Extreme Precipitation Indices over China in CMIP5 Models. Part I: Model Evaluation

Zhihong Jiang, Wei Li, Jianjun Xu, Laurent Li

Extreme Precipitation Indices over China in CMIP5 Models. Part I: Model Evaluation

ZHIHONG JIANG AND WEI LI

Key Laboratory of Meteorological Disaster of Ministry of Education, Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, Nanjing, China

JIANJUN XU

Environmental Science and Technological Center, College of Science, George Mason University, Fairfax, Virginia

LAURENT LI

Laboratoire de Météorologie Dynamique, CNRS, Université Pierre et Marie Curie, Paris, France

(Manuscript received 1 February 2015, in final form 26 July 2015)

ABSTRACT

Compared to precipitation extremes calculated from a high-resolution daily observational dataset in China during 1960–2005, simulations in 31 climate models from phase 5 of the Coupled Model Intercomparison Project (CMIP5) have been quantitatively assessed using skill-score metrics. Four extreme precipitation indices, including the total precipitation (PRCPTOT), maximum consecutive dry days (CDD), precipitation intensity (SDII), and fraction of total rainfall from heavy events (R95T) are analyzed. Results show that CMIP5 models still have wet biases in western and northern China. Especially in western China, the models' median relative error is about 120% for PRCPTOT; the 25th and 75th percentile errors are of 70% and 220%, respectively. However, there are dry biases in southeastern China, where the underestimation of PRCPTOT reach 200 mm. The performance of CMIP5 models is quite different between western and eastern China. The simulations are more reliable in the east than in the west in terms of spatial pattern and interannual variability. In the east, precipitation indices are more consistent with observations, and the spread among models is smaller. The multimodel ensemble constructed from a selection of the most skillful models shows improved behavior compared to the all-model ensemble. The wet bias in western and northern China and dry bias over southeastern China are all decreased. The median of errors for PRCPTOT has a decrease of 69% and 17% in the west and east, respectively. The good reproduction of the southwesterlies along the east coast of the Arabian Peninsula is revealed to be the main factor explaining the improvement of precipitation patterns and extreme events.

1. Introduction

It is well known that a global-scale warming dominates climate change over the past century. According to the Fifth Assessment Report (AR5) of the Intergovernmental Panel on Climate Change (IPCC; IPCC 2013), a changing climate significantly impacts the frequency, intensity, spatial extent, duration, and timing of weather and climate extremes. A warming climate has been shown to exacerbate and trigger certain climate extremes, including increases in severe and extreme precipitation events (Easterling et al. 2000; Zhai et al. 2005; Qian et al. 2007; Feng et al. 2011). China is especially vulnerable to extreme precipitation events, which can cause huge losses for society, the economy, and natural ecosystems. Public awareness of extreme events has risen sharply in recent years because of the increasing catastrophic influence of natural hazards. Meteorological disasters are estimated to cause a loss of

Corresponding author address: Zhi-Hong Jiang, Key Laboratory of Meteorological Disaster of Ministry of Education, Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing University of Information Science and Technology, 219 Ningliu Rd., Nanjing 210044, China.
E-mail: zhjiang@nuist.edu.cn

DOI: 10.1175/JCLI-D-15-0099.1

2.37% for the Chinese gross domestic product each year since 1990, and the frequency of extreme events is predicted to increase, accompanying climate warming (Wang and Zheng 2012). Thus, assessment and investigation of such events are of importance for developing adaptation strategies to reduce risks of precipitation extremes (Zhai et al. 2008; Ren et al. 2011; Wang et al. 2012).

Global climate models are the main tools used by the scientific community to reproduce the current climate and project future changes of extreme precipitation events. Given the cascade of uncertainties in models' projections of future change, it is essential to evaluate the models' performance in simulating extremes with respect to observations before assessing projection in the future. During recent years, considerable efforts have been made to assess the ability of models in simulating extreme changes in terms of spatial and temporal changes over Chinese regions based on models from the World Climate Research Programme's (WCRP) phase 3 of the Coupled Model Intercomparison Project (CMIP3). Jiang et al. (2009, 2012) showed that models have certain abilities to simulate both the spatial distribution and trend of extreme precipitation indices, but because of the limitation of coarse spatial resolution and other uncertainties, the simulation results still show many discrepancies. Xu et al. (2011) showed that models had limited skills in reproducing the interannual variation of extreme precipitation events. According to the study by Li et al. (2010), all the CMIP3 models underestimate extreme precipitation, especially during summer in eastern China, where the extreme precipitation is underestimated by around 50%. The multimodel ensemble (MME) method, defined as the average of simulations from multiple models, has been widely applied in the evaluation of models and projection of future climate because it can reduce the uncertainty from individual models and shows superior behaviors compared to any single model (Palmer et al. 2005; Thomson et al. 2006; IPCC 2007; Jiang et al. 2009). For future projections, an MME performing well in the present-day climate predicts more converged future changes. However, not all models have the ability to simulate specific variables over the target regions; the best ensemble mean cannot be achieved without looking into each member model's performance (Sun and Ding 2008). A suite of models performing substantially better than others should be selected and combined together to improve the credibility of projection. (Schmittner et al. 2005; Pierce et al. 2009; Knutti 2010; Chen et al. 2011; Seo and Ok 2013).

The improvement of climate models is always a challenge for the modeling community. Recently, models with more sophisticated physics and higher resolution from phase 5 of CMIP (CMIP5) are available. Compared to the previous phase (CMIP3), some 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), therefore hopefully resulting in better skill in representing current climate conditions. Sillmann et al. (2013a), who investigated the performance of CMIP5 models in simulating extremes indices on a global scale, reported that models tend to simulate more intense precipitation and fewer consecutive wet days. For China, Ou et al. (2013) showed that extreme precipitation is generally overestimated by most models, especially in western China and in the mountainous regions, while the simulated climatology of extreme precipitation in eastern China is fairly well simulated. However, this study only gives the distribution of errors for individual models in simulating extreme events, and there is no quantitative comparison on the spatial pattern and interannual variability between different CMIP5 models. How well does each model capture the observed changes of extreme events? Which models should be selected as the best performers based on a set of skill metrics? To our knowledge, such questions were rarely addressed in the available literature. Furthermore, we believe that a comparison between different ensembles comprising a selected set of good models, all models and a set of less-skillful models is useful to reduce models uncertainties and to enhance our confidence of projected climate change. To do so, we need to establish quantitative metrics measuring the performance in simulating the spatial pattern and temporal variability of extreme precipitations.

In addition, because of the difference of topography and atmospheric circulation affecting the extreme precipitation between eastern and western China, an evaluation of the models' abilities in simulating extreme precipitation should be done separately for eastern and western China, which can provide a more useful reference for the models' selection and allow us to make analyses concerning future climate change with greater confidence.

Therefore, the goal of this work is to quantitatively evaluate the capability of CMIP5 models in representing the present-day extreme precipitation over eastern and western China in terms of spatial and temporal variation using different skill-score metrics. We will also select the most skillful and less-skillful models in the two regions, based on the models' overall rankings. Furthermore, the possible cause of the models' biases will be discussed in different ensembles.

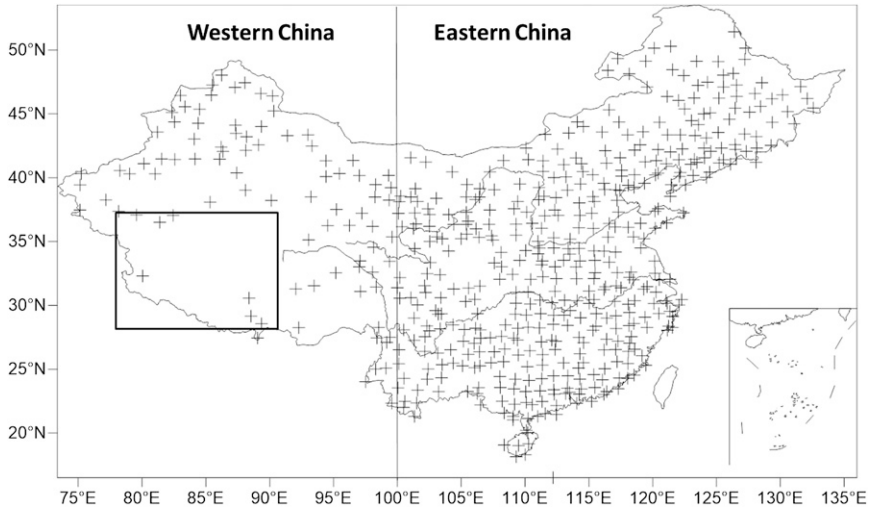

FIG. 1. Locations of the 592 stations with daily precipitation over mainland China. The vertical solid line indicates the demarcation between eastern and western China. The box over western China contains few observations and was excluded.

Our study is organized as follows. Section 2 introduces the data and methods. The model simulation of precipitation extremes over China is presented in section 3. The performance of selected models is described in section 4. Finally, section 5 gives a summary and discussion.

2. Data and diagnostic methodology

a. Observations and model datasets

Because of limitations of spatial inhomogeneity in station observations, gridded datasets with an appropriate homogenization and interpolation methodology are generally preferred for evaluating model simulations (Kiktev et al. 2003; Alexander et al. 2006). Ou et al. (2013) studied the sensitivity of extreme indices to horizontal resolution and concluded that calculating extreme indices from gridded precipitation is more appropriate than from station-based data. Therefore, the observational data used in this study are from a high-quality daily gridded precipitation dataset with 0.5° × 0.5° resolution deduced from 592 station observations in China (Fig. 1) during 1960–2005. The datasets are produced by Chen (2010). The location of stations is shown in Fig. 1. (This gridded dataset can be freely obtained by contacting Dr. Deliang Chen through http://rcg.gvc.gu.se/data/ChinaPrecip/prepdata/.)

The simulated daily precipitation accumulation (pr) data from 31 CMIP5 models from 18 different modeling institutes were retrieved through data portals of the Earth System Grid Federation. The historical experiment represents present-day climate, and only one realization is analyzed from each model. To be consistent with observations, the historical run used in this study is selected for the same period from 1961 to 2005. The institutions, model names, and resolution information of each model are listed in Table 1.

The horizontal resolution differs from model to model. We first calculate the extreme precipitation indices from different models at their native grids. Indices from observations are also calculated on the original grid of 0.5° × 0.5°. To facilitate model intercomparison and validation against observations, we interpolate all indices to a common 1° latitude by 1° longitude grid using a bilinear interpolation scheme.

Because the density of station distribution is different in western and eastern China (Fig. 1), and the atmospheric general circulation is also significantly different in those two regions (Zhang et al. 1984), we divide the gridded indices for models and observations into two parts: eastern China (east of 100°E) and western China (west of 100°E). Since there are few instrumental stations located in the western part of the Tibetan Plateau (south of 37°N and west of 92°E), this subregion is also excluded from our analysis. Finally, there are 525 grid cells of 1° × 1° within eastern China and 325 within western China.

The monthly geopotential height at 500 hPa and zonal (U) and meridional (V) components of the wind at 850 hPa are taken from the 31 CMIP5 models to be used in our study to search for possible relation with the models' biases. The reanalysis datasets are from the 40-yr European Centre for Medium-Range Weather Forecasts (ECMWF) Re-Analysis (ERA-40) and the ECMWF interim reanalysis (ERA-Interim) for the

TABLE 1. Model institution identification (ID), modeling center and country, model name, and atmospheric resolution of 31 CMIP5 global climate models. (Expansions of acronyms are available at http://www.ametsoc.org/PubsAcronymList.)

Institution ID	Modeling center and country	Model name	Atmospheric resolution (lat × lon)
CSIRO–BoM	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia	ACCESS1.0	1.25° × 1.875°
		ACCESS1.3	1.25° × 1.875°
BCC	Beijing Climate Center, China Meteorological Administration, China	BCC_CSM1.1	2.8125° × 2.8125°
		BCC_CSM1.1(m)	1.125° × 1.12°
GCESS	College of Global Change and Earth System Science, Beijing Normal University, China	BNU-ESM	2.8° × 2.8°
CCCma	Canadian Centre for Climate Modelling and Analysis, Canada	CanESM2	2.8° × 2.8°
NCAR	National Center for Atmospheric Research, United States	CCSM4	1.25° × 0.94°
National Science Foundation (NSF)–DOE–NCAR	Community Earth System Model contributors, United States	CESM1(CAM5)	1.25° × 0.94°
CMCC	Centro Euro-Mediterraneo per I Cambiamenti Climatici, Italy	CMCC-CM	0.75° × 0.75°
		CMCC-CMS	1.875° × 1.875°
CNRM–CERFACS	Centre National de Recherches Météorologiques–Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, France	CNRM-CM5	1.4° × 1.4°
EC-EARTH	EC-EARTH consortium	EC-EARTH	1.125° × 1.125°
LASG–CESS	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and Center for Earth System Science, Tsinghua University, China	FGOALS-g2	2.8° × 3°
LASG-IAP	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences, China	FGOALS-s2	2.8° × 1.4°
NOAA/GFDL	NOAA/Geophysical Fluid Dynamics Laboratory, United States	GFDL CM3	2.5° × 2.0°
		GFDL-ESM2G	2.5° × 2.0°
		GFDL-ESM2M	2.5° × 2.0°
MOHC	Met Office Hadley Centre, United Kingdom	HadCM3	2.5° × 3.75°
		HadGEM2-CC	1.875° × 1.25°
		HadGEM2-ES	1.875° × 1.25°
IPSL	L'Institut Pierre-Simon Laplace, France	IPSL-CM5A-LR	3.75° × 1.895°
		IPSL-CM5A-MR	2.5° × 1.27°
MIROC	National Institute for Environmental Studies, The University of Tokyo, Japan	MIROC4h	0.56° × 0.56°
		MIROC5	1.40625° × 1.40625°
		MIROC-ESM	2.8125° × 2.8125°
		MIROC-ESM-CHEM	2.8125° × 2.8125°
MPI-M	Max Planck Institute for Meteorology, Germany	MPI-ESM-LR	1.875° × 1.875°
		MPI-ESM-MR	1.875° × 1.875°
		MPI-ESM-P	1.875° × 1.875°
MRI	Meteorological Research Institute, Japan	MRI-CGCM3	1.125° × 1.125°
NCC	Norwegian Climate Centre, Norway	NorESM1-M	1.8725° × 2.5°

period 1960–2012. All models and reanalyses are interpolated to 2.5° × 2.5° resolution to facilitate the comparison.

b. *Extreme precipitation indices*

As already used in previous studies (Frich et al. 2002; Klein Tank and Konnen 2003; Zhang et al. 2011; Li et al. 2013), three indices representing the extreme events are used in this research (Table 2). The maximum consecutive dry days (CDD) represents the dry part of the precipitation spectrum, and the simple daily intensity index (SDII) indicates the wet part. The percentage of the total rainfall from events exceeding the long-term 95th percentile (R95T) represents strong precipitation events. In addition, the total precipitation (PRCPTOT) is also used to measure the climatology (Moberg et al. 2006), which may help to explain the variation of extreme precipitation. These indices are generally considered effective in extracting climate change information, are highly sensitive to global warming, and have been widely used to identify and monitor extreme precipitation for the IPCC AR5 (Xu et al. 2011;

TABLE 2. Indicator, acronym, and definition of four indices used in the study.

Indicator	Acronym	Definition
Total precipitation	PRCPTOT	Let R_{wj} be the daily precipitation amount for day w of period j. Then the total climatological precipitation in period j is $\text{PRCPTOT}_j = \sum_{w=1}^{W} R_{wj}$.
Fraction of total rainfall from events exceeding the long-term 95th percentile	R95T	Let R_j be the sum of daily precipitation amount for period j, R_{wj} be the daily precipitation amount for wet day($R > 1$ mm) of period j, and $R_{wn}95$ the 95th percentile of precipitation for wet days in the specified period. Then R95T_j is determined as $\text{R95T}_j = \sum_{w=1}^{W} R_{wj}/R_j$, $R_{wj} > R_{wn}95$.
Precipitation intensity	SDII	Let R_{wj} be the daily precipitation amount for wet day w ($R > 1$ mm) of period j. Then the mean precipitation amount for wet days is $\text{SDII}_j = \sum_{w=1}^{W} R_{wj}/W$.
Maximum consecutive dry days	CDD	Let R_{ij} be the daily precipitation amount for day i of period j. Then count the largest number of consecutive days where $R_{ij} < 1$ mm.

Sillmann et al. 2013a,b; Zhang et al. 2013; Zhou et al. 2014). All extreme indices are calculated with the Statistical and Regional Dynamical Downscaling of Extremes for European Regions (STARDEX) diagnostic extremes indices software (Haylock et al. 2006).

c. Model performance metrics

1) TAYLOR DIAGRAM

To evaluate the overall skill in reproducing the spatial pattern of the present-day extreme precipitation, the Taylor diagram (Taylor 2001) is used, which can provide a statistical summary of comparisons between simulations and observations in terms of their spatial correlation coefficient, their centered pattern root-mean-square (RMS) difference, and the ratio of spatial standard deviations of the model and observations. The spatial correlation coefficient is the quantity that measures the degree of phase agreement of two fields. The centered pattern RMS difference is the quantity that measures the degree of agreement in amplitude. The centered pattern RMS difference and the standard deviations of indices are normalized by the corresponding observations. A perfect simulation would be that the centered pattern RMS error is equal to 0 and both the spatial correlation and ratio of spatial standard deviations are close to 1.

2) INTERANNUAL VARIABILITY SKILL SCORE (IVS)

In terms of temporal variation, we know that it is not possible for models to reproduce the interannual variation of observations. But we can expect that models can reproduce the temporal standard deviation, which is evaluated by a measure of skill score IVS described by Chen et al. (2011) as follows:

$$\text{IVS} = \left(\frac{\text{STD}_m}{\text{STD}_o} - \frac{\text{STD}_o}{\text{STD}_m}\right)^2, \tag{1}$$

where STD_m and STD_o denote the interannual standard deviation of model simulations and observations, respectively. Smaller IVS values indicate a better agreement between the simulations and observations.

3) COMPREHENSIVE RATING METRICS (MR)

Each extreme precipitation index should allow us to rank models on the basis of their Taylor diagram and IVS. An overall ranking considering all extreme indices can be obtained with a comprehensive rating index MR, which is defined as

$$\text{MR} = 1 - \frac{1}{nm}\sum_{i=1}^{n} \text{rank}_i, \tag{2}$$

where m is the number of models, and n is the number of indices. The rank of the best-performing model is 1; the worst model has 31 for its rank. Therefore, the closer to 1 the value of MR is the higher the skill of the simulation.

3. Model evaluation

In this section, we quantitatively assess the performance of individual climate models in reproducing the climatological spatial pattern and interannual variability of indices over eastern and western China, respectively.

a. Regional averages of indices over the two regions

To have a general appreciation of models' capabilities in simulating precipitation extremes, annual mean values of the four indices in western and eastern China during the period 1961–2005 are first calculated. Relative errors of each model, with respect to the observation, are then evaluated and shown in Fig. 2. Figure 2 is a color-coded "portrait diagram" showing relative errors of each model for all indices, the color shading and the number within each box represents the corresponding values. Some striking features can be observed. First, compared to CMIP3 models (Jiang et al. 2012; Chen et al. 2011), large wet biases are still observed in CMIP5 models, with negative biases in CDD and positive biases in PRCPTOT for all models, especially over western China. Second, the relative errors of PRCPTOT and

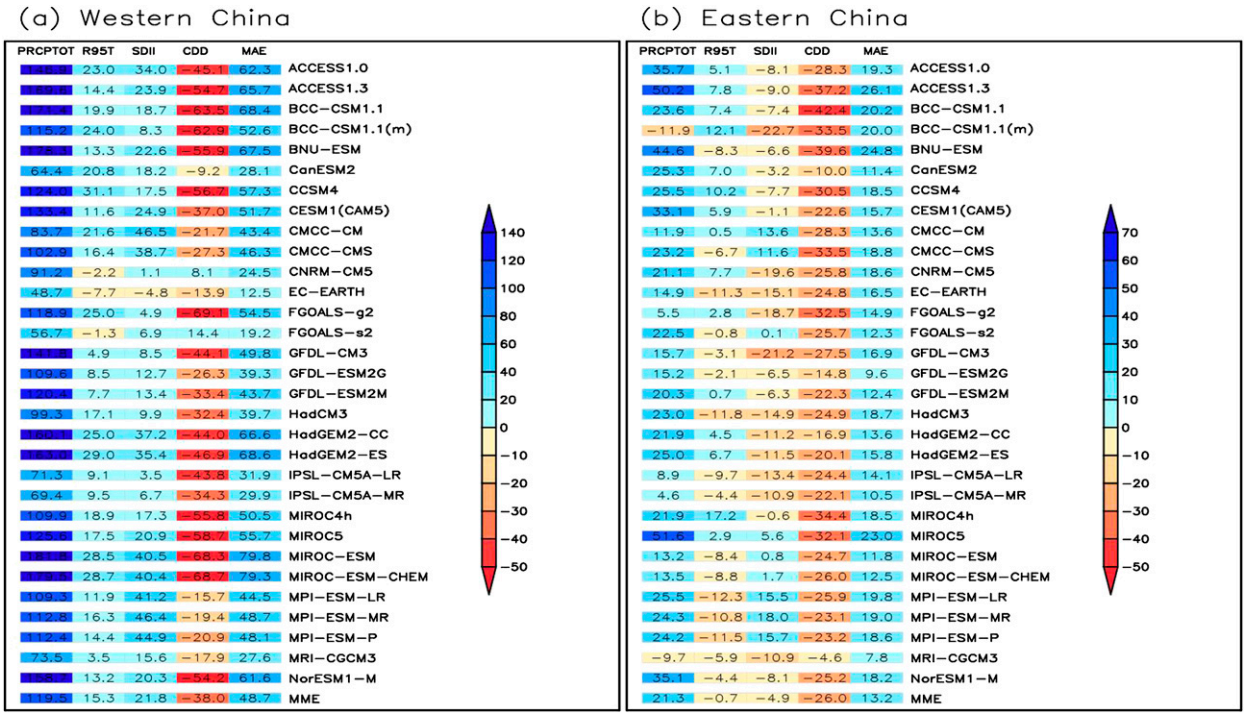

FIG. 2. Relative errors of simulated extreme precipitation indices averaged over China and the MAE during 1961–2000 for each model [(modeled − observed)/observed × 100]: (a) western and (b) eastern China. Numbers within the shaded boxes indicate the corresponding values.

CDD in western China are far larger than those in eastern China (Fig. 2a versus Fig. 2b). For example, biases of PRCPTOT vary from −11.9% to +50% in the east, while the counterparts are ranged from +48% to 180% in the west. These results are consistent with the finding in Su et al. (2013), who pointed out that CMIP5 models overestimate the precipitation in the Tibetan Plateau by 62%–183% in different model ensembles that they used. Third, the indices SDII and R95T are relatively well simulated compared to PRCPTOT and CDD in both eastern and western China, as the amplitudes of biases for SDII and R95T are smaller than those for PRCPTOT and CDD.

Furthermore, in eastern China (Fig. 2b), all models present positive biases for PRCPTOT except for BCC_CSM1.1(m) and MRI-CGCM3, but all models have negative biases for CDD. The underestimation of CDD indicates that the number of wet days in models is overestimated. Therefore, although PRCPTOT is overestimated, most models tend to underestimate SDII, and the multimodel mean R95T is also slightly underestimated. Unlike in eastern China, the other three indices except CDD are overestimated by all models in western China (Fig. 2a). The positive bias in SDII is caused by the fact that the overestimation of PRCPTOT has larger influence on SDII than does the overestimation of wet days. The multimodel mean overestimation of PRCPTOT and R95T can reach 120% and 15%, respectively. In addition, models from the same group exhibit a coherent performance for most indices, such as the models from the Max Planck Institute for Meteorology (MPI-M), L'Institute Pierre-Simon Laplace (IPSL), and the NOAA/Geophysical Fluid Dynamics Laboratory (NOAA/GFDL), but there exists a significant discrepancy between the models with different resolutions. For example, the three pairs of models—BCC_CSM1.1 and BCC_CSM1.1(m); CMCC-CM and CMCC-CMS; and MIROC4h and MIROC5—reveal that models with higher resolution have a better simulation skill of precipitation amount and wet days than models with lower resolution (Feng et al. 2011; Sillmann et al. 2013a). But models with higher resolution obtain a large amount of R95T in eastern China, which would come from a more active rainband of the East Asian summer monsoon in the higher-resolution models (Kusunoki et al. 2006, 2011).

In summary, most models reproduce mean indices more accurately in eastern China than in western China. To have an overall appreciation, we calculate the mean absolute value of the relative errors (MAE; last color-coded column in Figs. 2a,b) that consists of calculating the mean of the four relative errors without considering

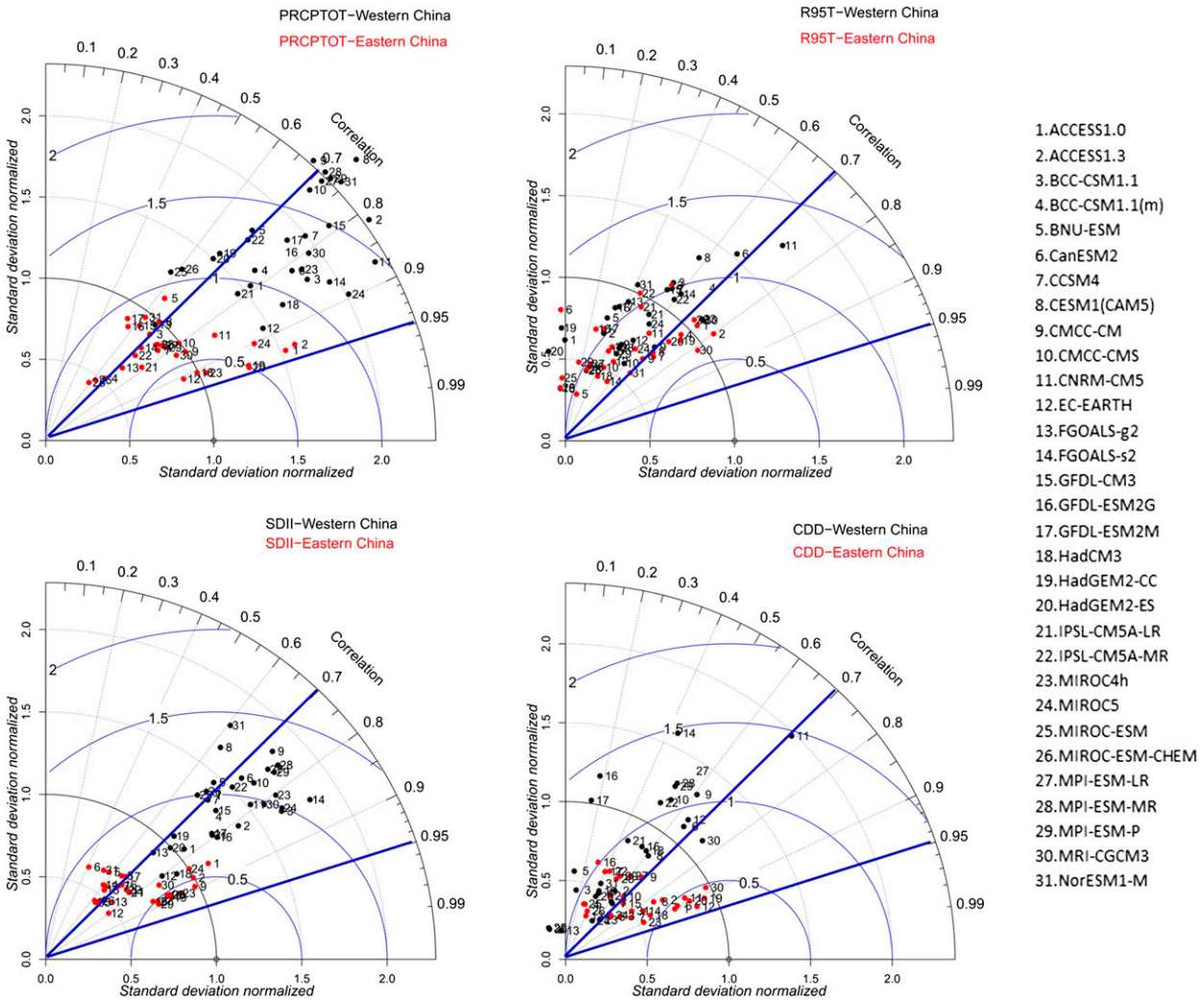

FIG. 3. Taylor diagrams of four indices between observation and 31 models over western (black) and eastern China (red). On the Taylor diagrams, angular axes show spatial correlations between modeled and observed fields; radial axes show spatial standard deviation (root-mean-square deviation), normalized against that of the observations. Each dot represents a model, identified by its number on the right. For the models located between the two blue lines, correlations are between 0.7 and 0.95.

the sign of their values. Based on MAE, MRI-CGCM3, IPSL-CM5A-MR, and GFDL-ESM2G seem to produce better results than other models in eastern China. However, indices in western China are poorly represented by most models except EC-EARTH.

b. Evaluation for spatial variation

We use Taylor diagrams to assess the performance of models in simulating the spatial pattern of indices. Figure 3 shows the Taylor diagram of the model simulations against observations, with red for the east and black for the west. The majority of models have spatial correlation (dotted radial lines) with the reference data between 0.7 and 0.95 (thick blue lines) in the two parts for PRCPTOT and SDII. This indicates that the coupled models have a reasonable performance in simulating the spatial distribution of PRCPTOT and SDII. However, similar to the distribution of PRCPTOT (Fig. 3, top left), the pattern of SDII (Fig. 3, bottom left) is clearly separated by the ratio of variance (dotted circular lines): most models have a ratio of variance between 0.5 and 1.0 in eastern China, but the majority of models have a ratio of variance larger than 1.5 in western China, indicating the simulated spatial variation is larger than observation in western China and smaller than observation in the east. All models have the centered normalized RMS difference expressed by the blue solid line below 1.0 in eastern China, but almost all models for PRCPTOT and half of models for SDII are larger than 1.0 in

western China, indicating the amplitude of biases for PRCPTOT and SDII are higher in the west than in the east.

In contrast, for heavy rainfall events (R95T), the spatial correlation significantly decreases (Fig. 3, top right). Only 8 of the 31 models have correlation coefficients between 0.7 and 0.95 in eastern China, while all models' coefficients are smaller than 0.7 in western China. The correlation coefficients of three models (HadGEM2-ES, HadGEM2-CC, and ACCESS1.0) in the west and three other models (MIROC-ESM, MIROC-ESM-CHEM, and CanESM2) in the east are even out of the range for the figure. This indicates that coupled climate models have less skill to simulate the extremes in the two regions compared to PRCPTOT and SDII. As for the ratio of variance, the simulated spatial variance of almost all models is smaller than for observations in the east, and those of half of the models are larger than for observations in the west. However, the centered normalized RMS difference has similar values between western and eastern China.

For CDD (Fig. 3, bottom right), the correlation coefficients in half of the models exceed 0.7 in the east, while all models' correlations are smaller than 0.7 in the west. In some models, such as MIROC-ESM and MIROC-ESM-CHEM, the correlations are of negative signs. The models are generally more skillful at simulating CDD in eastern China than in western China. Similar to R95T, the simulated spatial variance of all models is smaller than for observations in the east, and half of the models have larger spatial variance than for observations in the west with small correlation coefficients. Half of the models have a centered normalized RMS difference larger than 1 in western China.

Additionally, the centralized distribution in the east for all indices implies a small intermodel spread. In contrast, the much larger intermodel spread in the west is observed because of the loosely scattered distribution in the Taylor diagram. This indicates that the models differ widely in their simulation ability to reproduce the spatial variations of extreme indices in the west, and the difference between models is not too large in the east. In other words, there is a large uncertainty among models when simulating the spatial pattern in western China.

In summary, all models have a better performance simulating the spatial pattern of all indices in eastern China (compared to that in western China), as there is a small model spread and the ratio of variance is closer to 1.0 in the east. Generally speaking, models have less skill for simulating R95T and CDD in the east, which indicates that CMIP5 models have certain difficulties with simulating heavy precipitation and consecutive dry days. In contrast, all indices are poorly simulated in western China. To quantify the capability of individual models to reproduce observations, three aspects (pattern correlation, spatial standardized deviation ratio, and root-mean-square difference) of the Taylor diagram are first examined, and then the MR are calculated for the purpose of obtaining a comprehensive ranking for each model. Figure 4 displays the models' rankings in terms of pattern correlation (left), spatial standardized deviation ratio (center), and root-mean-square difference (right) for each index. The model's rank is assigned based on the MR defined before. The models' capabilities are ordered from top to bottom by model rank, with smaller numbers (red) indicating the model is more reliable. It is interesting to note that the three statistical metrics and four precipitation indices indicate a similar rank of models in the east. The five leading models are HadGEM2-CC, MIROC4h, HadGEM2-ES, MRI-CGCM3, and CMCC-CM. In contrast, in the west, the rank of individual models is sensitive to the three components of the Taylor diagram and the indices. Based on the MR, the best five models are EC-EARTH, IPSL-CM5A-LR, MRI-CGCM3, HadCM3, and MIROC4h. The performance of models in simulating spatial variation is quite different in the two regions.

c. *Evaluation for interannual variability*

The performance in simulating temporal variation is also a very important factor to measure the capability of models. Here, the IVS skill score defined in section 2c is used to quantify the similarity of interannual variability between modeled and observed variables. IVS of the extreme indices averaged over the two parts of China are calculated for 1960–2005. Results for each model are shown in Fig. 5.

Similar to the result of the Taylor diagram mentioned above, the interannual variability of all indices is better simulated in the east than in the west, as the IVS values in the east are closer to 0 than those in the west, and the values from models have a larger spread in the west than in the east. Take CDD as an example: IVS are ranged from 0 to 3.0 in the east but between 0 and 8.0 in the west.

The interannual variability of indices in the west is poorly expressed, with a large spread of IVS values among models (Fig. 5a). The indices are not equally well modeled in the west: PRCPTOT and CDD are the two indices with the highest IVS values, and R95T and SDII are relatively well reproduced. We then average the rank across all indices to get an overall measure of the interannual variability skill. Models with a higher rank perform better than those with a lower rank.

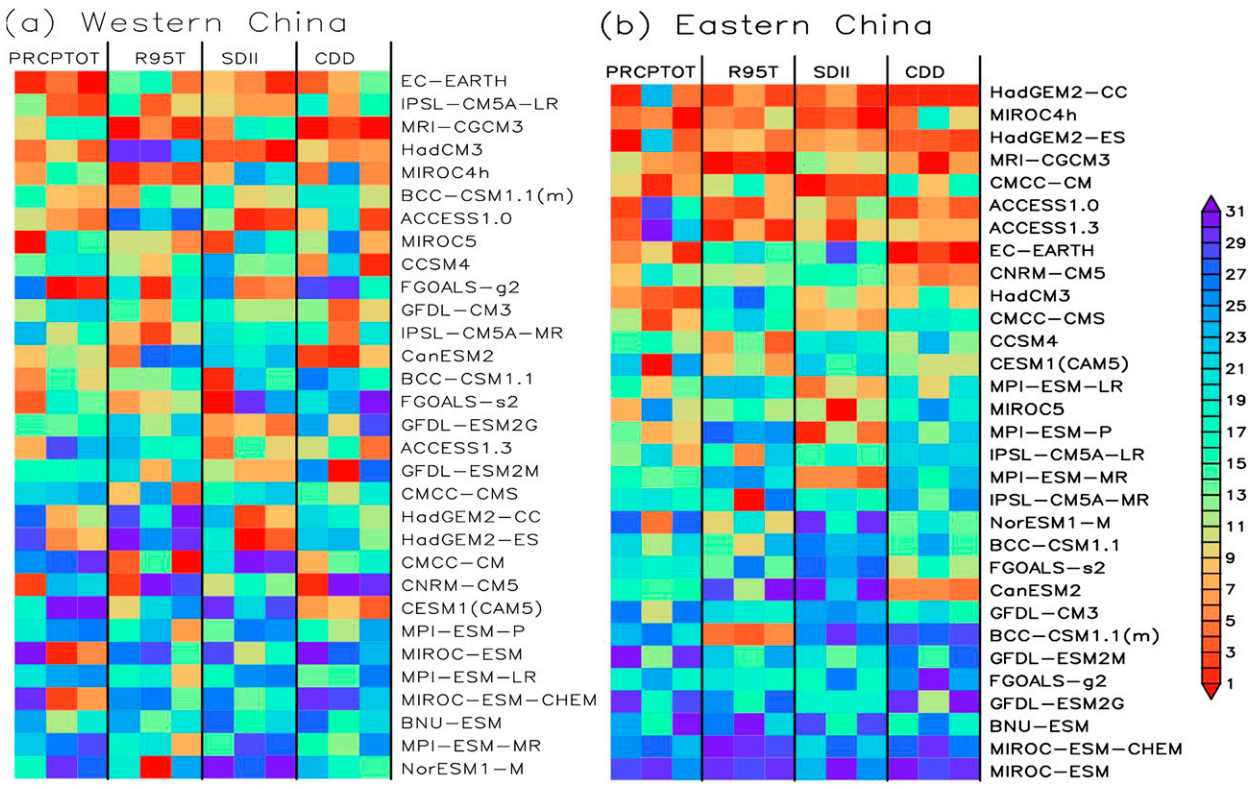

FIG. 4. The portrait diagram for each index of the rank of correlation (left), spatial standardized deviation ratio (center), and root-mean-square difference (right) over (a) western and (b) eastern China. The models names listed on the right in each panel follow their averaged rank. Colors as marked in the label bar indicate a model's rank for each item.

MRI-CGCM3, EC-EARTH, HadCM3, CMCC-CM, and IPSL-CM5A-MR are the five best models in simulating the interannual variability of the precipitation extreme events in the west.

In the east (Fig. 5b), all indices are relatively well simulated, but CDD is an exception and is poorly captured by most models, with a large discrepancy. This can be explained by the fact that models cannot reasonably reproduce the interannual variability of the monsoon. The average ranks for those four indices show that the models MRI-CGCM3, MIROC4h, CMCC-CM, HadGEM2-CC, and CCSM4 are the top five models. Note that the five models also fall into the top models with good performance in spatial pattern mentioned

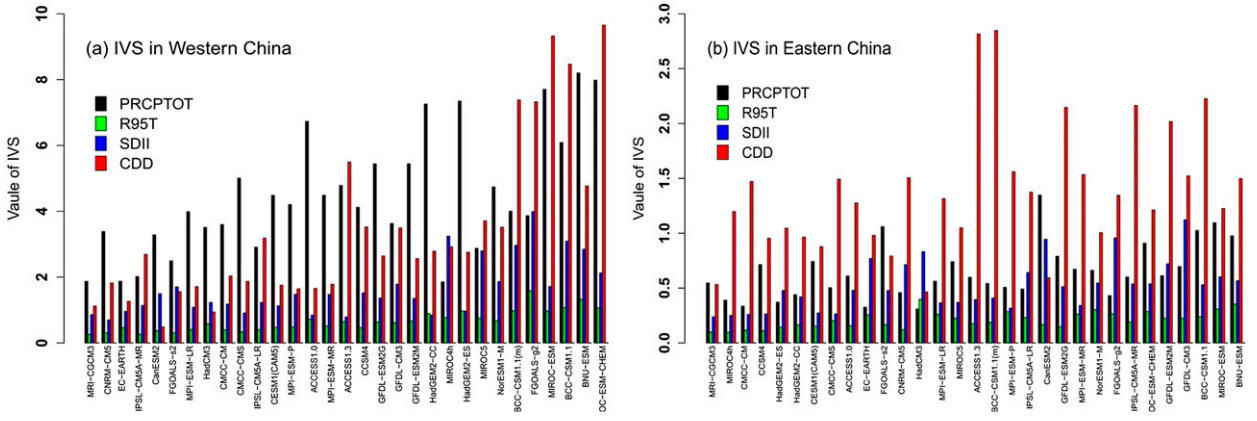

FIG. 5. Model skill scores of IVS for the four indices in (a) western and (b) eastern China. For IVS values close to 0, the greater the model's skill .

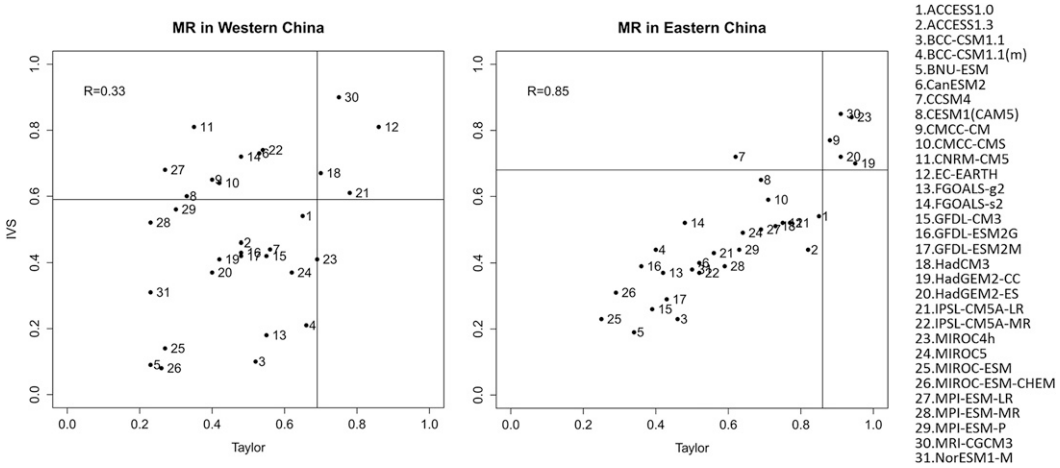

FIG. 6. Scatter diagrams showing models' MR based on Taylor diagrams (*x* axis) and IVS (*y* axis; interannual variability) in (left) western and (right) eastern China. Each dot represents a model, identified by its number on the right. The correlation coefficient between the ranking index from the Taylor diagrams and that from IVS is 0.33 and 0.85 in the west and east, respectively. Models in the top-right quadrant are of good performance for both criteria.

above. This indicates that performance in the simulation of spatial pattern is consistent with the capability to represent its interannual variability.

Overall, most models have difficulty simulating the interannual variability of extreme indices, and they present a better simulation of the interannual variability for all indices in the east compared to that in the west. Most of the models perform quite well for all extreme indices in eastern China with the exception of CDD. However, all indices are generally poorly simulated in western China, especially for PRCPTOT and CDD. An overall measure of interannual variability skill in two regions is calculated according to the averaged ranking across all indices. The five best models in western China are MRI-CGCM3, CNRM-CM5, EC-EARTH, IPSL-CM5A-MR, and CanESM2. The five models MRI-CGCM3, MIROC4h, CMCC-CM, CCSM4, and HadGEM2-ES are more skillful in eastern China.

d. Overall model ordering

Because the discrepancy of model ranking exists based on the Taylor diagram and IVS, a comprehensive assessment of models is undertaken in this study using MR (Fig. 6). The MR of the Taylor diagram and the MR of IVS are correlated significantly in eastern China. The correlation coefficient between the MR of the Taylor diagram and the MR of IVS is 0.85, which is statistically significant at the 95% level in the east, but a value of 0.33 in the west is not statistically significant at the 95% level. The result implies that the consistency between models in simulating spatial pattern and interannual variability in the east is better than that in the west. It is worth noting that, although the correlation between the MR of the Taylor diagram and the MR of IVS in the west is low, the models with the best performance are relatively centralized, with the MR of the Taylor diagram and the MR of IVS close to 1. MIROC4h, MRI-CGCM3, CMCC-CM, HadGEM2-CC, and HadGEM2-ES are revealed to be reliable models when spatial and temporal variation is comprehensively considered in eastern China, as those models are centralized with MR larger than 0.8 for the Taylor diagram and larger than 0.6 for IVS. The best models from the 31 GCMs in the west are MRI-CGCM3, EC-EARTH, HadCM3, and IPSL-CM5A-LR, as the MR of the Taylor diagram is close to 0.7 and the MR of IVS is close to 0.6. The models with the best performance have a considerable discrepancy between the east and west. Meanwhile, the correlation coefficient of 0.38 for models' ranks over western and eastern China indicates that a significant discrepancy of models' capabilities exists between the two subregions.

In summary, the models' skill in simulating extreme indices is quite different over the two parts of China. Five models (MIROC4h, MRI-CGCM3, CMCC-CM, HadGEM2-CC, and HadGEM2-ES) are selected for the best MME (BMME) composite; and MIROC-ESM, MIROC-ESM-CHEM, BNU-ESM, BCC-CSM1.1, and GFDL-ESM2M are included in the worst MME (WMME) composite in eastern China. In western China the BMME composite includes MRI-CGCM3, EC-EARTH, HadCM3, and IPSL-CM5A-LR. Note that the five best models in the east are obviously different from those in the west. Only MRI-CGCM3 is included in the best models both in eastern and western China. The five best models in the east all have a higher resolution, but this expectation is not extendable to the west. The

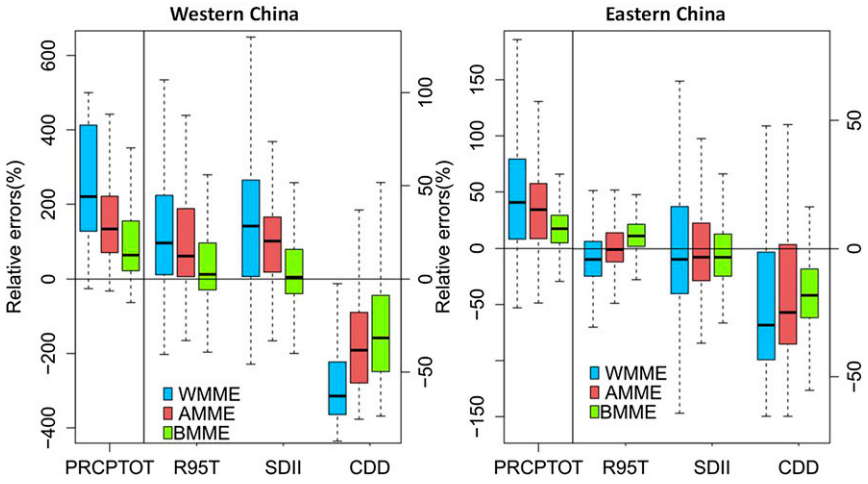

FIG. 7. Box-and-whisker plot for relative errors of extreme precipitation indices from WMME, AMME, and BMME over (left) western and (right) eastern China. The upper and lower limits of the box indicate the 75th and 25th percentile values; the horizontal line in the box indicates the ensemble median; and the whiskers show the error range of the ensemble (scale of PRCPTOT is on the left *y* axis, and scale of R95T, SDII, and CDD is on the right *y* axis).

WMME composites are the same for both eastern and western China.

4. Performance of optimal models

The most reliable models with good performance are selected over western and eastern China, respectively. We now examine how well the BMME simulates extreme indices and gives quantitative and visual improvements. Figure 7 illustrates the relative errors of the all-model ensemble mean (AMME), BMME, and WMME for each index during the period 1960–2005 for western and eastern China, respectively. The errors are presented in the form of a box-and-whisker plot showing the median interquartile range (box) spanned by the 25th and 75th quantiles and 5th and 95th percentiles (whiskers) across the grid cells over the two parts of China.

CMIP5 models still have wet biases in China, especially in western China, where the models' median relative error is about 120% for PRCPTOT, and the 25th and 75th percentile errors are 70% and 220%, respectively. For the east, PRCPTOT and CDD have a notable improvement in BMME compared to AMME and WMME, as the median is close to 0, the whiskers are shorter, and the interquartile model range is obviously smaller. But the improvement of SDII and R95T is not significant. The median of R95T is positive in BMME but negative in AMME and WMME, mainly because the best-performance models with higher resolution overestimate R95T in this region. However, the variability of R95T across the east region is smaller in BMME than that in AMME and WMME, as the interquartile range (box) and the 5th and 95th percentiles are smaller (whiskers are shorter). The improvement of indices for BMME is clearer in the west compared to that in the east. For example, the median of errors for PRCPTOT in the east and the west decreased by 17% and 69%, respectively. It is notable that BMME can remarkably reduce the biases in western China, but the variability of errors across this region represented by the box-and-whisker plot do not show significant improvement. Meanwhile, the standard deviation of PRCPTOT decreases by 17.8% in eastern China, whereas it decreases by 12.7% in the west. This implies that a relatively larger spatial difference still exists across western China in BMME, mainly because the best skillful models still cannot reproduce the extreme events related to the complex topography in western China.

Figure 8 further shows the spatial pattern of the differences between values simulated in the three ensembles and observed for all indices. AMME has its own deficiency, which is confirmed by the fact that there are wet biases in rare rainfall regions (western and northern China) and dry biases in frequent rainfall regions (southeastern China). For example, the overestimation of PRCPTOT over western and northern China can be up to 800 and 200 mm, respectively. The underestimation of PRCPTOT over southeastern China can reach 200 mm. AMME tends to overestimate R95T and SDII over western China and underestimate them over eastern China. The spatial pattern of CDD is the

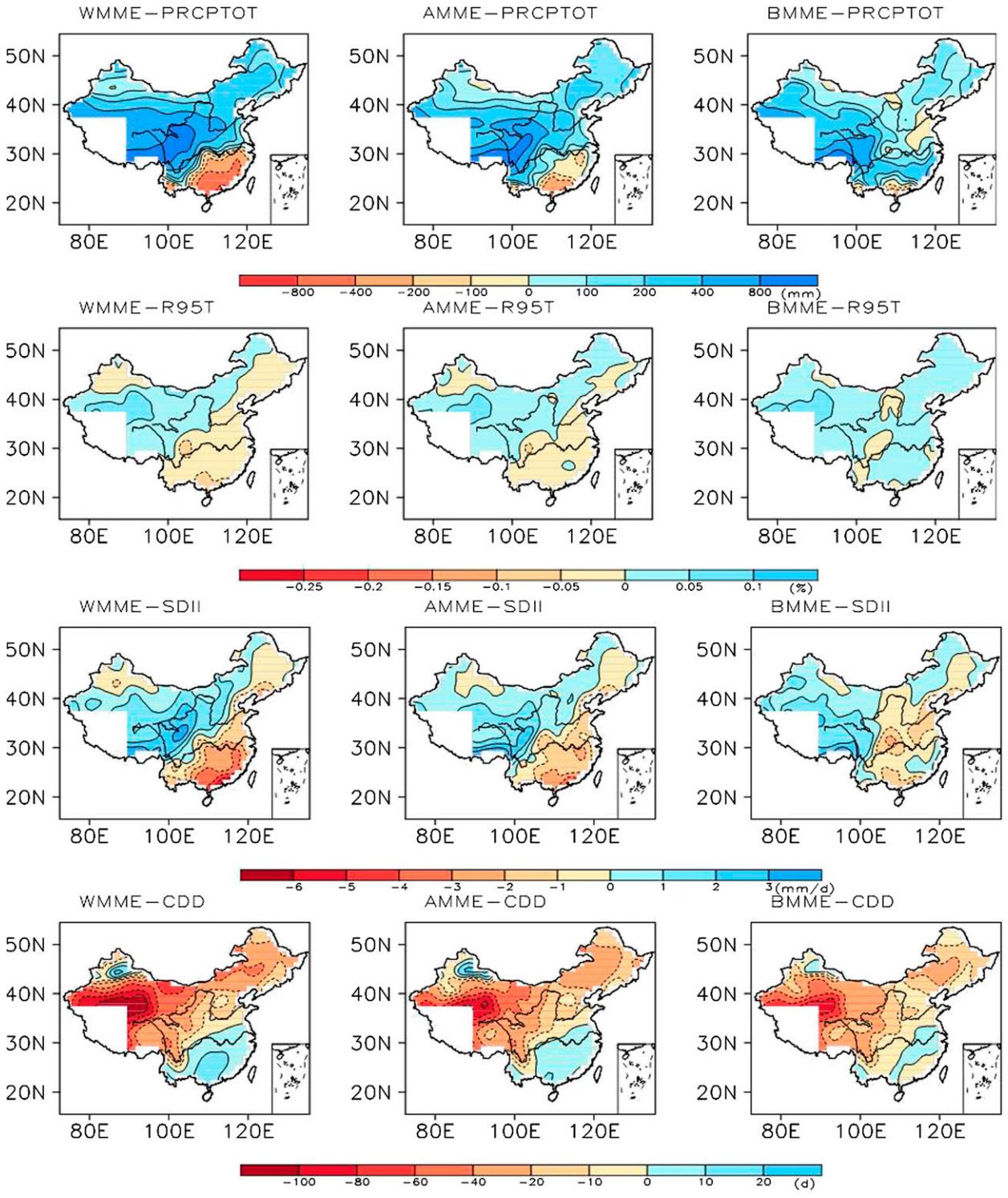

FIG. 8. Models' biases in terms of the four indices for (left) WMME, (center) AMME, and (right) BMME. BMME shows general improvements.

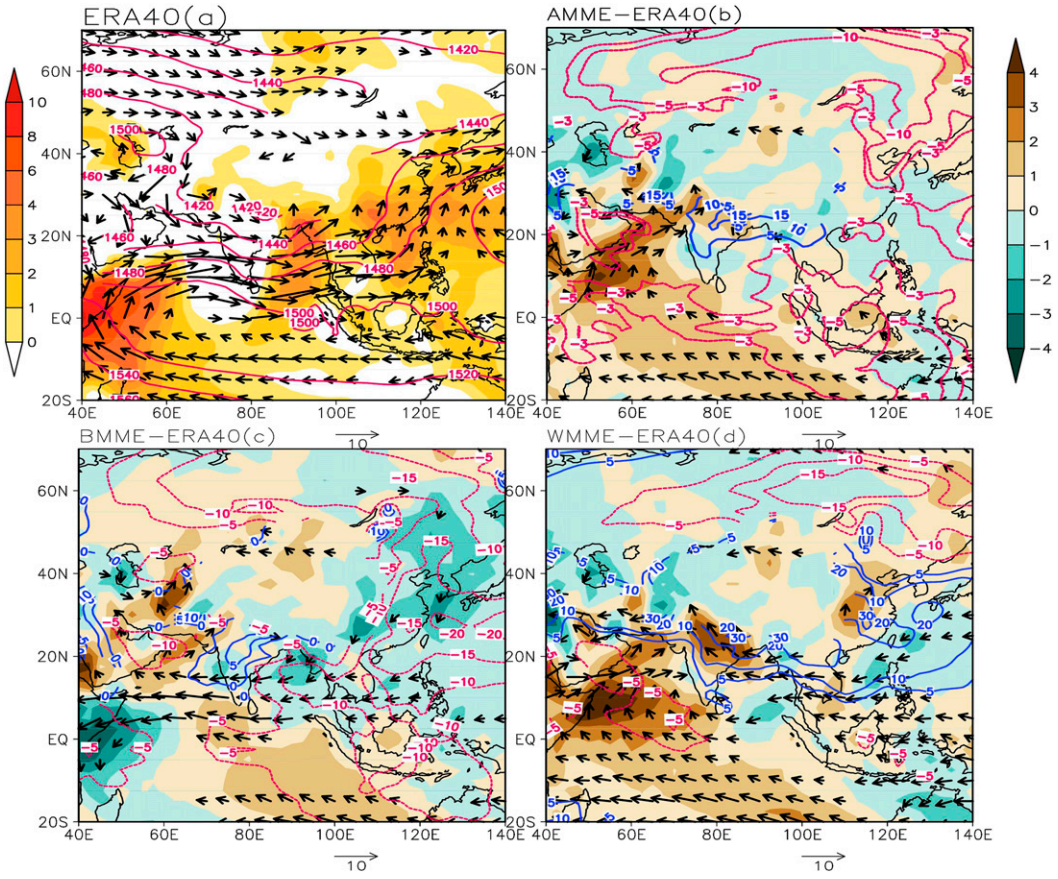

FIG. 9. June–August seasonal-mean geopotential height (contours; gpm), meridional wind speed (shaded; m s^{-1}), and wind vectors (m s^{-1}) at 850 hPa from (a) ERA-40 and (b)–(d) difference fields between the three ensembles and ERA-40. Only vectors more than 3 m s^{-1} are shown.

opposite of that for PRCPTOT, the underestimation of CDD over western China is up to 80 days. For BMME, the wet bias in western China and northern China is significantly reduced, especially over western China. The mean absolute errors for PRCPTOT and CDD over western China are decreased by 100 mm and 20 days, respectively.

To show the robustness of our results, we also calculated an additional index for extreme precipitation: the number of days with precipitation larger than the 95th percentile (R95N). It complements R95T as a measure of extreme precipitation. The mean absolute error for R95N over western China decreased from 5 days in AMME to 2 days in BMME.

The deficiency of AMME over southeastern China indicated by an underestimation of both total precipitation and precipitation intensity is also improved in BMME. For example, R95T in AMME is characterized by a significant underestimation. However, BMME shows more precipitation and extreme precipitation than observation in southeastern China. This is mainly because the higher resolution of models from BMME is capable of producing extreme precipitation.

To further investigate possible causes for those three ensembles' biases of extreme precipitation, the atmospheric circulations in the three different ensembles are compared with reanalysis data from ERA-40. Considering the fact that June–August (JJA) is the key season for extreme precipitation in China, Fig. 9 shows JJA geopotential height (contours), meridional wind speed (shaded), and wind vectors at 850 hPa from ERA-40 (Fig. 9a), together with difference fields between the three model ensembles and ERA-40 (Figs. 9b–d). The main components of the low-level circulation include the southwesterlies along the east coast of the Arabian Peninsula, southwest and west flows extending from the Arabian Sea to the South China Sea, and the southerlies over East Asia. They can be well captured by the three ensemble means. It is notable that the southwesterlies along the east coast of the Arabian Peninsula are overestimated in AMME and WMME, associated with a higher moisture transport, especially in WMME.

Previous studies show that the southwesterlies along the east coast of the Arabian Peninsula exert an important influence on the East Asian precipitation (Fan 2007; Zhu 2012). Meanwhile, the geopotential height at 850 hPa is overestimated over the western Pacific in WMME (the largest difference is approximately up to 30 gpm). The positive pressure errors over the western Pacific indicate a western Pacific subtropical high stronger in WMME than in the other two ensembles. This also suggests that the East Asian summer monsoon is overestimated. A strong southerly wind over eastern China is accompanied with strong southwesterlies over the Arabian Sea, which leads a northward extension of the East Asian summer monsoon (Wang and Zhou 2005; Zhang et al. 2008; You et al. 2011). As a result, WMME exhibits significant wet biases over western and northern China. However, in the case of BMME, the southwesterlies along the east coast of the Arabian Peninsula are relatively well simulated so that the wet bias is significantly decreased over western China and northern China. The underestimation of the western Pacific subtropical high and the southward shift over eastern China contribute to a weak East Asian summer monsoon, resulting in a decrease of wet biases over northern China and an increase of wet biases over southeastern China (Zhao and Zhou 2009).

In summary, AMME has its own deficiency, which is confirmed by the fact that there are wet biases in rare rainfall regions (western and northern China) and dry biases in frequent rainfall regions (southeastern China). Wet biases in western and northern China make a dominant contribution for ranking models. Reductions of biases in those regions are very important for the selection of BMME composite models. BMME can better simulate this southwesterly along the east coast of the Arabian Peninsula, relatively, thus showing a simultaneous improvement over western and northern China. But at the same time, BMME does present a too-wet bias in southeastern China. This overestimation of rainfall is reasonable with the weakening simulation of summer monsoon in BMME that is associated with the underestimation of the subtropical high and southward shift over eastern China. However, this too-wet bias in southeastern China may not influence the rank of models. AMME seems to benefit from a cancellation of errors among all models.

5. Conclusions and discussion

In the present study, we quantitatively evaluated and ranked the performance of the CMIP5 models for simulating extreme precipitation indices in eastern and western China, respectively. The assessment was done in terms of spatial patterns and temporal variability for the period 1960–2005. We used skill-score methods, such as the Taylor diagram representing the spatial pattern and IVS representing the interannual variability. According to the overall ranking of models, we selected the most skillful models over eastern and western China, respectively. We also compared results from different ensemble constructions: the ensemble from the best models against that from all models and from the less-skillful ones. This comparison helped us to understand models biases in precipitation indices in China and their relations with the large-scale atmospheric circulation. Our main findings can be summarized as follows.

1) CMIP5 models have wet biases in the region with an overestimation of PRCPTOT and an underestimation of CDD, especially in western China, where the systematic wet biases have a median value of 120% in PRCPTOT, and the 25th and 75th percentiles are 70% and 220%, respectively. Generally speaking, the CMIP5 models can well reproduce the basic characteristics of extreme precipitation indices, including both the spatial variation and interannual variability in eastern China. However, in western China, most models can hardly reproduce the basic characteristics of extreme indices.
2) The spatial patterns of PRCPTOT and SDII are better simulated than those of R95T and CDD by most of the studied models in eastern China, whereas in western China the spatial variations of all indices are poorly captured. For the interannual variability, except for CDD, the other three indices are relatively well simulated in eastern China. In western China, models can hardly reproduce the interannual variability of extreme indices. SDII and R95T are relatively better simulated than the other two indices.
3) The performance of individual models in the simulation of spatial pattern is more consistent with the temporal variability over eastern China. MRI-CGCM3, CMCC-CM, MIROC4h, HadGEM2-ES, and HadGEM2-CC are revealed to be the best models to represent both spatial pattern and temporal variability in eastern China. However, MRI-CGCM3, EC-EARTH, HadCM3, and IPSL-CM5A-LR are the leading models in western China.
4) AMME tends to simulate higher PRCPTOT but smaller CDD over all China, except southeastern China, where PRCPTOT is underestimated and CDD is overestimated. AMME tends to overestimate R95T and SDII in western China, especially at the periphery of the Tibetan Plateau, and underestimate these two indices in eastern China. The wet bias in BMME is significantly reduced over western and

northern China, with the median of errors for PRCPTOT in the west decreased by 69%. BMME is also shown to be capable of simulating extreme precipitation events over southern China.

5) The wet biases over western and northern China are tightly related to the overestimation of southwesterlies along the east coast of the Arabian Peninsula and overestimation of the western Pacific subtropical high, inducing a stronger East Asian summer monsoon. BMME can reproduce the southwesterlies along the east coast of the Arabian Peninsula relatively well, which thus shows a simultaneous improvement over western and northern China. But it shows an underestimation of the western Pacific subtropical high and a southward shift over eastern China, which indicates a weakening summer monsoonal flow in eastern China. Therefore, the BMME presents a too-wet bias in southeastern China.

The present study provides a reference for the performance of different CMIP5 models in simulating extremes of rainfall in China. It gives some useful indications for climate impact and improvement of model performance studies in the Chinese region. For example, the main aim of our study is to find the possible reasons for the models' performances based on model evaluation, which is more meaningful for the models' improvement; the biases of models can also be studied as another possible reason. In addition, each model can be assigned with a weighting factor according to its rank of performance in simulating the spatial pattern and interannual variability. Different models with different weights are useful for a more precise future projection of extreme precipitation changes (Santer et al. 2009; Knutti 2010), which will be the focus of our next work. Furthermore, regional climate downscaling studies using the best models in extreme precipitation, as we selected here, can be helpful to reach more accurate results for regional climate change (Seo and Ok 2013). It is worth noting that higher resolutions are expected to effectively improve models' capabilities for simulating precipitation in eastern China, but this expectation is not extendable to the western areas. This implies the important role of the complete and accurate representations of physical processes in the mountainous areas, which is still a challenge for current global climate models.

Acknowledgments. We thank the editor and anonymous reviewers for their constructive comments. We acknowledge the modeling groups listed in Table 1 of this paper for making their simulations available for analysis, the PCMDI for collecting and archiving the CMIP5 model output, and the World Climate Research Programme's Working Group on Coupled Modelling. For CMIP, the U.S. Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. This work is supported by the State Key Program of National Natural Science Foundation of China (41230528) and the National Basic Research Program ''973'' (Grant 2012CB955204). This study is also supported by the Special Research Program for Public Welfare (Meteorology) of China under Grant GYHY201306024, the Research and Innovation Project for College Graduates of Jiangsu Province (Grant KYLX_0843), and a project Funded by the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions.

REFERENCES

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

Chen, D., T. Ou, L. Gong, C. Y. Xu, and C. H. Ho, 2010: Spatial interpolation of daily precipitation in China: 1951–2005. *Adv. Atmos. Sci.*, **27**, 1221–1232, doi:10.1007/s00376-010-9151-y.

Chen, W. L., Z. Jiang, and L. Li, 2011: Probabilistic projections of climate change over China under the SRES A1B scenario using 28 AOGCMs. *J. Climate*, **24**, 4741–4756, doi:10.1175/2011JCLI4102.1.

Easterling, D. R., J. L. Evans, P. Y. Groisman, T. R. Karl, K. E. Kunkel, and P. Ambenje, 2000: Observed variability and trends in extreme climate events: A brief review. *Bull. Amer. Meteor. Soc.*, **81**, 417–425, doi:10.1175/1520-0477(2000)081<0417:OVATIE>2.3.CO;2.

Fan, K., 2007: Zonal asymmetry of the Antarctic Oscillation. *Geophys. Res. Lett.*, **34**, L02706, doi:10.1029/2006GL028045.

Feng, L., T. Zhou, B. Wu, T. Li, and J. J. Luo, 2011: Projection of future precipitation change over China with a high-resolution global atmospheric model. *Adv. Atmos. Sci.*, **28**, 464–476, doi:10.1007/s00376-010-0016-1.

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 climatic extremes during the second half of the twentieth century. *Climate Res.*, **19**, 193–212, doi:10.3354/cr019193.

Haylock, M. R., G. C. Cawley, C. Harpham, B. L. Wilby, and C. M. Goodess, 2006: Downscaling heavy precipitation over the United Kingdom: A comparison of dynamical and statistical methods and their future scenarios. *Int. J. Climatol.*, **26**, 1397–1415, doi:10.1002/joc.1318.

IPCC, 2007: *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 996 pp.

——, 2013: *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, 1535 pp., doi:10.1017/CBO9781107415324.

Jiang, D. B., Y. Zhang, and Y. Q. Sun, 2009: Ensemble projection of 1–3°C warming in China. *Chin. Sci. Bull.*, **54**, 3326–3334, doi:10.1007/s11434-009-0313-1.

Jiang, Z., J. Song, L. Li, W. L. Chen, Z. Wang, and J. Wang, 2012: Extreme climate events in China: IPCC-AR4 model evaluation and projection. *Climatic Change*, **110**, 385–401, doi:10.1007/s10584-011-0090-0.

Kiktev, D., D. M. H. Sexton, L. Alexander, and C. K. Folland, 2003: Comparison of modeled and observed trends in indices of daily climate extremes. *J. Climate*, **16**, 3560–3571, doi:10.1175/1520-0442(2003)016<3560:COMAOT>2.0.CO;2.

Klein Tank, A. M. G., and G. P. Konnen, 2003: Trends in indices of daily temperature and precipitation extremes in Europe 1964–1999. *J. Climate*, **16**, 3665–3680, doi:10.1175/1520-0442(2003)016<3665:TIIODT>2.0.CO;2.

Knutti, R., 2010: The end of model democracy? An editorial comment. *Climatic Change*, **102**, 395–404, doi:10.1007/s10584-010-9800-2.

Kusunoki, S., J. Yoshimura, H. Yoshimura, A. Noda, K. Oouchi, and R. Mizuta, 2006: Change of Baiu rain band in global warming projection by an atmospheric general circulation model with a 20-km grid size. *J. Meteor. Soc. Japan*, **84**, 581–611, doi:10.2151/jmsj.84.581.

——, R. Mizuta, and M. Matsueda, 2011: Future changes in the East Asian rain band projected by global atmospheric models with 20-km and 60-km grid size. *Climate Dyn.*, **37**, 2481–2493, doi:10.1007/s00382-011-1000-x.

Li, H., L. Feng, and T. Zhou, 2010: Multi-model projection of July–August climate extreme changes over China under CO_2 doubling. Part I: Precipitation. *Acta Meteor. Sin.*, **28**, 433–477, doi:10.1007/s00376-010-0013-4.

Li, J., Q. Zhang, Y. D. Chen, and V. P. Singh, 2013: GCMs-based spatiotemporal evolution of climate extremes during the 21st century in China. *J. Geophys. Res.*, **118**, 11 017–11 035, doi:10.1002/jgrd.50851.

Moberg, A., and Coauthors, 2006: Indices for daily temperature and precipitation extremes in Europe analyzed for the period 1901–2000. *J. Geophys. Res.*, **111**, D22106, doi:10.1029/2006JD007103.

Ou, T., D. Chen, H. W. Linderholm, and J. H. Jeong, 2013: Evaluation of global climate models in simulating extreme precipitation in China. *Tellus*, **65A**, 1393–1399, doi:10.3402/tellusa.v65i0.19799.

Palmer, T. N., F. J. Doblas-Reyes, R. Hagedorn, and A. Weisheimer, 2005: Probabilistic prediction of climate using multi-model ensembles: From basics to applications. *Philos. Trans. Roy. Soc. London*, **B360**, 1991–1998, doi:10.1098/rstb.2005.1750.

Pierce, D. W., T. P. Barnett, B. D. Santer, and P. J. Gleckler, 2009: Selecting global climate models for regional climate change studies. *Proc. Natl. Acad. Sci. USA*, **106**, 8441–8446, doi:10.1073/pnas.0900094106.

Qian, W., J. Fu, and Z. Yan, 2007: Decrease of light rain events in summer associated with a warming environment in China during 1961–2005. *Geophys. Res. Lett.*, **34**, 224–238, doi:10.1029/2007GL029631.

Ren, G., and Coauthors, 2011: Change in climatic extremes over mainland China based on an integrated extreme climate index. *Climate Res.*, **50**, 113–124, doi:10.3354/cr01023.

Santer, B. D., and Coauthors, 2009: Incorporating model quality information in climate change detection and attribution studies. *Proc. Natl. Acad. Sci. USA*, **106**, 14 778–14 783, doi:10.1073/pnas.0901736106.

Schmittner, A., M. Latif, and B. Schneider, 2005: Model projections of the North Atlantic thermohaline circulation for the 21st century assessed by observations. *Geophys. Res. Lett.*, **32**, 113–133, doi:10.1029/2005GL024368.

Seo, K. H., and J. Ok, 2013: Assessing future changes in the East Asian summer monsoon using CMIP3 models: Results from the best model ensemble. *J. Climate*, **26**, 1807–1817, doi:10.1175/JCLI-D-12-00109.1.

Sillmann, J. V., V. Kharin, X. W. Zhang, F. Zwiers, and D. Bronaugh, 2013a: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res. Atmos.*, **118**, 1716–1733, doi:10.1002/jgrd.50203.

——, ——, ——, ——, and ——, 2013b: Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections. *J. Geophys. Res.*, **118**, 2473–2493, doi:10.1002/jgrd.50188.

Su, F., X. Duan, D. Chen, Z. Hao, and L. Cuo, 2013: Evaluation of the global climate models in the CMIP5 over the Tibetan Plateau. *J. Climate*, **26**, 3187–3208, doi:10.1175/JCLI-D-12-00321.1.

Sun, Y., and Y. H. Ding, 2008: An assessment on the performance of IPCC AR4 climate models in simulating interdecadal variations of the East Asian summer monsoon. *Acta Meteor. Sin.*, **22**, 472–488.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, doi:10.1029/2000JD900719.

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

Thomson, M. C., F. J. Doblas-Reyes, S. J. Mason, R. Hagedorn, S. J. Connor, T. Phindela, A. P. Morse, and T. N. Palmer, 2006: Malaria early warnings based on seasonal climate forecasts from multi-model ensembles. *Nature*, **439**, 576–579, doi:10.1038/nature04503.

Wang, H. J., and Coauthors, 2012: Extreme climate in China: Facts, simulation and projection. *Meteor. Z.*, **21**, 279–304, doi:10.1127/0941-2948/2012/0330.

Wang, W. G., and G. G. Zheng, 2012: *Annual Report on Actions to Address Climate Change: Climate Finance and Low Carbon Development* (in Chinese). Social Science Academic Press, 306 pp.

Wang, Y. Q., and L. Zhou, 2005: Observed trends in extreme precipitation events in China during 1961–2001 and the associated changes in large-scale circulation. *Geophys. Res. Lett.*, **32**, L09707, doi:10.1029/2005GL022574.

Xu, C. H., Y. Luo, and Y. Xu, 2011: Projected changes of precipitation extremes in river basins over China. *Quat. Int.*, **244**, 149–158, doi:10.1016/j.quaint.2011.01.002.

You, Q., and Coauthors, 2011: Changes in daily climate extremes in China and their connection to the large scale atmospheric circulation during 1961–2003. *Climate Dyn.*, **36**, 2399–2417, doi:10.1007/s00382-009-0735-0.

Zhai, P. M., X. B. Zhang, and X. H. Pan, 2005: Trends in total precipitation and frequency of daily precipitation extremes over China. *J. Climate*, **18**, 1096–1108, doi:10.1175/JCLI-3318.1.

——, Z. W. Yan, and X. K. Zhou, 2008: Climate extremes and related disasters in China. *Regional Climate Studies of China*, C. B. Fu et al., Eds., Regional Climate Studies, Springer Berlin Heidelberg, 313–344, doi:10.1007/978-3-540-79242-0_8.

Zhang, H., K. Fraedrich, R. Blender, and X. Zhu, 2013: Precipitation extremes in CMIP5 simulations on different time scales. *J. Hydrometeor.*, **14**, 923–928, doi:10.1175/JHM-D-12-0181.1.

Zhang, J. J., Z. B. Sun, and S. J. Chen, 1984: An attempt to classify the natural synoptic seasons for each year over East Asia by EOF (in Chinese). *Acta Meteor. Sin.*, **42**, 46–56.

Zhang, Q., C. Xu, Z. Zhang, Y. D. Chen, C. Liu, and H. Lin, 2008: Spatial and temporal variability of precipitation maxima during 1960–2005 in the Yangtze River basin and possible association with large-scale circulation. *J. Hydrol.*, **353**, 215–227, doi:10.1016/j.jhydrol.2007.11.023.

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

Zhao, P., and Z. J. Zhou, 2009: An East Asian subtropical summer monsoon index and its relationship to summer rainfall in China. *Acta Meteor. Sin.*, **23**, 18–28.

Zhou, B., Q. H. Wen, Y. Xu, L. Song, and X. Zhang, 2014: Projected changes in temperature and precipitation extremes in China by the CMIP5 multimodel ensembles. *J. Climate*, **27**, 6591–6611, doi:10.1175/JCLI-D-13-00761.1.

Zhu, Y., 2012: Variations of the summer Somali and Australia cross-equatorial flows and the implications for the Asian summer monsoon. *Adv. Atmos. Sci.*, **29**, 509–518, doi:10.1007/s00376-011-1120-6.