

Assessing model performance of climate extremes in China: an intercomparison between CMIP5 and CMIP3

Huopo Chen · Jianqi Sun

Received: 20 March 2014 / Accepted: 29 December 2014 / Published online: 19 January 2015
© Springer Science+Business Media Dordrecht 2015

Abstract In this study, we present a brief analysis of the performances of global climate models participating in the Coupled Model Intercomparison Project Phase 5 (CMIP5) in simulating climate extreme events in China and compare the results with those of the previous model generation (CMIP3). The primary focus of this analysis is the climate mean and variability of each extreme index. Results show that the CMIP5 models are generally able to capture the mean climate extremes and trends compared with a new gridded observational dataset. The model spread for some extreme indices is reduced in CMIP5 when compared with CMIP3. Furthermore, the models generally show higher skills in simulating the temperature-based indices than the precipitation-based indices in terms of means and linear trends. Results from six reanalyses further reveal large uncertainties for these indices and it is difficult to say which reanalysis is better for comparison with the simulations of all indices. Based on the relative errors of the climatology, the model evaluation varies considerably from one index to another. However, some models appear to perform substantially better than the others when the average of all indices is considered for each model, and the median ensembles outperform the individual models in terms of all the extreme indices and their means. Additionally, a relationship is observed between the improved simulation of the climate mean and the improved performance of its variability, although this improvement is limited to particular models.

1 Introduction

In general, climate extreme events, including extreme temperature and extreme precipitation, have major impacts on society, the economy, ecosystems, and human health. Most previous studies have consistently concluded that global warming has begun to affect the frequency, intensity, and duration of extreme events, and some of these changes are projected to continue

Electronic supplementary material The online version of this article (doi:10.1007/s10584-014-1319-5) contains supplementary material, which is available to authorized users.

H. Chen (✉) · J. Sun
Nansen-Zhu International Research Centre, Institute of Atmospheric Physics, Chinese Academy of Sciences,
PO Box 9804, Beijing 100029, China
e-mail: chenhuopo@mail.iap.ac.cn

H. Chen
Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disasters, Nanjing
University of Information Science & Technology, Nanjing, China

into the future (IPCC 2007, 2012, 2013). Such results are also robust in China (Chen 2013; Chen and Sun 2014).

Climate extreme events are generally multifaceted phenomena and are extremely difficult to monitor. Thus, several international groups have made many attempts to develop valuable indices that facilitate the monitoring and analysis of such extremes, such as the 27 climate extreme indices proposed by the Expert Team on Climate Change Detection and Indices (ETCCDI) and the 54 indices proposed by the project of Statistical and Regional Dynamical Downscaling of Extremes for European Regions (STARDEX). These indices are substantially similar and are mainly calculated using daily minimum and maximum temperatures and daily precipitation. The 27 ETCCDI indices are analyzed in this study.

The ETCCDI climate indices are very popular and widely used in climate research and related fields due to their robustness and fairly straightforward calculation and interpretation. For example, these indices are used for investigating global changes in extreme events, in observational records (e.g., Frich et al. 2002; Alexander et al. 2006; Donat et al. 2013a), in various detection and attribution studies (e.g., Min et al. 2010; Morak et al. 2011), in model performance evaluations (e.g., Sillmann et al. 2013a), and future climate projections (e.g., Sillmann and Roeckner 2008; Orłowsky and Seneviratne 2012; Sillmann et al. 2013b; Chen et al. 2014). Indices have also been developed to enable regional analyses of changes in extreme temperature and precipitation in the Arabian region (Donat et al. 2013b), North America (Peterson et al. 2008), the Mediterranean Basin (Gao and Giorgi 2008), and other regions. In China, some of these indices are also widely used for revealing the extreme changes in gauged records (Zhai et al. 2005; Wang et al. 2012) and projections (Gao et al. 2002; Chen et al. 2012; Chen 2013). Additionally, the model performances in simulating the climate means and extremes have been also demonstrated by some previous studies (e.g., Seo et al. 2013; Sillmann et al. 2013a) from a global perspective and for several regions including China. However, some spatial variations have not been considered in China, and further analyses are needed, especially for evaluating CMIP5 (Coupled Model Intercomparison Project Phase 5) model performance against the performance of the model's previous version.

The climate in China is dominated by monsoonal variability, which strongly affects approximately one-fifth of the world's population. An accurate prediction of future climate changes is thus important and necessary to allow better planning and help mitigate impacts at the national level. To improve the accuracy of predictions, one important task is to evaluate model performances regarding the climate mean and extreme events. This study provides a comprehensive analysis of the performances of the CMIP5 models in simulating climate extreme aspects based on the 27 ETCCDI indices, the models, observations, and reanalysis datasets for China, and the works of previous studies (e.g., Sillmann et al. 2013a). Compared with CMIP3, the improvement of CMIP5 models is significant (Taylor et al. 2012); thus, the CMIP3 model performances will be also compared with the CMIP5 model performances to determine whether the simulation of regional extreme events improves with the improvement of the climate models.

2 Data and methods

2.1 Datasets

Daily minimum and maximum near-surface temperatures (TN and TX, respectively) and daily precipitation (PR) are downloaded from the Earth System Grid (ESG) data portal for 30 CMIP5 (cf. Table S1 in the supplementary material) and 19 CMIP3 (cf. Table S2) models. The

analysis in this study is based on the first ensemble member of each model, referred to as *r1i1p1* in CMIP5 experiment and *run1* in CMIP3. Compared with the CMIP3 models, the main improvements in the CMIP5 models include the following: (1) addition of interactive ocean and land carbon cycles with varying degrees of complexity, (2) more comprehensive modeling of the indirect effects of aerosols, (3) inclusion of time-evolving volcanic and solar forcing in most models, and (4) higher horizontal and vertical resolutions (Taylor et al. 2012). Additionally, to assess the model performance in simulating extreme events, a newly gridded daily dataset with a resolution of 0.25° latitude \times 0.25° longitude, which was constructed by Wu and Gao (2013, hereafter referred to as Gao), is employed. More information about this observed data please sees the online supplementary material.

2.2 Climate extreme indices and processing

The detailed definitions of the 27 ETCCDI climate extreme indices are described in several previous studies (e.g., Zhang et al. 2011; Donat et al. 2013b; Sillmann et al. 2013a), which are also summarized in Table S3. All indices are calculated for the CMIP5 and CMIP3 simulations and for the observations in China using a FORTRAN package, *FClimDex.f*, as documented at the ETCCDI climate change indices website (<http://etccdi.pacificclimate.org/software.shtml>). Additionally, to reduce the observational uncertainty, the indices from six widely used reanalyses are also employed: ERA40, ERA-Interim, NCEP-1, NCEP-2, CFSR, and JRA25. The indices for the first four reanalyses are available for download from the ETCCDI indices archive website hosted by the Canadian Centre for Climate Modeling and Analysis (<http://www.cccma.ec.gc.ca/data/climdex/climdex.shtml>), while CFSR and JRA25 are computed using the FORTRAN package. The main objective of this study is to evaluate the model performances of both CMIP5 and CMIP3. For convenience, all indices are resampled to a common 240×121 grids ($1.5^\circ \times 1.5^\circ$) using a first-order conservation remapping procedure using the Climate Data Operators (CDO). Ultimately, 416 gridpoints are obtained for China.

3 Results

3.1 Temperature extremes

3.1.1 Spatial performance

First, the model performance in simulating the spatial structure of the 1981–2000 climatology of the absolute temperature indices is analyzed and compared with the observations. The simulated maximum of the daily maximum temperature (TXx) and minimum of the daily minimum temperature (TNn), which describe the hottest days and coldest nights of a year, respectively, from the CMIP5 multimodel medians compare well with the observations (see auxiliary material, Figure S1). The temperature gradients from high to low altitudes and from north to south are also reasonably captured by the CMIP5 models. However, some discrepancies are also clear for the temperature indices when compared to the observations (Fig. 1). The simulated temperature on the hottest days is overestimated over most parts of eastern China and Xinjiang regions but underestimated over parts of the Tibetan Plateau. The temperature on the coldest days shows a negative bias over all of China, except over the north of Northeast China. These spatial characteristics are similar for other absolute indices, such as the numbers of summer days (SU), tropical nights (TR), frost days (FD), and ice days (ID) (Figures S2 and S3). Relative to the observations, the cold biases are obvious, although

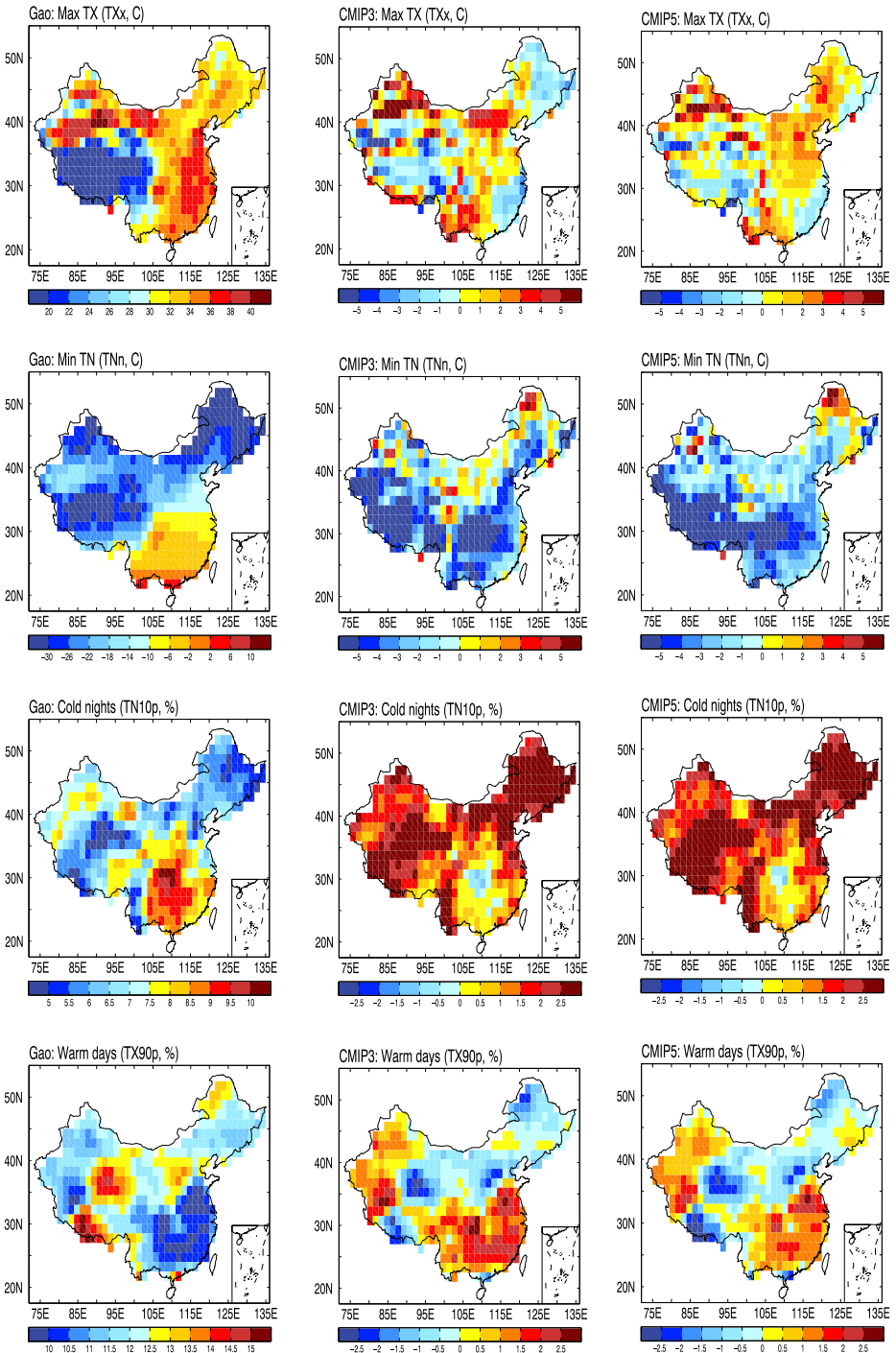


Fig. 1 Mean of observed (*left column*) annual maximum of TX (TXx), minimum of TN (TNn), cold nights (TN10p), and warm days (TX90p) during 1981–2000. The middle and right columns show the differences of the CMIP3 and CMIP5 multimodel median ensembles relative to the observations

significant improvements are observed for the CMIP5 models. Thus, the underestimation is clear for some absolute temperature indices, particularly over the Tibetan Plateau. For example, the simulated temperature on the coldest nights and the growing season length (GSL) are underestimated by 3.0 °C and 22.5 days, respectively, relative to the average observations over China. In contrast, the FD and ID are overestimated by 9.2 and 25.6 days, respectively. Meanwhile, the CMIP5 median ensemble shows a warm bias in TXx (0.6 °C), particularly over eastern China and the Tianshan Mountain regions, compared with the observations (Fig. 1).

Generally, it is difficult to evaluate model performances with percentile indices (some discussions about the evaluations of the percentile indices can be found in supplementary material) because the mean threshold exceedance rate in the base period is approximately the same for all models, reanalyses, and observations (Sillmann et al. 2013c). Keeping this in mind, the calculations of these percentile indices for the models and observations are implemented over the standard base period of 1961–1990 (except for some reanalyses, including ERA-Interim, NCEP-2, CFSR, and JRA25, which have inadequate time periods, and are instead calculated over the period of 1979–2008), and the climatic mean analyses are calculated over the period of 1981–2000. From Figure S1 and Fig. 1, a uniform overestimation in TN10p and a large regional difference in TX90p have been reported in China. Underestimation of TX90p in some parts of Tibet, North China, and Northeast China and overestimation in other regions, particularly in southern China, are observed (Fig. 1). Further, the simulated percentile indices of cold and warm spell durations (CSDI and WSDI, respectively) exhibit spatial structures that agree with those of the observations; however, a bias is obvious in which the CSDI (2.2 days) is uniformly overestimated and the WSDI (1.7 days) is uniformly underestimated in China.

The Taylor diagram provides a concise statistical summary of the model performances in simulating the extreme indices in terms of their spatial correlations, root-mean-square differences, and the ratio of variances. Fig. 2 shows the results of the temperature-related extreme indices for the individual CMIP5 models. Among the absolute indices, the poorest skills and the large model spread are observed in simulating the diurnal temperature range (DTR). However, a metric analysis of the Taylor diagrams consistently shows that the median ensemble outperforms the individual models for both the absolute and percentile indices. Further analysis of the individual CMIP5 models does not reveal a clear relationship between the model's spatial resolution and its simulation of temperature indices at a regional scale in this study or at a global scale in a previous study (Sillmann et al. 2013a). For example, for TXx and TNn, the high-resolution models generally present lower root-mean-square errors (RMSEs). However, a relatively large RMSE is observed for DTR (and for some percentile indices) with the high-resolution models in CMIP5, such as MIROC4h, CCSM4, and MRI-CGCM3.

The CMIP3 models can also adequately reproduce the spatial structures of the temperature indices in China but with relatively larger biases of extreme indices when compared with the CMIP5 models (Fig. 1, Figures S1–S3). For example, the regional averaged GSL is underestimated by 27.7 days, whereas the FD and ID are overestimated by 15.1 and 31.5 days, respectively; all of these values are much larger than those of the CMIP5 models. Similar features are observed for the percentile indices. Furthermore, a larger model spread is observed for most temperature indices in CMIP3 than in CMIP5, which implies relatively larger uncertainties in the CMIP3 results. In addition, the simulated skills of the models are generally lower in the CMIP3 median ensemble than in the CMIP5, except for CSDI (Fig. 2). Overall, with the improvement of the climate models, CMIP5 exhibits a better performance in capturing the climatic means of extreme indices, and the spreads among models are reduced for some indices.

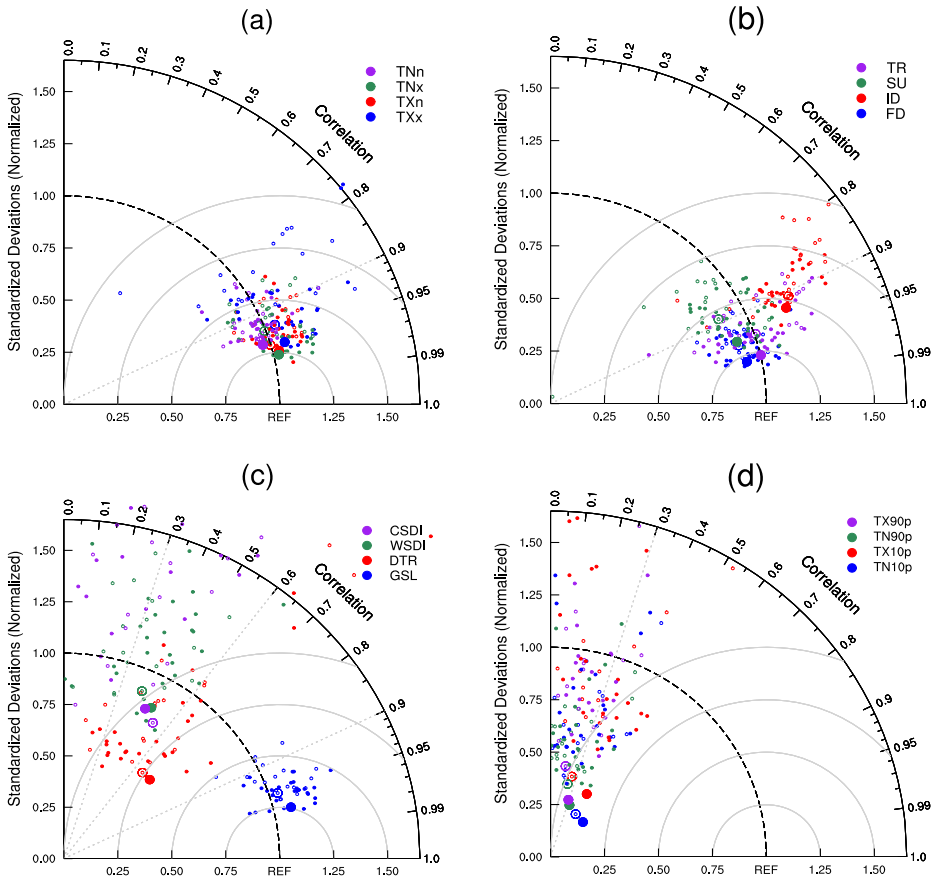


Fig. 2 Multivariable Taylor diagrams of the 20th century CMIP3 (*hollow*) and CMIP5 (*solid*) simulated climatic means (1981–2000) for 16 temperature extreme indices in China. Each colored dot represents an individual simulation by a particular model, whereas each larger dot represents a multimodel median ensemble

As observed from Figs. 1 and 2, models in both CMIP3 and CMIP5 can successfully capture the spatial structures of the climatology of some temperature indices, despite the relatively larger biases in CMIP3 than in CMIP5. Further evaluations of the spatial RMSEs of the temperature-related extreme indices indicate that the multimodel medians of CMIP3 are generally much larger than those of CMIP5 for most of the absolute indices, while they are very similar for the percentile indices (Figure S4). The interquartile model ranges for the absolute indices, which spanned the 25th to 75th quantiles of the multimodel ensemble, are often smaller in CMIP5 compared with CMIP3. Additionally, reanalyses also present large RMSEs of the temperature indices that are comparable to the values of CMIP3 and CMIP5 models. However, the spread between the six reanalyses is similar or even larger than the interquartile model spread. Thus, caution should be used when relying on the reanalysis datasets for model evaluations.

3.1.2 Temporal performance

The temporal evolutions of the regionally averaged indices in China for the models, reanalyses, and observations are shown in Fig. 3 for 1961–2010. It is clear that the interquartile

model spread of the absolute temperature indices, indicated by the shading, is much larger in CMIP3 than in CMIP5. However, the multimodel medians of the two CMIP ensembles are comparable. Due to the modeling cold bias in China, differences in the temporal evolutions of the indices between the observations and simulations are clearly visible, except for TXx (Fig. 3a) and TR. The multimodel median shows smaller values for TNn (Fig. 3b), TXn, SU (Fig. 3d), GSL, and DTR, but greater values for FD (Fig. 3c), ID, and TNx than the observations.

The observed tendencies of the temperature-related extreme indices to increase or decrease are successfully reproduced by both the models and reanalyses, but the differences are also clearly visible when the absolute values are considered (Fig. 3). Additionally, large differences between reanalyses are observed and greater than or equal to the interquartile model spread. Similar to Sillmann et al. (2013a), differences between reanalyses that are produced by the same research centers are generally smaller than differences between reanalyses of different centers (i.e., ERA40 and ERA-Interim versus NCEP-1 and NCEP-2). Furthermore, for indices, such as TXx and FD, both the CMIP3 and CMIP5 models better compare with the more recent reanalyses, i.e., ERA-Interim, NCEP-2, JRA25, and CFSR, than with the earlier ones, i.e., ERA40 and NCEP-1.

Due to the intrinsic designs of percentile indices, the temporal evolutions exhibit a fairly good agreement among models, reanalyses, and observations when compared with the absolute indices (Figure S5). Consistent with the observed changes, a decrease is clearly visible for cold nights and an increase is apparent for warm days for both the models and reanalyses. The frequency of cold nights is observed to decrease to 7 % in 2005 using CMIP5 and in 2000 using CMIP3 from the nominal level of approximately 10 % in the base period; the warm days approximately increase to 12 %.

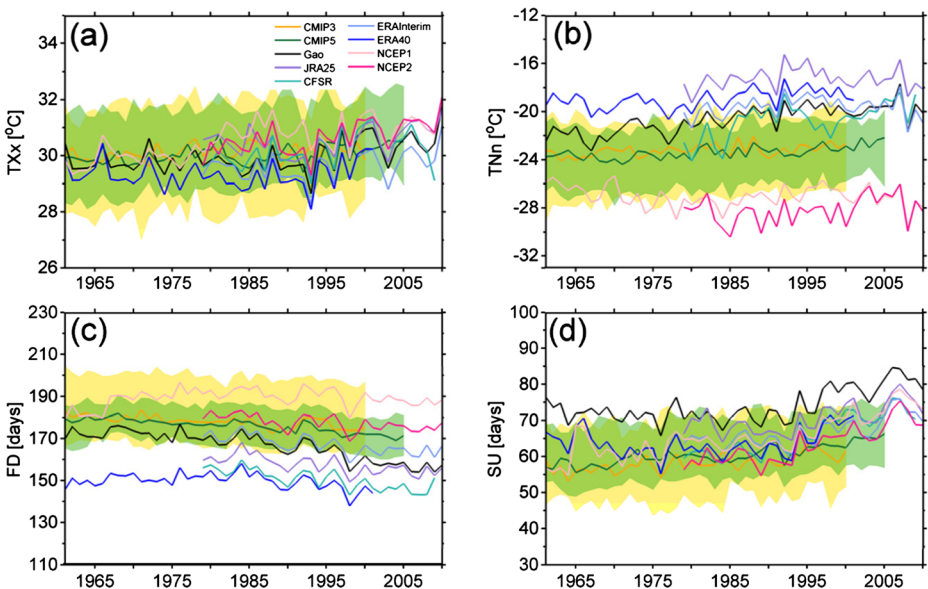


Fig. 3 Regional means of temperature indices in China from 1961 to 2005 based on the median ensemble of 30 CMIP5 and from 1961 to 2000 based on 19 CMIP3 models. The shading indicates the interquartile model spread (between the 25th and the 75th quartiles). The observed results (black) from 1961 to 2010 are also shown. Also shown are the reanalysis JRA25 from 1979 to 2010, CFSR from 1979 to 2009, ERA40 from 1961 to 2001, ERA-Interim from 1979 to 2010, NCEP1 from 1961 to 2010, and NCEP2 from 1979 to 2010

Although substantial differences are observed among the models, reanalyses, and observations, the spread is reduced for the anomalies with respect to the reference period of 1981–2000 (Figures not shown) when the absolute values of the indices are considered. Furthermore, the long-term trends in the historical evolution of these indices are more clearly visible for the anomalies. Generally, the warm extreme indices show an increase over time and the cold extreme indices show a decrease. For instance, FD exhibits a statistically significant decrease, whereas SU present an increase, particularly after the 1990s. The linear trend of each index is well captured for the individual models, and a high model agreement is observed for both the CMIP3 and the CMIP5 models, despite a large spread of the calculated RMSEs (Table S4).

3.2 Precipitation extremes

3.2.1 *Spatial performance*

The simulated climatology of each extreme precipitation index is compared with the observations for the reference period of 1981–2000. The main features of the climatological spatial patterns of precipitation-related indices, as represented by the annual maximum 5-day precipitation (RX5day), very wet days (R95p), and very heavy precipitation days (R20mm), are reasonably well captured by the CMIP5 models. In particular, the northwest–southeast contrast of the annual mean precipitation and its related extreme indices are well simulated. However, the regional difference in the simulated magnitudes of the precipitation extremes is clearly visible when compared with the observations in China (Figures S6 and S7).

The overestimation is pronounced in total wet-day precipitation (PRCPTOT) as well as in the simple daily precipitation intensity (SDII, Figures S6 and S7) in China. The high PRCPTOT values in southern China are reasonably well captured by the CMIP5 median, whereas they are overestimated in the other regions, particularly in northern China and Tibet. This difference may be attributed to the difficulty of coarse resolution models to represent the complex topography over these regions (e.g., Gao et al. 2012) despite the finer resolutions of the CMIP5 models. This result is also the case for the indices of RX5day, R95p, and R20mm. However, the underestimation of these extreme indices is also clear over southern China and has also been revealed by previous studies (Chen 2013; Chen and Sun 2013). The opposite case is observed for CDD (consecutive dry days) whose climatological structure is reasonably well captured by the CMIP5 median; however, it is overestimated over southern China and some parts of North Xinjiang and underestimated over the other regions (Figures S6 and S7).

A concise statistical analysis of all the precipitation indices is presented in Fig. 4. The model spread of most of the precipitation indices is much larger than that of the temperature indices in both the CMIP3 and CMIP5 models. An intercomparison of these models further indicates that there is almost no decrease of the model spread in CMIP5 compared to CMIP3 for most of the precipitation indices, unlike the results of the temperature indices. However, it can be observed that the multimodel median often outperforms the individual models. Generally, the CMIP5 models perform better. These features are also verified by the evaluation of spatial RMSEs of the precipitation indices (Figure S4b). The RMSE values of the CMIP5 models compare well with those of CMIP3, and there is almost no improvement from CMIP3 to CMIP5. The interquartile model spread in the CMIP5 models is as large as or even greater than that in CMIP3. Additionally, large RMSEs are also observed for reanalyses, and the values are comparable with the simulations; even the spread between these six reanalyses is much larger than the interquartile model spread for most of the precipitation indices.

As mentioned above, over the dry regions of China, the CMIP5 and CMIP3 models generally simulate shorter dry spells compared with the observations. However, over most

regions in China, both the CMIP5 and CMIP3 models simulate much longer wet spells (CWD, consecutive wet days) relative to the observations, and the CMIP5 models simulate longer wet spells than the CMIP3 models (figure not shown). Among the 11 precipitation indices, the largest spatial variance ratio is found in CWD, but a better model agreement is observed in CMIP5 than in CMIP3 (Fig. 4d).

Overall, both the CMIP5 and CMIP3 models have a relatively weaker ability to simulate the climatic means of precipitation and precipitation-related indices than to simulate the temperature-related indices in China. This is also the case for the multimodel median ensembles. Furthermore, no obvious improvement is observed from CMIP3 to CMIP5 in simulating the precipitation indices in China in terms of the RMSE. Thus, the further improvement of the model ability to simulate precipitation and its related extremes is still a large challenge for the next generation model.

3.2.2 Temporal performance

Temporal evolutions of the absolute values of the precipitation indices averaged over China are presented in Fig. 5. Generally, the interquartile model spread of CMIP5 is smaller than that of CMIP3, except for CDD. The case is opposite for median ensembles, which tend to be much

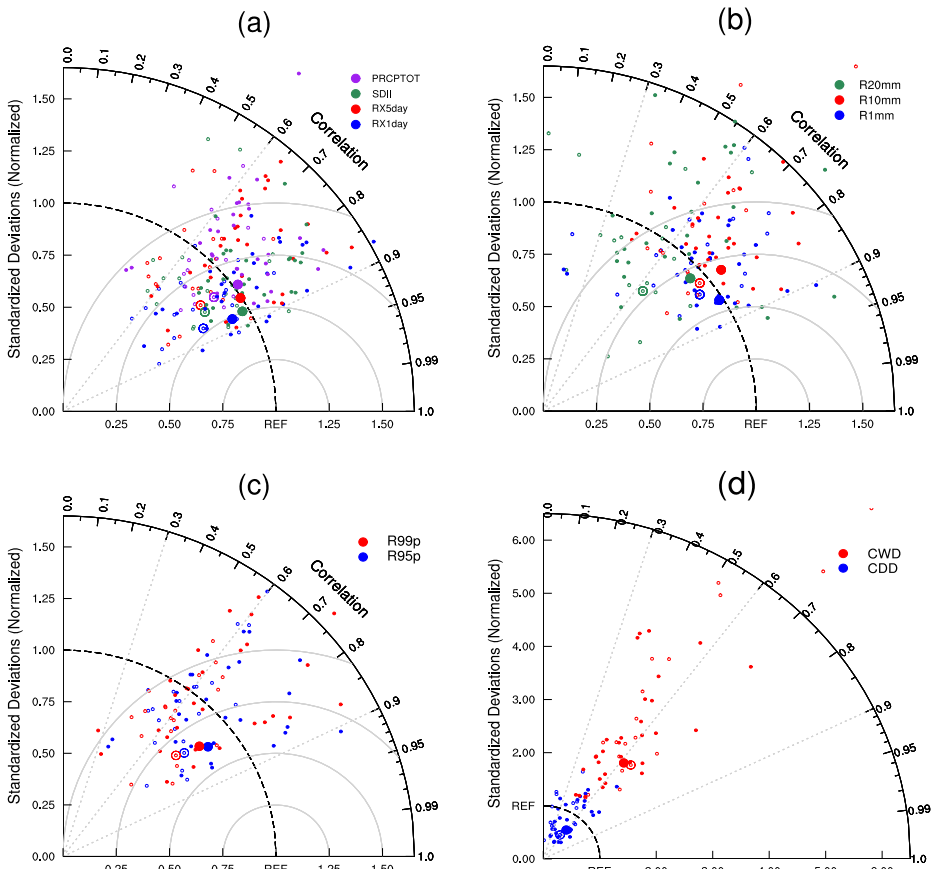


Fig. 4 Same as in Fig. 2 but for the 11 extreme precipitation indices

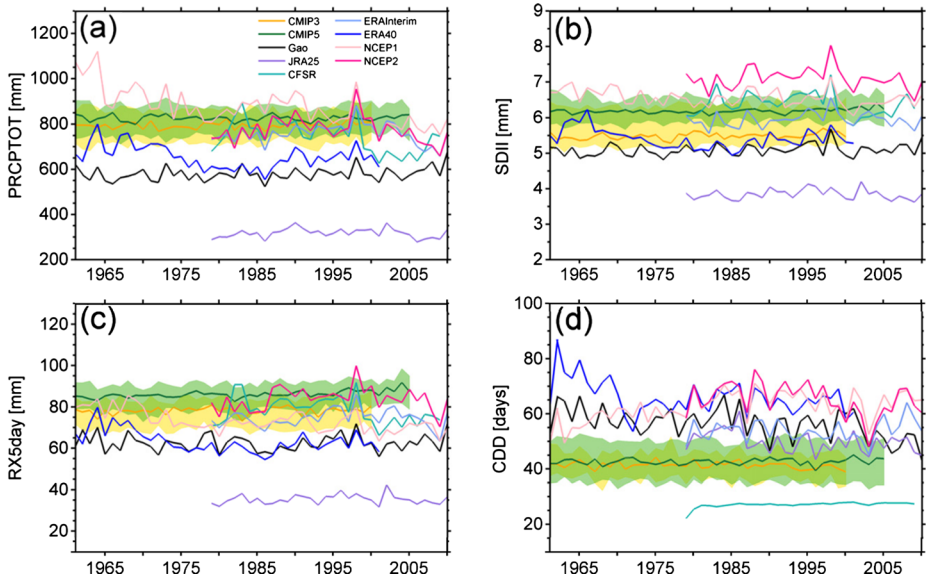


Fig. 5 Same as in Fig. 3 but for the extreme precipitation indices

higher in CMIP5. Sillmann et al. (2013a) partly attribute this finding to the improvement in the spatial resolutions of the CMIP5 models, where high-resolution models generally simulate greater values of precipitation extremes compared with low-resolution models.

Compared with the observations, both the CMIP5 and CMIP3 models simulate higher values of precipitation indices (excluding CDD), as discussed previously. Large differences are observed between the models, reanalyses, and observations for all the precipitation indices and are clearly visible from the RMSE evaluations (Table S4). The interquartile model spread of some indices is somewhat reduced in CMIP5 compared to CMIP3 (with exception of CDD), but the associated uncertainties are still considerable. The temporal evolutions of six reanalyses also present large differences, and the spread is much larger than the interquartile model spread for the precipitation indices. Additionally, discrepancies between CMIP5 and CMIP3 are found when simulating the trends of several indices (Table S4). The observed increase or decrease is well captured by the CMIP3 median ensemble but not by CMIP5. Taking PRCPTOT and R1mm as examples, the observed increases are well reproduced by the CMIP3 medians, whereas the CMIP5 medians reveal negative tendencies. These results are generally evaluated with a relatively low model consistency both in CMIP3 and CMIP5, although these trends are within the observed confidence interval of 5 to 95 % levels.

4 Relative model performances

From the above metric analysis, we can see that the model performance generally varies from one extreme index to another for individual CMIP5 and CMIP3 models. Thus, it is difficult to simply describe the model skill in simulating these extreme indices. To further understand the model evaluation, two additional metrics, i.e., an exploratory model climate performance index (MCPI) for climatology and an exploratory model variability index (MVI) for the variability of all fields proposed and constructed by Gleckler et al. (2008), are used. On the basis of these

two metrics, a brief discussion of climate model performance in simulating extreme indices in China is presented in the following paragraphs.

The MCPI is defined as the simple mean of each model’s relative errors across the 27 ETCCDI indices discussed in this study. A negative value of the MCPI generally indicates a higher skill than the typical model, and a positive value indicates a lower skill. The MVI is a metric analysis of how well a model simulates the interannual variability; a relatively small value generally indicates a better agreement with the reference data. For more detailed information on these two metric analyses, please refer to the study of Gleckler et al. (2008).

Figure 6a presents the “portrait” diagram that provides a summary of the relative performances of the individual models in simulating the extreme indices; they are evaluated on the

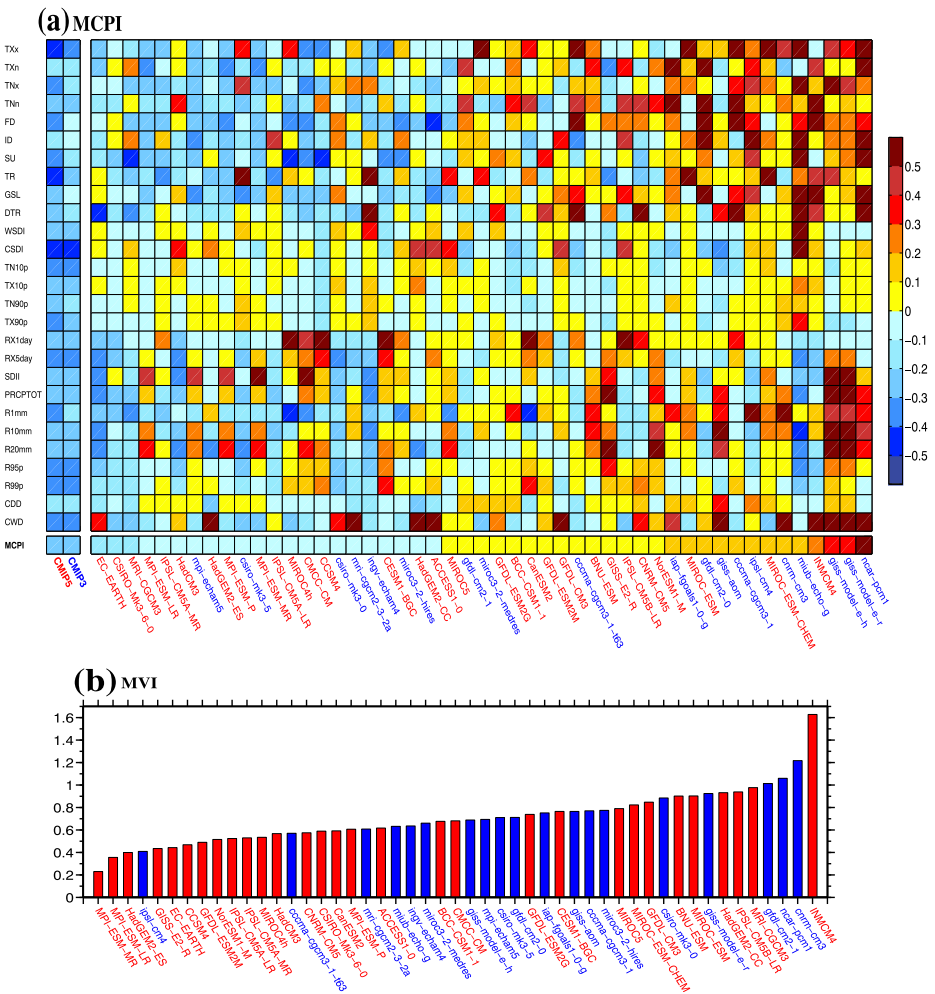


Fig. 6 **a** Portrait diagram displaying the relative error metrics for both CMIP5 (red) and CMIP3 (blue) of the annual cycle climatologies (1981–2000) with respect to the observations. **b** Display of model variability index, defined as the multivariable mean of the ratio of simulated to observed variance for both CMIP5 (red) and CMIP3 (blue) models. Smaller values in (a) and (b) indicate a better agreement with the observations, and the models are ranked according to the model climate performance index (averaged over relative error metrics of the 27 ETCCDI indices) in (a) and the model variability index in (b)

basis of the relative magnitudes of the spatially and temporally averaged RMSE. The portraits are arranged such that the rows are labeled by the extreme indices and the columns are the model names. The models are ranked in order of their performance according to the MCPI values. In this figure, the extent of the model errors are indicated with colors; the colder colors indicate the better performance of a model than the typical model, on average, and the warmer colors indicate the opposite.

For the 27 ETCCDI indices, a large difference is observed in the model performance. No model can match the observations reasonably well for any index, and even “better” models have a large spread in their variable-by-variable performance. For example, the model EC-EARTH presents relatively higher performances for most indices than the typical model, but with a large relative error in CWD. For temperature-based indices, some models, including EC-EARTH, CSIRO-Mk3-6-0, MPI-ESM-LR, MPI-ESM-P, MPI-ESM-MR, CMCC-CM, CCSM4, and CESM1-BGC, as well as some models in CMIP3 (*mpi-echam5*), reveal a relatively high performance. However, the CMIP5 models generally perform much better than the CMIP3 models; some CMIP3 models present larger relative errors for most indices than the typical models, such as *ncar-pcm1*, *giss-model-e-r*, *miub-echo-g*, and *ipsl-cm4*. Precipitation-based indices are also represented reasonably well by some models, such as EC-EARTH, CSIRO-Mk3-6-0, MRI-CGCM3, and IPSL-CM5A-LR in CMIP5, as well as by some models in CMIP3, including *csiro-mk3-5*, *csiro-mk3-0*, *ingv-echam4*, and *miub-echo-g*. However, higher relative errors are generally observed in CMIP3 models, including *ncar-pcm1*, *giss-model-e-r*, *giss-model-e-h*, and *giss-aom*.

A comparison of model performances in simulating temperature- and precipitation-based indices shows that no model can simulate all the indices better than the other models. Among the 49 models in CMIP5 and CMIP3, CSIRO-Mk3-6-0 performs better than the typical model regarding most indices, except for TXn, FD, ID, and SDII. For the other models, higher errors of most temperature-based indices generally correspond to lower errors of most precipitation-based indices, while relatively lower errors of most temperature-based indices generally correspond to higher errors of most precipitation-based indices. For example, the patterns of most of the precipitation indices are reasonably well captured by the model *miub-echo-g*, whereas relatively larger errors are observed for almost all the temperature indices compared to those of the typical model.

In addition to the individual models, the results of the multimodel median ensembles for both CMIP5 and CMIP3 are also presented on the far left of the panel. A median model is obtained by first calculating the multimodel median for each index, and then deriving its relative RMSE. Clearly, the median models generally outperform the individual models of both CMIP5 and CMIP3 because some of the systematic errors are canceled out in the multimodel median. A comparison between CMIP5 and CMIP3 further shows that the CMIP5 median presents a better simulation than the CMIP3 median for most indices.

The average relative error of each model (i.e., MCPI) is displayed at the bottom row of Fig. 6a. A total of 22 models (16 in CMIP5 and 6 in CMIP3) show relatively lower MCPI values than the typical model. According to the MCPI values, the model EC-EARTH shows the high performance in simulating the extreme indices in China, but the *ncar-pcm1* performs much poor. Of course, the model medians outperform the individual models (MCPI: -0.29 in CMIP5 and -0.23 in CMIP3, respectively). The result from Fig. 6a also indicates an unclear relationship between the model’s spatial resolution and its corresponding performance in simulating extreme indices at the regional scale. For example, the MCPI for the MIROC4h model, which has the finest resolution in CMIP5, is much larger than that for the CSIRO-Mk3-6-0 model, which has a coarse resolution.

Further intercomparisons between CMIP3 models and their updated versions in CMIP5 reveal that the updated models generally show a higher skill than the old models; for example, the CSIRO-Mk3-6-0 model corresponds to *csiro-mk3-5* and *csiro-mk3-0*, MRI-CGCM3 corresponds to *mri-cgcm2-3-2a*, IPSL-CM5A-MR and IPSL-CM5A-LR correspond to *ipsl-cm4*, MIROC4h corresponds to *miroc3-2-hires* and *miroc3-2-medres*, and so on. However, the *gfdl-cm2-1* model shows a relatively lower error than its updated version GFDL-CM3, although both errors are larger than those of a typical model.

The model performance in simulating the interannual variability based on the model variability index is depicted in Fig. 6b. Among the top 22 models that display a relatively lower MVI in simulating the temporal variability of extreme indices, 17 models are from CMIP5 and 5 are from CMIP3; thus, most CMIP5 models provide better performances than the CMIP3 models. Two models from MPI (MPI-ESM-MR and MPI-ESM-LR) perform well and INMCM4 performs relatively poor among the models. An intercomparison between CMIP3 and the corresponding updated models in CMIP5 also indicates the higher skills of the updated models in capturing interannual variability. However, some models show lower skills despite significant improvements in their new versions, for example, *mri-cgcm2-3-2a* corresponding to MRI-CGCM3.

Is there any relationship between how well a model simulates the climate mean and its ability to capture the interannual variability on a regional scale? For this purpose, we compare each model's climate and variability simulation skills in China (Figure S8). For both the MCPI and MVI, a smaller value corresponds to a better skill; thus, "better" models are found in the lower left corner of Figure S8. A higher density and shorter distance to the lower left corner of the figure are observed for most models in CMIP5, unlike those in CMIP3. Additionally, there seems to be a relationship between a model's relative skill in simulating the climate mean and its interannual variability performance; the calculated correlation is 0.45 for 49 models. Furthermore, this relationship appears to be much stronger in the CMIP5 models than in the CMIP3 models, with correlations of 0.57 and 0.37, respectively. This result is exciting for both model developers and users as the models show increased skills in simulating both the mean and variability following their modification, although this is not the case for several models.

5 Conclusions

The main purpose of this study is to provide a systematic analysis of the performances of the CMIP5 models in simulating climate extreme indices in China by comparing the results to CMIP3 models, new gridded observations and six reanalysis datasets. The 27 extreme indices defined by ETCCDI are calculated using a consistent methodology for the models, reanalyses, and observations.

Regarding the temperature-based indices, our results indicate that the spatial patterns and trends of the climate extremes in China are captured reasonably well by the CMIP5 models. However, there are some discrepancies due to the cold bias in model simulations compared to the observations. For example, an overestimation of FD and ID but an underestimation of TNn and GSL are apparent. Nevertheless, the increase observed over recent decades is well simulated by the models for certain temperature indices, such as TXx, TNn, GSL, and for the number of SU and TR. Decreases are simulated for the number of FD and ID; this result agrees with the observational results. These high skills are also well reflected in the temperature percentile indices for both spatial patterns and trends. Similar results are observed in the CMIP3 simulations, but a larger inter-model spread is observed among the CMIP3 models when compared with the CMIP5 models.

Both the CMIP3 and CMIP5 models generally show relatively higher skills to simulate temperature indices than precipitation indices and the median ensembles appear to outperform the individual models. The results from CMIP5 indicate that the overestimation is pronounced in PRCPTOT, SDII and some other precipitation-related indices in China but with regional differences. A large inter-model uncertainty is observed in simulating temporal variations; however, model agreement is apparent for some indices. Compared with the CMIP3 simulations, we also note that the CMIP5 models tend to simulate more intense precipitation events in China, which is consistent with the results on a global scale (Sillmann et al. 2013a). Additionally, the CMIP5 median ensembles show much larger precipitation than the observations and the CMIP3 ensembles. This discrepancy may partly be due to the improvement in the spatial resolution of the CMIP5 models such that more light precipitation events are simulated (Chen and Sun 2013; Sillmann et al. 2013a).

The variations of extreme indices in China that derived from the six reanalyses compare well with the observations, but a discrepancy is clearly visible. The difference between these reanalyses is greater than or equal to the interquartile model spread of both the CMIP3 and CMIP5 results for some indices (e.g., FD, SU, RX5day, and CDD). Furthermore, it is difficult to say which reanalysis is better for the comparison with the simulations of indices. Thus, the reference data should be carefully selected when evaluating model performances of simulating extreme indices patterns at the regional scale.

We also compare model performances on the basis of the average relative errors (MCPI) and the ratio of simulated to observed variance (MVI) for all extreme indices considered. A conclusion is that the updated CMIP5 models generally perform better than their corresponding old versions in CMIP3 when simulating extreme indices in China in terms of both climate and variability simulations. Furthermore, this improvement in the climate mean simulation is often accompanied by an improved skill in the variability simulation for some models of CMIP5. However, this result requires additional validation in the future. Based on this study, the projected changes in the climate extreme indices in China will be the focus of the next study.

Acknowledgments We sincerely acknowledge the two anonymous reviewers whose kindly and valuable comments that greatly improved the manuscript. This research was jointly supported by the National Basic Research Program of China (Grant No. 2012CB955401), National Natural Science Foundation of China (Grant No. 41305061), and the ‘Strategic Priority Research Program-Climate Change: Carbon Budget and Relevant Issues’ of Chinese Academy of Sciences (XDA05090306).

References

- Alexander L et al (2006) Global observed changes in daily climate extremes of temperature and extremes. *J Geophys Res* 111:D05109. doi:10.1029/2005JD006290
- Chen HP (2013) Projected change in extreme rainfall events in China by the end of the 21st century using CMIP5 models. *Chin Sci Bull* 58(12):1462–1472
- Chen HP, Sun JQ (2013) Projected change in East Asian summer monsoon precipitation under RCP scenario. *Meteorol Atmos Phys* 121:55–77
- Chen HP, Sun JQ (2014) Changes in climate extreme events in China associated with warming. *Int J Climatol*. doi:10.1002/joc.4168
- Chen HP, Sun JQ, Chen XL, Zhou W (2012) CGCM projections of heavy rainfall events in China. *Int J Climatol* 32:441–450

- Chen HP, Sun JQ, Chen XL (2014) Projection and uncertainty analysis of global precipitation-related extremes using CMIP5 models. *Int J Climatol* 34:2730–2748. doi:[10.1002/joc.3871](https://doi.org/10.1002/joc.3871)
- Donat MG et al (2013a) Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: the HadEX2 dataset. *J Geophys Res* 118:2098–2118
- Donat MG et al (2013b) Changes in extreme temperature and precipitation in the Arab region: long-term trends and variability related to ENSO and NAO. *Int J Climatol*. doi:[10.1002/joc.3707](https://doi.org/10.1002/joc.3707)
- Frich P, Alexander L, Della-Marta P, Gleason B, Haylock M, Klein Tank A, Peterson T (2002) Observed coherent changes in climate extremes during the second half of the twentieth century. *Clim Res* 19:193–212
- Gao XJ, Giorgi F (2008) Increased aridity in the Mediterranean region under greenhouse gas forcing estimated from high resolution simulations with a regional climate model. *Glob Planet Chang* 62:195–209
- Gao XJ, Zhao ZC, Giorgi F (2002) Changes of extreme events in regional climate simulations over East Asia. *Adv Atmos Sci* 19:927–942
- Gao XJ, Shi Y, Zhang DF, Giorgi F (2012) Climate change in China in the 21st century as simulated by a high resolution regional climate model. *Chin Sci Bull* 57(10):1188–1195
- Gleckler PJ, Taylor KE, Doutriaux C (2008) Performance metrics for climate models. *J Geophys Res* 113: D06104. doi:[10.1029/2007JD008972](https://doi.org/10.1029/2007JD008972)
- IPCC (2007) Climate change 2007: The physical science basis, contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge
- IPCC (2012) Changes in climate extremes and their impacts on the natural physical environment. In: Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups I and II of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, pp109–230
- IPCC (2013) Climate Change 2013: The physical science basis. Contribution of working group I to the fifth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp
- Min SK, Zhang X, Zwiers FW, Hegerl GC (2010) Human contribution to more-intense precipitation extremes. *Nature* 470:378–381
- Morak S, Hegerl GC, Kenton (2011) Detectable regional changes in the number of warm nights. *Geophys Res Lett* 38:L17703. doi:[10.1029/2011GL048531](https://doi.org/10.1029/2011GL048531)
- Orlowsky B, Seneviratne SI (2012) Global changes in extreme events: regional and seasonal dimension. *Clim Chang* 110:669–696
- Peterson TC, Zhang X, Brunet-India M, Vázquez-Aguirre JL (2008) Changes in North American extremes derived from daily weather data. *J Geophys Res* 113:D07113. doi:[10.1029/2007JD009453](https://doi.org/10.1029/2007JD009453)
- Seo KH, Ok J, Son JH (2013) Assessing future changes in the East Asian summer monsoon using CMIP5 coupled models. *J Clim* 26:7662–7675
- Sillmann J, Roeckner E (2008) Indices for extreme climate events in projections of anthropogenic climate change. *Clim Chang* 86:83–104
- Sillmann J, Kharin VV, Zhang X, Zwiers FW, Bronaugh D (2013a) Climate extremes indices in the CMIP5 multimodel ensemble: part 1. Model evaluation in the present climate. *J Geophys Res* 118:1716–1733. doi:[10.1002/jgrd.50203](https://doi.org/10.1002/jgrd.50203)
- Sillmann J, Kharin VV, Zwiers WF, Zhang X, Bronaugh D (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](https://doi.org/10.1002/jgrd.50188)
- Sillmann J, Kharin VV, Zwiers FW, Zhang X, Bronaugh D, Donat MG (2013c) Evaluating model-simulated variability in temperature extremes using modified percentile indices. *Int J Climatol*. doi:[10.1002/joc.3899](https://doi.org/10.1002/joc.3899)
- Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93:485–498
- Wang HJ et al (2012) Extreme climate in China: facts, simulation and projection. *Meteorol Z* 21(3):279–304
- Wu J, Gao XJ (2013) A gridded daily observation dataset over China region and comparison with the other datasets (in Chinese). *Chin J Geophys* 56(4):1102–1111
- Zhai PM, Zhang XB, Wan H, Pan XH (2005) Trends in total precipitation and frequency of daily precipitation extremes over China. *J Clim* 18:1096–1108
- Zhang X, Alexander L, Hegerl GC, Jones P, Tank AK, Peterson TC, Trewin B, Zwiers FW (2011) Indices for monitoring changes in extremes based on daily temperature and precipitation data. *WIREs Clim Chang* 2: 851–870