



HAL
open science

Multi-Model Ensemble Projection of Precipitation Changes over China under Global Warming of 1.5 and 2°C with Consideration of Model Performance and Independence

Tong Li, Zhihong Jiang, Lilong Zhao, Laurent Li

► **To cite this version:**

Tong Li, Zhihong Jiang, Lilong Zhao, Laurent Li. Multi-Model Ensemble Projection of Precipitation Changes over China under Global Warming of 1.5 and 2°C with Consideration of Model Performance and Independence. *Journal of Meteorological Research*, 2021, 35 (1), pp.184-197. 10.1007/s13351-021-0067-5 . hal-03195080

HAL Id: hal-03195080

<https://hal.science/hal-03195080>

Submitted on 10 Apr 2021

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.

Multi-model Ensemble Projection of Precipitation Changes over China under Global Warming of 1.5°C and 2°C with Consideration of Model Performance and Independence

Tong LI ¹, Zhihong JIANG ^{2*}, Lilong ZHAO ¹, and Laurent LI ³

1 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 210044

2 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 210044

3 Laboratoire de Météorologie Dynamique, IPSL, CNRS, Sorbonne Université, Ecole Normale Supérieure, Ecole Polytechnique, Paris 75005

(Submitted April 08, 2020; Accepted October 09, 2020)

Supported by the National Key Research and Development Program of China (2017YFA0603804, 2016YFA0600402 and 2018YFC1507704).

* Corresponding author: zhjiang@nuist.edu.cn. Tel: 025-58731135; 13813907836

Submission author: litong@nuist.edu.cn. Tel: 18851765368

©The Chinese Meteorological Society and Springer-Verlag Berlin Heidelberg 2020

ABSTRACT

A weighting scheme jointly considering model performance and independence (PI-based weighting scheme) is employed to deal with multi-model ensemble of precipitation over China from 17 global climate models. Four precipitation properties including mean and extremes are used to evaluate model performance and independence. The PI-based scheme is also compared to a Rank-based weighting scheme and to the simple arithmetic mean (AM) scheme. It is shown that the PI-based scheme achieves notable improvements in western China, with biases decreasing for all parameters. However, improvements are small and almost insignificant in eastern China. After calibration and validation, the scheme is used for future precipitation projection under the 1.5°C and 2°C global warming targets (above preindustrial level). There is a general tendency to wetness for most regions in China, especially in terms of extreme precipitation. The PI scheme shows larger inhomogeneity in spatial distribution. For total precipitation PRCPTOT (95th percentile extreme precipitation R95P), the land fraction for a change larger than 10% (20%) is 22.8% (53.4%) in PI, while 13.3% (36.8%) in AM, under 2°C global warming. Most noticeable increase exists in the central and east part of western China.

Key words: model performance and independence, multi-model ensemble, mean and extreme precipitation, future projection, 1.5°C and 2°C global warming

Citation: Li, T., Jiang, Z., Zhao, L. et al. Multi-Model Ensemble Projection of Precipitation Changes over China under Global Warming of 1.5 and 2°C with Consideration of Model Performance and Independence. *J Meteorol Res* 35, 184–197 (2021). <https://doi.org/10.1007/s13351-021-0067-5>

1. Introduction

Climate change under global warming is a major challenge for many natural ecosystems on the earth and for human societies (IPCC, 2013; WMO, 2019). The global average temperature is on increase and expected to continue increasing over the 21st century with a projected global temperature change about 1.4 °C to 4.8 °C under the medium and high emissions scenarios RCP 6.0 and RCP 8.5 at the end of the century (IPCC, 2013). Changes resulting from global warming may include rising sea levels due to the melting of glaciers and ice sheets, as well as an increase in severe climate events, such as extreme temperatures, extreme precipitation, and droughts etc. (Xu et al., 2013; Aslam et al., 2017; Guirguis et al., 2018). These extreme climate events provoke substantial economic losses and civilian casualties, which raises the urgency of searching adaptation and mitigation measures to combat climate change (Jones et al., 2014; Li et al., 2018; Zhan et al., 2018; Chen et al., 2020). The Paris Agreement on climate change sets an ambitious target of holding the increase of global average temperature to well below 2°C above preindustrial levels and recommends all efforts to limit the temperature increase to 1.5°C above preindustrial levels (UNFCCC, 2015), recognizing this would significantly reduce the risks and impacts of climate change (Schleussner et al., 2015; James et al., 2017; King and Karoly, 2017).

Climate models play a crucial role in studying the potential impacts of climate change, the ability of GCMs to reproduce observed features of current climate and past climate change increase our confidence to correctly make projections (Palmer et al., 2005; Zhou et al., 2007; Semenov and Stratonovitch, 2010; Wang et al., 2018; Ren et al., 2019). Multi-model arithmetic mean is widely used to reduce multi-model uncertainties and to make climate projections more reliable (Knutti, 2010; Knutti et al., 2010). The underlying assumption of the multi-model ensemble mean practice is that all models are reasonably independent, equally plausible, and distributed around the reality (Sanderson et al., 2015a, 2015b; Knutti et al., 2017). It also assumes that the range of models' projections is representative of what we believe is the uncertainty. But these are all strong assumptions, not always satisfied. And the reality is that some models are worse than others in how well they represent the observed mean climate and trend (Eyring et al., 2015; Baumberger et al., 2017). So weighting schemes were proposed based on the performance of climate models. The common practice is using the ability of model's simulated patterns against observations as a measure of model's skill (Perkins et al., 2009; Qi et al., 2017).

In the same line of weighting models based on their performance, the Rank-based scheme is widely applied in the evaluation of models and projection of future climate change (Chen et al., 2011). It constitutes simply of sorting models with well-defined criteria and using their rank to determine their weightings. This approach was shown to be quite efficient in reducing uncertainty from individual models and it usually outperforms any single model and the multi-model mean (Jiang et al., 2015; Li et al., 2016). Most methods formulate weights based on the performance of the model. They constitute an improvement of the simple arithmetic mean, but they ignore the problem of models' interdependence (Abramowitz and Bishop, 2015; Sanderson et al., 2015a, 2015b). Various studies have pointed out that some model pairs are closely related due to the increasing replication of code across institutions and even sharing common module of

model is natural (Knutti et al., 2013; Masson and Knutti, 2013). For example, some models are submitted by the same institution only with different resolutions (e.g., MPI-ESM-MR and MPI-ESM-LR), most institutions produce different models with similar configurations, or options for interactive atmospheric chemistry or carbon cycle (e.g., CMCC-CESM and CMCC-CM). So, it is important to consider the models interdependence in multi-model ensemble projections of climate change. However, few studies relative to this ensemble strategy are reported for future climate projection in China.

In this paper, a weighting scheme proposed by Knutti et al. (2017) which considers both model performance and independence (PI-based weighting scheme) is applied to construct projections of mean and extreme precipitation over China under the 1.5°C and 2°C warming targets. In Section 2, we describe the study area and datasets used. Section 3 introduces the PI-based weighting scheme and the Rank-based scheme. Followed in Section 4 are the performance comparison and projected changes in climatology and climate extremes indices. Finally, a summary of the major findings and conclusions are provided in Section 5.

2. Data

2.1 Observations and model datasets

The daily gridded precipitation dataset CN05.1, covering a period of 1961-2005 with $0.25^\circ \times 0.25^\circ$ resolution is used in this study for evaluation of ensemble performance of the fifth phase of the Coupled Model Intercomparison Project model (Taylor et al., 2012). The CN05.1 dataset is merged from 2416 weather observing stations in China. The station data is quality controlled by the China Meteorological Administration (CMA). More details concerning the data can be found in Xu et al. (2009) and Wu and Gao (2013).

Since climate in China is complex and atmospheric circulations between eastern and western China are quite different (Zhang et al., 1984; Jiang et al., 2015), the entire multi-model ensemble weighting procedure is carried out separately in eastern (east of 105°E) and western China (west of 105°E) in this work. In the western area, observation stations are scarce and models generally have difficulties to take into account the complex topography, one should be precautionary in using the results of reproduction and projection of climate over this area. The topography and location of surface weather observation stations across China are shown in Fig.1, the lines of 105°E and 40°N helping understanding better for the latter.

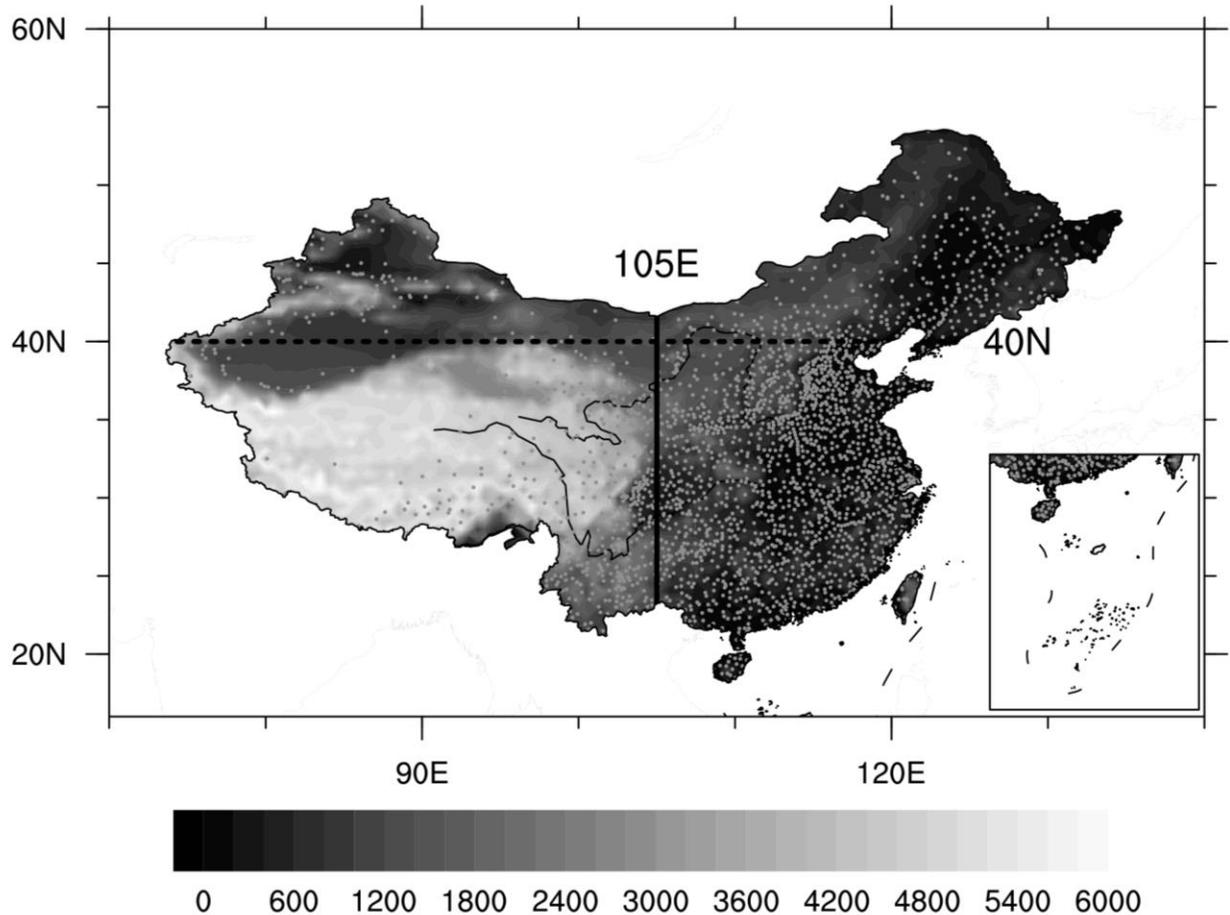


Fig. 1. Topography and location of surface weather stations in China.

17 CMIP5 models' historical simulations and future projections from Representative Concentration Pathways (RCPs) 8.5 are used to generate 1.5 and 2°C warming. These models were selected on the sole criterion of data availability for our purpose, especially with daily precipitation for our warming targets at 1.5 and 2 °C. All the datasets retrieved through data portals of the Earth System Grid Federation, which can be obtained from <https://esgf-node.llnl.gov/search/cmip5/>. Some basic characteristics of the used models are listed in Table 1. The situation is somehow heterogeneous, some institutions providing single model version while others multiple versions of their model with different resolutions, different physical packages, or different complexity of the earth system. We can thus expect to have a huge diversity for the CMIP5 multi-model ensemble simulations. And at the same time there is also a strong interdependence among them. For the 17 CMIP5 models that we selected following the sole criterion of data availability for our analysis, three models come from BCC, two models from IPSL, three models from MIROC, two models from MPI; and other seven models from seven different institutions. In our present work, we use only one realization from each model or model version. And we are aware that a full exploitation of the ensemble realizations would create a stronger interdependence.

Table 1. Model name, modeling center and country, Institution identification (ID), and atmospheric resolution of 17 CMIP5 global climate models (Expansions of acronyms are available at <http://www.ametsoc.org/PubsAcronymList>)

| Model name | Modeling center and country | Institution ID | Atmospheric resolution (lat × lon) |
|---------------|--|----------------|------------------------------------|
| BCC-CSM1-1 | | | 2.81258°×2.81258° |
| BCC-CSM1-1-M | Beijing Climate Center, China Meteorological Administration, China | BCC | 1.125°×1.12° |
| BNU-ESM | | | 2.8°×2.8° |
| CanESM2 | Canadian Centre for Climate Modelling and Analysis, Canada | CCCMA | 2.8°×2.8° |
| CCSM4 | National Center for Atmospheric Research, United States | NCAR | 1.25°×0.94° |
| CNRM-CM5 | Centre National de Recherches Météorologiques– Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique, France | CNRM–CERFACS | 1.4°×1.4° |
| CSIRO-Mk3-6-0 | Commonwealth Scientific and Industrial Research Organization, Australia | CSIRO-QCCCE | 1.875°×1.875° |
| GFDL-CM3 | NOAA/Geophysical Fluid Dynamics Laboratory, United States | NOA/GFDL | 2.5°×2.0° |
| IPSL-CM5A-LR | | | 3.75°×1.895° |
| IPSL-CM5A-MR | L’Institut Pierre-Simon Laplace, France | IPSL | 2.5°×1.27° |
| MIROC5 | | | 1.40625°×1.40625° |
| MIROC-ESM | National Institute for Environmental Studies, The | MIROC | 2.815°×2.815° |

| | | | |
|----------------|---|-------|---------------|
| MIROC-ESM-CHEM | University of Tokyo, Japan | | 2.815°×2.815° |
| MPI-ESM-LR | Max Planck Institute for Meteorology, Germany | MPI-M | 1.875°×1.875° |
| MPI-ESM-MR | | | 1.875°×1.875° |
| MRI-CGCM3 | Meteorological Research Institute, Japan | MRI | 1.125°×1.125° |
| NorESM1-M | Norwegian Climate Centre, Norway | NCC | 1.8725°×2.5° |

2.2 Climate extreme indices

Three extreme precipitation indices are investigated in this research, including the maximum 5-day precipitation (RX5DAY), maximum daily precipitation (RX1DAY), and strong precipitation events (R95P). In addition, the total precipitation (PRCPTOT) is used to represent the precipitation climatology. These indices are generally considered effective in extracting climate change information and have been widely used to identify and monitor extreme precipitation (Sillmann et al., 2013; Zhang et al., 2013; Zhou et al., 2016). All extreme indices of models were calculated with the Expert Team on Climate Change Detection and Indices (ETCCDI) (Zhang et al., 2011).

The extreme precipitation indices from different models and observation were first calculated at their original grids. To facilitate model intercomparing and evaluation against observations, all CMIP5 data and observation were interpolated onto the same $1^\circ \times 1^\circ$ grid, using bilinear interpolation.

3. Methodology

3.1 PI-based weighting scheme

The weighting scheme used here was first proposed by Knutti et al. (2017) considering both performance and independence of models (hereafter PI-based weighting scheme). It is based on two general considerations: models that agree poorly with observations for a selected set of diagnostics get less weight, and models that largely duplicate existing models also get less weight.

As proposed by Knutti et al. (2017), the single model weight w_i for model i is defined as

$$w_i = e^{-(D_i/\sigma_d)^2} / \left(1 + \sum_{j \neq i}^M e^{-(S_{ij}/\sigma_s)^2} \right). \quad (1)$$

where M is the total number of models, D_i is the distance of model i to observations, and S_{ij} is the distance between model i and model j . The parameters σ_d and σ_s determine how strongly the model performance and similarity are weighted, which can be determined

through a cross-validation procedure. A large σ_d would effectively make the ensemble converge to the situation of model democracy, whereas a small σ_d puts important weights on only a few models; σ_s determines the typical distance by which a model would be considered similar to another. More details of the methodology can be found in [Knutti et al. \(2017\)](#).

In this paper, model evaluation metrics include the assessment of both spatial pattern and temporal variability for precipitation, which allows us to deduce the distances D_i and S_{ij} in Eq. (1) by using the average of two normalized values: normalized root-mean-square error (RMSE) and normalized interannual variability score (IVS), the normalization being realized by their respective median values.

The spatial root-mean-square error is defined as:

$$\text{RMSE}(m, o) = \sqrt{\frac{1}{N} \sum_{k=1}^N (M_k - O_k)^2}. \quad (2)$$

where M_k and O_k denote the model pattern of a variable under investigation and the corresponding observed pattern, respectively. N indicates the number of spatial points. Smaller RMSE value represents a better agreement between the model and observation.

The performance in interannual variation can be evaluated by a measure of skill score IVS (Interannual Variability Skill Score) described by [Chen et al. \(2011\)](#) as follows:

$$\text{IVS}(m, o) = \left(\frac{\text{STD}_m}{\text{STD}_o} - \frac{\text{STD}_o}{\text{STD}_m} \right)^2. \quad (3)$$

where STD_m and STD_o denote the interannual standard deviation of model and observation respectively. Smaller IVS value indicates a better agreement between the simulation and observation. Eq. (2) and (3) are shown in the form of a measure between a model and the observation. They can be easily converted into inter-model measures between models i and j (not shown).

Since multiple (four in this study) precipitation indices are used in this study, we define a comprehensive distance D of performance to combine the four individual distances D_v as Eq. (4). Our choice was based on the fact that the reliability of future climate projection should rely on a comprehensive measure of the performance of climate models. The four precipitation indices PRCPTOT, RX1DAY, RX5DAY and R95P represent different characteristics of precipitation. Therefore, an integrated (or comprehensive, or combined) distance should be a good choice. Coefficients w_v are set as 1/4 to make their sum equal to 1. Different coefficients can be taken into account based on the practical applications in follow-up studies.

$$D = \sum_{v=1}^4 w_v D_v. \quad (4)$$

When the PI-based weighting scheme is used, its free parameters σ_d and σ_s can be determined through a cross-validation procedure. It is actually a perfect model test in which each model is, in turn, treated as the truth ([Brunner et al., 2019](#)). It is to be noted that our perfect model test uses two distinct periods in the past and in the future respectively: the calibration is done for the past period, but the cross-validation is

evaluated for the future period. In other words, for each model (considered as truth in turn), a set of weightings for the remaining models are calculated to mimic the past values of the true model. Once calibrated, the PI-based weighting scheme is used to deduce the future values of the true model. Such a cross-validation strategy allows us to test the future projection capability of the weighting scheme.

In practice, we followed the procedure of Knutti et al. (2017) and Brunner et al. (2019), and calculated the percentage of cases falling down to the 5th-95th percentiles range obtained from the true model. The good value of σ_d was then empirically determined as its minimum value making this percentage equal to 90%. Fig. 2 shows the percentage of cases, in function of prescribed σ_d values (with an incremental bin of 0.01), where the actual outcomes of the perfect model tests are in the 5th-95th percentiles range. This cross-validation is repeated for each of the four variables, as represented by the four curves in Fig. 2. By choosing the minimum value of the four curves when the threshold of 90% is reached, σ_d is estimated to be 0.54 in western China, and 0.46 in eastern China. These values are quite consistent with that obtained by Knutti et al. (2017) who also normalized their Euclidean distance by the median value. The dependence on the value of σ_s is rather weak relative to σ_d , so the value of σ_s is set as 0.5 as suggested in Knutti et al. (2017). The PI-based weighting scheme is now fully established after its free parameters are calibrated.

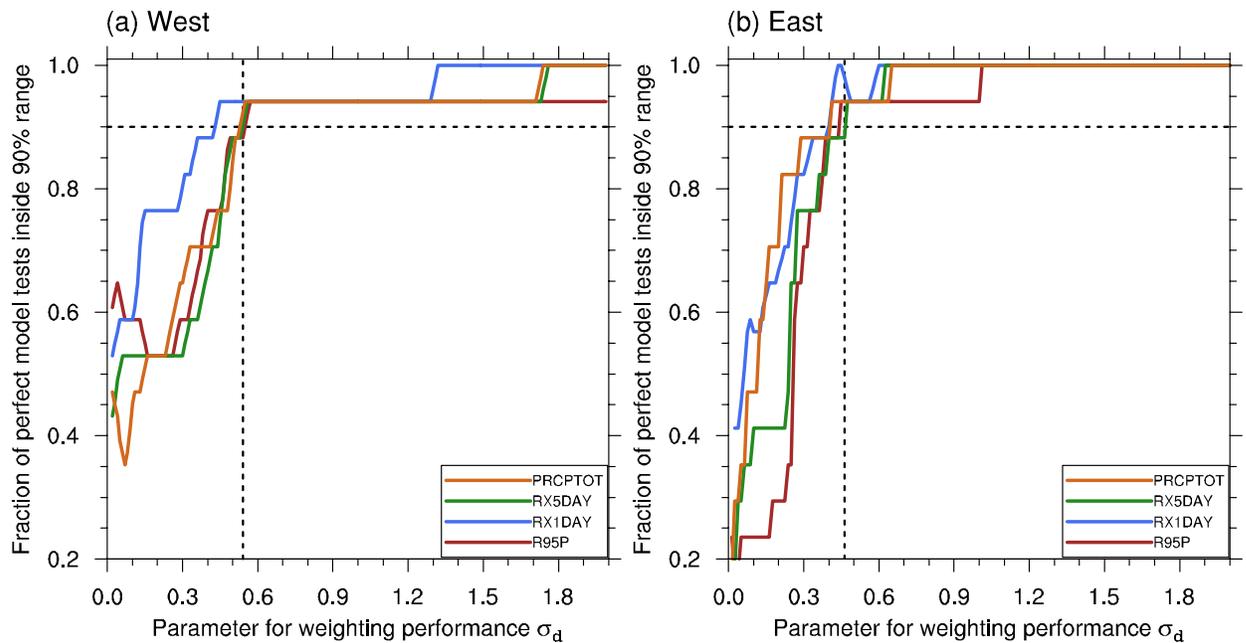


Fig. 2. The fraction of cases for four indices when the actual outcomes of the perfect model tests are in the 5th-95th range weighted by all other models. For all indices, the minimum value of σ_d for the fraction expected to exceed 90%.

3.2 Rank-based weighting scheme

In order to assess the proposed weighting scheme, we evaluate it against two largely-used algorithms, the Rank-based weighting scheme (Chen et al., 2011; Jiang et al.,

2015; Li et al., 2016) and arithmetic mean (hereafter called AM), the latter serving as a baseline. To ensure a fair comparison, the Rank-based weighting scheme uses the same evaluation metrics based on the spatial pattern and interannual variability of relevant precipitation indices. If R_i is the rank for model i after evaluation which can be based on multiple criteria, we can then calculate a performance indicator P_i and it can be ultimately converted into weight w_i after normalization:

$$P_i = \frac{\sum_{i=1}^M R_i}{R_i} . \quad (5)$$

and

$$w_i = \frac{P_i}{\sum_{i=1}^M P_i} . \quad (6)$$

Eq. (6) can guarantee the sum of all weights is equal to 1.

3.3 Periods for calibration and evaluation, and determination of 1.5°C and 2°C global warming

For all the three weighting schemes (PI-based, Rank-based and AM scheme), their calibration period is from 1961 to 1985. The validation is performed with independent samples from 1986 to 2005. After the calibration and validation, the three weighting algorithms are applied to future precipitation projection under the framework of 1.5°C and 2°C global warming targets.

To determine the timing of the two warming targets, a 40-year period from 1861 to 1900 is firstly selected as the pre-industrial reference. For each of the models in our investigation, the time series of annual global mean temperature is smoothed by a 21-year running-average operation, and we can then find the year when 1.5°C (2°C) warming is reach. A 21-year window centered on this year is then used to define the 1.5°C (2°C) world. The reference period 1985-2005 are used to represent nowadays to assess the changes in precipitation climate.

4. Results

4.1 Model skill and weights

The final weights obtained by the two weighting schemes, sorted by descending order of PI-based scheme are shown in Fig. 3. For both western and eastern China, the PI-based weighting shows a larger range from strong weights to weak ones, which means models with a good performance and a large independence from others are more prominent in the PI algorithm.

Although σ_d in western China is larger than its counterpart in eastern China, models closer to observation get higher weights, which indicates that the inconsistency among models in the west is more significant. For both areas, we can see that models from a same institution have generally closer weights, but they are not among the highest,

which may be the direct consequence of the PI algorithm with interdependence of models taken into account. We can observe that the models IPSL-CM5A-LR and IPSL-CM5A-MR, both from IPSL, have a better performance above the average. However, the models BCC-CSM1-1 has a similar weight as BCC-CSM1-1-M in the east, but a quite different weight, compared to BNU-ESM from the same institution. In an idealized condition, if a second member happens to be identical to the first (that is, $s_{ij} = 0$), the two members would have half their original weight. We can generalize the example into the case of N identical members, each would have a weight scaled by $1/N$. In reality, CMIP5 members are not perfectly independent and our work presented here uses only the first run from each model, which makes the members in our ensemble have a relative weak repeatability from each other.

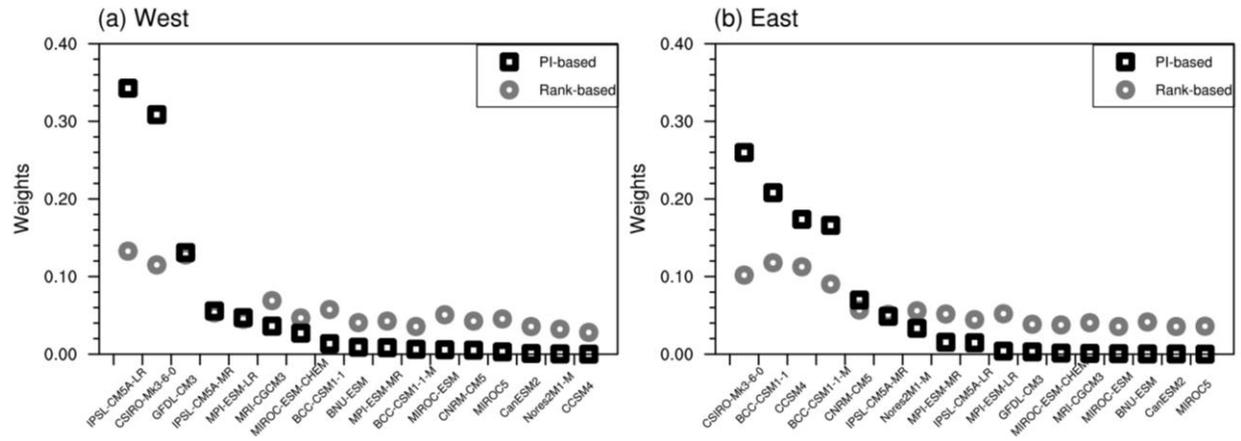


Fig. 3. The models' weights obtained by PI-based weighting scheme (black squares) and Rank-based weighting scheme (gray circles) over (a) western and (b) eastern China.

4.2 Evaluation of the weighting strategies

Fig. 4 shows relative biases (areal mean in shaded bar, first and third quartiles in bar with whiskers) of the four indices reproduced by three weighting strategies, PI-based, Rank-based and AM, against observations during the evaluation period 1986-2005 in western and eastern China, respectively. We can see that areal-mean biases in the western area are (about 2 to 3 times) larger than those in the eastern area for all precipitation indices. This can be explained by the fact that most models have mediocre performance in reproducing extreme events related to the complex topography in western China (Jiang et al., 2015; Chen et al., 2017).

In the western area, the areal-mean biases of PI-based scheme for the four indices PRCPTOT, RX5DAY, RX1DAY and R95P are 32, 79, 46 and 150%, respectively. Biases of the PI-based scheme are much lower than those from the other two schemes, and the regional interquartile bias ranges are also smaller. Take the most noticeable index R95P as example, on the basis of AM, the areal-mean relative bias decreases by 101%, and decreases by 78% if compared to the Rank scheme. The 25th and 75th percentile errors of AM are 81 and 327%, while only 32 and 195% reproduced by PI-based scheme. However, biases' differences among indices in eastern China are not very prominent.

From the 25th and 75th percentile, we can see the higher dry and wet biases of extreme precipitation are better reduced in the weighting schemes.

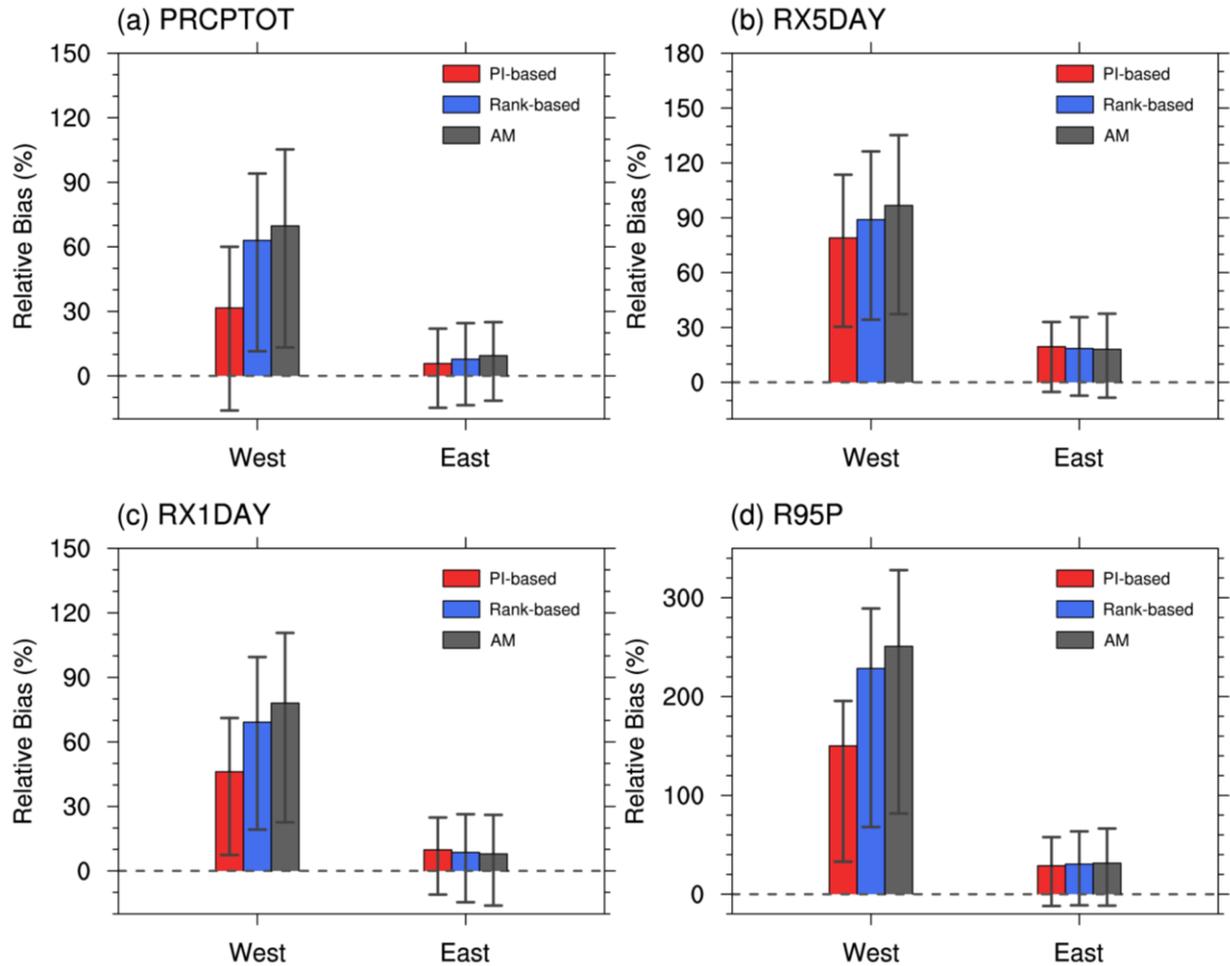


Fig. 4. The regional relative bias simulated by PI-based, Rank-based and AM schemes against observations for four precipitation indices during the period of 1986-2005. Color bars are the mean biases and whiskers represent the first and third quartiles (unit: %).

Among the four precipitation indices, the relative bias for R95P is the most noticeable, even the best performing PI scheme shows a relative bias of 150 and 29% in western and eastern China, respectively. So, the spatial distribution of R95P's biases is illustrated in Fig. 5 for the three weighting strategies respectively to find more spatial details. It can be seen that large biases (>50%) are located in western China to the south of 40°N, especially in the periphery of the Tibetan Plateau where biases are about 150 to 300%. In eastern China, an underestimate lower than 50% exists in south of the Yangtze River. Most areas in the north of the Yangtze River have wet biases about 100%. Comparing the three weighting schemes, we can find that areas with significant improvement in the PI-based weighting scheme are concentrated in the north and central Tibetan Plateau where we see the largest biases. They are largely reduced to a level

below 50%. In other parts of China, especially in eastern areas, the improvement of PI-based scheme is relatively modest. Similar results are also found in other indices (not shown), with the largest biases concentrated in the periphery of the Tibetan Plateau. They are nevertheless smaller compared to the case of R95P. Biases in the western area are the mostly reduced. Although the relative bias reduction is large over the Tibetan Plateau, it should be noted that observation stations are sparse and unevenly distributed over the Plateau, which can generate spurious results with improperly-designed interpolation algorithm. One should be precautionary in using the results of reproduction and projection of climate over this area.

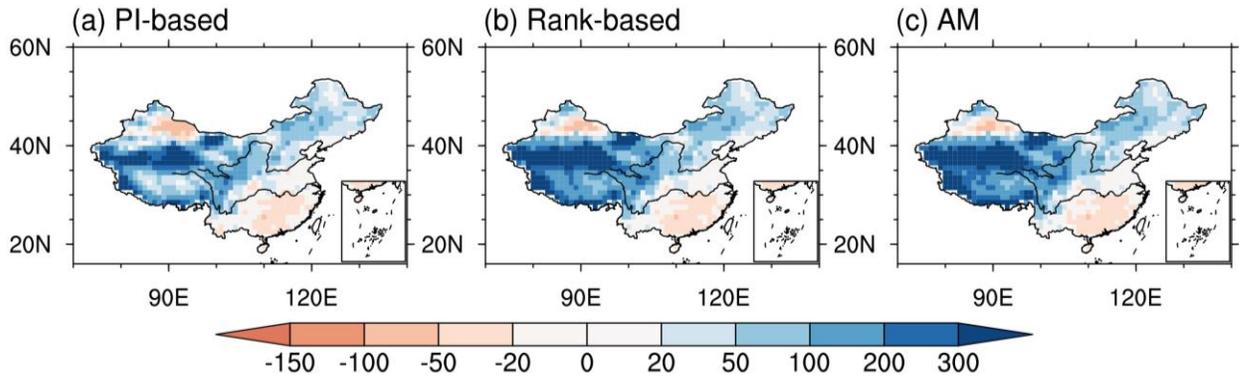


Fig. 5. Spatial distribution of relative bias for R95P from (a) PI-based (b) Rank-based and (c) AM schemes against observations during the period 1986-2005 (unit: %). Warm and cool colors indicate dry and wet bias respectively.

Taylor diagram is now used to present a concise statistical analysis of the three weighting schemes in the evaluation period 1986-2005 (Fig. 6). It can display three pieces of information: pattern correlation coefficient, ratio of the centered standard deviations and root mean square error, two of them being independent and allowing to deduce the third one. There is generally a weak performance with the AM (gray markers). In western China, the spatial correlation coefficients of major indices between the simulations and observations are less than 0.8 and the maximum does not exceed 0.85; the standard deviations are also large. In eastern China, all markers representing the AM are the most distant from the ref-point. It is easy to see that the performance reproducing the spatial pattern by the weighted schemes is improved, and PI (red markers) is the best with pattern correlation coefficients higher than 0.85 and ratios of spatial variance closer to 1 for all the four precipitation indices in western China. In eastern China, the performance of three strategies are similar, all markers concentrating together with spatial correlation coefficients between 0.85 and 0.9, and standard deviation ratios between 0.5 and 0.6. Again, R95P shows its weakest score compared to other indices.

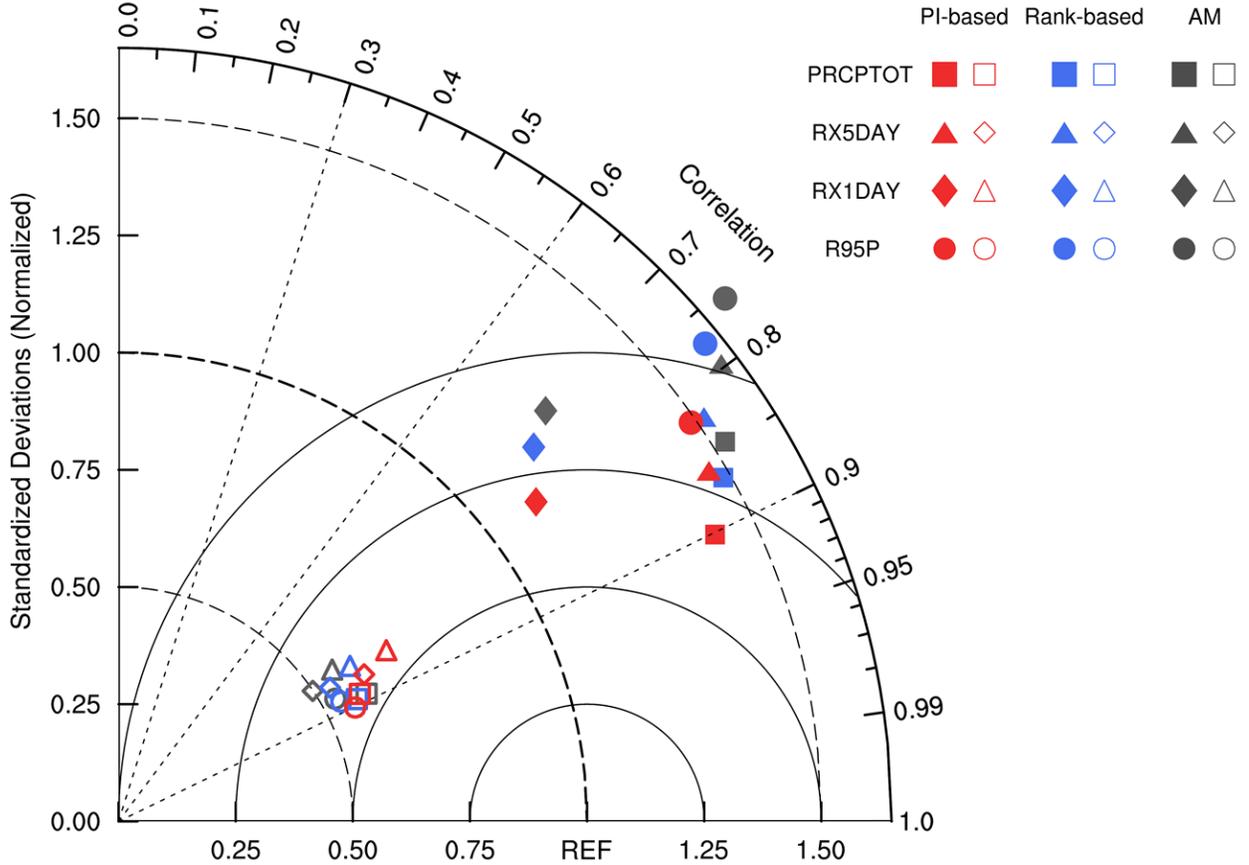


Fig. 6. Taylor diagram showing the four precipitation indices under the three weighting schemes during the period 1986-2005. The solid and hollow markers represent western and eastern China respectively.

In brief, the PI-based weighting strategy effectively enhances our capability of reproducing the precipitation spatial characteristic in China, especially in western China, all precipitation indices with the PI-based scheme show significant improvement. In eastern China, however, the improvements are small and insignificant.

4.3 Projections under the 1.5 °C and 2°C warming

In view of the added value of combining multi-model results, especially by the PI-based scheme, a projection of future climate corresponding to 1.5°C and 2°C warming targets (relative to preindustrial) under RCP8.5 emission scenario was produced with the above weighting strategies. Fig. 7 displays boxplots showing the spatial distributions of relative changes for the four precipitation indices and for the three weighting schemes, respectively. Each boxplot allows to present the 10%, 25%, 50%, 75%, and 90% percentiles of relevant field across all spatial grids, and for the two warming targets at 1.5 °C (blank boxes) and 2°C (hatched boxes), respectively. As a conventional practice, the plotted change is relative to 1985 – 2005, a common reference period, while the target warming levels (1.5°C and 2°C) are relative to preindustrial. Compared with the Rank-

based scheme (blue box and whisker) and the AM scheme (gray box and whisker), the PI-based scheme (red box and whisker) shows larger dispersions (boxes and whiskers are longer), which means the projected fields have larger inhomogeneity in their spatial distribution, with more pronounced wet and dry conditions. The areal-mean values (gray dots) are relatively close from each other with those from the PI-based scheme slightly higher.

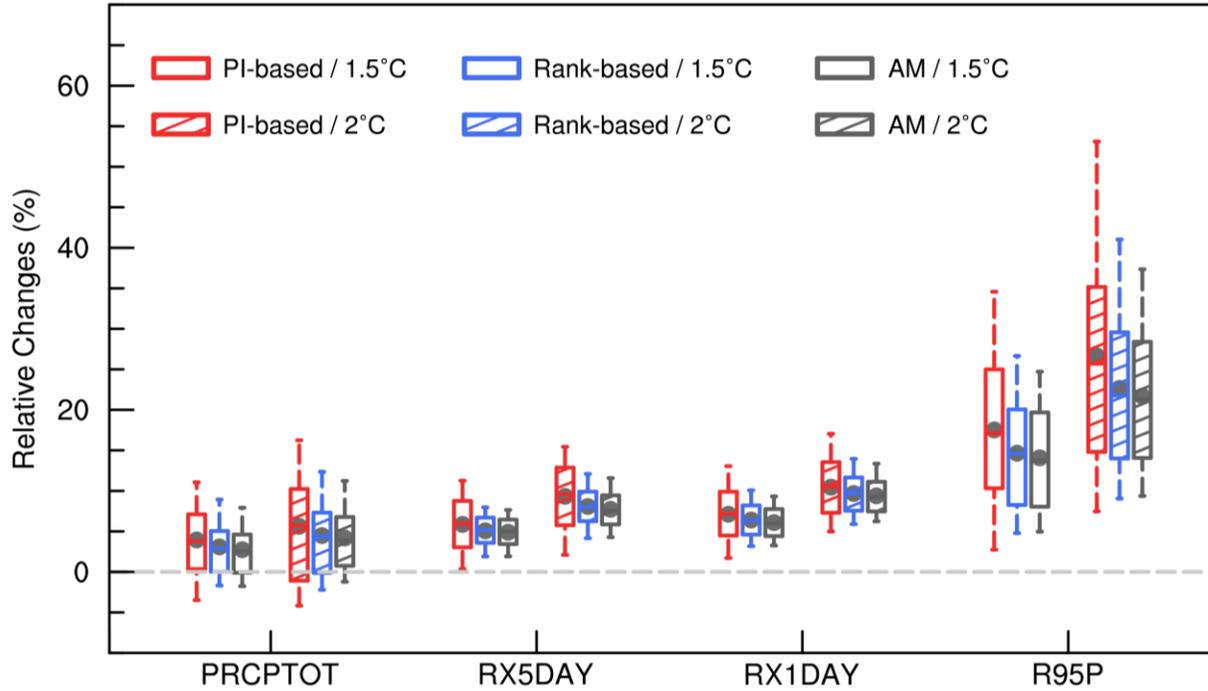


Fig. 7. Boxplot for relative changes of total precipitation and extreme indices (compared to the period of 1985-2005) from PI-based, Rank-based weighting schemes and AM over China under 1.5°C and 2°C global warming (unit: %).

To better illustrate the impact of weighting schemes on the projection of precipitation, results can also be displayed in the form of a cumulative distribution function (CDF) across China, as shown in Fig. 8. Such a graphic presentation is very useful since the CDF can be directly interpreted as the fraction of the national territory where changes are inferior to the corresponding value of the abscissa. On the right side of the ordinate, we also indicated the complementary values (1-CDF) of what shown on the left side. It can be interpreted as the fraction of the national territory where changes are superior to the corresponding value of the abscissa. Each panel in Fig. 8 represents one of the four precipitation indices with curves corresponding to the three weighting schemes and to the 1.5°C and 2°C warming targets, respectively. For all indices, the Rank-based CDFs are almost congruent with the basic AM's, while obvious distinctions can be detected for PI.

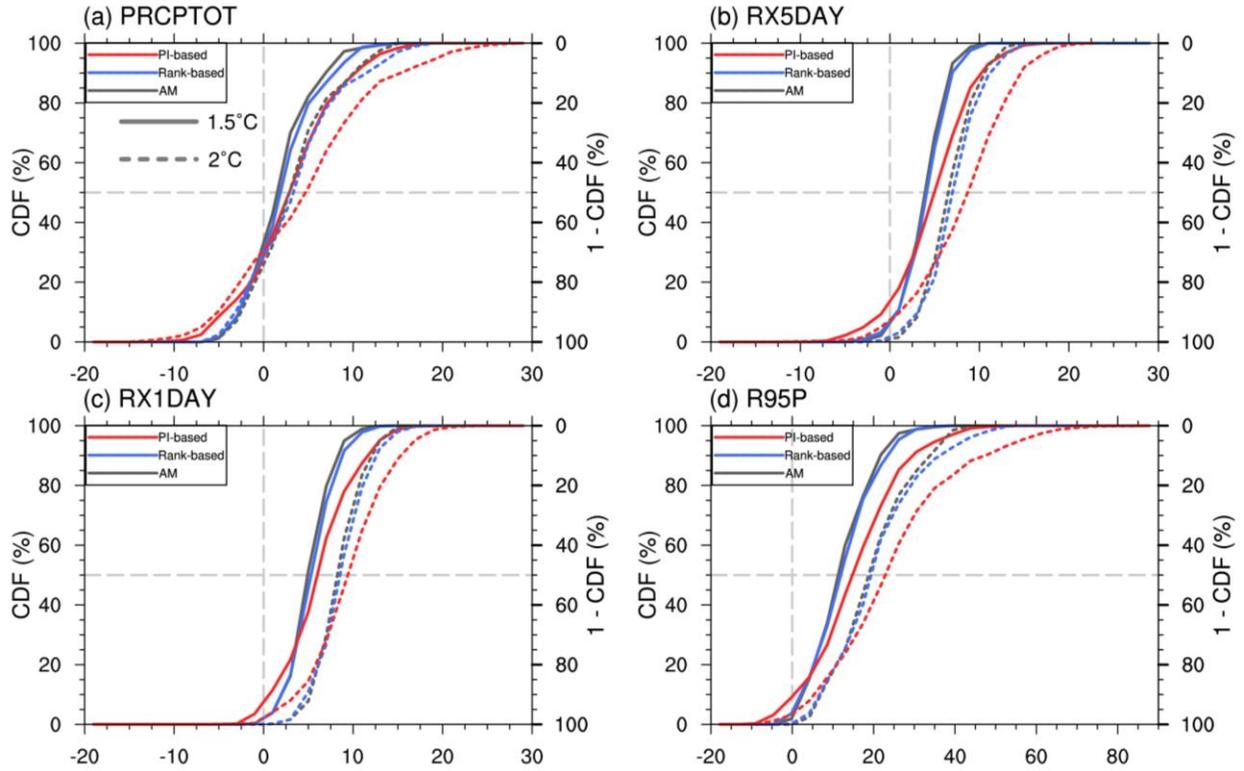


Fig. 8. The spatial fraction CDFs of changes weighted by PI-based, Rank-based and AM under 1.5°C and 2°C global warming for the indices over China. The abscissa displays the changes of precipitation indices while the left ordinate shows the CDF values, the complementary values (1-CDF) shown on the right side. The changes are expressed by percentages (%) compared to the period of 1985-2005.

Firstly, we examine the changes of PRCPTOT, relative to present day (1985-2005), the PI-based scheme shows that half of China would experience an increase larger than 3.8 and 5.7% in the 1.5°C and 2°C warmer climates, respectively. Under the same conditions, a quarter of China would experience an increase larger than 7.1% and 10.2%. Compared with the PI-based scheme, the increases in PRCPTOT are slightly smaller in AM, projecting that half of China would increase by 2.5 and 3.9%, and quarter of China by 4.6 and 6.8%, respectively. These characteristics show some similarity among the four precipitation indices, with a trend to larger values in the order of PRCPTOT, RX5DAY, RX1DAY and R95P. We remark large increases and large dispersions for R95P, under the 1.5°C and 2°C warming target, half of China may experience changes larger than 17.1% (25.7%) in PI's projection while 13.7% (21.3%) in AM's.

The CDF curves in Fig. 8 can also be examined from the vertical perspective, which gives the land fraction for a given change threshold. If we use a threshold of 10% for PRCPTOT, we can see that the fraction of national territory where PRCPTOT changes exceed the threshold of 10%, under the 1.5°C (2°C) global warming target, is 10.6% (22.8%) with the PI-based scheme, while around 3.9% (12.5%) with the Rank-

based scheme, and 2.8% (13.3%) with AM. Similarly, for R95P with a 20% threshold, the fraction of national territory is 26.7% (53.4%), 13.2% (37.8%), 9.7% (36.8%) for the three weighting schemes (PI, Rank and AM), respectively. If we use the threshold 0 for PRCPTOT, we can then evaluate the fraction of the national territory with mean rainfall increasing (or decreasing). It is certainly fortuitous to observe that the three weighting schemes and the two warming targets all converge to a situation of two thirds of China with rainfall increased and the remaining third with rainfall decreased. If we apply the threshold 0 to extreme indices, the majority of the national territory shows increases with RX5DAY by 86.4% (92.7%), RX1DAY by 92.5% (97.7%) and R95P by 91% (95.8%), respectively, for the warming target 1.5°C (2°C).

The spatial distributions of PRCPTOT changes weighted by the three schemes are depicted in Fig. 9. Consistent with what shown by the boxplots and CDF curves, the spatial structure has a larger range with the PI-based scheme. Robust increases are detected in the central and east part of western China in all three projections, but the increase amount projected by the PI-based model exceeds 20% in the 1.5 °C global warming. Especially in the northern Tibetan Plateau, it is twice bigger than that projected by AM (about 10%). Under 2°C global warming, the increase with the PI-based model exceeds 25%, while it is about 10% in AM. For the half-a-degree additional warming (right column), the difference among the three weighting models is not significant except for the center and east Tibetan Plateau and Inner Mongolia area, where the change amplitude is more noticeable with the PI-based model. Similar results can be found in RX5DAY, RX1DAY and R95P, but R95P shows more significant increases especially in the Tibetan Plateau and in northeastern China.

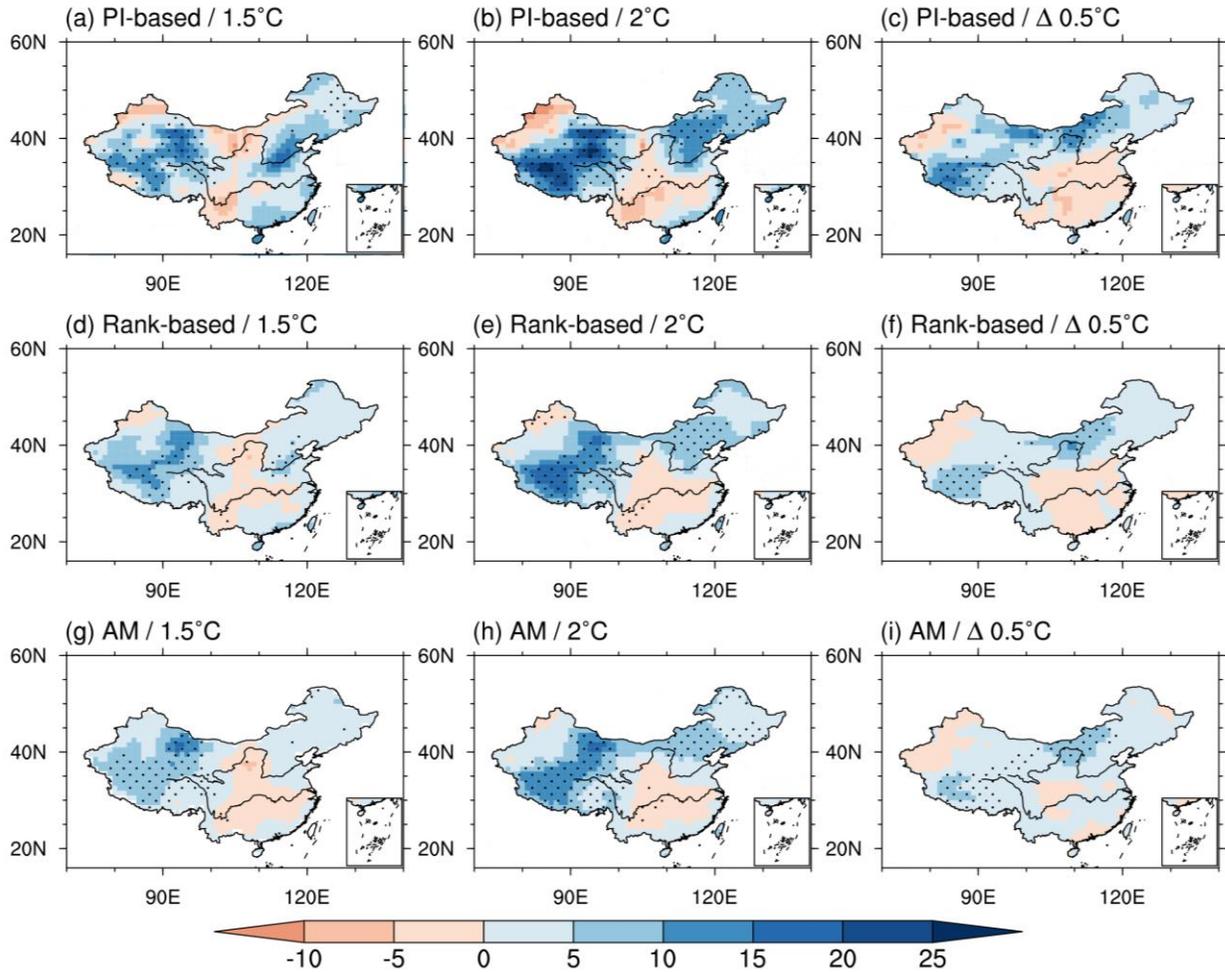


Fig. 9. Percent changes of PRCPTOT for the 1.5°C (first column), 2°C (second column) global warming under RCP 8.5 scenario compared to the period 1985–2005 and the additional 0.5 °C warming (third column) from simulations of PI-based (top), Rank-based (middle) weighting schemes and AM (bottom). Areas with significant changes above 95% confidence level are marked with black dots (unit: %).

The reason resides certainly in the fact that the PI-based scheme has sharper transitions in the distribution of weights among models. That is, the final results are basically dominated by a few models having larger weights. In other words, “good” localized characteristics from these well-behaved models are less affected by other models. With the same argument, we can expect that the arithmetic mean algorithm (AM), which performs simple average over many models, would underestimate high-value regions. Sanderson et al. (2017) pointed out that some noticeable differences of the uncertainty range also exist with weighting of models, the underlying causes being the same as we evoked here. Similar conclusion is found in other researches on the weighting strategy, the projected changes made by weighted ensembles are generally comparable

from each other on the spatial patterns, but more pronounced amplitude in uncertainty regions (Langenbrunner and Neelin, 2017; Massoud et al., 2019).

The projected changes with the PI-based strategy under the 1.5 °C and 2°C warming targets are depicted in Fig. 10 for the three extreme precipitation indices. It is obvious that all extreme precipitations increase across almost whole China. The mean increase for RX5DAY is by 5.8 and 9.3% under 1.5 °C and 2°C global warming. Similarly, RX1DAY increases by 7.1 and 10.4%. The increase of strong precipitation R95P is the most noticeable, the national mean being by 17.5 and 26.6%, respectively. For all extreme indices, the Tibetan Plateau and northeastern China are areas with significant wetter conditions, especially in the 2°C warming, higher than other areas about 5%. Meanwhile, there is no significant increase found in the middle reaches of the Yangtze River and the Yellow River, even a slight decrease can be observed. An additional 0.5 °C warming may lead to increases in heavy precipitation over most areas, the national mean increases of RX5DAY and RX1DAY are 3.5 and 3.4%, respectively, the most significant response of R95P can reach 9.3%. The center and east of the Tibetan Plateau and the north of Inner Mongolia have been shown to be more significantly associated to the additional global warming. Although a precise comparison with previous works is out of scope of the present paper, most our findings are consistent with what reported in the current literature (Wu et al., 2015; Chen et al., 2017; Shi et al., 2018) showing reliable features, such as the increase of extreme precipitation indices in the Tibetan Plateau and northeast China, and their insignificant change in central China.

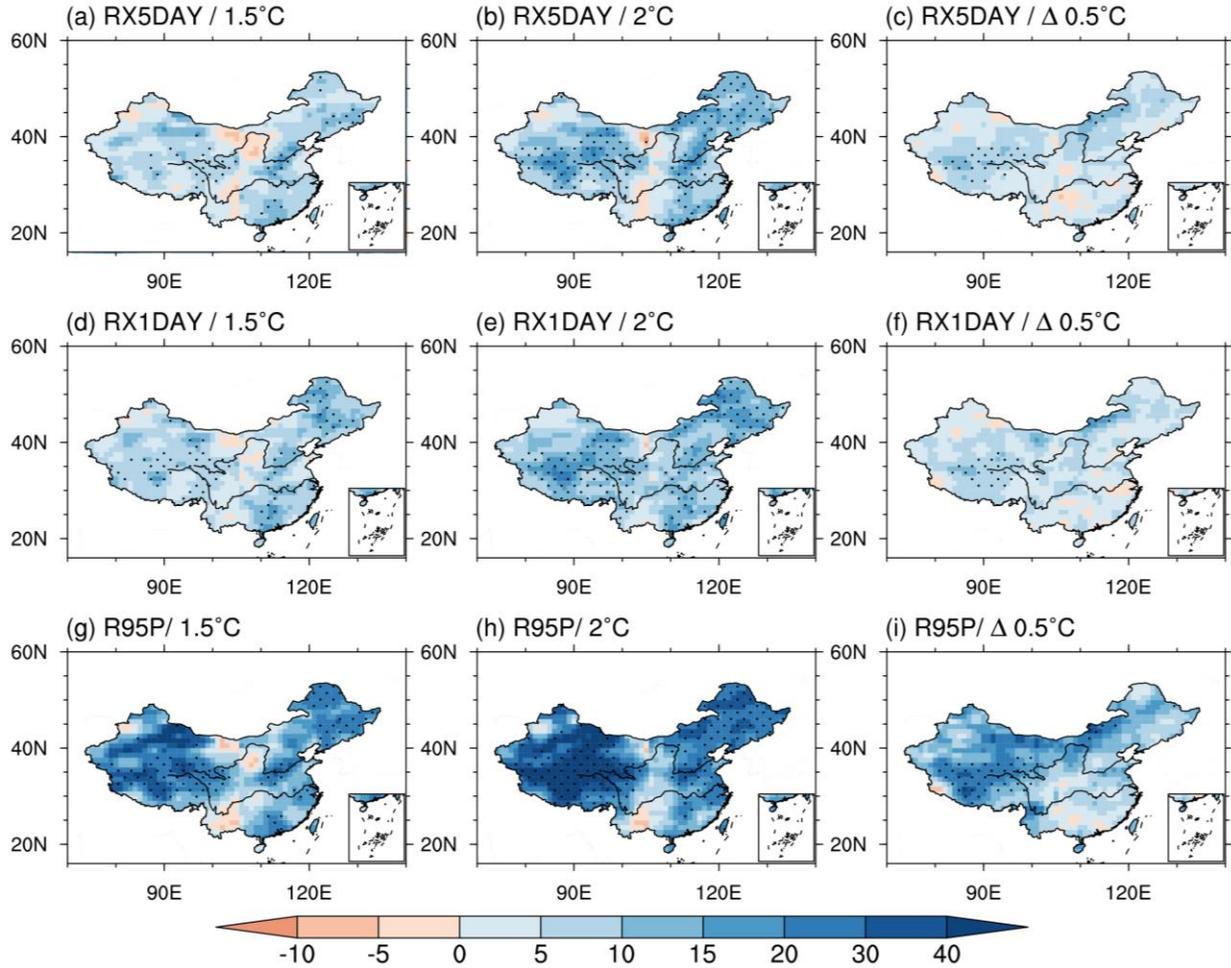


Fig. 10. Percentage changes of the three extreme precipitation indices projected by PI-based weighting scheme for the 1.5°C (first column), 2°C (second column) global warming under RCP 8.5 scenario compared to the period 1985-2005 and the additional 0.5 °C warming (third column). Areas with significant changes above 95% confidence level are marked with black dots (unit: %).

5. Conclusions and discussion

A weighting scheme considering both model performance and independence (PI-based weighting scheme) is used to deal with multi-model ensemble of precipitation over China under the warming targets of 1.5°C and 2°C. To better appreciate its performance, a work of inter-comparison with the performance-only (Rank-based) scheme and a full-model-democracy scheme (arithmetic mean, AM) was carried out. Main conclusions are as the following.

(1) Compared to the Rank-based and AM strategies, the PI-based weighting scheme produces a better simulation of spatial patterns for the total precipitation and extreme indices, especially in western China over the Tibetan Plateau. From the baseline of AM, the western mean bias decreases about 38% for PRCPTOT, 18% for RX5DAY,

32% for RX1DAY and 101% for R95P. For all the four indices, the spatial pattern correlation coefficients are higher than 0.8, and the ratios of spatial variations are closer to 1. Nevertheless, no significant improvements are found in eastern China.

(2) An inter-comparison of future climate projection among the three schemes shows that their spatial patterns are highly correlated but PI has larger inhomogeneity in spatial distribution. A few results under the 2°C global warming target can be detailed here for the total precipitation PRCPTOT and strong precipitation R95P. Their critical change values, allowing to divide whole China area into two equal halves under the 2°C global warming target, are 5.7% and 25.7% in PI, while 3.9% and 21.3% in AM. The land fraction of whole China's territory with a change of PRCPTOT (R95P) larger than 10% (20%) is 22.8% (53.4%) in PI, while 13.3% (36.8%) in AM. The difference shows an increase of 9.5% (16.6%). In the central and east part of western China, the increase for both PRCPTOT and R95P is the most noticeable in PI, which can exceed 20% and 40% for PRCPTOT and R95P respectively, twice bigger than in AM.

Our results show an obvious improvement in western China for the PI-based ensemble scheme. However, both observation and simulation are less reliable in western China due to limited observation stations and the presence of complex topography including high mountains and the Tibetan Plateau. So, results over this area should be regarded with precautions.

It is to be noted that any weighting scheme of models are inevitably related to the choice of distance metrics and variables used, which implies intrinsic uncertainties for the methodology. In the present work, we used a combined set of diagnostics (PRCPTOT, RX5DAY, RX1DAY, R95P) and took the spatial and temporal features simultaneously into account to determine model performance and independence. Other performance metrics, not only for the mean state and extreme, but also for trends, or key physical process should be explored in future.

The basic idea of PI-based scheme is generic and could be applied to a wider range of climate change issues by integrating a larger set of GCM simulations. It is particularly promising in the upcoming CMIP6 framework which includes larger members of a same model.

Acknowledgments. 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 simulations and making their model outputs available.

References

- Abramowitz, G., and C. H. Bishop, 2015: Climate model dependence and the ensemble dependence transformation of CMIP projections. *J. Climate*, **28**, 2332-2348, doi: 10.1175/JCLI-D-14-00364.1.
- Aslam, A. Q., S. R. Ahmad, I. Ahmad, et al., 2017: Vulnerability and impact assessment of extreme climatic event: A case study of southern punjab, pakistan. *Science of the Total Environment*, **580**, 468-481, doi: 10.1016/j.scitotenv.2016.11.155.
- Baumberger, C., R. Knutti, and G. H. Hadorn, 2017: Building confidence in climate model projections: An analysis of inferences from fit. *Wires. Clim. Change.*, **8**, doi: 10.1002/wcc.454.
- Brunner, L., R. Lorenz, M. Zumwald, et al., 2019: Quantifying uncertainty in European climate projections using combined performance-independence weighting. *Environ. Res. Lett.*, **14**, doi: 10.1088/1748-9326/ab492f.
- Chen, H. P., J. Q. Sun, and H. X. Li, 2017: Future changes in precipitation extremes over China using the NEX-GDDP high-resolution daily downscaled data-set. *Atmos. Oceanic Sci. Lett.*, **10**, 403-410, doi: 10.1080/16742834.2017.1367625.
- Chen, H. P., J. Q. Sun, and H. X. Li, 2020: Increased population exposure to precipitation extremes under future warmer climates. *Environ. Res. Lett.*, doi: 10.1088/1748-9326/ab751f.
- Chen, W. L., Z. H. Jiang, and L. Li, 2011: Probabilistic projections of climate change over China under the SRES A1B scenario using 28 AOGCMs. *J. Climate*, **24**, 4741-4756, doi: 10.1175/2011JCLI4102.1.
- Eyring, V., S. Bony, G. A. Meehl, et al., 2015: Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, **9**, 1937-1958, doi: 10.5194/gmd-9-1937-2016.
- Guirguis, K., A. Gershunov, D. R. Cayan, et al., 2018: Heat wave probability in the changing climate of the southwest US. *Climate Dyn.*, **50**, 3853-3864, doi: 10.1007/s00382-017-3850-3.
- 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*. Stocker, T. F., D. Qin, G. -K. Plattner, et al., Eds., United Kingdom and New York, NY, USA, Cambridge University Press, 356 pp.
- James, R., R. Washington, C. Schleussner, et al., 2017: Characterizing half - a - degree difference: A review of methods for identifying regional climate responses to global warming targets. *Wires. Clim. Change.*, **8**, e457, doi: 10.1002/wcc.457.
- Jiang, Z. H., W. Li, J. J. Xu, et al., 2015: Extreme precipitation indices over China in CMIP5 models. Part I: Model evaluation. *J. Climate*, **28**, 8603-8619, doi: 10.1175/jcli-d-15-0099.1.
- Jones, B., B. O'Neill, C. Tebaldi, et al., 2014: Population exposure to heat-related extremes: Demographic change vs climate change. *AGU Fall Meeting Abstracts*.

- King, A. D., and D. J. Karoly, 2017: Climate extremes in Europe at 1.5 and 2 degrees of global warming. *Environ. Res. Lett.*, **12**, 114031, doi: 10.1088/1748-9326/aa8e2c.
- Knutti, R., 2010: The end of model democracy? *Climatic Change*, **102**, 395-404, doi: 10.1007/s10584-010-9800-2.
- Knutti, R., D. Masson, and A. Gettelman, 2013: Climate model genealogy: Generation CMIP5 and how we got there. *Geophys. Res. Lett.*, **40**, 1194-1199, doi: 10.1002/grl.50256.
- Knutti, R., R. Furrer, C. Tebaldi, et al., 2010: Challenges in combining projections from multiple climate models. *J. Climate*, **23**, 2739-2758, doi: 10.1175/2009JCLI3361.1.
- Knutti, R., J. Sedláček, B. M. Sanderson, et al., 2017: A climate model projection weighting scheme accounting for performance and interdependence. *Geophys. Res. Lett.*, **44**, 1909-1918, doi: 10.1002/2016gl072012.
- Langenbrunner, B., and J. D. Neelin, 2017: Pareto-Optimal estimates of California precipitation change. *Geophys. Res. Lett.*, **44**, 12,436-412,446, doi: 10.1002/2017gl075226.
- Li, D. H., T. J. Zhou, L. W. Zou, et al., 2018: Extreme high - temperature events over east asia in 1.5° C and 2° C warmer futures: Analysis of NCAR CESM low - warming experiments. *Geophys. Res. Lett.*, **45**, 1541-1550, doi: 10.1002/2017gl076753.
- Li, W., Z. H. Jiang, J. J. Xu, et al., 2016: Extreme precipitation indices over China in CMIP5 models. Part II: Probabilistic projection. *J. Climate*, **29**, 8989-9004, doi: 10.1175/jcli-d-16-0377.1.
- Masson, D., and R. Knutti, 2013: Predictor screening, calibration, and observational constraints in climate model ensembles: An illustration using climate sensitivity. *J. Climate*, **26**, 887-898, doi: 10.1175/JCLI-D-11-00540.1.
- Massoud, E. C., V. Espinoza, B. Guan, et al., 2019: Global climate model ensemble approaches for future projections of atmospheric rivers. *Earth's Future*, **7**, 1136-1151, doi: 10.1029/2019ef001249.
- Palmer, T. N., G. J. Shutts, R. Hagedorn, et al., 2005: Representing model uncertainty in weather and climate prediction. *Annu. Rev. Earth Planet. Sci.*, **33**, 163-193, doi: 10.1146/annurev.earth.33.092203.122552.
- Perkins, S. E., A. J. Pitman, and S. A. Sisson, 2009: Smaller projected increases in 20-year temperature returns over Australia in skill-selected climate models. *Geophys. Res. Lett.*, **36**, doi: 10.1029/2009GL037293.
- Qi, Y. J., C. Qian, and Z. W. Yan, 2017: An alternative multi-model ensemble mean approach for near-term projection. *Int. J. Climatol.*, **37**, 109-122, doi: 10.1002/joc.4690.
- Ren, H. L., Y. J. Wu, Q. Bao, et al., 2019: The China multi-model ensemble prediction system and its application to flood-season prediction in 2018. *J. Meteorol. Res.*, **33**, 540-552, doi: 10.1007/s13351-019-8154-6.

- Sanderson, B. M., R. Knutti, and P. Caldwell, 2015a: Addressing interdependency in a multimodel ensemble by interpolation of model properties. *J. Climate*, **28**, 5150-5170, doi: 10.1175/JCLI-D-14-00361.1.
- Sanderson, B. M., R. Knutti, and P. Caldwell, 2015b: A representative democracy to reduce interdependency in a multimodel ensemble. *J. Climate*, **28**, 5171-5194, doi: 10.1175/JCLI-D-14-00362.1.
- Sanderson, B. M., M. Wehner, and R. Knutti, 2017: Skill and independence weighting for multi-model assessments. *Geosci. Model Dev.*, **10**, 2379-2395, doi: 10.5194/gmd-10-2379-2017.
- Schleussner, C., T. Lissner, E. M. Fischer, et al., 2015: Differential climate impacts for policy-relevant limits to global warming: The case of 1.5 °C and 2 °C. *Earth System Dynamics Discussions*, **7**, 327-351, doi: 10.5194/esd-7-327-2016.
- Semenov, M. A., and P. Stratonovitch, 2010: Use of multi-model ensembles from global climate models for assessment of climate change impacts. *Climate Research*, **41**, 1-14, doi: 10.3354/cr00836.
- Shi, C., Z. H. Jiang, W. L. Chen, et al., 2018: Changes in temperature extremes over China under 1.5 °C and 2 °C global warming targets. *Advances in Climate Change Research*, **9**, 120-129, doi: 10.1016/j.accre.2017.11.003.
- Sillmann, J., V. V. Kharin, X. Zhang, et al., 2013: 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.
- Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485-498, doi: 10.1175/BAMS-D-11-00094.1.
- UNFCCC: United nations framework convention on climate change (2015). Decision 1/cp.21. The Paris agreement. [Available online at <http://unfccc.int/resource/docs/2015/cop21/eng/109r01.pdf>.]
- Wang, B., L. H. Zheng, D. L. Liu, et al., 2018: Using multi-model ensembles of CMIP5 global climate models to reproduce observed monthly rainfall and temperature with machine learning methods in Australia. *Int. J. Climatol.*, **38**, 4891-4902, doi: 10.1002/joc.5705.
- WMO, 2019: Wmo statement on the state of the global climate in 2018. Wmo-no.1233. [Available online at https://library.wmo.int/doc_num.php?explnum_id=5789.]
- Wu, J., and X. J. Gao, 2013: A gridded daily observation dataset over China region and comparison with the other datasets (in chinese). *Chin. J. Geophys.*, **56**, 1102-1111, doi: 10.6038/cjg20130406.
- Wu, J., B. T. Zhou, and Y. Xu, 2015: Response of precipitation and its extremes over China to warming: CMIP5 simulation and projection. *Chin. J. Geophys.*, **58**, 461-473, doi: 10.1002/cjg2.20187.

- Xu, J. Y., Y. Shi, X. J. Gao, et al., 2013: Projected changes in climate extremes over China in the 21st century from a high resolution regional climate model (RegCM3). *Sci. Bull.*, **058**, 1443-1452, doi: 10.1007/s11434-012-5548-6.
- Xu, Y., G. X. Jie, S. Yan, et al., 2009: A daily temperature dataset over China and its application in validating a RCM simulation. *Adv. Atmos. Sci.*, **26**, 763-772, doi: 10.1007/s00376-009-9029-z.
- Zhan, M. J., X. C. Li, H. M. Sun, et al., 2018: Changes in extreme maximum temperature events and population exposure in China under global warming scenarios of 1.5 and 2.0°C: Analysis using the regional climate model cosmo-clm. *J. Meteorol. Res.*, **32**, 99-112, doi: 10.1007/s13351-018-7016-y.
- Zhang, J. J., Z. B. Sun, and S. J. Chen, 1984: An attempt to classify the natural synoptic seasons for each year over east asia by eof (in chinese). *Acta Meteor. Sinica*, **42**, 46-56.
- Zhang, X., H. Wan, F. W. Zwiers, et al., 2013: Attributing intensification of precipitation extremes to human influence. *Geophys. Res. Lett.*, **40**, 5252-5257, doi: 10.1002/grl.51010.
- Zhang, X., L. Alexander, G. C. Hegerl, et al., 2011: Indices for monitoring changes in extremes based on daily temperature and precipitation data. *Wires. Clim. Change.*, **2**, 851-870, doi: 10.1002/wcc.147.
- Zhou, B., Y. Xu, J. Wu, et al., 2016: Changes in temperature and precipitation extreme indices over China: Analysis of a high-resolution grid dataset. *Int. J. Climatol.*, **36**, 1051-1066, doi: 10.1002/joc.4400.
- Zhou, T. J., Y. Q. Yu, H. L. Liu, et al., 2007: Progress in the development and application of climate ocean models and ocean-atmosphere coupled models in China. *Adv. Atmos. Sci.*, **24**, 1109-1120, doi: 10.1007/s00376-007-1109-3.