Comparison of Monthly Temperature Extremes Simulated by CMIP3 and CMIP5 Models

YAO YAO, YONG LUO, JIANBIN HUANG, AND ZONGCI ZHAO

Center for Earth System Science, and Ministry of Education Key Laboratory for Earth System Modeling, Tsinghua University, Beijing, China

(Manuscript received 30 July 2012, in final form 20 March 2013)

ABSTRACT

The extreme monthly-mean temperatures simulated by 28 models in the fifth phase of the Coupled Model Intercomparison Project (CMIP5) are evaluated and compared with those from 24 models in the third phase of the Coupled Model Intercomparison Project (CMIP3). Comparisons with observations and reanalyses indicate that the models from both CMIP3 and CMIP5 perform well in simulating temperature extremes, which are expressed as 20-yr return values. When the climatological annual cycle is removed, the ensemble spread in CMIP5 is smaller than that in CMIP3. Benefiting from a higher resolution, the CMIP5 models perform better at simulating extreme temperatures on the local gridcell scale. The CMIP5 representative concentration pathway (RCP4.5) and CMIP3 B1 experiments project a similar change pattern in the near future for both warm and cold extremes, and the pattern is in agreement with that of the seasonal extremes. By the late twenty-first century, the changes in monthly temperature extremes projected under the three CMIP3 (B1, A1B, and A2) and two CMIP5 (RCP4.5 and RCP8.5) scenarios generally follow the changes in climatological annual cycles, which is consistent with previous studies on daily extremes. Compared with the CMIP3 ensemble, the CMIP5 ensemble shows a larger intermodel uncertainty with regard to the change in cold extremes in snow-covered regions. Enhanced changes in extreme temperatures that exceed the global mean warming are found in regions where the retreat of snow (or the soil moisture feedback effect) plays an important role, confirming the findings for daily temperature extremes.

1. Introduction

A growing number of extreme climate events on various spatial–temporal scales have been observed in the past, and these events have significantly affected human society and the natural environment (Field et al. 2012). The reason for the increase in the frequency and intensity of extreme climate events and whether these events are related to anthropogenic global warming has attracted much attention from scientists. Currently, accumulating evidence suggests that the increase in temperature extremes is closely related to human activities (Rahmstorf and Coumou 2011; Stott et al. 2011).

Climate models have been widely used to study anthropogenic effects on climate extremes, as these models can, to some extent, reproduce these climate extremes (Tebaldi et al. 2006). The third phase of the Coupled Model Intercomparison Project (CMIP3) has opened up countless investigations, including studies on extreme temperatures, based on CMIP3 model outputs (Meehl et al. 2007). Because of the perceptible change in the shape of the tail of the daily temperature distribution (Donat and Alexander 2012), most studies are performed using the extreme daily temperatures based on commonly used climate indices (Meehl and Tebaldi 2004; Kodra et al. 2011; Zhang et al. 2011) and extremes with long return periods (Kharin et al. 2007; Zwiers et al. 2011) calculated from the daily outputs of models. In contrast, the extreme temperature on the monthly scale has seldom been examined in previous studies. There is no doubt that anomalously cold or warm months have a profound impact on human society, particularly in relation to health (Goklany 2008), agriculture (Dreessen et al. 2012), and economics (Halsnaes et al. 2007). As an extreme example, consecutive daily high-temperature events in Moscow resulted in the hottest month on record in July 2010 (Grumm 2011). As a result of both the magnitude and duration of the Russian heatwave, more
than 50,000 people died, and Russian grain production was reduced by 20%–30% relative to the 2009 levels, causing an economic loss of more than $15 billion (U.S. dollars). Thus, there is an urgent social need to learn more about the risks of extreme months. Accordingly, the present study focuses on the extreme monthly temperatures and attempts to assess their simulation in climate models.

Compared with previous studies of model-simulated monthly temperature extremes (e.g., Räisänen and Ruokolainen 2008; Räisänen and Ylhäisi 2011), we take a more comprehensive approach and analyze the extremes on multispatial scales. Indeed, the evaluation of simulations across diverse scales will help us better understand the performance of these models. Because of the lack of homogeneous time series of global daily temperatures, reanalysis data are typically used when evaluating the daily extremes simulated by models; however, uncertainties in the daily temperature extremes also exist in such reanalyses (Kharin et al. 2007). Therefore, in this study we validate the model simulations with two station-based datasets and two reanalyses to better estimate the observational uncertainty and strengthen the confidence in validations for model simulations.

The latest climate models participating in the fifth phase of the Coupled Model Intercomparison Project (CMIP5) are now available, and components such as terrestrial and marine carbon cycles, dynamic vegetation, and indirect effects of aerosols are included in most of the models for the first time (Taylor et al. 2012). Representative concentration pathways (RCPs) have replaced the Special Report on Emission Scenarios (SRESs) as the latest generation of scenarios for long-term experiments. The comparison between the two generations of climate models is important not only for researchers from various fields who will use the model outputs but also for modelers by facilitating further model development. An objective of the present paper is to identify whether there are improvements in the new models in simulating the monthly temperature extremes in the recent past climate; the projected future changes under the new scenarios are also addressed. The historical simulations of warm and cold months in the 24 CMIP3 models and 28 CMIP5 models are compared, and the projected future changes under the SRES and RCP scenarios are evaluated.

2. Datasets

Two station-based datasets are used to evaluate the performance of the models in this study: University of East Anglia’s Climate Research Unit (CRU) CRU TS3.1 (Mitchell and Jones 2005) and the Global Historical Climate Network (GHCN) + Climate Anomaly Monitoring System (CAMS; Fan and van den Dool 2008) from the Climate Prediction Center (CPC). Both datasets are gridded monthly-mean surface temperature datasets with a spatial resolution of $0.5^\circ \times 0.5^\circ$, and they cover the global land area. However, only the land area north of $60^\circ$S is considered for the observations and model output in this study because the CRU data in Antarctica are missing.

On the monthly scale, the temperatures in the reanalyses agree well with those from station observation data and can provide added value in poorly observed regions, such as high latitudes and parts of Africa (Dee et al. 2011a). In this study, the monthly-mean 2-m temperatures from the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Re-Analysis (ERA-Interim; Dee et al. 2011b) and NCEP Climate Forecast System Reanalysis (CFSR) (Saha et al. 2010) are also used for validation. Note that because the 2-m temperature is not assimilated in CFSR the monthly-mean value of the 6-hourly forecast is used instead.

The monthly-mean surface air temperatures from the 24 CMIP3 models are provided by the Program for Climate Model Diagnosis and Intercomparison. The twentieth-century climate simulations (20C3M) and the SRES B1, A1B, and A2 experiments are used for the present-day study and for future projections. However, the results from the SRES B1 and A2 experiments are only available for 21 and 19 CMIP3 models, respectively. GISS-EH, INGV-SXG, and HadGEM1 lack the SRES B1 simulations, whereas CGCM3.1-T63, FGOALS-g1.0, GISS-AOM, GISS-EH, and MIROC3.2 (hires) lack the SRES A2 simulations. The 28 CMIP5 model results are obtained through data portals of the Earth System Grid Federation. The historical RCP4.5 and RCP8.5 experiments from the 28 CMIP5 models are used in this study. Both the 20C3M and the historical experiments are driven by anthropogenic and natural forcings, and the simulation is also forced by a time-evolving land cover for most of the CMIP5 models (Taylor et al. 2012). Only one realization from each model is used here so that all models are weighted equally in the considered multimodel statistics. The institution and resolution information of each model and full model expansions are listed in Table 1, showing a generally higher horizontal resolution for the CMIP5 models.

3. Methods

There are many approaches to describing climate extremes. Thus, differing conclusions may be drawn. For extremes on the daily scale, such climate indices as warm nights (TN90P) and the diurnal temperature range (DTR)
<table>
<thead>
<tr>
<th>Sponsor, country</th>
<th>CMIP3 models</th>
<th>CMIP3 model expansion</th>
<th>CMIP5 models</th>
<th>CMIP5 model expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing Climate Center (BCC), China</td>
<td>—</td>
<td>—</td>
<td>BCC_CSM1.1 (128 x 64)</td>
<td>Beijing Climate Center, Climate System Model, version 1.1</td>
</tr>
<tr>
<td>Bjerknes Centre for Climate Research (BCCR), Norway</td>
<td>BCCR-BCM2.0 (128 x 64)</td>
<td>Bjerknes Centre for Climate Research Bergen Climate Model, version 2.0</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Beijing Normal University (BNU), China</td>
<td>—</td>
<td>—</td>
<td>BNU-ESM (128 x 64)</td>
<td>Beijing Normal University - Earth System Model</td>
</tr>
<tr>
<td>Canadian Centre for Climate Modelling and Analysis (CCCMA), Canada</td>
<td>CGCM3.1-T47 (96 x 48)</td>
<td>Canadian Centre for Climate Modelling and Analysis (CCCma) Coupled Global Climate Model, version 3.1-T47</td>
<td>CanESM2 (128 x 64)</td>
<td>Second Generation Canadian Earth System Model</td>
</tr>
<tr>
<td>CNRM-CM3 (128 x 64)</td>
<td>Centre National de Recherches Météorologiques Coupled Global Climate Model, version 3.1-T63</td>
<td>CNRM-CM5 (256 x 128)</td>
<td>Centre National de Recherches Météorologiques Coupled Global Climate Model, version 5</td>
<td></td>
</tr>
<tr>
<td>Norwegian Climate Center (NCC), Norway</td>
<td>—</td>
<td>—</td>
<td>NorESM1-M (144 x 96)</td>
<td>Norwegian Earth System Model, version 1 (intermediate resolution)</td>
</tr>
<tr>
<td>National Institute of Meteorological Research (NIMR), Korea</td>
<td>—</td>
<td>—</td>
<td>HadGEM2-AO (192 x 145)</td>
<td>Hadley Centre Global Environment Model, version 2 - Atmosphere and Ocean</td>
</tr>
<tr>
<td>National Center for Atmospheric Research (NCAR), United States</td>
<td>CCSM3 (256 x 128)</td>
<td>Community Climate System Model, version 3</td>
<td>CCSM4 (288 x 192)</td>
<td>Community Climate System Model, version 4</td>
</tr>
<tr>
<td>Institute of Numerical Mathematics (INM), Russia</td>
<td>PCM (128 x 64)</td>
<td>Parallel Climate Model</td>
<td>INM-CM4 (180 x 120)</td>
<td>Institute of Numerical Mathematics Coupled Model, version 4.0</td>
</tr>
<tr>
<td>Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia</td>
<td>CSIRO-Mk3.0 (192 x 96)</td>
<td>Commonwealth Scientific and Industrial Research Organisation Mark, version 3.0</td>
<td>ACCESS1.0 (192 x 145)</td>
<td>Australian Community Climate and Earth-System Simulator, version 1.0</td>
</tr>
<tr>
<td>Istituto Nazionale di Geofisica e Vulcanologia (INGV), Italy</td>
<td>INGV-SXG (320 x 160)</td>
<td>Istituto Nazionale di Geofisica e Vulcanologia, SINTEX-G</td>
<td>ACCESS1.3 (192 x 145)</td>
<td>Australian Community Climate and Earth-System Simulator, version 1.3</td>
</tr>
<tr>
<td>Max Planck Institute (MPI), Germany</td>
<td>ECHAM5 (192 x 96)</td>
<td>MPI-ESM-LR (192 x 96)</td>
<td>Max Planck Institute Earth System Model, low resolution</td>
<td></td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>MPI-ESM-MR (192 x 96)</td>
<td>Max Planck Institute Earth System Model, medium resolution</td>
</tr>
<tr>
<td>Sponsor, country</td>
<td>CMIP3 models</td>
<td>CMIP3 model expansion</td>
<td>CMIP5 models</td>
<td>CMIP5 model expansion</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------------</td>
<td>--------------------</td>
<td>------------------------------------------------------------</td>
<td>--------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Meteorological Institute of the University of Bonn (MIUB)/Korea Meteorological Administration (KMA), Germany/Korea</td>
<td>ECHO-G (96 x 48)</td>
<td>ECHAM and the global Hamburg Ocean Primitive Equation</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>IAP, China</td>
<td>—</td>
<td>—</td>
<td>FGOALS-g2 (128 x 60)</td>
<td>Flexible Global Ocean–Atmosphere–Land System Model gridpoint, second spectral version</td>
</tr>
<tr>
<td>Geophysical Fluid Dynamics Laboratory (GFDL), United States</td>
<td>FGOALS-g1.0 (128 x 60)</td>
<td>Flexible Global Ocean–Atmosphere–Land System Model gridpoint, version 1.0</td>
<td>FGOALS-s2 (128 x 108)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>GFDL-CM2.0 (144 x 90)</td>
<td>Geophysical Fluid Dynamics Laboratory Climate Model, version 2.0</td>
<td>GFDL-CM3 (144 x 90)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>GFDL-CM2.1 (144 x 90)</td>
<td>Geophysical Fluid Dynamics Laboratory Climate Model, version 2.1</td>
<td>GFDL-ESM2G (144 x 90)</td>
<td>—</td>
</tr>
<tr>
<td>Goddard Institute for Space Studies (GISS), United States</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>GISS-AOM (90 x 60)</td>
<td>Goddard Institute for Space Studies, Atmosphere–Ocean Model</td>
<td>GISS-E2-R (144 x 90)</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>GISS-EH (72 x 46)</td>
<td>Goddard Institute for Space Studies Model E, coupled with the HYCOM ocean model</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>GISS-ER (72 x 46)</td>
<td>Goddard Institute for Space Studies Model E-R</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>L’Institut Pierre-Simon Laplace (IPSL), France</td>
<td>IPSL-CM4 (96 x 72)</td>
<td>L’Institut Pierre-Simon Laplace Coupled Model, version 4</td>
<td>IPSL-CM5A-LR (96 x 96)</td>
<td>L’Institut Pierre-Simon Laplace Coupled Model, version 5, coupled with NEMO, low resolution</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>IPSL-CM5A-MR (144 x 143)</td>
<td>L’Institut Pierre-Simon Laplace Coupled Model, version 5, coupled with NEMO, mid resolution</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>IPSL-CM5B-LR (96 x 96)</td>
<td>L’Institut Pierre-Simon Laplace Coupled Model, version SB, coupled with NEMO, low resolution</td>
</tr>
</tbody>
</table>
have been widely used to indicate “moderate extremes” (Zhang et al. 2011). For a rare climate event, such as a 1-in-20-yr daily maximum temperature, the extreme value theory (EVT) has been applied to estimate the event’s return value or return level (Kharin et al. 2007). Given a rather small sample size (one value per year), the monthly-mean temperature extremes favored in this study are difficult to quantify nonparametrically. Because we define the extreme in terms of 20-yr return values of annual temperature extremes, the extreme can be estimated from the tail of a fitted parametric distribution. For warm months (the warmest month in each year), a generalized extreme value (GEV) distribution is fitted, and the 95th percentile is considered the warm extreme. For cold months (the coldest month in each year), the temperatures are multiplied by \(2^{1}\) before fitting a distribution; the sign of the 95th percentile of the fitted GEV distribution is then reversed again, thereby generating the cold extreme. The method of L-moments is applied to estimate the parameters of the GEV distribution (Hosking 1990), which is a feasible approach that can be universally performed for all the grid boxes in this study. An alternative to the application of EVT, and an easier way to define the extreme, is the mean plus (or minus) 1.645 standard deviation for the annual maxima (or minima) of the monthly temperatures (e.g., Becker et al. 2013), which is approximately equal to the 20-yr return value if the monthly-mean temperature for each grid box is assumed to be from a Gaussian distribution. However, such a definition can cause a biased estimation for extremes in regions where the monthly-mean temperatures are markedly nonnormal (von Storch and Zwiers 2002). To evaluate whether the two hypotheses in distribution are appropriate here, the Kolmogorov–Smirnov tests are performed on the null hypothesis that the annual extremes of the monthly means in a grid box are drawn from GEV or from a Gaussian distribution, with the parameters estimated by the L-moment method. The tests show that there is no need to reject either of the two distributions. At the 0.05 level, 5.18% and 5.67% of the global land fails the test for warm and cold extremes, respectively, when fitted using the GEV distribution. Bootstrapping, generated by resampling the global fields to preserve the spatial dependencies, is employed to quantify the sampling error, which may be large when the distribution is estimated from a small sample (Efron 1987). The variance of sampling errors generated from fitting the GEV distribution is slightly larger than from fitting the Gaussian distribution due to the estimation of more parameters in the former.

The time period 1980–99 is considered for the current climate simulation because it is covered by both the
reanalyses and CMIP model simulations. The time periods 2016–35 and 2080–99 are chosen for near-future and long-term projections, respectively. All the temperature data are preprocessed as anomalies to the annual mean climatology in 1980–99 to eliminate the possible influence caused by the systematic bias of model climatology because the extreme, rather than the mean climate, is considered here. The observation and model data are remapped onto a Gaussian N64 grid (256 × 128) using bilinear interpolation prior to the analysis. To study the extreme temperatures at different latitudes, the global land area is divided into four zonal bands in this study: the north frigid zone (NFZ; 66.5°–90°N), the north temperate zone (NTZ; 23.5°–66.5°N), the torrid zone (TZ; 23.5°S–23.5°N), and the south temperate zone (STZ; 66.5°–23.5°S). For regional study purposes, the temperature extremes are spatially averaged over 21 regions, similar to the approach employed by Giorgi and Francisco (2000; Fig. 1). This region definition has been widely adopted in previous evaluations based on CMIP3 models (e.g., Weisheimer and Palmer 2005; Stott et al. 2011).

4. Historical simulation of monthly temperature extremes

The spatially averaged 20-yr return values of monthly-mean temperatures are evaluated and presented in the left panel of Fig. 2. The red and blue boxes represent the CMIP3 and CMIP5 models, respectively, and the five bands in each box indicate the lower and upper bounds, the median, and the 25th and 75th percentiles of the multimodel ensemble (MME). The corresponding statistics for the observations and reanalyses are calculated from the bootstrap samples to illustrate the sampling error. Globally (land only and excluding Antarctica) and zonally averaged, the model-simulated that extreme monthly temperatures agree well with both the observations and reanalyses. The CFSR cold extremes in the NFZ are warmer than those in the observations and ERA-Interim; however, further investigation is required to determine whether this is because the 2-m temperatures are not analyzed in CFSR.

The good performance of the models for simulation on a large spatial scale is unsurprising because previous studies have indicated that globally averaged daily (Kharin et al. 2007) and seasonal (Anderson 2011) temperature extremes can be well simulated using CMIP3 models. Moreover, spatial smoothing reduces the sampling error at the expense of the loss of local information (Masson and Knutti 2011). Regardless, most models tend to simulate too high warm and too low cold extremes of monthly-mean temperature, and t tests performed on the global land mean extremes show that such bias is significant at the 5% level for both the CMIP3 and CMIP5 ensembles. However, the bias is much lower when the annual cycle is removed (Fig. 2, right). As a first-order approximation, the simulation of an overly strong annual cycle may account for most of the bias in extreme monthly-mean temperatures.

The sampling error is small on global and zonal scales (Kharin et al. 2007; Stott et al. 2011) and can be negligible when compared with the intermodel uncertainty interval. The difference between the two station-based datasets is larger than the sampling error, particularly for the TZ, a region for which the station spatial coverage is poor. The intermodel differences are larger at high-latitude regions, where the interannual temperature variability is greater. F tests are performed to determine whether the CMIP5 ensemble has a smaller intermodel variance than the CMIP3 ensemble; at the 5% level, this difference is only significant for the cold extremes in the TZ. As the first-order approximation, the uncertainty in the simulated seasonal cycle contributes to most of the intermodel variance. In a snow-covered
region such as the NFZ, the strength of seasonal cycle is correlated with the snow–albedo feedback (Hall and Qu 2006), and the large intermodel spread in the snow–albedo feedback found in the CMIP3 models (Qu and Hall 2007) partly explains the relatively greater intermodel difference in the temperature extremes in the NFZ. Removing the annual cycles, the extreme monthly-mean temperature mainly depends on the magnitude of the interannual variability. If we eliminate the bias on the seasonal cycle, the intermodel differences become smaller but are still larger than the sampling error. Consistent with previous research on the interannual variability (Scherrer 2011), the intermodel differences are greater over high latitudes. F tests are again performed between the CMIP3 and CMIP5 ensembles, with the results showing that CMIP5 has a significantly smaller intermodel spread than CMIP3 at a 5% level for the global and zonal mean extreme temperatures, with the exception of the warm extreme in the NFZ. It is worth mentioning that recent studies also suggest that improvements in the CMIP5 models can be found in the simulation of ENSO (Kim and Yu 2012; Zhang and Jin 2012), which greatly influences the global climate interannual variability.

To evaluate the overall performance in simulating extreme monthly-mean temperatures on the gridbox scale, the root-mean-square error (RMSE) between the global fields (land only and excluding Antarctica) is calculated with reference to the observations and reanalyses. The uncentered RMSE used here is defined as follows:

$$\frac{1}{W} \sum_i \sum_j w_{ij} (X_{ij} - Y_{ij})^2.$$  

The simulated and observed extremes correspond to X and Y, with the longitude and latitude dimensions denoted by i and j, respectively. The term W is the sum of the weights (w_{ij}) proportional to the gridbox area. In the interest of brevity, the individual outcomes obtained from the two station-based datasets and two reanalyses are not shown here (the differences among them are in fact small). Instead, the average of the four RMSEs with respect to the different station-based datasets and reanalyses is provided to represent the overall competence (Fig. 3). Note that the results in this study are obtained from temperatures on a Gaussian N64 grid (approximately 1.4° × 1.4°) and that the choice of spatial scale would have an impact on the outcome (e.g., Sakaguchi et al. 2012). On larger scales, the models tend to show better performance, and their RMSEs become smaller.

![Boxplots of spatially averaged (left) 20-yr return values for (top) warm and (bottom) cold extremes in 1980–99 and (right) those with climatological annual cycles removed.](image-url)
The RMSEs of the warm extremes of most of the models are approximately 2°C–3°C, which is greater than the global average of sampling errors at each grid box. In our study, the global averages of the warm and cold extreme sampling errors in CRU are 0.31°C and 0.59°C, respectively, similar to those (0.34°C and 0.54°C) estimated from the daily temperatures (Kharin et al. 2007). The RMSEs of the cold extremes are generally larger than those of the warm extremes, as is the sampling error. The MME means have smaller RMSEs than any of the individual models, as some biases may be canceled out in the ensemble (Räisänen 2007). The RMSE of the CMIP5 ensemble mean (blue) is smaller than for the CMIP3 (red), although the difference is minor. Some improvements can also be found in the simulation from the individual models. As shown in Fig. 3, of the 13
modeling groups employed in both CMIP3 and CMIP5, 10 modeling groups have smaller RMSEs for the cold extreme simulations in their latest versions compared with the older versions. Chi-squared tests show that such a difference is significant at the 10% level. For the warm extreme simulations, 8 of the 13 modeling groups show improvements. Because of the generally higher horizontal resolutions in the CMIP5 models, the topography is more detailed and thus benefits the temperature simulation on the gridbox scale (Masson and Knutti 2011). For example, the latest models from the Institute of Atmospheric Physics (IAP; FGOALS-g2 and FGOALS-s2) perform markedly better than their previous version (FGOALS-g1.0), according to the decrease in RMSE.

We also evaluate the simulations on the gridbox scale after the climatological annual cycles are removed from the extreme monthly-mean temperatures (not shown). Consistent with the above results, the RMSEs become smaller but are still larger than the sampling errors. Improvements in the CMIP5 ensemble mean and individual models are also detectable, implying a better representation of the interannual variability in the latest models.

To evaluate the simulations of the extreme monthly-mean temperatures in different regions, the relative error defined by Gleckler et al. (2008) is applied as a metric in this study. The relative error is a normalized RMSE, and thus a lower value indicates a better performance. Note that to identify the differences between the two ensembles, the normalization for RMSEs of both the CMIP3 and CMIP5 models are with respect to the multimodel ensemble median errors in the CMIP3 ensemble. According to this definition, a relative error less than zero would indicate a better performance than the median of the CMIP3 ensemble. Figure 4 shows the relative error of the extreme monthly-mean temperatures in the 24 CMIP3 and 28 CMIP5 models for the 21 regions compared using both observations and reanalyses. Each grid cell that represents a single model for a certain region is separated into four smaller boxes. The top corners represent the outcome of the CRU- and CPC-gridded station observations, and the bottom corners represent those for the ERA-Interim and CFSR reanalyses. In general, the differences between the observations and reanalyses play a minor role here, and the biases are systematic compared with the different reference data. Consistent with the findings above, the ensemble mean and ensemble median outperform most of the individual models in almost every region: the errors can be partly canceled through the multimodel mean or median, resulting in a better performance (Tebaldi and Knutti 2007).

Because the observational uncertainty discussed above does not exert much influence on our results, the relative errors with respect to the different station-based datasets and reanalyses are then averaged for the purposes of comparison. The improvements in the CMIP5 models on the regional scale can be found in some regions, particularly for cold extremes. The relative errors of the CMIP5 ensemble mean are smaller than those of the CMIP3 ensemble mean in 9 regions for warm extremes and 16 regions for cold extremes. The individual model simulations also indicate greater improvements for cold extremes than warm extremes. Chi-squared tests are performed on the number of 28 CMIP5 models having relative errors less than zero at each region, with significant improvements at the 5% level found in 4 and 6 regions for warm and cold extremes, respectively. Consistent with the above findings, the improvements still exist after the climatological annual cycles are removed.

5. Future projection of monthly temperature extremes

In this section, we begin by examining the projected future changes of the extreme monthly-mean temperatures in the near future (2016–35) under two scenarios (SRES B1 and RCP4.5 in the CMIP3 and CMIP5 experiments, respectively). Considering that scenario uncertainty does not play an important role in the next one to two decades (Hawkins and Sutton 2009), inter-scenario differences can be largely ignored when comparing the CMIP3 and CMIP5 ensembles. The change relative to that in the recent (1980–99) climate is often not significant due to the low signal-to-noise ratio. Thus, a method similar to that suggested by Tebaldi et al. (2011) is adopted here to determine the agreement of the models. For each model and its bootstrap samples, a Student’s t test (or Wilcoxon’s signed rank test) is performed for the derived 20-yr return value (or waiting period) grid point by grid point. The corresponding region is masked out in white if less than half the models show a significant change at the 10% level through t tests; otherwise, the corresponding region is stippled if 90% of the models agree on the direction of the change.

Figure 5 shows the MME mean change in the temperature extremes simulated by the CMIP3 (top) and CMIP5 (bottom) models with the intermodel agreements measured from the method discussed above. Except for several grid boxes, nearly no value in the global land area is masked out, as over half the models project a significant change in temperature extremes. Additionally, a positive change is projected by the ensemble mean for almost all the land areas around the globe, although the direction of the change lacks sufficient confidence at locations over the global land, as indicated by the un-stippled grid boxes in Fig. 5. The most substantial
change can be found in the snow-covered regions at the high latitudes of the Northern Hemisphere during cold months, a result that has also been found for extreme daily minimum temperatures and is attributed to snow retreat (Kharin et al. 2007). As a result, the global (land only and excluding Antarctica) mean cold extremes will warm faster than the warm extremes in the near future.

Despite the different scenarios applied in CMIP3 and CMIP5, the patterns of the warm and cold extreme changes are similar between the two ensemble means. The spatial correlations between the two fields are 0.63 for warm extremes and 0.89 for cold extremes.

Figure 6 shows the MME median of the near-future (2016–35) return periods for the temperature extremes under the current climate (1980–99). The waiting period for warm extremes is significantly shortened, whereas...
the cold extremes show a tendency to occur much more rarely. In the tropics and southern Eurasia, regions where the signal-to-noise ratio of warming is generally higher, many areas are projected to experience 20-yr return values of warm extremes in less than 2 years, indicating that the present-day extremes will become the norm in those regions. This result is in accordance with previous studies on extreme warm seasons from CMIP3 outputs (Anderson 2011; Diffenbaugh and Scherer 2011). The change is less significant at high latitudes, where the magnitude of interannual variability is larger, but it still shows a reduction in the waiting period by a factor of 2 to 4.

For cold extremes, the changes are most significant in tropical South America, Africa, and South Asia. Consistent with the findings of Räisänen and Ylhäisi (2011), the cold months in areas where the interannual variability is small become much less frequent given a modest warming. However, in parts of the Northern Hemisphere midlatitudes, the probability projected by the ensemble medians for the occurrence of cold extremes is only slightly less than under the current climate and not all the models agree on the direction of the change, which is in accordance with the discussion regarding Fig. 5. Yang and Christensen (2012) reported on the European cold winter months simulated by 13 CMIP5 models and suggest that the chance of cold winters occurring in Europe is only slightly reduced in the near future. In their study, the pattern of cold Europe/warm Arctic is detected, potentially indicating a connection between cold European winters and the decrease in Arctic sea ice. Note that cold winters differ from the cold extremes defined in our study, and further investigation is required to determine whether the proposed mechanism also applies here.

The average spatial change in extreme monthly-mean temperatures by the end of the twenty-first century (2080–99) under different scenarios in the CMIP3 and CMIP5 ensembles is shown in the left panel of Fig. 7. The change is greater under higher greenhouse gas emission scenarios for both the warm and cold global (land only and excluding Antarctica) and zonal mean (land only) extreme temperatures, with the greatest warming projected under the RCP8.5 scenario and the least under the B1 scenario. The CMIP5 RCP4.5 simulations project a lesser warming compared with A1B and A2, but it is slightly greater than B1. Globally averaged, the cold extremes will warm faster than the warm extremes under all five scenarios, and t tests show that the difference is significant at the 10% level under the B1, RCP4.5, and A1B scenarios. Changes in the snow cover and Arctic sea ice are mainly responsible for these differences, as the cold extremes in the NFZ warm much faster than in other zonal bands.

The projected changes in the extreme monthly-mean temperatures are similar to those for the extreme daily
temperatures (e.g., Kharin et al. 2007), as the contribution is mostly attributed to the change in the climatological maxima and minima of the annual cycle. The results after removing the changes in the annual cycles are shown in the right panel of Fig. 7. Roughly speaking, the changes in extremes with the climatological annual cycles removed are not significant, indicating that the interannual variability will most likely remain unchanged. This is also the case for the daily temperatures, with the exception of cold temperatures at high latitudes where the change in cold extremes exceeds the warming of the mean temperatures in the climatologically coldest months due to changes in the short-term temperature variability.

Consistent with Hawkins and Sutton (2009), the scenario uncertainty becomes dominant with time. As illustrated in Fig. 7, there is little overlap between the intermodel uncertainty ranges under scenarios B1 and RCP8.5 during 2080–99, and the intermodel differences are larger for the cold extremes in the NFZ than elsewhere, which may be related to the uncertainties in snow cover feedback. The uncertainty range of cold extremes in the NFZ is larger in the CMIP5 experiments than in the CMIP3 experiments, and F tests performed on the CMIP5 RCP8.5 and CMIP3 simulations show that the difference is significant at the 5% level. As the interannual variability changes little, the intermodel variability in the changes in extremes with the annual cycles removed is small.

By the late twenty-first century, the simulated changes in the regionally averaged extreme monthly-mean temperatures follow the mean changes in climatology, as discussed above, and the uncertainties are mainly derived from the scenarios. Thus, the change in extremes in proportion to the global mean warming may make additional sense. Figure 8 shows the change in the regionally averaged extremes scaled by the change in the global mean temperature. To indicate the intermodel differences, each grid cell is divided into five parts, representing the 5%, 25%, 50%, 75%, and 95% quantiles of MME. Scaled by the global mean temperature change, the changes in both the warm and cold extremes vary little under different scenarios, rather the geographical features become notable. The ratio is close to one for most of the regions at low latitudes [e.g., eastern Africa (EAF), western Africa (WAF), and Southeast Asia (SEA)], indicating that the changes in the warm and cold extremes in these regions approximate the magnitude of the global mean warming. Moreover, this result does not contradict previous studies on extreme seasonal mean temperatures that used similar regional definitions (Weisheimer and Palmer 2005; Stott et al. 2011). These studies suggested that low-latitude regions are expected to experience the largest increase in the frequency of extreme seasonal temperatures. Consistent with the above findings, the interannual variability is lower in these regions, and a small shift in the mean climatology would make the extreme become the norm. The change is generally greater for cold extremes than warm extremes at the moderate to high latitudes. In snow-covered regions ([North Asia (NAS), Northern
Europe (NEU), Greenland (GRL), and Alaska (ALA)), the large changes in cold extremes are most likely due to snow retreat, which is also the case for the extreme daily temperatures (e.g., Orlowsky and Seneviratne 2012). The change in warm extremes is greater in regions that become generally drier and is largest in Mediterranean (MED), a region where the heat stress risks are also projected to increase dramatically (Diffenbaugh et al. 2007), perhaps due to soil moisture effects (Hirschi et al. 2011).

6. Summary and discussion

The simulations of extreme temperatures in 24 CMIP3 models and 28 CMIP5 models are evaluated in this paper. Different from previous studies on extreme daily or seasonal temperatures, our study focuses on the annual extremes of monthly means, as extreme events on this time scale may exert a profound impact on both the natural environment and human society.

Validated through observations (CRU and CPC) and reanalyses (ERA-Interim and CFSR), the performance of the two MMEs in simulating the monthly temperature extremes under the current climate (1980–99) is evaluated. In general, both the CMIP3 and CMIP5 MMEs can simulate warm and cold extremes reasonably well. Note that the temperature extremes are calculated from the anomalies of the annual mean climatology of the individual models. Globally and zonally averaged, the models tend to overestimate the magnitude of temperature extremes; as first-order approximations, the biases in the climatological annual cycles contribute to most of this overestimation. When the annual cycles are removed, improvements in the CMIP5 models can be found, with a narrower ensemble spread in the CMIP5 ensemble than in CMIP3. This result may imply a better representation in interannual variability by the CMIP5 models. Improvements in simulating the regional temperature extremes can also be found, based on the overall performance, compared with the two station-based datasets and the two reanalyses.

The CMIP5 MME simulates a similar pattern in the change in the monthly extreme temperatures in the near future (2016–35), as does the CMIP3 ensemble. The patterns are also in agreement with the CMIP3 simulations for seasonal extremes (e.g., Diffenbaugh and Scherer 2011). The projections of temperature extremes in the late twenty-first century under the three SRES
FIG. 8. Regional-averaged change in (top) warm and (bottom) cold extremes scaled by global mean temperature change simulated by the CMIP3 and CMIP5 models. Within each scenario, the five grid cells from left to right represent the 5%, 25%, 50%, 75%, and 95% quantiles of the MME.
(B1, A1B, and A2) and the two RCP scenarios (RCP4.5 and RCP8.5) are compared. Consistent with previous studies on daily extremes (Kharin et al. 2007), the changes in the extreme monthly temperatures mainly follow the changes in the mean temperatures of the climatologically warmest (or coldest) months. However, the uncertainty range is still large for the change in cold extremes in snow-covered regions. The change in the extreme monthly mean temperatures scaled by global mean temperature changes is examined to determine the response in regional temperature extremes, regardless of the scenario uncertainty. Robust regional changes are projected in regions possibly affected by snow retreat or soil moisture feedback, which is in accordance with studies on daily extremes (e.g., Orlowski and Seneviratne 2012).

Although the sampling errors for the spatial means of temperature extremes are negligible compared with the intermodel differences and scenario uncertainty, the local sampling errors are large for some regions, thus reducing confidence in the near-future predictions. In addition, the signal-to-noise ratio is lower for the near-future change in temperature extremes than for the long-term projections due to the weaker signal. As a new endeavor for CMIP5, the decadal experiment assimilates the initial conditions and may provide the possibility for a more skillful prediction of the mean (Fyfe et al. 2011) and extreme temperatures (Eade et al. 2012). Our future studies will be conducted on the initialized decadal predictions simulated by the CMIP5 models.

Acknowledgments. This work is supported by the National Basic Research Program of China (2010CB950501-03 and 2009CB421407) and the National Natural Science Foundation of China (41275076 and 40905046). We acknowledge the World Climate Research Programme’s Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modeling groups (listed in Table 1 of this paper) for producing and making available their model output. For CMIP, the U.S. Department of Energy’s Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led the development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We thank the three anonymous reviewers for constructive suggestions.

REFERENCES


