



**HAL**  
open science

## Does CMIP6 Inspire More Confidence in Simulating Climate Extremes over China?

Huanhuan Zhu, Zhihong Jiang, Juan Li, Wei Li, Cenxiao Sun, Laurent Li

► **To cite this version:**

Huanhuan Zhu, Zhihong Jiang, Juan Li, Wei Li, Cenxiao Sun, et al.. Does CMIP6 Inspire More Confidence in Simulating Climate Extremes over China?. *Advances in Atmospheric Sciences*, Springer Verlag, 2020, 37 (10), pp.1119-1132. 10.1007/s00376-020-9289-1 . hal-03047353

**HAL Id: hal-03047353**

**<https://hal.archives-ouvertes.fr/hal-03047353>**

Submitted on 8 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Does CMIP6 inspire more confidence in simulating climate extremes over China?**

**Huanhuan Zhu<sup>1</sup>, Zhihong Jiang<sup>\*1,2</sup>, Juan Li<sup>1,2</sup>, Wei Li<sup>1,2</sup>, Cenxiao Sun<sup>1</sup>, Laurent Li<sup>3</sup>**

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

*2. Joint International Research Laboratory of Climate and Environment Change, Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disaster, Nanjing University of Information Science and Technology, Nanjing, China*

*3. Laboratoire de Météorologie Dynamique, CNRS, Sorbonne Université, Ecole Normale Supérieure, Ecole Polytechnique, Paris, France*

\*Corresponding author : Zhihong Jiang

Email: zhjiang@nuist.edu.cn

## Abstract

Based on climate extreme indices calculated from a high-resolution daily observational dataset in China during 1961-2005, the performance of 12 climate models from phase 6 of the Coupled Model Intercomparison Project (CMIP6), and 30 models from the phase 5 of CMIP (CMIP5), are assessed in terms of spatial distribution and interannual variability. CMIP6 multi-model ensemble mean (CMIP6-MME) can well simulate the spatial pattern of annual mean temperature, maximum daily maximum temperature and minimum daily minimum temperature. However, CMIP6-MME has difficulties to reproduce cold nights and warm days, and has large cold biases over the Tibetan Plateau. Its performance in simulating extreme precipitation indices is generally lower than in simulating temperature indices. Compared to CMIP5, CMIP6 models show improvements in the simulation of climate indices over China. This is particularly true for precipitation indices for both the climatological pattern and the interannual variation, except for the consecutive dry days. The areal-mean bias for total precipitation has been reduced from 127% (CMIP5-MME) to 79% (CMIP6-MME). The most striking feature is that the dry biases in southern China, very persistent and general in CMIP5-MME, are largely reduced in CMIP6-MME. A stronger ascent together with more abundant moisture can explain this reduction of dry biases. Wet biases for total precipitation, heavy precipitation and precipitation intensity in the eastern Tibetan Plateau are still present in CMIP6-MME, but smaller, compared to CMIP5-MME.

**Keywords:** CMIP6; CMIP5; Intercomparison; Climate Extremes

### Article Highlights:

CMIP6 models, as CMIP5 models, generally perform better in simulating  $T_{av}$ ,  $TX_x$  and  $TN_n$  than in simulating extreme precipitation indices.

The persistent dry biases in southern China in CMIP5-MME are largely reduced in CMIP6-MME.

CMIP6 models show obvious improvements in simulating precipitation extremes compared with CMIP5 models.

## 1. Introduction

General circulation models (GCMs) are important tools for understanding the climate system, reproducing its past, predicting and projecting its future changes. To make models comparable to each other, the Working Group on Coupled Modelling (WGCM) in the framework of World Climate Research Programme (WCRP) established the Coupled Model Intercomparison Project (CMIP). CMIP is also devoted to providing standardized climate simulations and outputs (Meehl et al., 2007), and to facilitating the use of such simulations among different scientific communities. In the past years, CMIP made great contributions to the last assessment reports produced by the Intergovernmental Panel on Climate Change (IPCC) (IPCC 2007, 2012, 2013).

According to the IPCC Fifth Assessment Report, CMIP5 models exhibit improvements in the simulation of surface temperature and large-scale precipitation compared to the previous exercise of CMIP (IPCC 2013). They generally have a better skill in simulating surface air temperature than in simulating precipitation (IPCC 2007, 2012, 2013; Sillmann et al. 2013; Flato et al., 2014; Koutroulis et al., 2016). It is also shown that CMIP5 models are more skillful than CMIP3 models in simulating various aspects of the Asian summer monsoon (Sperber et al., 2013), as well as the climatological spatial pattern and the dominant mode of summer precipitation in the Pan-Asian monsoon region (Gao et al., 2015).

Recently, several studies focused on the capability of CMIP models in reproducing climate over China. The same conclusion was reached with a better performance for CMIP5 models, compared to their precedent generation CMIP3 models (Committee of Chinese National Assessment Report on Climate Change, 2015). However, there are some discrepancies for both temperature and precipitation. For instance, cold biases are generally present in western China (Guo et al., 2013; Sun et al., 2015), and overestimated precipitation in the eastern part of the Tibetan Plateau (Xu et al., 2010; Su et al., 2013). Extreme precipitation is also generally

overestimated, especially in western China and in mountainous regions, while precipitation in southern China is always underestimated (Jiang et al., 2009; Jiang et al., 2012, 2015; Ou et al., 2013; Chen et al., 2014; Chen and Sun, 2014; Chen and Frauenfeld, 2014; Kusunoki and Arakawa, 2015).

Nowadays, CMIP enters into its phase 6 (CMIP6). CMIP6 models have higher spatial resolution and improved parameterization schemes for major physical and biogeochemical processes of the climate system (Eyring et al., 2016). Most modeling groups are releasing their new simulations with the published documentation of their model evolution from CMIP5 to CMIP6 (Park et al., 2019; Kawai et al., 2019; Wu et al., 2019; Gusain et al., 2020). But few works were devoted to assessing the ensemble behaviors of CMIP6 models, especially in simulating climate extremes over China. This is actually what we want to address in this study. Two questions constitute our main motivation: How is the performance of CMIP6 models in terms of MME (multi-model ensemble) in simulating current climate extremes over China? What is the improvement from CMIP5 to CMIP6 for this regard?

In order to answer these questions, simulations from 12 CMIP6 climate models have been quantitatively assessed with skill-score metrics. We took all CMIP6 simulations available on ESGF when we started this work in August 2019. Their performance is compared with that of 30 existing CMIP5 models. We are aware that the 12 CMIP6 models used here are not necessarily the successor of the 30 CMIP5 models, but our goal is to assess the ensemble behaviors of the two phases of CMIP. This objective is quite distinct from that of each individual model to document changes of model constitution and performance. In this work, we want to provide a reliable scientific basis for end-users of CMIP6 simulations who are interested in the projection of future climate changes in China.

## **2. Data and methods**

### **2.1 Data**

As a reference from observation, we used the daily gridded dataset, CN05.1, with

a resolution of  $0.5^\circ$  in both latitude and longitude, including four basic variables: temperature (TM), daily-maximum temperature (TX), daily-minimum temperature (TN), and precipitation. It was provided by the National Climate Center of the China Meteorological Administration based on 2416 observation stations, unevenly covering whole China (Wu and Gao, 2013). This dataset has been widely used in researches on climate change over China (Dong et al., 2015; Xu et al., 2018).

The simulated daily maximum temperature and daily minimum temperature and precipitation from 12 CMIP6 models and 30 CMIP5 models were retrieved through data portals of the Earth System Grid Federation. Only the first historical realization was analyzed for each model. For both models and observations, we used in this study the same period from 1961 to 2005. The used models with their basic information are listed in Table 1 for CMIP5 and Table 2 for CMIP6, respectively.

Different climate indices from different models and observations were firstly calculated at their native grids. To facilitate the inter-comparison, a bilinear interpolation scheme was used to interpolate all indices to a common  $1^\circ \times 1^\circ$  grid.

Monthly vertical velocity together with meridional wind (V) and specific humidity were also used in our study to search possible causes for the better performance of CMIP6-MME to reproduce precipitation in south China. The corresponding variables from the reanalysis National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP/NCAR) with a resolution of  $2.5^\circ$  were used as a reference for the same period 1961-2005. All models were interpolated to the  $2.5^\circ \times 2.5^\circ$  grid to facilitate the inter-comparison.

## **2.2 Climate indices**

In this work, we consider 10 climate indices, including the annual average temperature and nine extreme indices defined by the Expert Team on Climate Change Detection and Indices (ETCCDI; <http://etccdi.pacificclimate.org/>; see Table 3 for details). These indices have been widely used in climate change researches and considered as representative for model performance (Frich et al., 2002; Zhang et al.,

2011; Zhou et al., 2014; Akinsanola et al., 2019).

## 2.3 Evaluation method

### 2.3.1 Taylor diagram

To evaluate the overall skill in reproducing the spatial pattern of the present-day climate indices, the Taylor diagram and Taylor skill scores (TS) (Taylor 2001; Wang et al., 2018) were used. Taylor diagram provides a concise statistical summary of the degree of correlation (PCC; pattern correlation coefficient), centered root mean square error (RMSE), and the ratio of spatial standard deviation (RSD). The similarity between observations and simulations can be quantified by their correlation and the amplitude of the variability. A perfect simulation would be that with the centered RMSE equal to 0 and both the PCC and RSD close to 1. Taylor skill score (TS) is a combined measure and calculated as:

$$TS = \frac{4(1+R)^2}{\left(\frac{\sigma_{so}}{\sigma_{sm}} + \frac{\sigma_{sm}}{\sigma_{so}}\right)^2 (1+R_0)^2} \quad (1)$$

where  $R$  is the spatial correlation coefficient between the simulation and observation,  $R_0$  is the maximum correlation coefficient attainable (here we use 0.999).  $\sigma_{sm}$  and  $\sigma_{so}$  are the standard deviations of the simulated and observed spatial patterns, respectively. The score equals 1 for a perfect match between the model and observation, and 0 for an inverse model performance.

### 2.3.2 Interannual variability skill scores (IVS)

The method to quantitatively express interannual variability skill score (IVS) is the same as employed in Chen et al. (2011) :

$$IVS = \left(\frac{\sigma_{to}}{\sigma_{tm}} - \frac{\sigma_{tm}}{\sigma_{to}}\right)^2 \quad (2)$$

where  $\sigma_{tm}$  and  $\sigma_{to}$  denote the interannual standard deviation of model simulations and observations, respectively. Smaller IVS values indicate better performance of the

simulation.

### **3. The evaluation of CMIP6 models and comparison with CMIP5 models**

#### **3.1 Temperature Indices**

##### **3.1.1 Climatology**

In order to evaluate the capability of models in reproducing temperature indices over China, Fig. 1 shows the box-and-whisker plots and the spatial distribution of biases between simulations (CMIP6-MME and CMIP5-MME) and observation from 1961 to 2005. All the box-and-whisker plots use the commonly-used convention with upper and lower limits of the box indicating the 75th and 25th percentile values, the horizontal line in the box indicating the median, the dot in the box indicating the mean, and the whiskers showing the 90th and 10th percentile values. It can be seen that CMIP6-MME has a general cold bias throughout the country. The areal-mean bias of CMIP6-MME for annual mean temperature ( $T_{av}$ ), maximum daily maximum temperature ( $TXx$ ) and minimum daily minimum temperature ( $TNn$ ) in whole China is  $-1.64^{\circ}\text{C}/-0.45^{\circ}\text{C}/-4.51^{\circ}\text{C}$ . Large cold biases are located in the Tibetan Plateau, where the local bias is more than  $4^{\circ}\text{C}$  for  $T_{av}$  and  $8^{\circ}\text{C}$  for  $TNn$ . Compared with CMIP5-MME, CMIP6-MME does not present obvious differences for  $T_{av}/TNn$ . Except for  $TXx$ , the general warm bias ( $0.93^{\circ}\text{C}$ ) in CMIP5-MME becomes a general cold bias ( $-0.45^{\circ}\text{C}$ ) in CMIP6-MME (Fig. 1e and f). Regions, where biases in CMIP6-MME are lower than that in CMIP5-MME and their difference is statistically significant at the 5% level are dotted in the middle panels. The dotted areas in Fig. 1e represent the simulated bias for  $TXx$  from CMIP6-MME is significantly reduced in north China and parts of Xinjiang. The dotted areas are mainly located in Xinjiang for  $T_{av}$  and  $TNn$ , which indicates the bias over Xinjiang also is reduced in CMIP6-MME (Fig. 1b and h).

Generally speaking, it is difficult to evaluate model performances with percentile



indices because the mean threshold exceedance rate in the base period is approximately the same for all models and observations (Sillmann et al. 2014; Chen and Sun, 2015). The calculations of these percentile indices for the models and observations are implemented over the standard base period of 1961-1990, and the climatic mean analyses are calculated over the period of 1961-2005. There is a systematic overestimation from CMIP6-MME for cold nights (TN10p), but a systematic underestimation for warm days (TX90p). CMIP5-MME shows similar spatial characteristics for TN10p. But the simulated bias is basically distributed as a north-south dipole for TX90p, with negative bias in the north and positive bias in the south.

Apart from biases measuring the deviation of models from observation, the inter-model spreading is also considered as an important assessment for CMIP models, since it indicates the degree of consensus across the climate modelling community, and it ultimately serves as a surrogate to measure uncertainty of climate models, especially for projection of future climate. With this idea in mind, we now evaluate the standard deviation (SD) among models for CMIP5 and CMIP6 respectively. We followed what was done in Jiang D. et al. (2016) and display the areal-mean SD over whole China in the upper side of each relevant panel in Fig. 1. The inter-model SD for all CMIP6 models (CMIP6-twelve) is generally smaller than that of all CMIP5 models (CMIP5-thirty) for most temperature indices (except for TNn). We think that such a difference is not significant, since CMIP5 (30 members) is more diverse than CMIP6 (12 members). Results (not shown) from a subset of six CMIP6 models (CMIP6-six) and their CMIP5 predecessors (CMIP5-six) do not permit to conclude significant differences, neither, which confirms that the inter-model spreading is quite comparable between CMIP5 and CMIP6.

Taylor diagrams and TS are also used to further evaluate the overall skills of models in reproducing the spatial pattern of temperature indices. Fig. 2 shows the Taylor diagrams for the 12 CMIP6 and 30 CMIP5 models and their MME against observations, combined with the TS histograms of their MME. The majority of

models (both CMIP5 and CMIP6) have PCC greater than 0.85 for  $T_{av}$ ,  $TX_x$ , and  $TN_n$ . This indicates that the coupled models have a good performance in simulating the spatial distribution of these indices, with RSD between 0.75 and 1.25 and RMSE less than 0.5. The result from MME is furthermore much better than individual models. The PCC and TS for each index of both CMIP5-MME and CMIP6-MME are above 0.94, the RSD close to 1, and the RMSE close to 0.25. Compared with CMIP5, CMIP6 models also show some improvements for  $TX_x$ , with larger PCC and TS. If we examine  $TN_{10p}$  and  $TX_{90p}$ , presented in the lower panels in Fig. 2, all models (both CMIP5 and CMIP6) show relatively poor performances, with PCC below 0.5. Even the MME is not good enough, with PCC below 0.6 and TS below 0.3. This feature has also been revealed by previous studies (e.g. Chen and Sun 2015).

We also compared the subset CMIP6-six models with their CMIP5-six predecessors by using similar Taylor diagrams (results not shown), and the performance of models for temperature indices is generally similar between CMIP6 and CMIP5. We may even notice that, for  $TN_n$ , CMIP6 models presents some degradation, mainly due to the poor performance of IPSL-CM6A-LR. The capability of their ensemble mean (MME) to reproduce the climatological temperature indices is also close to each other. For the subset of six affiliation-identified simulations, we also performed a further analysis as what presented in Sillmann et al. (2013), providing a compact graphical overview of models' performance relative to each other. The basic calculation is the root mean square error (RMSE) relative to the observed climatology. RMSE is then subtracted and normalized by the median value among the models to compare and for each parameter. Such a processing allows half of the models with positive values and the second half with negative values. It is clear that models with negative values have a better performance compared to those with positive values. Results for temperature indices are shown in Fig. 3. Colors are used to show magnitude of the normalized relative RMSE. Warm colors indicate models in the bad half, and cold colors indicate models in the good half. From a visual inspection, we can easily conclude that the six affiliated CMIP6 simulations are not

distinguishable from their counterpart from CMIP5.

Generally speaking, both CMIP6-MME and CMIP5-MME exhibit good capabilities to simulate the spatial distribution of  $T_{av}$ ,  $TX_x$  and  $TN_n$ , but have difficulties to reproduce  $TN_{10p}$  and  $TX_{90p}$ . CMIP6-MME still has cold biases over the Tibetan Plateau for  $T_{av}$  and  $TN_n$ , which may be related to the complex physical characteristics of the underlying surface of the plateau. Climate models, with a relatively coarse resolution, cannot depict the steep terrain of the plateau and the complex characteristics of the underlying surface (Gao et al., 2008; IPCC, 2007, 2013). However, compared with CMIP5-MME, the warm biases for  $TX_x$  in some regions of northwest China have significantly reduced in CMIP6-MME. The capability of CMIP6-MME to simulate the spatial distribution of  $TN_{10p}$  and  $TX_{90p}$  is still relatively poor but shows some improvement in comparison with CMIP5-MME. Consider all temperature indices as a whole, we can conclude that the simulation of their spatial pattern is of similar performance in CMIP6 and CMIP5, and the inter-model spread is at a comparable level in CMIP5 and CMIP6.

### **3.1.2 Interannual variability**

The performance in simulating the temporal variation is also a very important factor to measure the capability of models. The IVS defined in section 2.3.2 is here used to quantify the similarity of interannual variability between simulated and observed indices. Fig. 4 gives the IVS histogram of models for five temperature indices in China. Both CMIP6 and CMIP5 models can realistically simulate the interannual variation of temperature indices with the mean IVS less than 1. The mean IVS of  $T_{av}$ ,  $TN_{10p}$  and  $TX_{90p}$  from CMIP6 models are 0.26, 0.33 and 0.25, respectively. The mean IVS of  $T_{av}$ ,  $TN_{10p}$  and  $TX_{90p}$  from CMIP5 models are 0.35, 0.28 and 0.31, respectively. CMIP6 models perform better than CMIP5 models for  $TX_x$ , the mean IVS from CMIP6 and CMIP5 models is 0.51 and 0.91, respectively. Consider all temperature indices, we can again conclude that the simulation of interannual variation of different temperature indices is of similar performance between CMIP6 and CMIP5.

## 3.2 Precipitation Indices

### 3.2.1 Climatology

The simulation of precipitation is more challenging for models as it depends closely on the parameterization of complex physical processes and their implementation in models. Major challenges are linked to surface properties (topography, coastline, vegetation) that lead to much greater spatial heterogeneity at regional scales (IPCC, 2007, 2013). Fig. 5 shows the box-and-whisker plots and the spatial distribution of biases between simulation and observation for different precipitation indices for the common period from 1961 to 2005. The precipitation indices simulated by CMIP6-MME are generally overestimated, except for consecutive dry days (CDD). The areal-mean relative bias of CMIP6-MME for total precipitation (Prcptot), heavy precipitation (R95p) and precipitation intensity (Sdii) in whole China is 79%/85%/21%, respectively. The largest wet bias is located in the Himalayan mountains and the eastern part of the Tibetan Plateau, where local bias is more than 600 mm for Prcptot, 150 mm for R95p and 3 mm/day for Sdii, respectively (Fig. 5b, e and h). The subset ensemble mean from six models with identified affiliation, CMIP5-six and CMIP6-six, perform in a very similar way as in Fig. 5.

CMIP6-MME shows significant improvements in the simulation of precipitation indices compared with CMIP5-MME, except for CDD, which has similar performance in the two ensembles. For Prcptot (Fig. 5a), the areal-mean bias has been reduced from 127% in CMIP5-MME to 79% in CMIP6-MME. Similarly, the median relative bias has also been reduced from 51% (CMIP5-MME) to 30% (CMIP6-MME). Although the high skewness of precipitation distribution leads to different behaviors, both the mean and median show a clear improvement from CMIP5-MME to CMIP6-MME. CMIP5-MME shows obvious dry biases in the south of the Yangtze River for Prcptot, R95p and Sdii (negative bias for extremely heavy rain days, R20mm), while dry (negative) biases are significantly decreased in CMIP6-MME (Fig. 5b, e, h and n dotted areas). This feature is also observed with the subset ensemble results. The biases of Prcptot (exceed -400 mm) and R95p (exceed -100 mm)

in CMIP5-MME over south China are reduced to less than -200 mm and -50 mm in CMIP6-MME, respectively. The Tibet Plateau also is within the dotted area in Fig. 5 (middle column), which indicates the large wet bias here is reduced in CMIP6-MME.

CDD does not show obvious improvement in CMIP6-MME, with overestimation in Xinjiang and in southern China, but underestimation in other regions, especially in the Kunlun Mountains (north part of the Tibetan Plateau), where the bias is over 60 days (Fig. 5k). The areal-mean bias of CDD over whole China is -15.7 days. The spatial distribution from CMIP5-MME is similar, and the areal-mean bias over China is -13.1 days.

Unlike temperature indices showing inconclusive variation between CMIP5 and CMIP6 in terms of inter-model spreading, precipitation indices display a clearer trend of reduced inter-model spreading from CMIP5 to CMIP6. This is true for the total ensembles (Fig. 5) and the subset ensembles (results not shown).

Fig. 6 shows Taylor diagrams and Taylor scores which are a concise performance representation of all precipitation indices. The majority of models (both CMIP5 and CMIP6) have PCC between 0.6 and 0.9 for Prcptot, R95p and Sdii. This indicates that the coupled models have certain capability in simulating the spatial distribution of these indices. Especially for Sdii, the PCC is all larger than 0.7 and the RMSE is all smaller than 0.75. But they are generally less skillful in simulating CDD and R20mm. Especially for CDD, the PCC of almost all models (both CMIP5 and CMIP6) is smaller than 0.75, and the RSD is further away from 1. The model spread of most precipitation indices is much larger than that of temperature indices (Tav, TXx and TNn) in both CMIP5 and CMIP6. MME is generally better than individual models, with larger PCC and smaller RMSE.

The performance of CMIP6 models is clearly improved compared with that of CMIP5 models, especially for R95P and Sdii. The PCC for R95p/Sdii has increased from 0.80/0.87 (CMIP5-MME) to 0.86/0.92 (CMIP6-MME), respectively. And the TS for R95p/Sdii also increases from 0.79/0.83 to 0.86/0.89, respectively. The RSD from

CMIP6-MME is also closer to 1, and the RMSE closer to 0, which indicates that the performance in simulating the spatial pattern of R95p and Sdii is more significantly improved than for other indices from CMIP6-MME. CDD also presents certain improvements from CMIP5 to CMIP6, with TS increasing from 0.61 (CMIP5-MME) to 0.69 (CMIP6-MME). The subset ensemble CMIP6-six also shows significant improvements compared to its counterpart CMIP5-six for most precipitation indices, including the ensemble average. A few models (e.g., BCC-CSM2-MR, GFDL-CM4 and GFDL-ESM4) even systematically outperform their CMIP5 predecessors for all precipitation indices. The capability of climate models to reproduce climatological precipitation indices over China rises from CMIP5 to CMIP6. This improvement is also visible in Fig. 7, which presents the portrait diagram of multiple precipitation indices from the subset ensembles, with dominant blue colors for CMIP6-six and dominant orange colors for CMIP5-six.

### 3.2.2 Interannual variability

Fig. 8. shows the mean IVS histogram of models for the five precipitation indices averaged over China. For all indices, the mean IVS from CMIP6 models is lower than that of CMIP5 models. Taking Sdii as an example, the mean score from CMIP6 models is 0.91, and the mean score of CMIP5 models is 1.18. The IVS of other indices are larger, but CMIP6 models have smaller mean values than CMIP5 models.

Generally speaking, both CMIP6-MME and CMIP5-MME have certain capabilities of simulating precipitation indices, but not as good as simulating temperature indices (Tav, TXx and TNn). The models are more skillful to reproduce temperature than precipitation, which is also consistent with previous studies (IPCC, 2013; Kusunoki and Arakawa, 2015). For different precipitation indices, the performance of the models is different, with the best for Sdii and the worst for CDD. Consider all precipitation indices, we can conclude that CMIP6 models are superior to CMIP5 models in simulating both the spatial distribution and interannual variability, and have smaller inter-model spreads than CMIP5 models. In particular, the dry bias

of mean and extreme precipitation in southern China is largely reduced in CMIP6-MME. However, CMIP6-MME still has wet biases for Prcptot, R95p and Sdii in the eastern Tibetan Plateau, but these wet biases are smaller than those of CMIP5-MME. This improvement may be partly attributed to the higher model resolution and the better description of physical and chemical processes in CMIP6 models (Eyring et al., 2016). CMIP6 models generally improve the convective parameterization schemes and the cloud physics process (Cao et al., 2018; Wu et al., 2019) and append the indirect effects posed by aerosols onto the formation of clouds and precipitation (Wu et al., 2019; Voltaire et al., 2019).

To further investigate possible causes for the better performance of CMIP6-MME to reproduce precipitation in south China, the regional atmospheric circulation patterns from the two MMEs are compared with the NCEP reanalysis. Considering the fact that June-August (JJA) is the main season for precipitation occurred in China, Fig. 9 shows the difference of meridional circulation and specific humidity (shaded) zonally averaged within 110°E-120°E in summer. As shown in Fig. 9 (middle column), compared with NCEP, CMIP5-MME presents strong descent around 30°N. CMIP5-MME also shows less water vapor over south China, so there is a large dry bias over south China (Fig. 5b and e). For CMIP6-MME (Fig. 9 left column), however, the zone of strong descent is northward shifted, compared to CMIP5-MME. CMIP6-MME also shows a little more water vapor and a weak ascent around 22°N, compared to NCEP. So, the dry bias is reduced in south China in CMIP6-MME (Fig. 5a and d). Furthermore, compared with CMIP5-MME, CMIP6-MME presents stronger ascent between 22°N to 30°N. CMIP6-MME also shows more water vapor over south China (Fig. 9 right column). Hence, the stronger ascent accompanied with more moisture over south China in CMIP6-MME contributes to the decrease of dry biases over there.

#### **4. Conclusions and discussion**

In this study, we quantitatively evaluated the performance of 12 CMIP6 and 30

CMIP5 models for simulating climate extremes in China, through an inter-comparison against a gridded daily observation dataset. The assessment was done in terms of spatial patterns and temporal variability for the period 1961-2005. We used some commonly used skill-score methods, such as the Taylor diagram and TS representing the spatial pattern and IVS representing the interannual variability. We generally privileged the analysis on multi-model ensemble (MME) mean, except for the temporal variability that needs examination in each individual member. It is noticed that, besides the general ensemble, we used a subset of six simulations from CMIP5 and CMIP6 belonging to same institution. Results from the subset ensemble corroborate those from the general ensemble. The main findings of our study can be summarized as follows.

Both CMIP6 and CMIP5 models show good performance in simulating the climatological pattern and the interannual variation for temperature indices over China. The spreads among models for most temperature indices are at a comparable level in CMIP5 and CMIP6. CMIP6-MME and CMIP5-MME have very good capabilities of reproducing  $T_{av}$ ,  $TX_x$  and  $TN_n$ , with PCC and TS larger than 0.94. But they have difficulties to well reproduce  $TN_{10p}$  and  $TX_{90p}$ . CMIP6-MME generally shows slight improvements in comparison with CMIP5-MME. The large warm biases of CMIP5-MME for  $TX_x$  in parts of northwest China are significantly reduced in CMIP6-MME. However, the large cold bias of CMIP5-MME over the Tibetan Plateau still exists in CMIP6-MME for  $T_{av}$  (exceeding  $-4^{\circ}C$ ) and  $TN_n$  (exceeding  $-8^{\circ}C$ ).

The performance of the CMIP6 and CMIP5 models in simulating precipitation indices is not as good as that in simulating temperature indices ( $T_{av}$ ,  $TX_x$  and  $TN_n$ ). For different precipitation extremes, the capability of the models is also different, with the best for  $S_{dii}$ , the worst for CDD. CMIP6 models have a smaller inter-model spread for most precipitation indices than CMIP5 models do, which implies that a slightly larger consensus seems in CMIP6 historical simulations. We hope that there will be also smaller uncertainties when we deal with projection of future climate within CMIP6. Compared with CMIP5-MME, CMIP6-MME shows significant



improvements in simulating both the climatological pattern and the interannual variation for precipitation indices, except for CDD which remains at the same level. The areal-mean bias for Prcptot is reduced from 127% (CMIP5-MME) to 79% (CMIP6-MME). The dry biases for mean and extreme precipitation in southern China are also largely reduced in CMIP6-MME. However, CMIP6-MME still has wet biases for Prcptot, R95p and Sdii in the eastern Tibetan Plateau, although these wet biases are smaller than those of CMIP5-MME.

The dry biases for precipitation indices (Prcptot, R95p and Sdii) over south China are tightly related to the strong descent and weak water vapor content over there. CMIP6-MME, with ascent and moist air over south China, is able to produce more precipitation, and to reduce the dry biases over there. CMIP6 models generally have higher horizontal resolution and improved physical parameterizations (convection, radiation, cloud, land surface, etc.). All together certainly contribute to the revealed improvement, although we are not able now to determine the precise cause. In eastern China, convective precipitation is the main type of precipitation, so deficiencies in convection parameterization are likely to cause the systematic errors in precipitation simulation, especially for extreme precipitation (Li et al., 2012; Rosa and Collins, 2013; Mehran et al., 2014; Jiang et al., 2015). CMIP6 models generally improved the convection parameterization scheme and cloud physics (Cao et al., 2018; Wu et al., 2019), which can also partly explain the significant improvement of precipitation simulation in southern China.

It should be pointed out that CMIP6-MME still has large biases for temperature indices over the Tibetan Plateau and in Northwest China. The precipitation indices in Western China are also not well simulated, especially there are large wet biases in the eastern Tibetan Plateau. Topographic forcing has an important influence on the simulation of climate (Zhou et al., 2009; Song et al., 2013). It would be relatively difficult for models to reproduce the effects of the complex topography in Western China (Chen et al., 2012).

Due to the limited number of temporarily available CMIP6 models, the

evaluation of more CMIP6 models still needs to be carried out in the future. However, based on the results of the 12 CMIP6 models in this paper, the improvement of precipitation simulation compared with CMIP5 models is of great significance to provide more comprehensive climate information to end-users or policy-makers. The climate projection results of the Scenario Model Intercomparison Project (Scenario MIP) for CMIP6 (O'Neill et al., 2016) are also gradually released, which will allow us to assess future climate change projections from CMIP6 models. Based on this assessment, we would recommend, with more confidence, to use CMIP6 results for future climate projection in China.

## **Acknowledgments**

We wish to thank the three anonymous reviewers whose valuable comments and suggestions helped us to improve our manuscript. We would like to acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP. We thank the climate modeling groups for producing and making their model outputs available. This research is supported by the National Key Research and Development Program of China (2017YFA0603804 and 2018YFC1507704), the Natural Science Foundation of China (Grant 41805048).

## **References**

- Akinsanola, A., and W. Zhou, 2019: Projections of West African summer monsoon rainfall extremes from two CORDEX models. *Climate Dynamics*, **52**, 2017-2028.
- Cao, J., and Coauthors, 2018: The NUIST Earth System Model (NESM) version 3: description and preliminary evaluation. *Geoscientific Model Development*, **11**, 2975-2993.
- Chen, H., and J. Sun, 2014: Robustness of precipitation projections in China: comparison between CMIP5 and CMIP3 models. *Atmospheric and Oceanic Science Letters*, **7**, 67-73.
- Chen, H., and J. Sun, 2015: Assessing model performance of climate extremes in China:

- an intercomparison between CMIP5 and CMIP3. *Climatic Change*, **129**, 197-211.
- Chen, L., and O. W. Frauenfeld, 2014: A comprehensive evaluation of precipitation simulations over China based on CMIP5 multimodel ensemble projections. *Journal of Geophysical Research: Atmospheres*, **119**, 5767-5786.
- Chen, L., Z. G. Ma, and X. G. Fan, 2012: A comparative study of two land surface schemes in WRF model over Eastern China. *Journal of Tropical Meteorology*, **18**, 445.
- Chen, W., Z. Jiang, and L. Li, 2011: Probabilistic projections of climate change over China under the SRES A1B scenario using 28 AOGCMs. *Journal of Climate*, **24**, 4741-4756.
- Chen, X., Y. Xu, C. Xu, and Y. Yao, 2014: Assessment of precipitation simulations in China by CMIP5 multi-models. *Adv Clim Change Res*, **10**, 217-225(in Chinese).
- Committee of Chinese National Assessment Report on Climate Change, 2015: The third China's National Assessment Report on Climate Change, China Science Press (in Chinese).
- Dong, S., Y. Xu, B. Zhou, and Y. Shi, 2015: Assessment of indices of temperature extremes simulated by multiple CMIP5 models over China. *Advances in Atmospheric Sciences*, **32**, 1077-1091.
- Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, **9**, 1937-1958.
- Field, C. B., V. Barros, T. F. Stocker, and Q. Dahe, 2012: Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change. ed. Cambridge University Press.
- Flato, G., and Coauthors, 2014: Evaluation of climate models. 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, 741-866.

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

Gao, X., Y. Shi, R. Song, F. Giorgi, Y. Wang, and D. Zhang, 2008: Reduction of future monsoon precipitation over China: Comparison between a high resolution RCM simulation and the driving GCM. *Meteorology and Atmospheric Physics*, **100**, 73-86.

Gao, Y., H. Wang, and D. Jiang, 2015: An intercomparison of CMIP5 and CMIP3 models for interannual variability of summer precipitation in Pan-Asian monsoon region. *International Journal of Climatology*, **35**, 3770-3780.

Guo, Y., W. Dong, F. Ren, Z. Zhao, and J. Huang, 2013: Assessment of CMIP5 simulations for China annual average surface temperature and its comparison with CMIP3 simulations. *Adv Clim Change Res*, **9**, 181-186(in Chinese).

Gusain, A., S. Ghosh, and S. Karmakar, 2020: Added value of CMIP6 over CMIP5 models in simulating Indian summer monsoon rainfall. *Atmospheric Research*, **232**.

Huang, D. Q., J. Zhu, Y. C. Zhang, and A. N. Huang, 2013: Uncertainties on the simulated summer precipitation over Eastern China from the CMIP5 models. *Journal of Geophysical Research: Atmospheres*, **118**, 9035-9047.

IPCC 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, United Kingdom and New York, NY, USA.

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 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.

Jiang, D., Y. Zhang, and J. Sun, 2009: Ensemble projection of 1–3 C warming in China. *Chinese Science Bulletin*, **54**, 3326-3334.

Jiang, D., Z. Tian, and X. Lang, 2016: Reliability of climate models for China through the IPCC Third to Fifth Assessment Reports. *International Journal of Climatology*, **36**, 1114–1133.

Jiang, Z., J. Song, L. Li, W. Chen, Z. Wang, and J. Wang, 2012: Extreme climate events in China: IPCC-AR4 model evaluation and projection. *Climatic Change*, **110**, 385-401.

Jiang, Z., W. Li, J. Xu, and L. Li, 2015: Extreme Precipitation Indices over China in CMIP5 Models. Part I: Model Evaluation. *Journal of Climate*, **28**, 8603-8619.

Kawai, H., S. Yukimoto, T. Koshiro, N. Oshima, T. Tanaka, H. Yoshimura, and R. Nagasawa, 2019: Significant improvement of cloud representation in the global climate model MRI-ESM2. *Geoscientific Model Development*, **12**, 2875-2897.

Koutroulis, A., M. Grillakis, I. Tsanis, and L. Papadimitriou, 2016: Evaluation of precipitation and temperature simulation performance of the CMIP3 and CMIP5 historical experiments. *Climate Dynamics*, **47**, 1881-1898.

Kusunoki, S., and O. Arakawa, 2015: Are CMIP5 models better than CMIP3 models in simulating precipitation over East Asia? *Journal of Climate*, **28**, 5601-5621.

Li, F., D. Rosa, W. D. Collins, and M. F. Wehner, 2012: “Super-parameterization”: A better way to simulate regional extreme precipitation? *Journal of Advances in*

*Modeling Earth Systems*, **4**.

- Meehl, G. A., and Coauthors, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bulletin of the American Meteorological Society*, **88**, 1383-1394.
- Mehran, A., A. AghaKouchak, and T. J. Phillips, 2014: Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations. *Journal of Geophysical Research: Atmospheres*, **119**, 1695-1707.
- Miao, C., H. Ashouri, K. L. Hsu, S. Sorooshian, and Q. Duan, 2015: Evaluation of the PERSIANN-CDR daily rainfall estimates in capturing the behavior of extreme precipitation events over China. *Journal of Hydrometeorology*, **16**, 1387-1396.
- O'Neill, B. C., and Coauthors, 2016: The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, **9**, 3461-3482.
- Ou, T., D. Chen, H. W. Linderholm, and J. H. Jeong, 2013: Evaluation of global climate models in simulating extreme precipitation in China. *Tellus A: Dynamic Meteorology and Oceanography*, **65**, 1393-1399.
- Park, S., J. Shin, S. Kim, E. Oh, and Y. Kim, 2019: Global Climate Simulated by the Seoul National University Atmosphere Model Version 0 with a Unified Convection Scheme (SAM0-UNICON). *Journal of Climate*, **32**, 2917-2949.
- Rosa, D., and W. Collins, 2013: A case study of subdaily simulated and observed continental convective precipitation: CMIP5 and multiscale global climate models comparison. *Geophysical Research Letters*, **40**, 5999-6003.
- Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *Journal of Geophysical Research: Atmospheres*, **118**, 1716-1733.
- Sillmann, J., V. V. Kharin, F. W. Zwiers, X. Zhang, D. Bronaugh, and M. G. Donat, 2014: Evaluating model-simulated variability in temperature extremes using

- modified percentile indices. *International journal of climatology*, **34**, 3304-3311.
- Song, Y., F. Qiao, Z. Song, and C. Jiang, 2013: Water vapor transport and cross-equatorial flow over the Asian-Australia monsoon region simulated by CMIP5 climate models. *Advances in Atmospheric Sciences*, **30**, 726-738.
- Sperber, K. R., and Coauthors, 2013: The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Climate Dynamics*, **41**, 2711-2744.
- 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. *Journal of Climate*, **26**, 3187-3208.
- Sun, Q., C. Miao, and Q. Duan, 2015: Comparative analysis of CMIP3 and CMIP5 global climate models for simulating the daily mean, maximum, and minimum temperatures and daily precipitation over China. *Journal of Geophysical Research: Atmospheres*, **120**, 4806-4824.
- Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research: Atmospheres*, **106**, 7183-7192.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, **93**, 485-498.
- Voldoire, A., and Coauthors, 2019: Evaluation of CMIP6 DECK Experiments With CNRM-CM6-1. *Journal of Advances in Modeling Earth Systems*, **11**, 2177-2213.
- Wang, B., L. Zheng, D. Liu, F. Ji, A. Clark, and Q. Yu, 2018: Using multi-model ensembles of CMIP5 global climate models to reproduce observed monthly rainfall and temperature with machine learning methods in Australia. *International Journal of Climatology*, **38**, 4891-4902.
- Wu, J., X. Gao, 2013: A gridded daily observation dataset over China region and comparison with the other datasets. *Chin. J. Geophys.*, **56**: 1102-1111(in Chinese).
- Wu, T., and Coauthors, 2019: The Beijing Climate Center Climate System Model

- (BCC-CSM): the main progress from CMIP5 to CMIP6. *Geoscientific Model Development*, **12**, 1573-1600.
- Xu, Y., X. Gao, and F. Giorgi, 2010: Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections. *Climate Research*, **41**, 61-81.
- Xu, Y., X. Gao, F. Giorgi, B. Zhou, Y. Shi, J. Wu, and Y. Zhang, 2018: Projected changes in temperature and precipitation extremes over China as measured by 50-yr return values and periods based on a CMIP5 ensemble. *Advances in Atmospheric Sciences*, **35**, 376-388.
- Zhang, X., and Coauthors, 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wiley Interdisciplinary Reviews: Climate Change*, **2**, 851-870.
- 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. *Journal of Climate*, **27**, 6591-6611.
- Zhou, L., A. Dai, Y. Dai, R. S. Vose, C. Z. Zou, Y. Tian, and H. Chen, 2009: Spatial dependence of diurnal temperature range trends on precipitation from 1950 to 2004. *Climate Dynamics*, **32**, 429-440.



Table 1. Model name, modeling center and country, and atmospheric resolution of 30 CMIP5 global climate models.

Model number	Model name	Modeling center and country	Atmospheric resolution(lat × lon)
1	ACCESS1.0	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology(Australia)	1.24°×1.875°
2	BCC-CSM1.1	Beijing Climate Center, China Meteorological Administration	2.8°×2.8°
3	BCC-CSM1.1-m	(China)	1.125°×1.125°
4	BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University (China)	2.8°×2.8°
5	CanCM4	Canadian Centre for Climate Modelling and Analysis (Canada)	2.8°×2.8°
6	CanESM2		2.8°×2.8°
7	CCSM4	National Center for Atmospheric Research (USA)	0.94°×1.25°
8	CMCC-CM	Centro Euro-Mediterraneo per I Cambiamenti Climatici(Italy)	0.75°×0.75°
9	CMCC-CMS		1.875°×1.875°
10	CSIRO-Mk3-6-0	Commonwealth Scientific and Industrial Research Organization in collaboration with Queensland Climate Change Centre of Excellence (Australia)	1.875°×1.875°
11	CNRM-CM5	Centre National de Recherches Météorologiques–Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique (France)	1.4°×1.4°
12	FGOALS-g2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and Center for Earth System Science, Tsinghua University(China)	3°×2.8°

---

13	FGOALS-s2	LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences(China)	1.67°×2.8°
14	GFDL-CM3		2.0°×2.5°
15	GFDL-ESM2G	NOAA Geophysical Fluid Dynamics Laboratory (USA)	2.0°×2.5°
16	GFDL-ESM2M		2.0°×2.5°
17	HadCM3		2.5°×3.75°
18	HadGEM2-CC	Met Office Hadley Centre(United Kingdom)	1.24°×1.875°
19	HadGEM2-ES		1.24°×1.875°
20	IPSL-CM5A-LR		1.875°×3.75°
21	IPSL-CM5A-MR	L'Institut Pierre-Simon Laplace(France)	1.26°×2.5°
22	MIROC4h		0.56°×0.56°
23	MIROC5	National Institute for Environmental Studies,	1.4°×1.4°
24	MIROC-ESM	The University of Tokyo (Japan)	2.8125°×2.8125°
25	MIROC-ESM-CHEM		2.8125°×2.8125°
26	MPI-ESM-LR		1.875°×1.875°
27	MPI-ESM-MR	Max Planck Institute for Meteorology (Germany)	1.875°×1.875°
28	MPI-ESM-P		1.875°×1.875°
29	MRI-CGCM3	Meteorological Research Institute (Japan)	1.125°×1.125°
30	NorESM1-M	Norwegian Climate Centre (Norway)	1.8725°×2.5°

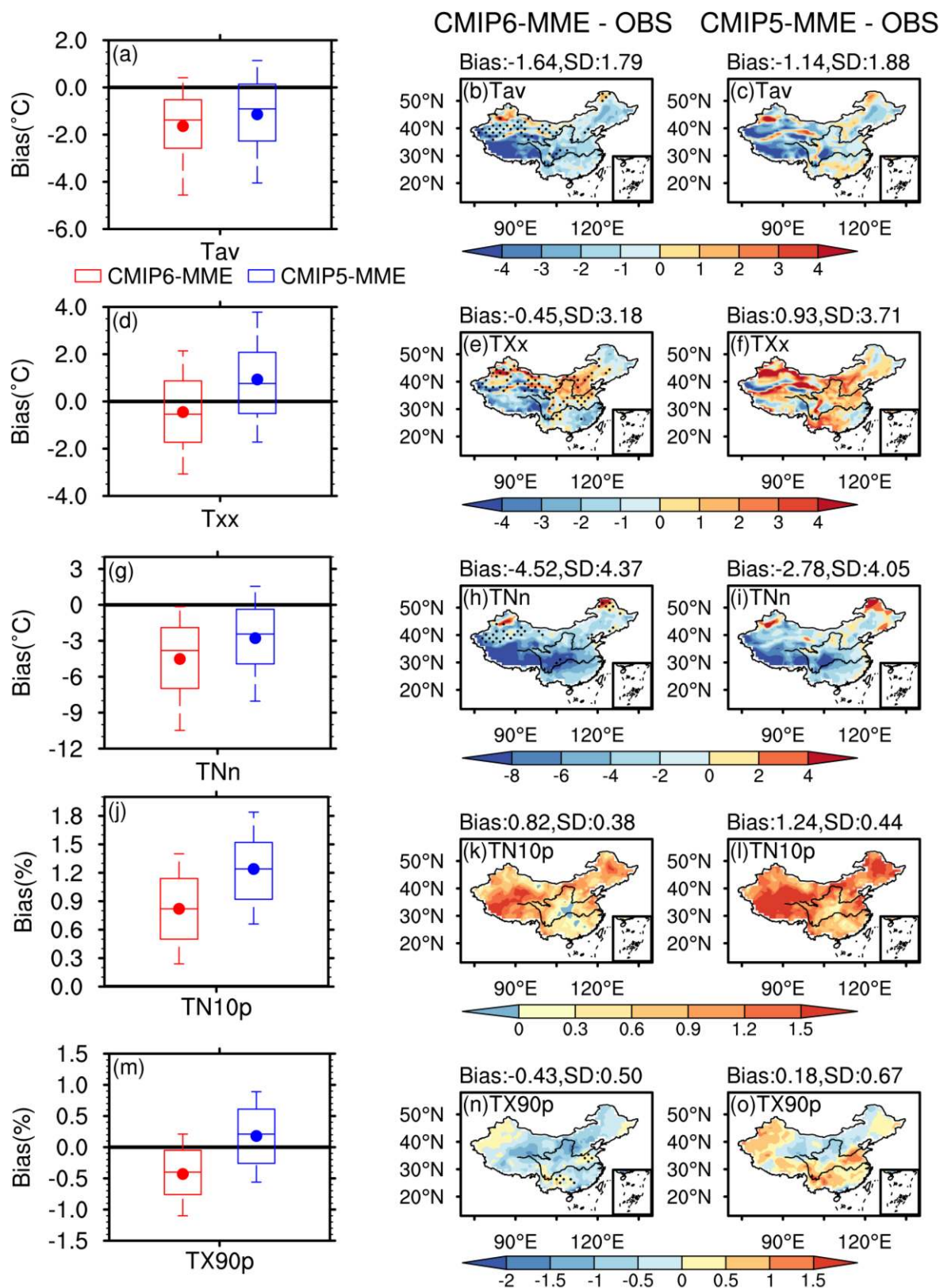
---

Table 2. Model name, modeling center and country, and atmospheric resolution of 12 CMIP6 global climate models.

Model number	Model name	Modeling center and country	Atmospheric resolution(lat × lon)
1	BCC-CSM2-MR	Beijing Climate Center, China Meteorological Administration	1.125°×1.125°
2	BCC-ESM1	(China)	2.8°×2.8°
3	CNRM-CM6-1	Centre National de Recherches Météorologiques–Centre Européen de	1.4°×1.4°
4	CNRM-ESM2-1	Recherche et de Formation Avancée en Calcul Scientifique (France)	1.4°×1.4°
5	EC-Earth3-Veg	EC-EARTH consortium	0.7°×0.7°
6	GFDL-CM4	NOAA Geophysical Fluid Dynamics Laboratory (USA)	1°×1.25°
7	GFDL-ESM4		1°×1.25°
8	IPSL-CM6A-LR	L’Institut Pierre-Simon Laplace(France)	1.26°×2.5°
9	MRI-ESM2-0	Meteorological Research Institute (Japan)	1.125°×1.125°
10	NESM3	Nanjing University of Information Science and Technology(China)	1.875°×1.875°
11	SAM0-UNICON	Seoul National University(Republic of Korea)	0.94°×1.25°
12	UKESM1-0-LL	Met Office Hadley Centre(UK)	1.25°×1.875°

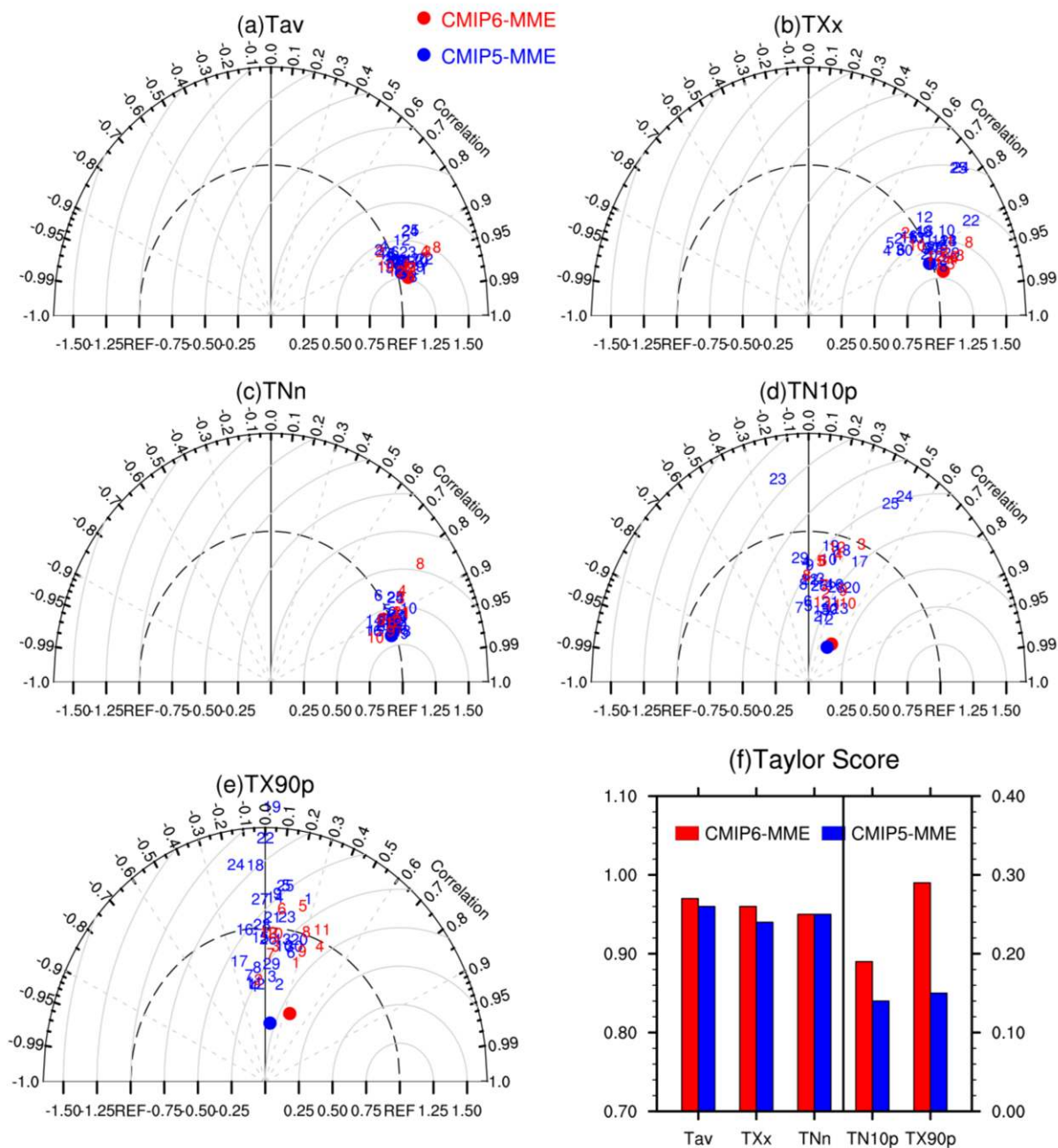
Table 3. Name, acronym, definition and unit of climate indices used in the study.

Name	Acronym	Definition	Units
Avg TM	Tav	Annual average value of daily temperature(TM)	°C
Max TX	TXx	Annual maximum value of daily maximum temperature(TX)	°C
Min TN	TNn	Annual minimum value of daily minimum temperature(TN)	°C
Cold nights	TN10p	Percentage of days when TN < 10th percentile	%
Warm days	TX90p	Percentage of days when TX > 90th percentile	%
Total precipitation	Preptot	Annual total precipitation in wet days (RR ≥ 1 mm)	mm
Heavy precipitation	R95p	Annual total precipitation from days > 95th percentile	mm
Precipitation intensity	Sdii	Total wet days precipitation divided by the number of wet days	mm/day
Consecutive dry days	CDD	Maximum number of consecutive days with RR <1 mm	days
Extremely heavy rain days	R20mm	Annual count of days with RR ≥ 20mm	days

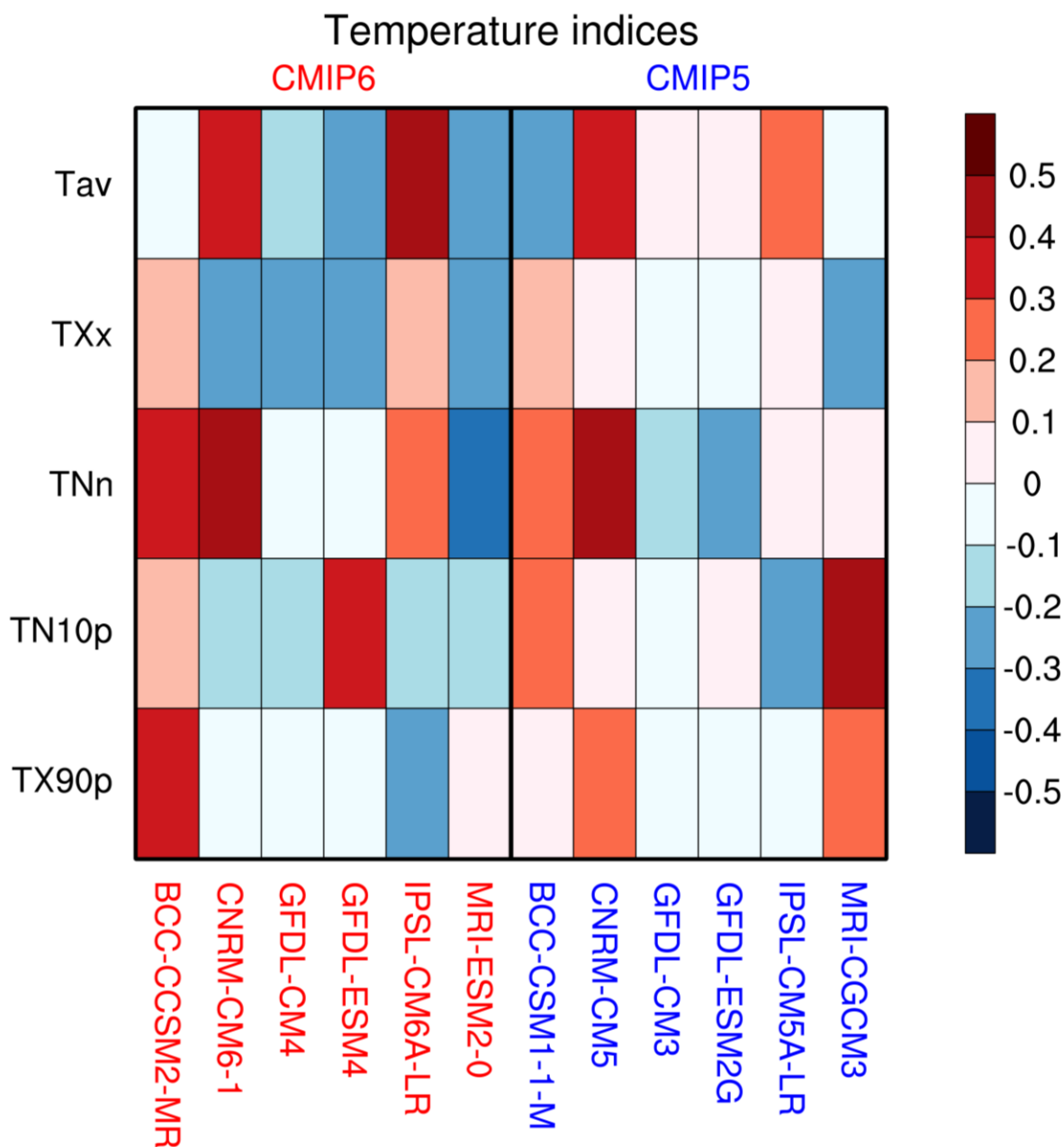


**Fig. 1.** Box-and-whisker plots (left column) and the spatial pattern of biases (simulation minus observation) in CMIP6-MME (middle column, red) and CMIP5-MME (right column, blue) of temperature indices for the historical period

1961-2005 (Units: °C, °C, °C, %, %). The areal-mean bias (Bias) over China and the inter-model standard deviation (SD) of the difference averaged over the country (middle and right column) are given on the top of each panel. From top to bottom are (a-c) Tav, (d-f) TXx, (g-i) TNn, (j-l) TN10p and (m-o) TX90p, respectively. The dotted areas in the middle panels represent regions where biases in CMIP6-MME are lower than in CMIP5-MME and the difference is statistically significant at the 5% level.

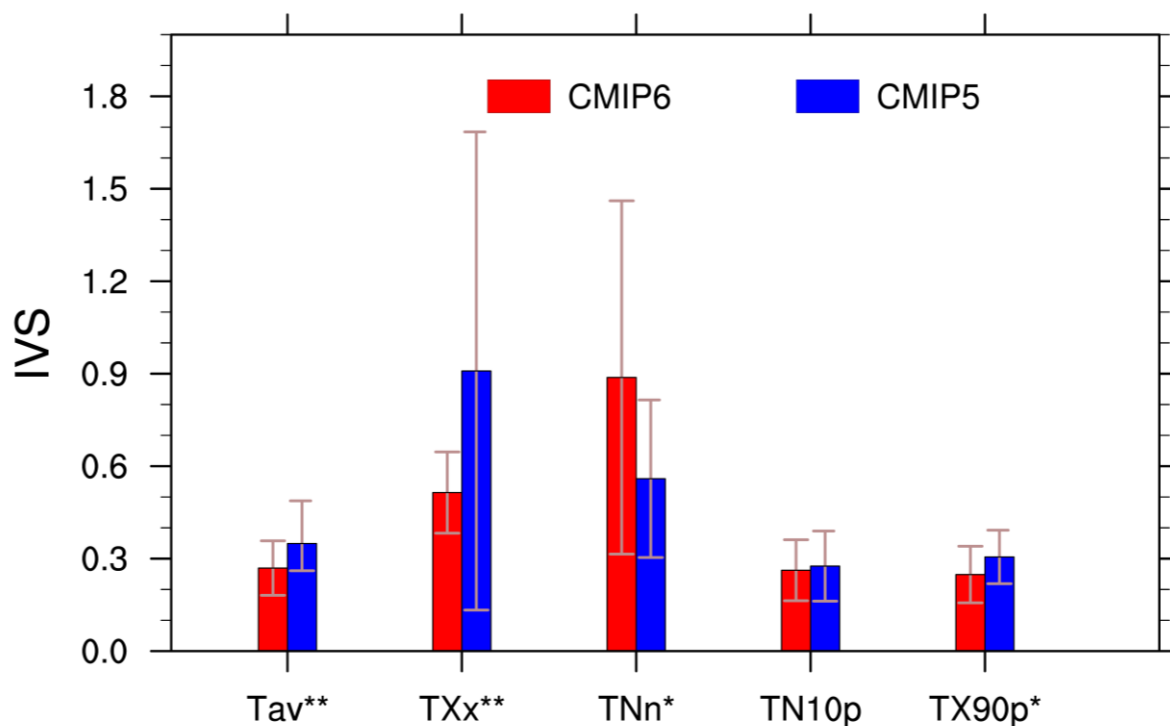


**Fig. 2.** Taylor diagrams (a-e) and skill scores in terms of TS (f) showing the performance of models (CMIP6 in red, CMIP5 in blue) in simulating climatological fields over China for five temperature indices. Angular axes show pattern correlation coefficients between simulated and observed fields; radial axes show the spatial centered root mean square error (normalized against the observed). Blue and red numbers indicate CMIP5 and CMIP6 models listed in Tables 1 and 2. The larger solid circles represent the MME. Scale of Tav, TXx, TNn is on the left y axis, and scale of TN10p and TX10p is on the right y axis.

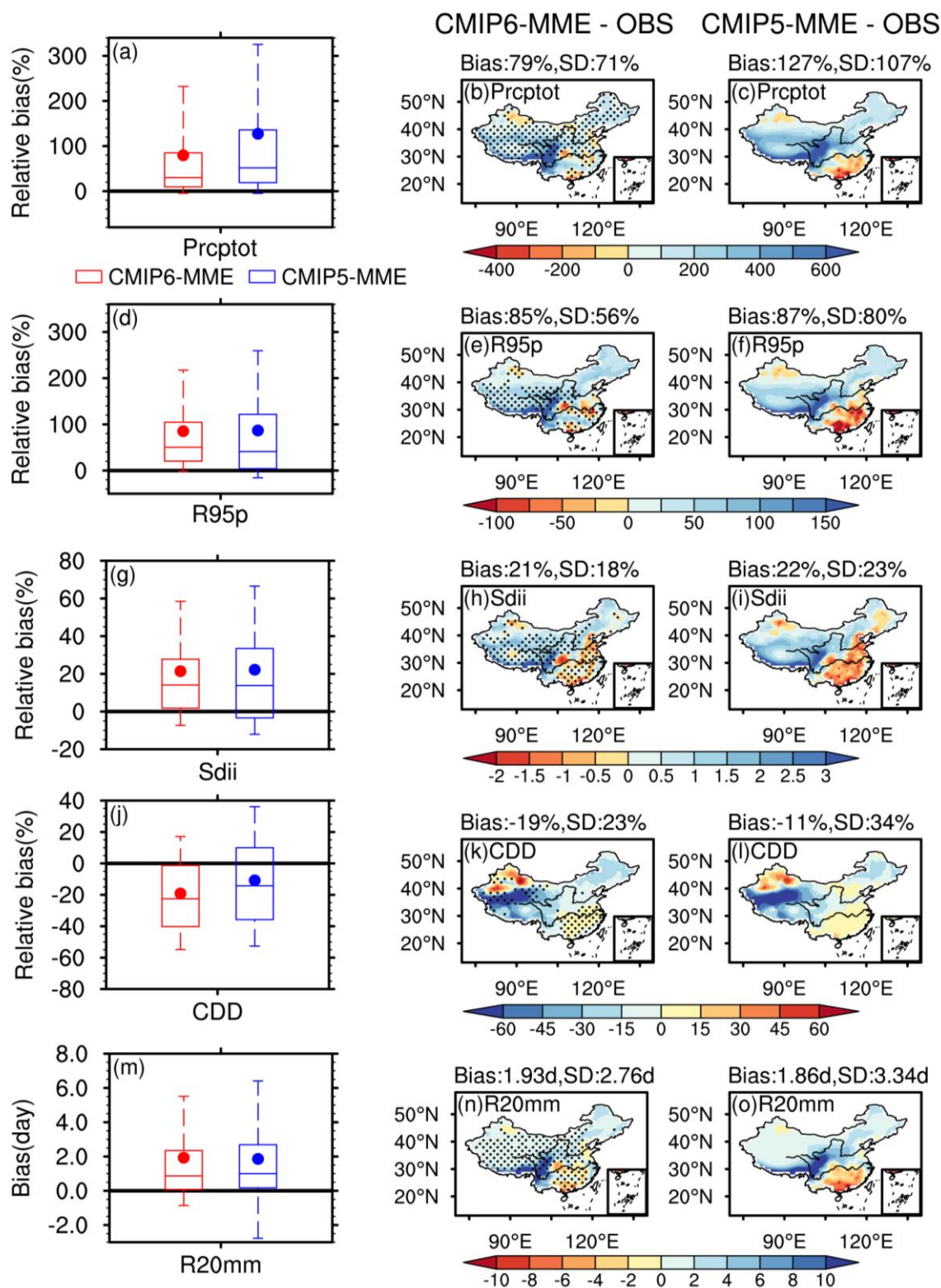


**Fig. 3.** The portrait diagram of relative spatially averaged root mean square errors (RMSEs) in the 1961-2005 climatologies of temperature indices simulated by the CMIP6 (red) and CMIP5 (blue) models from the same institution with respect to the observation.

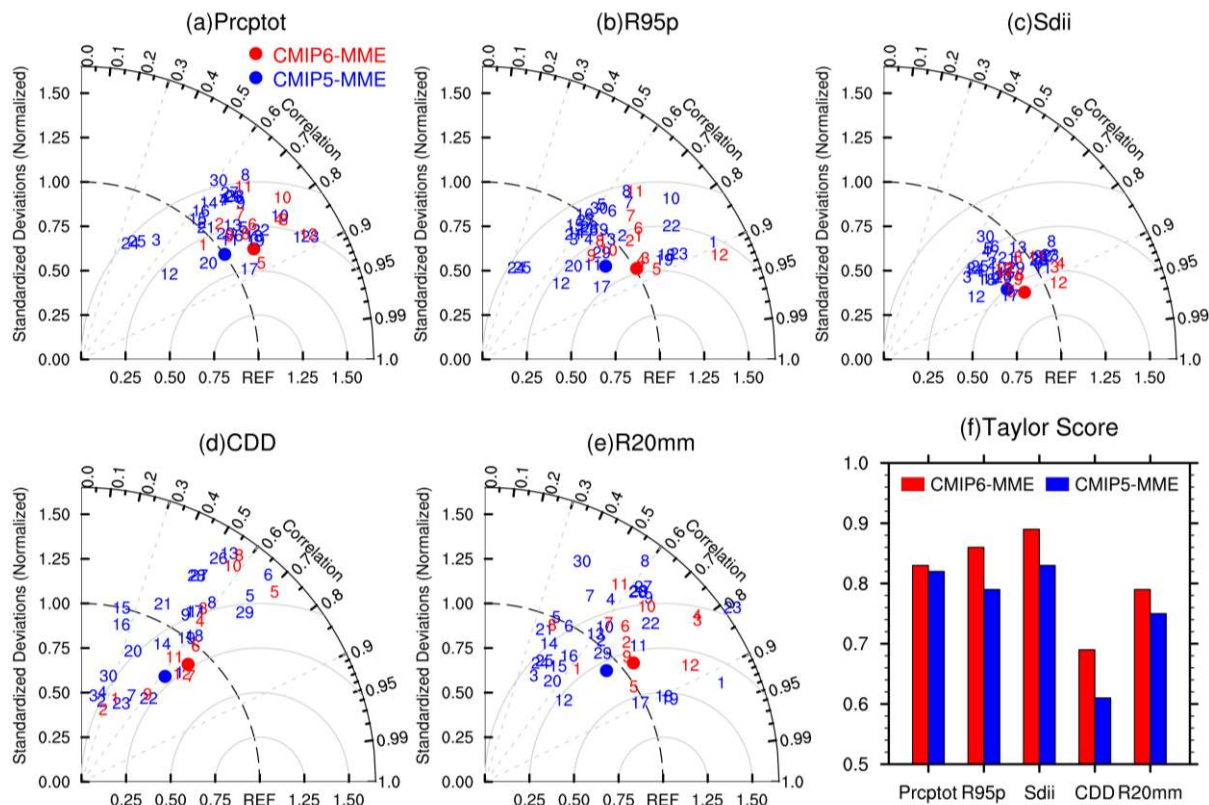




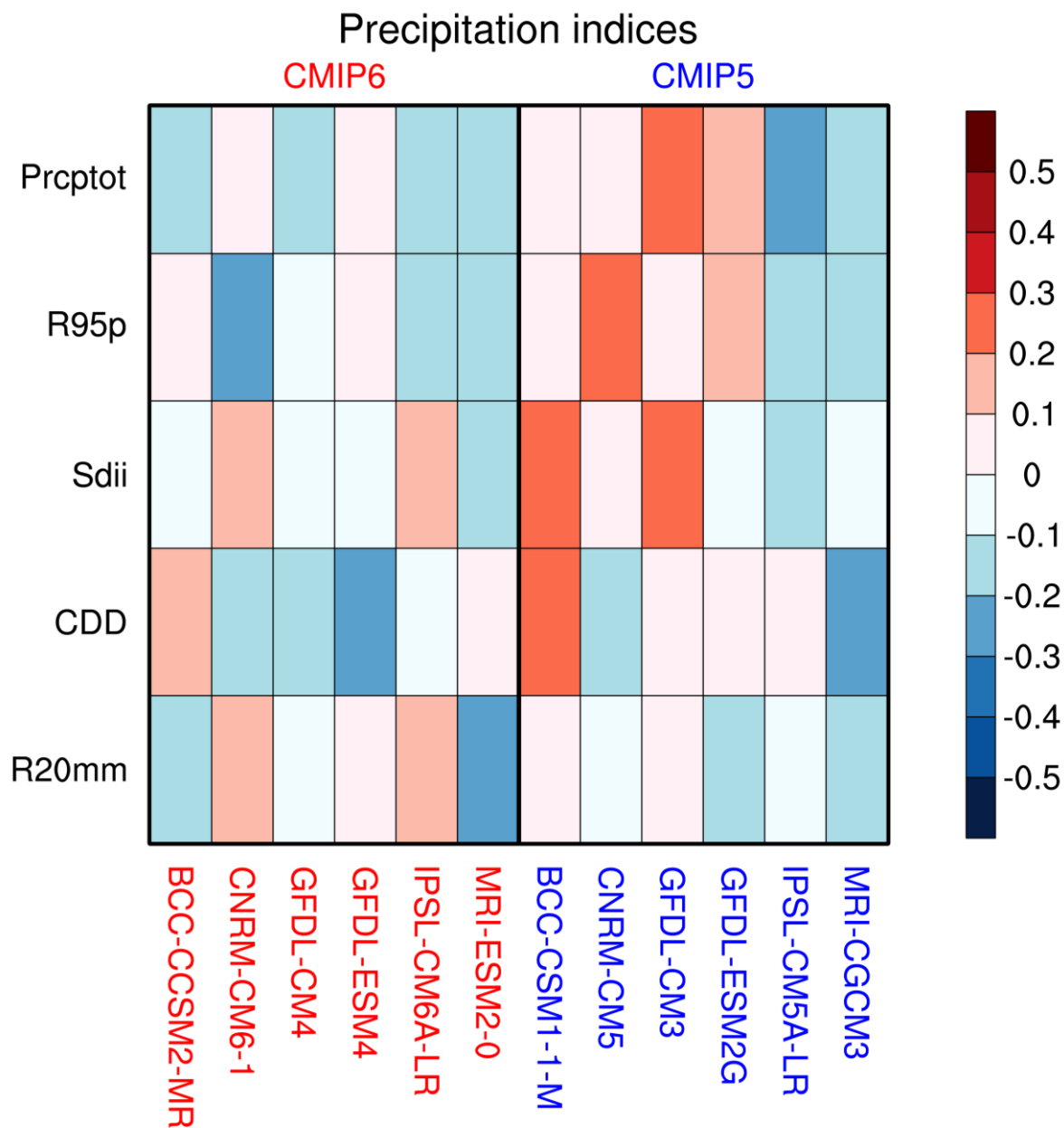
**Fig. 4.** Skill scores in terms of IVS for the five temperature indices in CMIP6 (red) and CMIP5 (blue) models over China. The filled bars show the ensemble mean, and the error bars represent ranges of one standard deviation ( $1\sigma$ ) among models. Asterisks (\*\*) indicates that the differences between CMIP6 and CMIP5 models are significant at the 95% confidence level based on the t-test, with an asterisk (\*) for 90%.



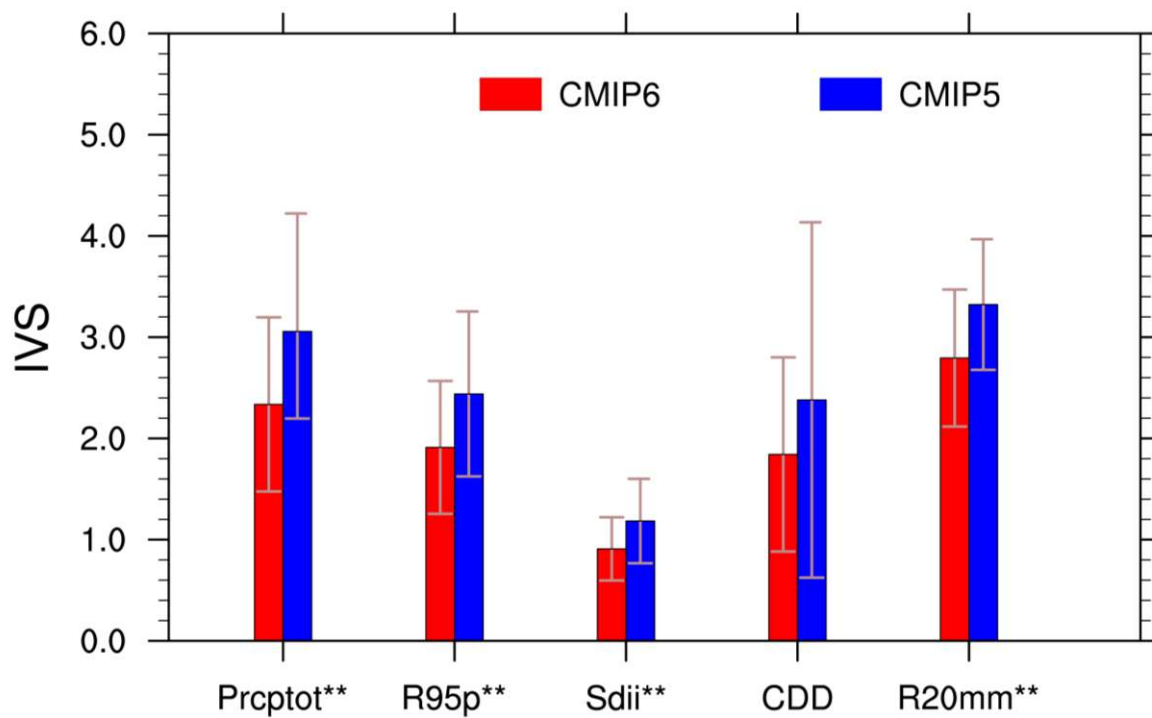
**Fig. 5.** Same as **Fig. 1.**, but for (a-c) Prcptot, (d-f) R95p, (g-i) Sdii, (j-l) CDD and (m-o) R20mm (Units: mm, mm, mm/day, days, days). The areal-mean percentage bias (Bias) over China and the inter-model standard deviation (SD) of the difference in percentage averaged over the country (middle and right column) are given on the top of each panel (but with bias and inter-model standard deviation of the difference for R20mm, Unit: days).



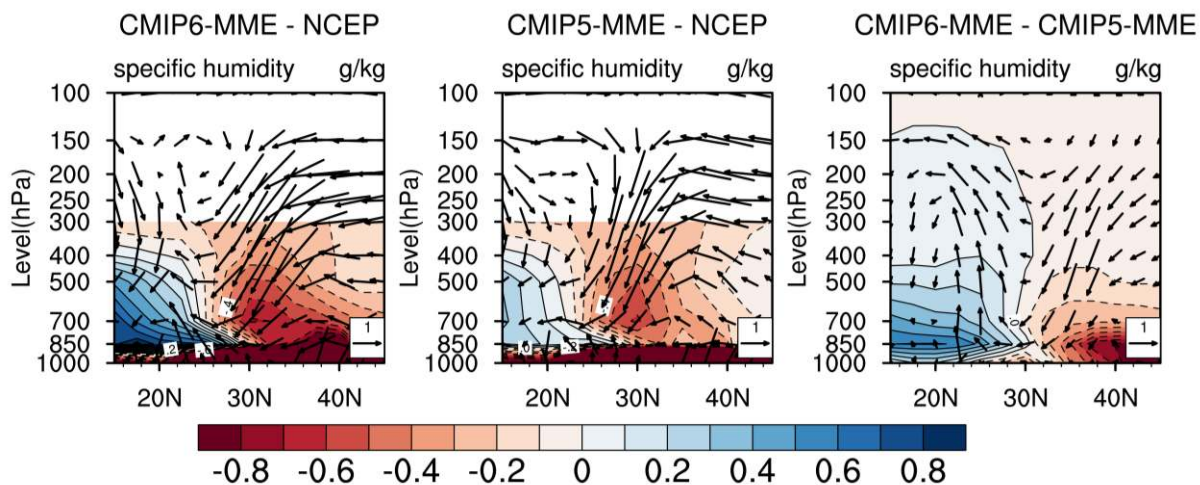
**Fig. 6.** Same as Fig. 2., but for five precipitation indices.



**Fig. 7.** Same as **Fig. 3.**, but for five precipitation indices.



**Fig. 8.** Same as **Fig. 4.**, but for five precipitation indices.



**Fig. 9.** Differences of meridional overturning circulation (vectors in  $\text{m s}^{-1}$ ) and specific humidity (shading, in  $\text{g kg}^{-1}$ , increase in blue and decrease in red) zonally averaged within  $110^{\circ}\text{E}$ - $120^{\circ}\text{E}$  for the historical period 1961-2005 in summer (JJA). From left to right are CMIP6-MME minus NCEP, CMIP5-MME minus NCEP and CMIP6-MME minus CMIP5-MME. The abscissa is the latitude and the ordinate is the pressure level ( hPa).(specific humidity has less levels(only to 300mb) from NCEP)