: CCCMA CGCM3.1/T63, 3: Centre National de Recherches Météorologiques Coupled Global Climate Model version 3 (CNRM CM3), 4: ECHAM and the global Hamburg Ocean Primitive Equation (ECHO G), 5: Geophysical Fluid Dynamics Laboratory Climate Model version 2.0 (GFDL CM2.0), 6: GFDL CM2.1, 7: GISS AOM, 8: GISS ER, 9: INM CM3.0, 10: IPSL CM4, 11: MIROC3.2(hires), 12: MIROC3.2(medres), 13: Max Planck Institute (MPI) ECHAM5, 14: MRI CGCM2.3.2, 15: NCAR CCSM3, and 16: NCAR PCM1.

larger increase in extreme precipitation amplitude of about 40%.

On the other hand, models developed by different modeling groups do not necessarily confirm this tendency. A typical example is given by the reanalyses. Extratropical precipitation extremes are about 20% stronger in NCEP2 than in ERA-40 although the atmospheric component of NCEP2 has lower resolution (T62 L28) than ERA-40 (T159 L60). Note, however, that ERA-40 data were available on a lower-resolution 144 × 73 regular grid that was obtained by a bilinear interpolation from its higher-resolution "reduced" Gaussian 320 × 160 grid. This reduction in resolution is unlikely to explain the differences between the two reanalyses. We verified this by bilinearly interpolating NCEP2 precipitation from the 192 × 94 NCEP2 grid onto the 144 × 73 ERA-40 grid; return values were reduced just by a few percent.

Overall, it does appear that the amplitude of extreme precipitation increases with resolution, particularly, in models with similar representations of dynamical and physical processes. However, this dependence is not very robust across different models. In particular, there is no statistically significant dependence of precipitation extremes on the model resolution simulated by different models in the Tropics where the details of the deep convection parameterizations seem to be of dominant importance at the spatial resolutions considered (Scinocca and McFarlane 2004).

Figure 8 offers an alternative way to summarize the degree of disagreement between the models in simulating extreme precipitation in the Tropics and extratropics. It shows the empirical cumulative distribution functions of the regional estimates of 10-, 20-, and 50-yr return values of annual 24-h precipitation extremes in the northern extratropics (35°–90°N, left-hand diagram) and Tropics (10°S–10°N, right-hand diagram) simulated by 16 models. The empirical cumulative distribution function of *x* is defined as the fraction of models that simulate return values less than, or equal to, *x*. The multimodel cumulative distributions of return values for different return periods are better separated in the extratropics than in the Tropics, indicating better intermodel consensus on the exceedance probability of a specified level in the extratropics than in the Tropics. However, the overlap between the distributions is still fairly large, even in the extratropics, indicating that the exceedance probability of specified precipitation events is presently not very reliably determined by the models.

5. Future changes in extreme values

In this section we document future changes in temperature and precipitation extremes as simulated by the IPCC AR4 multimodel ensemble. A particular aspect of this analysis is that we compare changes in extremes to the corresponding changes in time mean climatologies. Some previous studies (e.g., Kharin and Zwiers

2005) indicate that changes in return values of simulated temperature extremes on a global scale are mainly associated with changes in the location of the distribution of annual extremes. Here, we compare changes in warm and cold temperature extremes to the corresponding changes in the maxima and minima of the annual cycle, that is, to changes in mean temperatures of the climatologically warmest and coldest seasons, respectively. Relative changes in extreme precipitation are compared to changes in annual mean precipitation.

Simulated changes in years 2046–65 and 2081–2100 are calculated relative to the 1981–2000 baseline period. To evaluate the statistical significance of changes in the multimodel ensemble, the climate change anomalies obtained for individual models are treated as a sample of random and independent realizations from a "population of models." We then performed two types of statistical tests: the Student's *t* test and its nonparametric alternative, the Wilcoxon signed–rank test (Wilcoxon 1945) that does not require assumptions about the form of the distribution. Both tests produced very similar results. Results on statistical significance presented below are based on the Wilcoxon test performed at the 10% significance level. Note that these statistical tests account both for sampling uncertainties of the extreme value statistics and for intermodel uncertainties.

a. Changes in temperature extremes

Figure 9 displays multimodel mean differences between 2046–65 and 1981–2000 20-yr return values of annual warm and cold extremes as simulated by the IPCC AR4 models in the SRES A1B experiment. The upper panels show absolute changes in $T_{\max,20}$ and $T_{\min,20}$. The middle panels display changes in extreme temperatures relative to the corresponding changes in the maxima and minima of the annual cycle, that is, $\Delta(T_{\max,20} - \max\overline{T}^{\mathrm{ac}}_{\max})$ and $\Delta(T_{\min,20} - \min\overline{T}^{\mathrm{ac}}_{\min})$. Positive values in these diagrams indicate that changes in extreme warm or cold temperatures exceed changes in the corresponding mean temperature of the warmest or coldest month of the year. The bottom two panels display the estimated probability in 2046–65 of exceeding the late-twentieth-century 20-yr return levels of annual warm and cold temperature extremes expressed in terms of waiting times. Only those changes that are significant at the 10% significance level according to the nonparametric Wilcoxon test are displayed in color.

Changes in warm and cold extremes are comparable over ice free oceans. The models tend to simulate somewhat larger increases in warm extremes than in cold extremes over subtropical land regions, most notably over the Iberian Peninsula and North Africa but also in South Africa, southwestern Australia, Central America, and central South America (Fig. 10). These are regions that become generally drier. Larger increases in warm extremes are presumably attributed to reduced moderation by evaporative cooling from the land surface. Cold extremes warm significantly faster over extratropical landmasses and over high-latitude oceans. The enhanced warming of cold extremes is apparently attributed to the positive snow and sea ice albedo feedback effect in these regions (see also, e.g., Zwiers and Kharin 1998; Kharin and Zwiers 2000, 2005). It is also evident that changes in warm extremes closely follow changes in the mean summertime temperature virtually everywhere over the globe. Globally averaged, warm extremes increase only a few hundredths of a degree Celsius more than the mean temperature in the climatologically warmest month. On the other hand, changes in cold extremes substantially exceed changes in the mean temperature in the climatologically coldest month in regions where snow and sea ice retreat with global warming.

Not surprisingly, there are substantial projected changes in the exceedance probability of warm and cold events that are considered as extreme at the end of the twentieth century. In particular, the exceedance probability of 20-yr return values of 1981–2000 annual warm extremes doubles in high latitudes and more than triples in more moderate latitudes over land (i.e., waiting times are reduced by a factor of 2–4). Late-twentieth-century warm extremes are exceeded virtually every year in 2046–65 in the lower latitudes. On the other hand, late-twentieth-century cold extremes become less frequent and are practically never exceeded over most of the globe by the middle of the twenty-first century under the A1B forcing scenario.

Figure 11 displays the multimodel mean changes in zonally averaged $T_{\max,20}$ and $T_{\min,20}$ as simulated with the B1 (blue curves), A1B (green curves), and A2 (red curves) emission scenarios. The 2046–65 changes are indicated by the dashed-line curves while 2081–2100 changes are displayed as the solid-line curves. The upper two panels show absolute changes in warm and cold temperature extremes, while the lower two panels display their changes relative to the changes in mean temperature of the climatologically warmest or coldest months of year, respectively. The central 50% intermodel range of zonally averaged changes is also shown for the A1B experiment in both time periods by light green cross-hatching.

As expected, the smallest warming is simulated in the B1 experiment, which has the slowest growth of the greenhouse gas concentrations. The midcentury responses in the A1B and A2 experiments are compa-

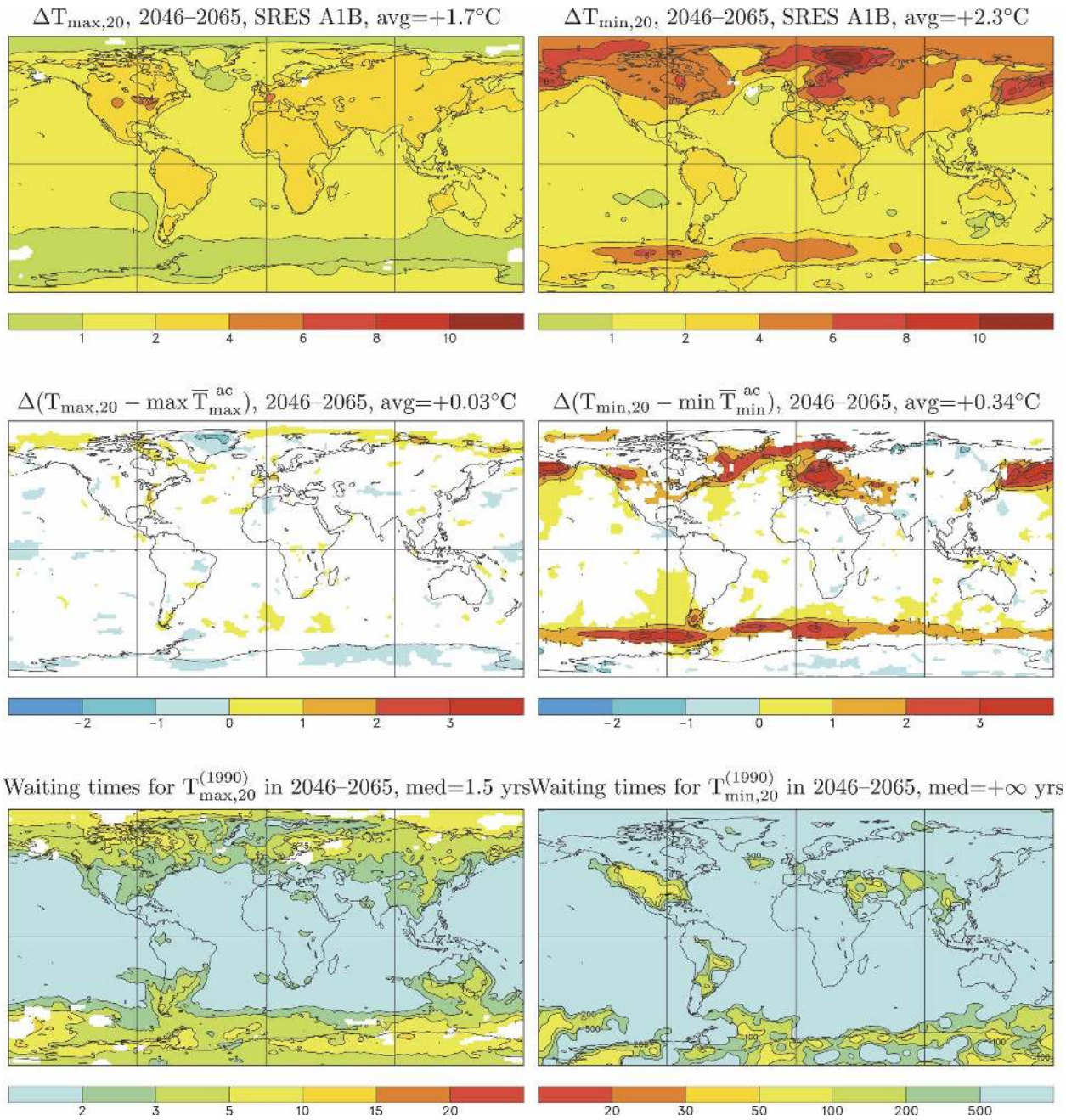

FIG. 9. (top) The multimodel mean change in 20-yr return values of annual (left) warm temperature extremes and (right) cold temperature extremes as simulated by 12 IPCC AR4 models in 2046–65 relative to 1981–2000 in the SRES A1B experiment. (middle) The corresponding changes in temperature extremes relative to the changes in the maximum (for $T_{max,20}$) or minimum (for $T_{min,20}$) of the annual cycle. Units are °C. (bottom) Waiting times (yrs) for late-twentieth-century temperature extremes $T_{max,20}$ and $T_{min,20}$ in 2046–65. Changes that are not statistically significant at the 10% level are masked out in white. Global averages (or global medians for the waiting times) are indicated in the titles.

rable, as expected from the similar magnitudes of the greenhouse forcing in these two scenarios leading up to this period. Warming in the A1B scenario tends to be slightly stronger than in the A2 scenario in the middle of the century, but not very significantly so, consistent with the larger sulfate aerosol loadings in the A2 scenario (see Fig. 1). The greatest warming is simulated in 2081–2100 under the A2 scenario as a result of the strong greenhouse forcing in this period. The late-twenty-first-century warming with the B1 emission sce-

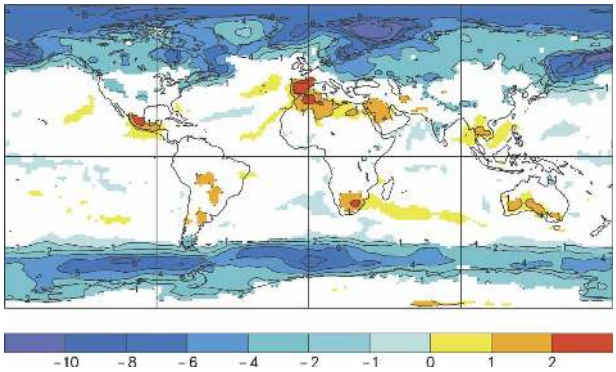

FIG. 10. The difference between the multimodel mean changes in $T_{max,20}$ and $T_{min,20}$ as simulated by 12 IPCC AR4 models in 2081–2100 relative to 1981–2000 in the SRES A1B experiment. Units are °C. Changes that are not statistically significant at the 10% level are masked out in white.

nario is comparable to that simulated in 2046–65 in the other two scenarios.

Figure 12 displays boxplot summaries of regionally averaged projected changes in temperature extremes for 2081–2100 relative to 1981–2000 when the three scenarios diverge in their degree of anthropogenic forcing. The multimodel mean change and the central 50% intermodel range of globally and land-averaged changes in temperature extremes are documented in Table 5. Cold extremes warm faster than the warm extremes by about 30%–40%, on average over the globe, and by about 25% over land. The warming of cold extremes is more than twice as large as that of warm extremes in the Arctic, about 50% larger in the North American region and about 30% larger in the European region. Changes in snow cover and sea ice are likely responsible for the greater warming of cold extremes in these regions.

The uncertainty of changes in temperature extremes simulated by individual models tends to be larger for cold extremes than for warm extremes. Over the oceans, the larger spread is confined to areas adjacent to sea ice and is likely associated with uncertainty in simulating sea ice changes under global warming. Overall uncertainty in warm extreme changes is dominated by intermodel differences in mid and high latitudes, while forcing uncertainty dominates in tropical and subtropical regions. This is, for example, evident in the plot of zonally averaged responses shown in Fig. 11 (top

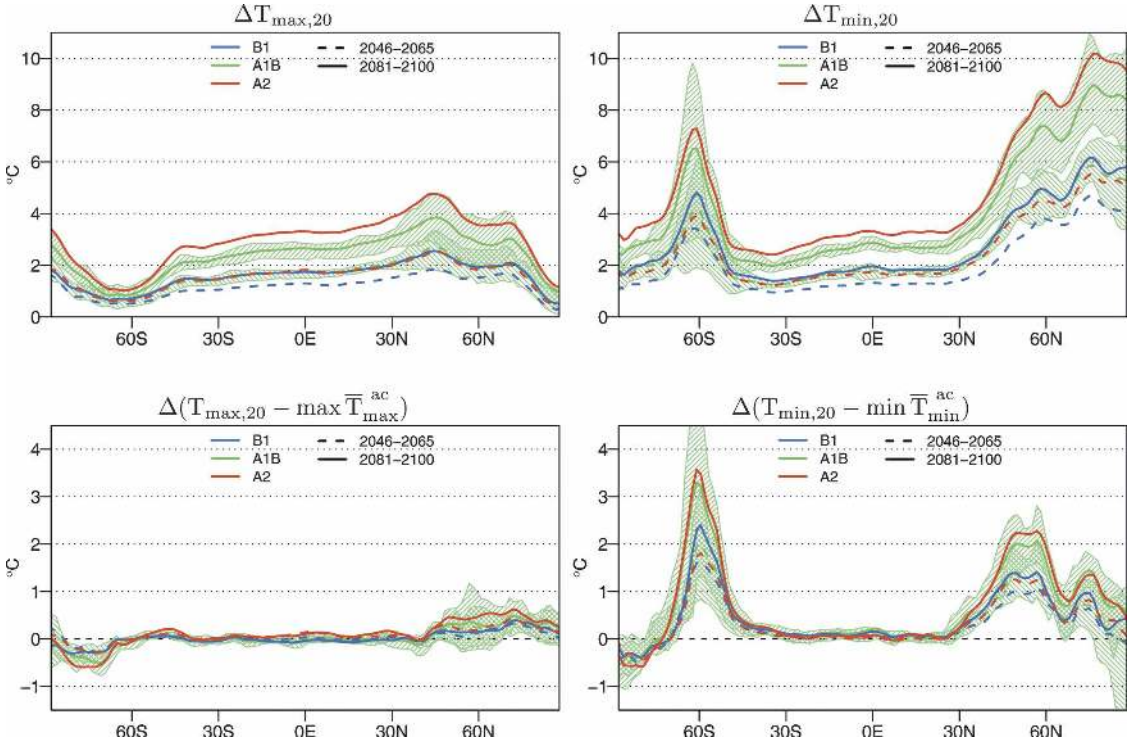

FIG. 11. Multimodel mean changes in zonally averaged (left) $T_{max,20}$ and (right) $T_{min,20}$ simulated by 12 IPCC AR4 models in 2046–65 (solid lines) and 2081–2100 (dashed lines) relative to 1981–2000 in the SRES B1 (blue curves), A1B (green curves), and A2 (red curves) experiments. The upper panels display absolute changes in temperature extremes. The lower panels display changes in temperature extremes relative to the corresponding changes in the maximum (for $T_{max,20}$) or minimum (for $T_{min,20}$) of the annual cycle. Light green hatching indicates the central 50% intermodel range for the A1B scenario in the two time periods. Units are °C.

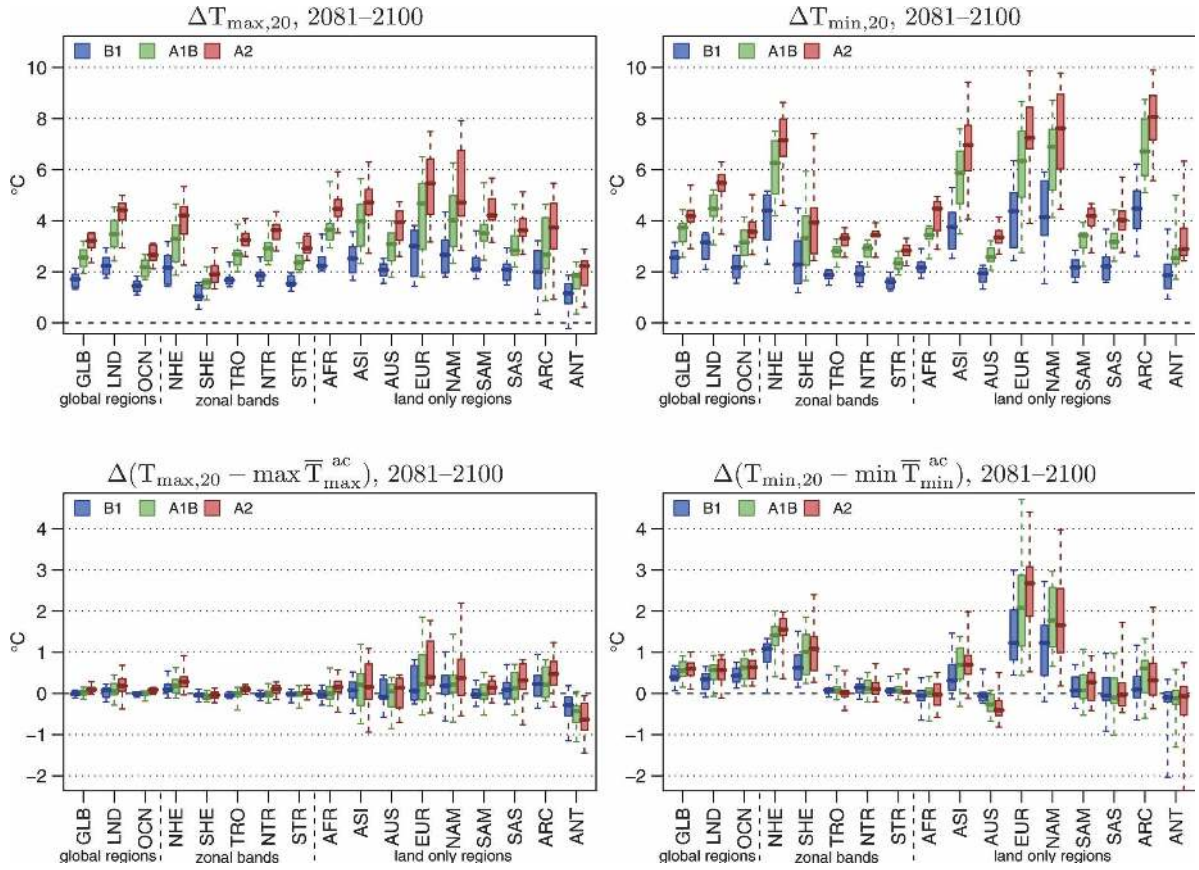

FIG. 12. Boxplots of changes in regionally averaged (left) warm temperature extremes $T_{max,20}$ and (right) cold temperature extremes $T_{min,20}$ as simulated by 12 IPCC AR4 models in 2081–2100 relative to 1981–2000. (top) Absolute changes in temperature extremes. (bottom) Changes in temperature extremes relative to the corresponding changes in the maximum (for $T_{max,20}$) or minimum (for $T_{min,20}$) of the annual cycle.

left). The late-twenty-first-century zonally averaged multimodel mean $T_{max,20}$ responses in the B1 and A2 experiments are well outside of the central 50% intermodel range in the A1B experiment between 45°S and 45°N but are generally within the intermodel range at higher latitudes. The boxplots of regionally averaged changes in warm extremes also confirm this tendency (Fig. 12, top left). There is only little overlap between the typical intermodel ranges of responses in warm extremes for different emission scenarios in Africa, South America, and in tropical and subtropical zonal bands, while the intermodel differences are comparable to, or larger than, interscenario differences in North America, Europe, the Arctic, and Antarctica. A similar tendency is found for uncertainty in changes in cold extremes except that the regions of comparatively larger intermodel differences, as compared to interscenario differences, are less uniformly distributed in northern mid and high latitudes but mainly confined to extratropical oceans and land regions in the vicinity of the retreating snow cover line.

TABLE 5. The multimodel ensemble mean and interquartile range of changes in 20-yr return values of annual warm and cold extremes ($\Delta T_{max,20}$ and $\Delta T_{min,20}$, °C) averaged over the globe and land as simulated by 10 IPCC AR4 models in 2046–65 and 2081–2100 relative to 1981–2000 in the SRES B1, SRES A1B, and SRES A2 experiments.

	2046–65			2081–2100		
	B1	A1B	A2	B1	A1B	A2
$\Delta T_{max,20}$ (°C) globe	$1.2^{1.4}_{1.1}$	$1.7^{1.9}_{1.5}$	$1.7^{1.8}_{1.6}$	$1.7^{1.9}_{1.4}$	$2.5^{2.9}_{2.2}$	$3.2^{3.5}_{2.9}$
$\Delta T_{max,20}$ (°C) land	$1.7^{2.0}_{1.5}$	$2.3^{2.6}_{2.0}$	$2.3^{2.4}_{2.1}$	$2.3^{2.6}_{1.9}$	$3.5^{4.0}_{3.0}$	$4.3^{4.7}_{4.0}$
$\Delta T_{min,20}$ (°C) globe	$1.7^{2.0}_{1.4}$	$2.3^{2.5}_{2.1}$	$2.1^{2.4}_{1.9}$	$2.4^{2.8}_{2.0}$	$3.5^{3.9}_{3.2}$	$4.1^{4.4}_{4.0}$
$\Delta T_{min,20}$ (°C) land	$2.1^{2.4}_{1.8}$	$2.9^{3.2}_{2.8}$	$2.8^{3.1}_{2.6}$	$2.9^{3.5}_{2.5}$	$4.5^{5.0}_{4.2}$	$5.4^{5.8}_{5.2}$

b. Changes in precipitation extremes

The multimodel ensemble change in precipitation extremes is displayed as the multimodel *median* response instead of the ensemble mean response. Both the mean

and the median are measures of the central tendency in the ensemble response. The two measures are generally very similar for changes in temperature extremes presented in the previous section, indicating that the distributions of responses in extreme temperatures simulated by individual models are reasonably symmetric about the central value. The situation is somewhat different for ensemble changes in extreme precipitation. The distribution of tropical changes (not shown) is skewed toward relatively larger responses. In particular, the two GFDL models simulate very large responses in the tropical extreme precipitation by approximately doubling the magnitude of 20-yr return values by the end of the twenty-first century in the SRES A2 experiment. The tendency toward generally wider upper tails and shorter lower tails in the distributions of ensemble responses is also present in the extratropics. A general skewness to the right is not surprising considering the fact that precipitation is a nonnegatively defined quantity. Therefore, possible outliers seem more likely to occur at the upper end of the distribution. The median response appears to be less sensitive to such outliers than the mean response.

The median value may also be a more appropriate choice as a measure of central tendency when the statistic in question depends on the location of the parent distribution in a highly nonlinear fashion. A typical example is the probability of exceedance above some large threshold. A positive shift of the overall distribution would result in a relatively larger increase in exceedance probability than would a negative shift of the same amplitude. For example, shifting the mean of a normal distribution one standard deviation to the right will result in a 30% increase in the probability of exceeding the original 90th percentile, while a negative change of the same amplitude will decrease the exceedance probability by only about 9%. This asymmetry in probability response will generally result in a positive bias of the ensemble mean probability response when averaged across a mixture of positive and negative responses. The median value seems to be less prone to such biases.

A similar effect is also expected for changes in waiting times of extreme precipitation events exceeding a specified threshold. By definition, the waiting time is $1/p$, where p is the probability of an extreme event and is bounded from below by one but unbounded from above, occasionally resulting in very large waiting times when the probability of an event approaches zero. Thus the arithmetic mean of estimated waiting times will likely be positively biased and thus the median value again seems to be a more appropriate measure of the central tendency of changes in waiting times. This is also true when calculating regional estimates of waiting times. Thus, in the following we use the spatial median value instead of the spatial mean value when reporting regional estimates of changes in waiting times for the late-twentieth-century extreme precipitation events.

The top two panels in Fig. 13 display the multimodel median response in annual mean precipitation as simulated by the IPCC AR4 models in 2046–65 (left panel) and 2081–2100 (right panel) in the SRES A1B experiment. The middle two panels display the corresponding changes in 20-yr return values of annual 24-h precipitation extremes. The changes are expressed as a percentage of 1981–2000 values. Mean precipitation increases in the Tropics and in the mid- and high latitudes, while it decreases in the subtropics. Negative changes in extreme precipitation occur over much smaller regions, as compared to those for mean precipitation, and are generally not statistically significant. There are extensive subtropical areas where the IPCC models predict an increase in the intensity of precipitation extremes, while mean precipitation decreases. The multimodel median globally averaged change in mean precipitation is 1.9% in 2046–65 and 3.4% in 2081–2100 in the SRES A1B experiment in the considered models. The corresponding changes in extreme precipitation are 7.7% and 12.3%. These findings are consistent with the results from a recent study by Emori and Brown (2005), who also found comparatively larger increases in extreme precipitation as compared to changes in mean precipitation in an ensemble of six climate models.

The bottom two panels of Fig. 13 display the multimodel ensemble median of waiting times for annual 24-h precipitation extremes of the size of 1981–2000 20-yr return values. Except for a few small subtropical regions where the amplitude of extreme precipitation decreases, waiting times for the late-twentieth-century extreme precipitation events are reduced almost everywhere over the globe. Not surprisingly, the changes in waiting times are consistent with changes in the amplitude of extreme precipitation. Roughly speaking, the waiting times are reduced by a factor of 2 with a 10% increase in the amplitude of P_{20}. Waiting times decrease almost everywhere over landmasses, except for North Africa where waiting times tend to increase. The spatial median value of the waiting times over land is reduced from 20 yr to about 12 yr in 2046–65 and less than 10 yr in 2081–2100 in the SRES A1B experiment (see Table 6). The greatest reductions in waiting time occur in tropical regions and high latitudes.

Interscenario differences are illustrated in Fig. 14. The change in mean precipitation is small or negative in the 45°–10°S and 10°–40°N zonal bands, whereas the magnitude of extreme precipitation increases by up to

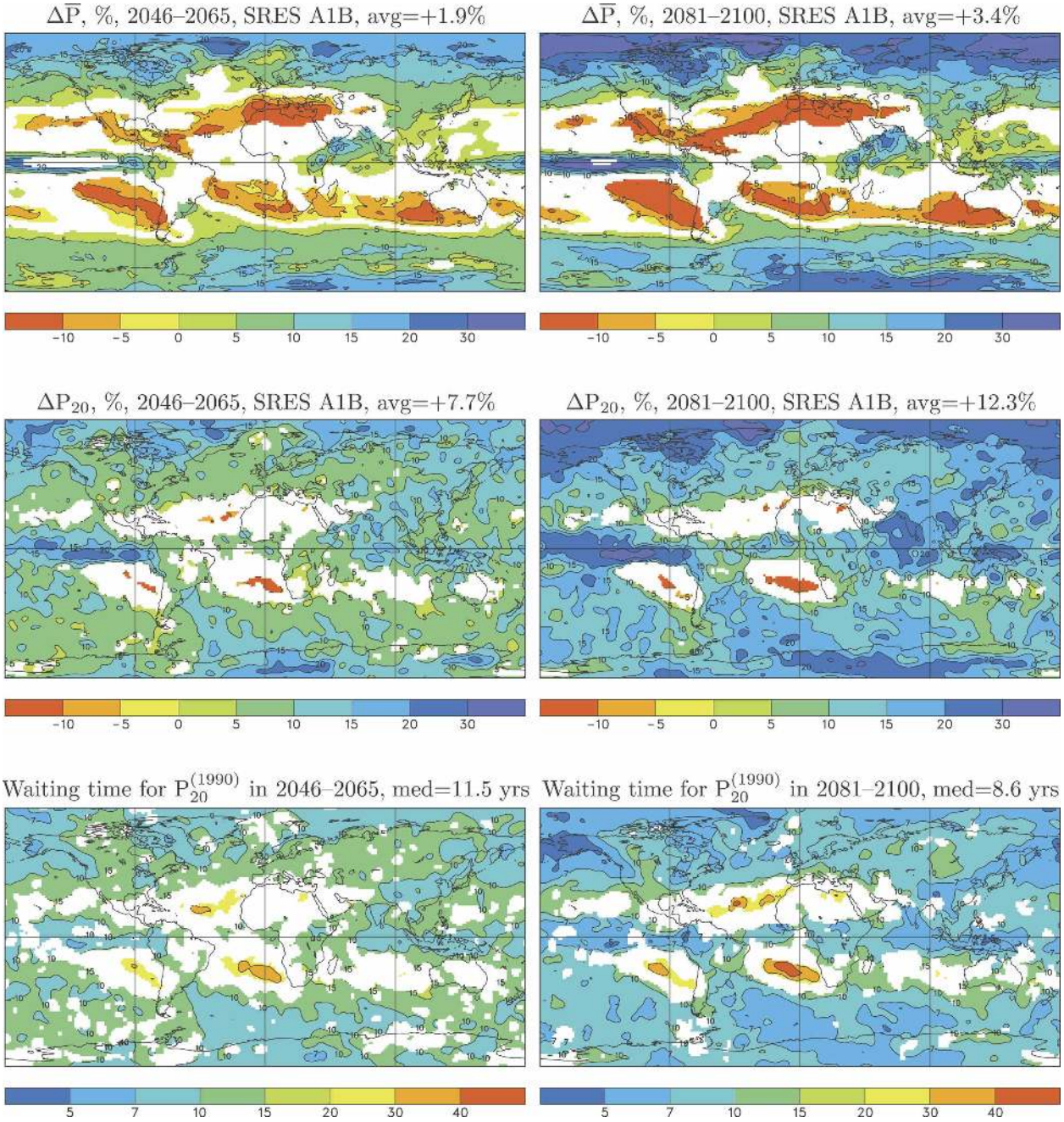

FIG. 13. The multimodel median relative change (%) in the (top) annual mean precipitation rate and (middle) in 20-yr return values of annual extremes of daily precipitation as simulated by 14 IPCC AR4 models in (left) 2046–65 and (right) 2081–2100 relative to 1981–2000 in the SRES A1B experiment. The lower panels display the corresponding median of waiting times (yr) for late-twentieth-century P_{20}. Changes that are not statistically significant at the 10% level are masked out in white. Global averages (or global medians for the waiting times) are indicated in the titles.

15%–20% in these regions, on average. Elsewhere, both mean and extreme precipitation increase. The Arctic appears to be the only region where the projected relative changes in mean precipitation exceed those in extreme precipitation.

Regional changes in mean and extreme precipitation and in waiting times are summarized in boxplot diagrams in Fig. 15 and in Table 6. It is evident from the boxplots that intermodel uncertainties in extreme precipitation changes are much larger than in mean pre-

TABLE 6. Multimodel ensemble median and interquartile range of relative changes in regional estimate of mean precipitation $\Delta\overline{P}$ (%), 20-yr return values of annual 24-h precipitation extremes ΔP_{20} (%), and the waiting times for present-day P_{20} ($\mathcal{T}$, yr) averaged over the globe, land, NHE (35°–90°N) and TRO (10°S–10°N) as simulated by 12 IPCC AR4 models in 2046–65 A2 experiments. Regional estimates of waiting times are defined as the area weighted median values.

	2046–65			2081–2100		
	B1	A1B	A2	B1	A1B	A2
$\Delta\overline{P}$ (%) globe	$2.2^{2.5}_{1.9}$	$2.9^{3.2}_{2.3}$	$2.4^{3.2}_{1.7}$	$3.6^{3.9}_{3.1}$	$4.6^{5.7}_{3.4}$	$5.3^{6.1}_{3.8}$
$\Delta\overline{P}$ (%) land	$2.4^{3.4}_{1.9}$	$3.7^{4.3}_{3.0}$	$3.4^{4.3}_{2.1}$	$4.2^{5.1}_{3.0}$	$5.7^{6.6}_{4.4}$	$6.8^{7.7}_{5.9}$
$\Delta\overline{P}$ (%) NHE	$4.0^{4.4}_{3.5}$	$5.0^{6.2}_{4.1}$	$4.3^{5.6}_{3.6}$	$6.1^{6.8}_{5.4}$	$8.2^{10.3}_{5.7}$	$8.8^{11.3}_{6.5}$
$\Delta\overline{P}$ (%) TRO	$3.9^{4.8}_{3.6}$	$5.2^{6.0}_{4.0}$	$4.7^{5.8}_{3.9}$	$5.9^{7.2}_{5.5}$	$7.7^{9.4}_{6.3}$	$9.7^{10.8}_{6.6}$
ΔP_{20} (%) globe	$7.4^{10.0}_{4.7}$	$10.5^{15.5}_{6.2}$	$9.6^{13.7}_{7.0}$	$10.8^{13.6}_{6.7}$	$16.3^{22.3}_{10.0}$	$19.4^{32.8}_{12.0}$
ΔP_{20} (%) land	$7.3^{11.1}_{4.6}$	$10.3^{16.0}_{5.9}$	$10.0^{14.7}_{5.5}$	$10.3^{15.1}_{6.9}$	$16.2^{24.3}_{9.8}$	$19.5^{33.8}_{11.6}$
ΔP_{20} (%) NHE	$7.7^{10.0}_{5.8}$	$10.6^{13.8}_{8.7}$	$10.6^{12.2}_{7.0}$	$10.7^{14.3}_{8.1}$	$17.9^{22.0}_{13.2}$	$21.8^{24.1}_{16.0}$
ΔP_{20} (%) TRO	$10.5^{14.9}_{5.9}$	$15.2^{22.1}_{7.9}$	$13.7^{18.9}_{9.3}$	$15.5^{19.0}_{8.9}$	$24.9^{29.8}_{10.4}$	$29.2^{47.9}_{17.8}$
$\mathcal{T}$(yr) globe	$13.2^{14.8}_{12.2}$	$11.5^{13.2}_{10.2}$	$11.5^{13.0}_{10.9}$	$12.2^{13.3}_{9.8}$	$9.5^{10.7}_{7.2}$	$7.5^{9.3}_{6.2}$
$\mathcal{T}$(yr) land	$13.2^{14.5}_{11.9}$	$10.7^{12.2}_{10.1}$	$11.6^{12.8}_{10.4}$	$11.8^{12.4}_{9.7}$	$9.0^{10.3}_{7.1}$	$7.2^{8.9}_{5.8}$
$\mathcal{T}$(yr) NHE	$12.6^{13.6}_{10.8}$	$10.6^{11.7}_{9.2}$	$10.7^{12.1}_{10.0}$	$10.8^{11.7}_{9.0}$	$8.2^{9.4}_{6.4}$	$6.6^{7.5}_{5.8}$
$\mathcal{T}$(yr) TRO	$11.4^{14.7}_{10.7}$	$9.4^{13.5}_{8.0}$	$9.5^{12.0}_{8.8}$	$9.4^{12.1}_{8.3}$	$6.9^{11.3}_{5.2}$	$5.7^{9.5}_{4.2}$

cipitation and that these uncertainties increase quite substantially with the increased anthropogenic forcing by the end of the twenty-first century. The largest spread in extreme precipitation changes simulated by individual models occurs in the Tropics, especially over the tropical Pacific. Globally averaged relative changes in mean and extreme precipitation are comparable to those averaged over landmasses only. Mean precipitation in the northern extratropics north of 35°N and in the Tropics (10°S–10°N) increases by about equal percentages (Table 6). However, simulated relative increases of the intensity of extreme precipitation in the Tropics exceed those in the northern extratropics by a factor of approximately 1.5, on average.

Allen and Ingram (2002) and Trenberth et al. (2003), among others, argue that, while global mean precipitation is primarily constrained by the energy budget, the intensity of heavy precipitation events should increase with the availability of moisture at a rate close to the Clausius–Clapeyron rate of about 6%–7% per kelvin. To verify this contention in the present ensemble of global climate models, the left panel of Fig. 16 displays the relative changes (%) in globally averaged P_{20} as a function of global annual mean temperature changes as simulated by the IPCC AR4 models in 2046–65 and 2081–2100 under the three emission scenarios. Daily mean temperature is approximated by the average of daily $T_{\max}$ and $T_{\min}$. A histogram of the "hydrological sensitivities" for extreme precipitation, $\Delta P_{20}(\%)/\Delta\overline{T}(\mathrm{K})$, is shown in the right-hand panel of Fig. 16.

The median sensitivity of about 6% K^{-1} is consistent with the projected Clausius–Clapeyron rates cited above. However, it is also evident that there is a great deal of intermodel variability. Four models particularly stand out. The two GFDL models simulate extremely large increases in globally averaged extreme precipitation, mostly in the Tropics, and are well outside the range of the remaining collection of models on the upper end. The lowest sensitivity of extreme precipitation to changes in mean temperature of 2%–3% K^{-1} is found in the INM-CM3.0 and GISS-ER models. The remaining models simulate the sensitivities in the range 4%–10%. It is also interesting to note that there seems to be no apparent relationship between the hydrological sensitivity for extreme precipitation and global temperature response. For example, the GFDL CM2.0 and INM-CM3.0 models have comparable global temperature changes but vastly different responses in extreme precipitation.

6. Summary

The present study documents the performance of global coupled climate models that participated in the IPCC diagnostic exercise for the Fourth Assessment Report in simulating annual extremes of surface temperature and daily precipitation rates and their changes as simulated by the models under the three emission scenarios, SRES B1, A1B, and A2. Among these three scenarios, B1 envisions the slowest growth of anthropogenic greenhouse forcing while A2 projects the fastest growing forcing.

Climate extremes are evaluated in terms of 20-yr return values of annual extremes for three time periods, 1981–2000, 2046–65, and 2081–2100. The 1981–2000 period serves as the baseline for future changes. Late-

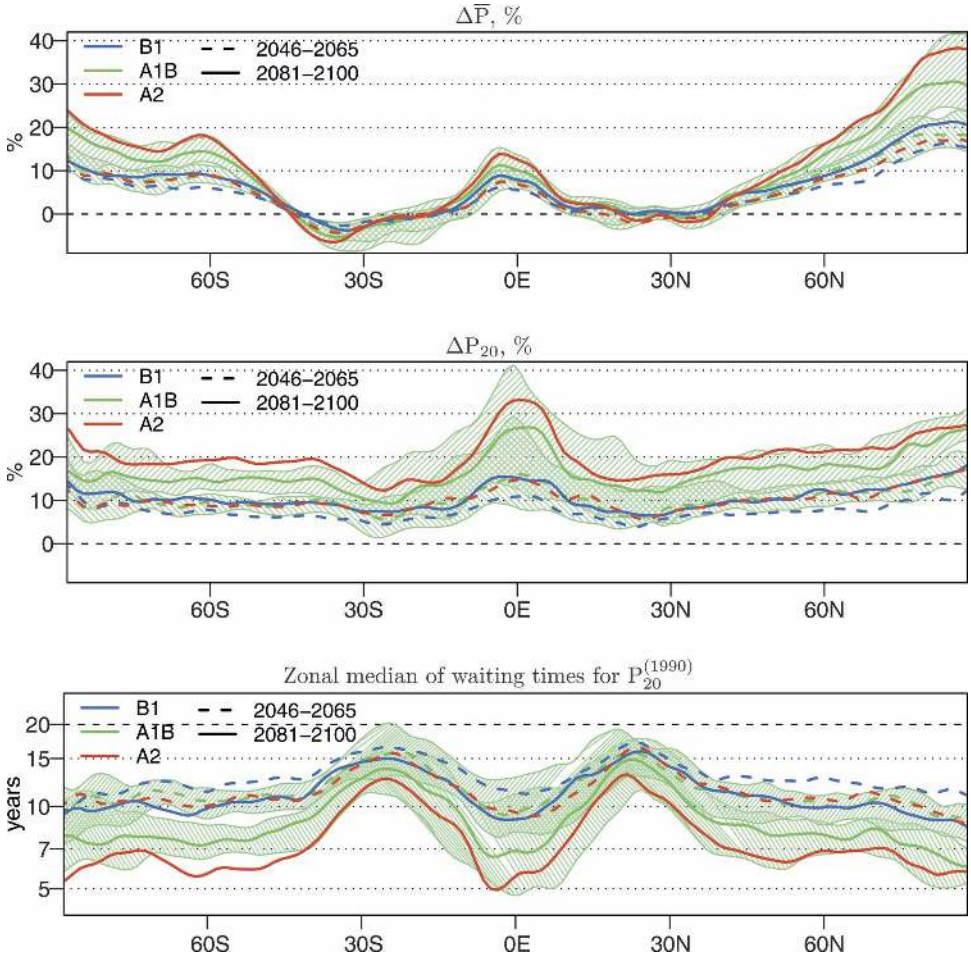

FIG. 14. Multimodel median relative change (%) in (top) the zonally averaged annual mean precipitation rate, (middle) 20-yr return values of annual extremes of 24-h precipitation rates, and (bottom) the zonal median of the waiting times for present-day P_{20} (yr) as simulated by 14 IPCC AR4 models in 2046–65 (dashed lines) and 2081–2100 (solid lines) relative to 1981–2000 in the SRES B1 (blue), A1B (green), and A2 (red) experiments. Light green hatching indicates the central 50% intermodel range for the A1B scenario in the two time periods.

twentieth-century temperature and precipitation extremes are evaluated for 14 and 16 models, respectively. Changes in extremes are evaluated only for models for which daily output was available for all three scenarios, resulting in a 12-model ensemble for temperature extremes and a 14-model ensemble for precipitation extremes. The analysis of changes was also repeated for all available models but without any substantial modifications in the results.

The simulated late-twentieth-century extremes are compared to those estimated from four reanalyses, the older NCEP–NCAR and ERA-15 products and the more recent NCEP–DOE AMIP-II and ERA-40 products. Model-simulated extremes of precipitation pentads (nonoverlapping 5-day means) were also compared to those estimated from the CMAP pentad dataset. Changes in the amplitude of warm and cold temperature extremes are compared to the corresponding mean changes for the climatologically warmest and coldest calendar months, respectively. Changes in the exceedance probabilities are also examined and expressed in terms of changes in waiting times for late-twentieth-century extreme events.

The results of the analysis are summarized as follows.

- Warm temperature extremes of the late-twentieth-century climate are plausibly simulated by the IPCC AR4 models. The multimodel mean globally averaged 20-yr return value of annual warm extremes $T_{\max,20}$ corresponds closely to estimates derived from the NCEP2 and ERA-40 reanalyses. The 50% intermodel range of globally averaged $T_{\max,20}$ estimates is

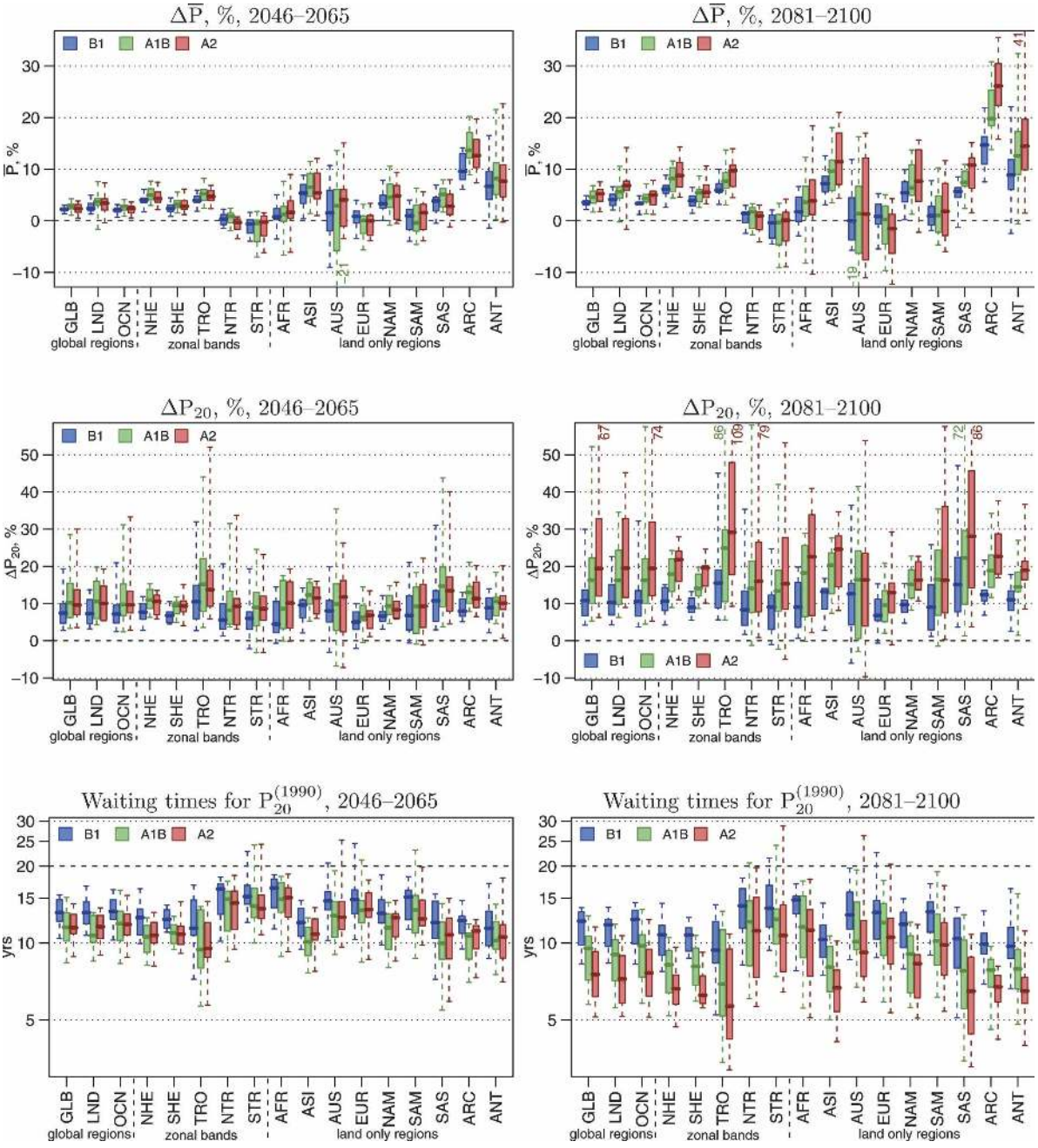

FIG. 15. Boxplots of relative changes (%) in (top) the regionally averaged annual mean precipitation rate ($\overline{P}^5_{20}$), (middle) 20-yr return values of annual extremes of 24-h precipitation rates (ΔP_{20}), and (bottom) the waiting times for late-twentieth-century P_{20} as simulated by 14 IPCC AR4 models in (left) 2046–65 and (right) 2081–2100 relative to 1981–2000 with the SRES B1 (blue), A1B (green), and A2 (red) emission scenarios. The boxes indicate the central 50% intermodel range and the median. The whiskers extend to the lower and upper model extremes. The regions are defined in Table 2.

fairly narrow indicating that most of the models perform well on a global scale. Model differences are generally larger over land than over oceans.

- Uncertainties in model-simulated cold extremes for the late-twentieth-century climate are larger than those for warm extremes. The available estimates from reanalyses are also less consistent. For example, the NCEP2 and ERA-40 estimates of 20-yr return values of annual cold extremes $T_{\min,20}$ disagree significantly over land, with the former being colder by more than 8°C than the latter on average. The multimodel ensemble mean of $T_{\min,20}$ estimates, aver-

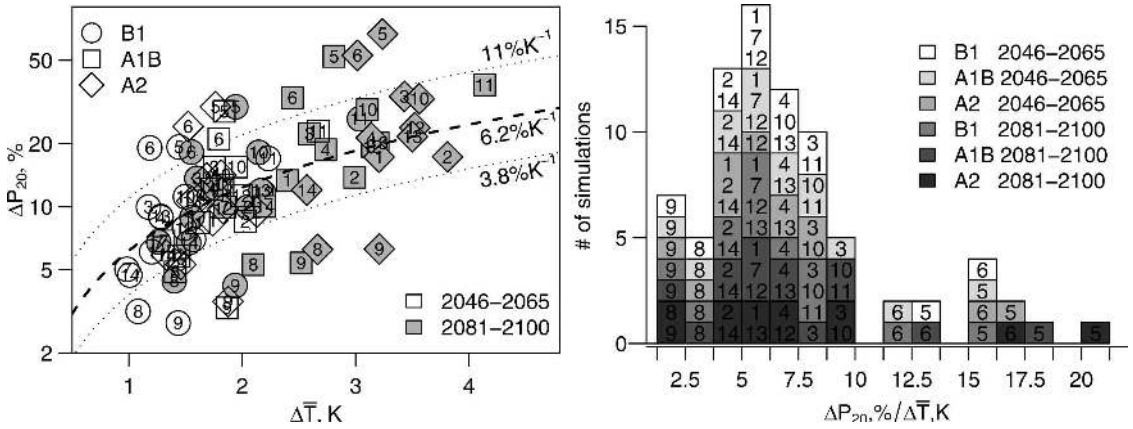

FIG. 16. (left) Relative changes (%) in globally averaged 20-yr return values of annual 24-h precipitation extremes (ΔP_{20}) plotted on a log scale as a function of globally averaged changes in mean surface temperature $\Delta\overline{T}$ (K) simulated by the IPCC AR4 models in the SRES B1 (circles), A1B (squares), and A2 (diamonds) experiments in 2046–65 (empty symbols) and 2081–2100 (gray symbols). The median slope of 6.2% K^{-1} and the 15th and 85th percentiles are indicated by the bold dashed and dotted lines, respectively. (right) Histogram of $\Delta P_{20}/\Delta\overline{T}$ (% K^{-1}) simulated by the IPCC AR4 models in the three scenarios and two time periods. The models are indicated by numbers in both panels as 1: CCCMA CGCM3.1(T47), 2: CCCMA CGCM3.1(T63), 3: CNRM CM3, 4: ECHO G, 5: GFDL CM2.0, 6: GFDL CM2.1, 7: GISS AOM, 8: GISS ER, 9: INM CM3.0, 10: IPSL CM4, 11: MIROC3.2(hires), 12: MIROC3.2(medres), 13: MPI ECHAM5, and 14: MRI CGCM2.3.2.

aged over the globe or over land, falls between the estimates from the two reanalyses. The central 50% intermodel range of simulated $T_{\min,20}$ averaged globally or over land coincides approximately with the range between the reanalyses.

- The performance of the IPCC AR4 CGCMs in simulating extreme precipitation is comparable to that of the AMIP2 atmospheric models reported by Kharin et al. (2005). The coupled models agree satisfactorily on the amplitude of 20-yr return values of annual precipitation extremes (P_{20}) in the extratropics but exhibit very large differences in the Tropics. For example, the intermodel standard deviation of P_{20} in temperate regions is about 20% of the ensemble mean value, whereas it is more than 60% of the ensemble mean value in tropical regions. The models compare reasonably well with the observational evidence in the extratropics. But, the available reanalyses do not provide a reliable consensus on the magnitude of the observed precipitation extremes in the Tropics where differences between reanalyses are comparable to those between the models. In the Tropics, models simulate more moderate precipitation extremes than in the ERA-40, NCEP2, and ERA-15 reanalyses. There is a tendency for stronger precipitation extremes in models with higher horizontal resolution, but the dependence on resolution is not robust across different models, particularly in the Tropics.
- Changes in warm extremes more or less follow changes in the mean temperature of the climatologically warmest month. That is, changes in the summertime mean temperature are reasonably good predictors of changes in warm extremes, at least on a global scale. The ensemble mean changes in $T_{\max,20}$ over land range from 1.7°C in 2046–65 in the SRES B1 experiment to 4.2°C in 2081–2100 in the SRES A2 experiment. The spread in the responses generally increases with the strength of the anthropogenic forcing, which is an indication of different climate sensitivities of the models. For example, the globally averaged root-mean-square (rms) difference from the ensemble mean response is about 0.7°C in 2046–65 in the B1 scenario but increases to 1.2°C in 2081–2100 in the A2 scenario when the forcing is substantially stronger. Multimodel mean increases in warm extremes averaged over land are about 35%–40% larger than those averaged over the globe in the considered scenarios and time periods.
- Cold extremes warm faster than warm extremes by about 30%–40% over the globe, on average, or about 25% over landmasses. Most of this excessive warming is confined to regions where snow and sea ice retreat under the global warming. For example, the magnitude of the warming in cold extremes in the Arctic regions north of 65°N is more than a factor of 2 greater than the corresponding warming in warm extremes. Cold extremes warm faster than warm extremes by about 30% in the European region and up to 50% in the North American region. Changes in the cryosphere are also likely responsible for faster warming of cold extremes as compared to changes in

the mean wintertime temperature. The ensemble mean change in $T_{\min,20}$ over land ranges from 2.1°C in 2046–65 in the SRES B1 experiment to 5.4°C in 2081–2100 in the A2 experiment. The intermodel uncertainty of these changes tends to be somewhat larger than for warm extremes. To a large degree, the larger spread of the model responses for cold extremes is confined to regions adjacent to the sea ice line and is likely associated with uncertainties in simulating sea ice changes under the global warming.

- Fractional increases in the intensity of precipitation extremes generally exceed those for mean precipitation. Extreme precipitation decreases in only a small fraction of the subtropical area where mean precipitation decreases. Globally averaged, the multimodel P_{20} increase is about 10% in the SRES B1 experiment, 16% in the A1B experiment, and 20% in the A2 experiment by the end of the twenty-first century. When compared to the corresponding changes in global mean temperature, changes in extreme precipitation translate roughly to a consensus of about 6% K^{-1} of global warming, which is consistent with the predicted change implied by the Clausius–Clapeyron relation, with most of the models simulating values in the range of 4%–10% K^{-1}. Consistent with changes in the magnitude of extreme precipitation, waiting times for late-twentieth-century extreme precipitation events are reduced almost everywhere over the globe. Waiting times in the Tropics and in mid and high latitudes are reduced by a factor of about 2 in 2046–65 and by a factor of about 3 in 2081–2100 in the SRES A1B and A2 experiments, with more moderate changes in the B1 experiment. The very large intermodel disagreements in the Tropics suggest that some physical processes associated with extreme precipitation are not well represented in models. This reduces our confidence in the projected changes in extreme precipitation.
- Model differences, rather than sampling error, appear to be the main source of uncertainty in the simulated late-twentieth-century temperature and precipitation extremes in the considered multimodel ensemble. The overall uncertainty in local changes in temperature extremes is dominated by intermodel differences and sampling errors in mid and high latitudes while forcing uncertainty becomes important in the Tropics and subtropical regions. Intermodel differences generally dominate the uncertainty in changes in precipitation extremes.

Acknowledgments. We acknowledge the international modeling groups for providing their data for analysis, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) for collecting and archiving the model data, the JSC/CLIVAR Working Group on Coupled Modelling (WGCM) and their Coupled Model Intercomparison Project (CMIP) and Climate Simulation Panel for organizing the model data analysis activity, and the IPCC WG1 TSU for technical support. The IPCC Data Archive at the Lawrence Livermore National Laboratory is supported by the Office of Science, U.S. Department of Energy. We are thankful to three anonymous reviewers for their helpful comments and constructive recommendations.

REFERENCES

Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature,* **419,** 224–232.

Coles, S. G., and M. J. Dixon, 1999: Likelihood-based inference for extreme value models. *Extremes,* **2,** 5–23.

Collins, W. D., and Coauthors, 2006: The Community Climate System Model version 3 (CCSM3). *J. Climate,* **19,** 2122–2143.

Cubasch, U., and Coauthors, 2001: Projections of future climate change. *Climate Change 2001: The Scientific Basis,* J. T. Houghton et al., Eds., Cambridge University Press, 525–582.

Delworth, T. L., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate,* **19,** 643–674.

Diansky, N. A., and E. M. Volodin, 2002: Simulation of present-day climate with a coupled atmosphere-ocean general circulation model. *Izv. Atmos. Oceanic Phys.,* **38,** 732–747.

Dupuis, D. J., and M. Tsao, 1998: A hybrid estimator for the Generalized Pareto and Extreme-Value Distributions. *Comm. Stat. Theory Methods,* **27,** 925–941.

Efron, B., and R. Tibshirani, 1993: *An Introduction to the Bootstrap.* Chapman and Hall, 436 pp.

Emori, S., and S. J. Brown, 2005: Dynamic and thermodynamic changes in mean and extreme precipitation under changed climate. *Geophys. Res. Lett.,* **32,** L17706, doi:10.1029/2005GL023272.

Frich, P., L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. M. G. Klein Tank, and T. Peterson, 2002: Observed coherent changes in climate extremes during the second half of the twentieth century. *Climate Res.,* **19,** 193–212.

Gibson, J. K., P. Kalberg, S. Uppala, A. Hernandes, A. Nomura, and E. Serrano, 1997: ERA Description. ECMWF Reanalysis Report Series 1, ECMWF, Reading, United Kingdom, 72 pp.

Gnanadesikan, A., and Coauthors, 2006: GFDL's CM2 global coupled climate models. Part II: The baseline ocean simulation. *J. Climate,* **19,** 675–697.

Gregory, J. M., and J. F. B. Mitchell, 1995: Simulation of daily variability of surface temperature and precipitation over Europe in the current and 2xCO2 climate using the UKMO high-resolution climate model. *Quart. J. Roy. Meteor. Soc.,* **121,** 1451–1476.

Hasumi, H., and S. Emori, 2004: K-1 coupled model (MIROC) description. K-1 Tech. Rep. 1, Center for Climate System Research, University of Tokyo, 34 pp.

Hosking, J. R. M., 1990: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc. (Ser. A),* **B52,** 105–124.

——, 1992: Moments or L-moments? An example comparing the

two measures of distributional shape. *Amer. Stat.,* **46,** 186–189.

——, J. R. Wallis, and E. F. Wood, 1985: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics,* **27,** 251–261.

International Ad Hoc Detection and Attribution Group, 2005: Detecting and attributing external influences on the climate system: A review of recent advances. *J. Climate,* **18,** 1291–1314.

Jenkinson, A. F., 1955: The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quart. J. Roy. Meteor. Soc.,* **81,** 158–171.

Jungclaus, J. H., and Coauthors, 2006: Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM. *J. Climate,* **19,** 3952–3972.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.,* **83,** 1631–1643.

Kharin, V. V., and F. W. Zwiers, 2000: Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *J. Climate,* **13,** 3760–3788.

——, and ——, 2005: Estimating extremes in transient climate change simulations. *J. Climate,* **18,** 1156–1173.

——, ——, and X. Zhang, 2005: Intercomparison of near surface temperature and precipitation extremes in AMIP-2 simulations, reanalyses, and observations. *J. Climate,* **18,** 5201–5223.

Martins, E. S., and J. R. Stedinger, 2000: Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrological data. *Water Resour. Res.,* **36,** 737–744.

Meehl, G. A., and Coauthors, 2006: Climate change projections for the twenty-first century and climate change commitment in the CCSM3. *J. Climate,* **19,** 2597–2616.

Min, S.-K., S. Legutke, A. Hense, and W.-T. Kwon, 2005: Internal variability in a 1000-year control simulation with the coupled climate model ECHO-G. Part I: Near surface temperature, precipitation, and mean sea level pressure. *Tellus,* **57A,** 605–621.

Palutikof, J. P., B. B. Brabson, D. H. Lister, and S. T. Adcock, 1999: A review of methods to calculate extreme wind speeds. *Meteor. Appl.,* **6,** 119–132.

Ramanathan, V., P. J. Crutzen, J. T. Kiehl, and D. Rosenfeld, 2001: Atmosphere—Aerosols, climate, and the hydrological cycle. *Science,* **294,** 2119–2124.

Russell, G. L., J. R. Miller, and D. Rind, 1995: A coupled atmosphere–ocean model for transient climate change studies. *Atmos.–Ocean,* **33,** 683–730.

——, ——, ——, R. A. Ruedy, G. A. Schmidt, and S. Sheth, 2000: Comparison of model and observed regional temperature changes during the past 40 years. *J. Geophys. Res.,* **105,** 14 891–14 898.

Schmidt, G. A., and Coauthors, 2006: Present-day atmospheric simulations using GISS ModelE: Comparison to in situ, satellite, and reanalysis data. *J. Climate,* **19,** 153–192.

Scinocca, J. F., and N. A. McFarlane, 2004: The variability of modeled tropical precipitation. *J. Atmos. Sci.,* **61,** 1993–2015.

Semenov, V. A., and L. Bengtsson, 2002: Secular trends in daily precipitation characteristics: Greenhouse gas simulation with a coupled AOGCM. *Climate Dyn.,* **19,** 123–140.

Simmons, A. J., and J. K. Gibson, 2000: The ERA-40 project plan. ECMWF ERA-40 Project Report Series 1, 63 pp.

Stephens, M. A., 1970: Use of the Kolmogorov-Smirnov, Cramer-von-Mises and related statistics without extensive tables. *J. Roy. Stat. Soc. (Ser. A),* **32B,** 115–122.

Tebaldi, C., K. Hayhoe, J. M. Arblaster, and G. A. Meehl, 2006: An intercomparison of model-simulated historical and future changes in extreme events. *Climatic Change,* **79,** 185–211.

Trenberth, K. E., A. Dai, R. M. Rasmussen, and D. B. Parsons, 2003: The changing character of precipitation. *Bull. Amer. Meteor. Soc.,* **84,** 1205–1217.

Voss, R., W. May, and E. Roeckner, 2002: Enhanced resolution modelling study on anthropogenic climate change: Changes in extremes of the hydrological cycle. *Int. J. Climatol.,* **22,** 755–777.

Washington, W. M., and Coauthors, 2000: Parallel climate model (PCM) control and transient simulations. *Climate Dyn.,* **16,** 755–774.

Wehner, M. F., 2004: Predicted twenty-first-century changes in seasonal extreme precipitation events in the Parallel Climate Model. *J. Climate,* **17,** 4281–4290.

Wilby, R. L., and T. M. L. Wigley, 2002: Future changes in the distribution of daily precipitation totals across North America. *Geophys. Res. Lett.,* **29,** 1135, doi:10.1029/2001GL013048.

Wilcoxon, F., 1945: Individual comparisons by ranking methods. *Biometrics,* **1,** 80–83.

Xie, P., J. E. Janowiak, P. A. Arkin, R. Adler, A. Gruber, R. Ferraro, G. J. Huffman, and S. Curtis, 2003: GPR pentad precipitation analyses: An experimental dataset based on gauge observations and satellite estimates. *J. Climate,* **16,** 2197–2214.

Yukimoto, S., and Coauthors, 2001: The new Meteorological Research Institute coupled GCM (MRI-CGCM2)—Model climate and variability. *Pap. Meteor. Geophys.,* **51,** 47–88.

——, A. Noda, T. Uchiyama, and S. Kusunoki, 2006: Climate change of the twentieth through twenty-first centuries simulated by the MRI-CGCM2.3. *Pap. Meteor. Geophys.,* **56,** 9–24.

Zhang, X., F. W. Zwiers, and G. Li, 2004: Monte Carlo experiments on the detection of trends in extreme values. *J. Climate,* **17,** 1945–1952.

Zwiers, F. W., and V. V. Kharin, 1998: Changes in the extremes of the climate simulated by CCC GCM2 under CO_2 doubling. *J. Climate,* **11,** 2200–2222.