

Intercomparison of multi-model ensemble-processing strategies within a consistent framework for climate projection in China

Huanhuan Zhu, Zhihong Jiang, Laurent Li, Wei Li, Sheng Jiang, Panyu Zhou, Weihao Zhao

2023

Pacific Climate Impacts Consortium (PCIC)

PCIC Publications

© 2023 Science China Press. This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature's AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: <https://doi.org/10.1007/s00382-017-3634-9>

The final publication is available at:

Zhu, H., Jiang, Z., Li, L., Li, W., Jiang, S., Shou, P., & Zhao, W. (2023). Intercomparison of multi-model ensemble-processing strategies within a consistent framework for climate projection in China. *Sci. China Earth Sci.*, 66, 2125–2141. <https://doi.org/10.1007/s11430-022-1154-7>

Downloaded from UVicSpace Research & Learning Repository

dspace.library.uvic.ca



**University
of Victoria**

Libraries

Intercomparison of multi-model ensemble-processing strategies within a consistent framework for climate projection in China

Huanhuan ZHU¹, Zhihong JIANG^{1*}, Laurent LI², Wei LI¹, Sheng JIANG¹, Panyu ZHOU¹ & Weihao ZHAO³

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 210044, China.

² Laboratoire de Météorologie Dynamique, CNRS, Sorbonne Université, Ecole Normale Supérieure, Ecole Polytechnique, Paris 75005, France.

³ School of Mathematics and Statistics, Nanjing University of Information Science and Technology, Nanjing 210044, China.

*Corresponding author. E-mail: zhjiang@nuist.edu.cn

Abstract

Climate change adaptation and relevant policy-making need reliable projections of future climate. Methods based on multi-model ensemble are generally considered as the most efficient way to achieve the goal. However, their efficiency varies and inter-comparison is a challenging task, as they use a variety of target variables, geographic regions, time periods, or model pools. Here, we construct and use a consistent framework to evaluate the performance of five ensemble-processing methods, i.e., multi-model ensemble mean (MME), rank-based weighting (RANK), reliability ensemble averaging (REA), climate model weighting by independence and performance (ClimWIP), and Bayesian model averaging (BMA). We investigate the annual mean temperature (Tav) and total precipitation (Prcptot) changes (relative to 1995–2014) over China and its seven subregions at 1.5 and 2 °C warming levels (relative to pre-industrial). All ensemble-processing methods perform better than MME, and achieve generally consistent results in terms of median values. But they show different results in terms of inter-model spread, served as a measure of uncertainty, and signal-to-noise ratio (SNR). ClimWIP is the most optimal method with its good performance in simulating current climate and in providing credible future projections. The uncertainty, measured by the range of 10th–90th percentiles, is reduced by about 30% for Tav, and 15% for Prcptot in China, with a certain variation among subregions. Based on ClimWIP, and averaged over whole China under 1.5/2 °C global warming levels, Tav increases by about 1.1/1.8 °C (relative to 1995–2014), while Prcptot increases by about 5.4/11.2%, respectively. Reliability of projections is found dependent on investigated regions and indices. The projection for Tav is credible across all regions, as its SNR is generally larger than 2, while the SNR is lower than 1 for Prcptot over most regions under 1.5 °C warming. The largest warming is found in

northeastern China, with increase of 1.3 (0.6–1.7)/2.0 (1.4–2.6) °C (ensemble's median and range of the 10th–90th percentiles) under 1.5/2 °C warming, followed by northern and northwestern China. The smallest but the most robust warming is in southwestern China, with values exceeding 0.9 (0.6–1.1)/1.5 (1.1–1.7) °C. The most robust projection and largest increase is achieved in northwestern China for Prcptot, with increase of 9.1 (–1.6–24.7)/17.9 (0.5–36.4)% under 1.5/2 °C warming. Followed by northern China, where the increase is 6.0 (–2.6–17.8)/11.8 (2.4–25.1)%, respectively. The precipitation projection is of large uncertainty in southwestern China, even with uncertain sign of variation. For the additional half-degree warming, T_{av} increases more than 0.5 °C throughout China. Almost all regions witness an increase of Prcptot, with the largest increase in northwestern China.

Keywords: Multi-model ensemble simulation; Ensemble-processing strategy; Global warming targets; Climate projection uncertainty assessment; Regional climate change in China

1. Introduction

Global mean surface temperature (GMST) in the first two decades of the 21st century (2001–2020) was 0.99 (0.84 to 1.10) °C higher than 1850–1900 (IPCC, 2021). Global warming exacerbates the threat of climate change to natural ecosystem and human societies. China with its vast territory is sensitive and vulnerable to climate change. Disasters related to meteorological and climatic events cause huge socio-economic losses each year (Blue Book on Climate Change in China (2021), 2021). Scientific projections and reliable information about future climate changes in China are crucial for assessing impacts, identifying risks, and making adaptation policies.

Global climate models (GCMs) are important tools for understanding the climate system, reproducing its past evolution, predicting and projecting its future changes. The Coupled Model Intercomparison Project phase 6 (CMIP6) is the last great effort of the international community for climate modelling. Compared with its previous phases, CMIP6 models have higher spatial resolution and more complete physical and biogeochemical processes (Zhou et al., 2019).

With a variety of numerical formulations and physical parameterization schemes, GCMs provide a range of climate change projections, implying that climate projection is inevitably accompanied by uncertainties (Knutti et al., 2013, Hidalgo and Alfaro, 2015). The multi-model ensemble approach is useful to explore the advantage of the ensemble and to avoid the weakness of individual models, and ultimately to better reduce uncertainty and improve reliability. The arithmetic mean is the simplest and mostly-used method to deal with a multi-model ensemble (Knutti et al., 2010, Zhou et al., 2014; Sanderson et al., 2015; Li et al., 2018; Zhu et al., 2020; IPCC, 2021). However, it has been noted that poorly selected models can degrade the overall skill of climate projections (Murphy et al., 2004). Thus, more sophisticated methods have been developed to address the challenge of combining multi-model ensemble projections (Giorgi and Mearns, 2002, 2003; Xu et al., 2010; Chen et al., 2011; Tan et al., 2016;

Knutti et al., 2017; Li et al., 2021a). Such methods often include the assumption that a model's ability to simulate past climates could be an indicator of its ability to simulate future climates (Smith and Chandler, 2010).

For the Chinese region, numerous attempts based on multi-model ensemble were reported in the literatures. For example, Zhou et al. (2014) using 24 CMIP5 models to project total precipitation changes over China, showed that the rainfall would increase by 8% and 14%, relative to 1986–2005, by the end of the 21st century under the Representative Concentration Pathway 4.5 and 8.5 (RCP4.5 and RCP8.5) scenarios. However, based on an optimal model ensemble applied to CMIP6 models, the increase (relative to 1995–2014) in annual precipitation in China by the end of the 21st century reaches 24% and 45% under the Shared Socioeconomic Pathway 2/4.5 and 5/8.5 (SSP2-4.5 and SSP5-8.5) scenarios (Yang et al., 2021). Li et al. (2021b) used an algorithm jointly considering model performance and independence, and found that the mean increase for heavy precipitation over China is by 17.5% and 26.6% for 1.5 and 2 °C global warming under RCP8.5.

The magnitude of projected climate change in China is generally not consistent among different researches reported in the literature. Their different setups such as target variables, time periods, models and ensemble members included, and future scenarios (Zhou et al., 2014; Li et al., 2016; Li et al., 2021b; Yang et al., 2021), lead to confused situations, and make their inter-comparison almost impossible. In addition, differences of these results may be caused by the projection uncertainties, as they usually include very large ranges. Thus, a reliable projection of future change is much needed. Questions naturally arise regarding the constrained climate projections in China of different ensemble methods and the agreement and uncertainty of these methods, as well as the different response for geographic regions, which represents the motivation of this study.

Here our work mainly develops a common experimental setup, with a common set of variables, spatial scales, time period, and model pools to provide a consistent and rigorous test-bed as far as possible. This common framework allows us to make an objective and fair evaluation of agreement and uncertainty across the five ensemble-processing methods (i.e., multi-model ensemble mean, rank-based weighting method, reliability ensemble averaging, climate model weighting by independence and performance, and Bayesian model averaging) included in this paper. We focus on the recommended targets of 1.5 and 2 °C global warming levels, following the Paris Agreement (UNFCCC 2015). We also aim to achieve a more reliable projection for climate change in China and its subregions and to provide useful information for assessing impacts, identifying risks, and making adaptation decisions.

2. Data and methods

2.1 Observations and model datasets

As a reference from observation, we use the high-resolution ($0.25^\circ \times 0.25^\circ$) daily gridded dataset, CN05.1, providing daily temperature and precipitation. CN05.1 was provided by the National Climate Center of the China Meteorological Administration,

and elaborated with data from 2416 observation stations across China (Wu and Gao, 2013). It covers the period from 1961 to the most recent day. For convenience, an interpolation of the high resolution CN05.1 to the 1-degree grid was applied to better match models. Simulations from 25 CMIP6 models (Table S1) are used in this study as primary information to elaborate different ensemble-processing schemes. For each model, the historical simulation and future scenario under high-emission pathway (SSP5-8.5) are used. SSP5-8.5 represents the combined scenario of a fossil-fueled growth pathway (i.e., SSP5) with strong radiative forcing which peaks at 8.5 W m^{-2} by 2100 (O'Neill et al., 2016). Models cover the period 1850–2014 for historical and 2015–2100 for future period. Only the first realization member was analyzed for each model to allow equity and fairness among models.

Two climate indices are used in this work, i.e., the annual average temperature (Tav) and the annual total amount of precipitation (Prcptot). Climate indices from different models and observations were firstly calculated on their native grids. To facilitate the intercomparison, a bilinear interpolation scheme was used to interpolate all indices onto a common $1^\circ \times 1^\circ$ grid.

2.2 Multi-model ensemble-processing strategy

Five ensemble-processing methods are involved in this study. The first one is the simple multi-model ensemble mean (MME). It is the easiest to realize and based on the intuitive idea that errors among models can be simply cancelled out when they are averaged together. Other four schemes are more-elaborated ones, with more or less physically-based criteria and different degrees of objectivity. They are the rank-based weighting method (RANK; Chen et al., 2011), reliability ensemble averaging (REA; Giorgi and Mearns, 2002, 2003; Zhu et al., 2022), climate model weighting by independence and performance (ClimWIP; Knutti et al., 2017; Li et al., 2021b), and Bayesian model averaging (BMA; Raftery et al., 2005; Tan et al., 2016), respectively.

2.2.1 Rank-based weighting method

The rank-based weighting method uses two evaluation metrics including the spatial pattern and interannual variability to quantify model capabilities. They are root-mean-square error (RMSE) and interannual variability score (IVS). RMSE is defined as:

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

where M_k and O_k denote the model pattern of a climate variable under investigation and the observed counterpart, respectively; P indicates the number of spatial points. Smaller RMSE implies a better agreement between the model and observation.

IVS is defined as:

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

where STD_m and STD_o denote the interannual standard deviation of model and

observation, respectively. IVS is firstly calculated at each grid point and then regionally averaged. Smaller IVS indicates better performance of the simulation.

According to Chen et al. (2011), the performance indicator R_i of an individual model i can be obtained as a function of its ranking position among others, as described by:

$$R_i = \frac{\sum_{i=1}^N S_i}{S_i} \quad (3)$$

where S_i is the rank of model i , which is a simple sum of its ranks for the two metrics; N indicates the number of models. Higher R_i indicates better model skill. The weight of model i (W_i) is then its normalized R_i , the sum of weights for all models being equal to 1.

2.2.2 Reliability ensemble averaging

We upgraded the original reliability ensemble averaging method (Zhu et al., 2022), and the reliability factor R_i of our REA methods can be expressed as:

$$R_i = [f_1(\mathbf{T}_{\text{bias}})]^{m1} \times [f_2(\mathbf{T}_{\text{abs, var}})]^{m2} \times [f_3(\mathbf{T}_{\text{rel, var}})]^{m3} \\ = \left\{ \frac{\varepsilon_{\text{T}}}{\text{abs}(\mathbf{T}_m - \mathbf{T}_o)} \right\}^{m1} \times \left\{ \frac{\varepsilon_{\text{STD}}}{\text{abs}(\text{STD}_m - \text{STD}_o)} \right\}^{m2} \times \left\{ \frac{1.0}{1.0 + \left(\frac{\text{STD}_m}{\text{STD}_o} - \frac{\text{STD}_o}{\text{STD}_m} \right)^2} \right\}^{m3} \quad (4)$$

where the function f_1 measures the model performance in reproducing the observed climatological state (geographic pattern). The functions f_2 and f_3 measure the model performance in reproducing the observed temporal (inter-annual) variability in terms of absolute and relative deviation, respectively. The latter is a ‘‘symmetric’’ statistic, ensuring that two models would have the same performance measure if their model-over-observation ratio and observation-over-model ratio are equal (Jiang et al., 2015). The parameter ε is a measure of the natural variability, computed as the difference between maximum and minimum values of 20-year moving averages of the observed variables. The inter-annual variability of climate variables or indices is represented by the temporal standard deviation (STD). The value of the function f_1/f_2 is set to 1 when $\text{abs}(\mathbf{T}_m - \mathbf{T}_o)/\text{abs}(\text{STD}_m - \text{STD}_o)$ is smaller than $\varepsilon_{\text{T}}/\varepsilon_{\text{STD}}$. In other words, a climate model is considered to be reliable if its bias is within the natural variability. The parameters $m1$, $m2$ and $m3$ in Eq. 4 can be used to adjust each criterion. We set them to 1 in this research. Finally, the weight of model i (W_i) is normalized by its R_i .

2.2.3 Climate model weighting by independence and performance

Knutti et al. (2017) proposed a weighting scheme which considers both performance and independence of models. The single model weight (W_i) for model i is defined as:

$$W_i = e^{-\frac{D_i^2}{\sigma_D^2}} / \left(1 + \sum_{j \neq i}^N e^{-\frac{S_{ij}^2}{\sigma_S^2}} \right) \quad (5)$$

where D_i is the distance between model i and the observation, and S_{ij} is the distance

between model i and model j , N is the total number of models. The distances D_i and S_{ij} is deduced from the average of two normalized measures, RMSE and IVS, as defined in section 2.2.1. σ_D and σ_S are skill and uniqueness parameters, respectively, which can be determined through a cross-validation procedure. A small σ_D puts important weights on only a few models, whereas a large σ_D would effectively make the ensemble converge to the situation of equity of all models (full democracy!). σ_S determines the typical distance by which a model would be considered similar to another. More details of the scheme can be found in Li et al. (2021b).

2.2.4 Bayesian model averaging

Bayesian model averaging (BMA) is a method that updates a prior distribution in light of new information provided by observations. The output of BMA is a weighted average of probability density functions (PDF) which are centred on the bias-corrected forecast. The BMA weights reflect the relative contributions of the component models to the prediction over the training samples and can be used to assess the usefulness of each ensemble member. In the case of a variable y to be forecasted on the basis of training data y_{obs} using N models M_1, \dots, M_N , the forecast PDF $p(y)$ can be given by

$$p(y) = \sum_{i=1}^N p(y | M_i) p(M_i | y_{\text{obs}}) \quad (6)$$

where $p(y|M_i)$ is the forecast PDF based on model i alone; $p(M_i|y_{\text{obs}})$ is the posterior probability of model i being correct given the training data and indicates how well the model i represents the training data. The total sum of the posterior model probabilities is 1, that is $\sum_{i=1}^N p(M_i | y_{\text{obs}}) = 1$, and they can thus be treated as the weights of models.

More detailed information on BMA can be referred to Tan et al. (2016). What calls for special attention is that the construction of BMA is based on the monthly scale in order to increase the sample size and make the model converge better, rather than the yearly scale (as in other methods). In other words, BMA further uses the metric considering the seasonal cycle of geophysical variables or indices to quantify model capabilities. To facilitate comparison with other methods, the results are projected to the annual scale (the annual average temperature and the annual total amount of precipitation).

2.3 Time windows for 1.5 and 2 °C global warming thresholds

The specific warming targets in this study are relative to the pre-industrial level, which is defined as 1861–1900 as in previous researches (Shi et al., 2018; Zhu et al., 2021). In order to determine a relatively stable climatology, a 21-year moving average is firstly employed to smooth the time series of GMST. Specific thresholds are then defined as the first year when GMST reaches 1.5/2 °C above their pre-industrial level for individual models. A time window of 20 years, comprising 9 years before and 10 years after the nominative year, is used to deduce the climate statistics. The difference of climate indices between the specific warming level and the reference period (1995–

2014) is used to assess future changes of climate indices. It is to be noted that the observed global temperature in the reference period is about 0.8 °C warmer than the pre-industrial level (IPCC 2021), the 1.5 and 2 °C global warming targets (which are relative to pre-industrial, 1861–1900) imply a global warming level of 0.7 and 1.2 °C relative to the reference period (1995–2014), respectively. This note applies to the whole paper when the 1.5 and 2 °C warming targets are invoked.

Table S2 shows the threshold crossing times (TCTs) of 1.5 and 2 °C global warming. It can be seen that there are large discrepancies in the TCTs among individual models. Some models (e.g., CanESM5, EC-Earth3-Veg and IPSL-CM6A-LR) cross the 1.5 °C warming threshold as early as 2010s whereas some other models (e.g., GFDL-ESM4, MIROC6 and NorESM2-LM) do not reach the mark until 2040s. This disagreement may be attributed to the mismatch in simulated and observed variability at decadal timescales (Schmidt et al., 2014), as well as differences in model climate sensitivity (Hu et al., 2017; Zhu et al., 2021). Models with higher transient climate response may reach the warming thresholds earlier than those with lower transient climate response (Hu et al., 2017).

2.4 Strategies in processing multi-model ensemble

Historical simulations, along with observation, are divided into calibration period 1961–1994 (34 years) and validation period 1995–2014 (20 years). The validation period also serves as the reference for future projection. Once the calibration and validation accomplished, the five multi-model ensemble-processing methods are subsequently applied to future climate projection under 1.5 and 2 °C global warming targets.

To explore the changing characteristics of climate indices in detail, we divided China into seven subregions, namely, northwestern China (NW, 36°–50°N, 74°–106°E), the Tibetan Plateau (TB, 28°–36°N, 78°–106°E), southwestern China (SW, 20°–28°N, 95°–106°E), northeastern China (NE, 42°–55°N, 106°–134°E), northern China (NC, 35°–42°N, 106°–126°E), Yangtze river basin (YR, 28°–35°N, 106°–123°E) and southeastern China (SE, 18°–28°N, 106°–120°E) (Figure 1).

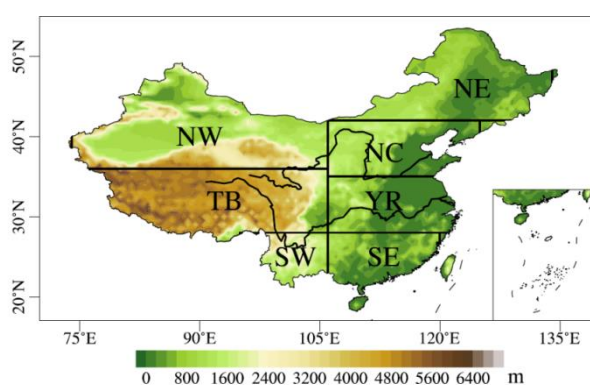


Figure 1 Domains and topography (shading, units: m) of seven subregions in China.

2.5 Skill evaluation metrics

Taylor diagram and Taylor skill scores (TSS) (Taylor, 2001) are used to evaluate the overall skill in reproducing the spatial pattern of climate indices during the reference period. Taylor diagram provides a concise statistic summary of the degree of correlation (PCC; pattern correlation coefficient), centered root mean square error (RMSE), and the ratio of spatial standard deviation (RSD). A perfect simulation would be that with the centered RMSE equal to 0 and both PCC and RSD close to 1. The TSS is a combined measure and calculated as:

$$\text{TSS} = \frac{4(1 + R_m)^2}{\left(\frac{\sigma_o}{\sigma_m} + \frac{\sigma_m}{\sigma_o}\right)^2 (1 + R_0)^2} \quad (7)$$

where R_m is the spatial correlation coefficient between the simulation and observation, R_0 is the maximum correlation coefficient attainable (here we use 0.999). σ_m and σ_o 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. This skill score has been widely used in many previous researches (Zhu et al., 2020; Li et al., 2021a).

As practiced in Li et al (2016), the signal-to-noise ratio (SNR) is used to quantitatively measure the robustness of the estimated results. The SNR in the ensemble methods can be calculated by the following equation:

$$S = \left| \sum_i W_i \Delta T_i \right| \quad (8)$$

$$N = \sqrt{\sum_i W_i (\Delta T_i - \sum_i W_i \Delta T_i)^2} \quad (9)$$

$$\text{SNR} = \frac{S}{N} \quad (10)$$

where ΔT_i denotes the i_{th} individual model simulation, and W_i is its weight. The signal (S) is the absolute value of the weighted projection. The noise (N) is the spreading among models, and is measured by the standard deviation. $\text{SNR} > 1$ implies the signal is greater than the noise, hence the projection is credible or reasonable to a certain extent. The larger the SNR is, the greater the confidence level becomes.

3. Results

3.1 Comparison of ensemble-processing methods

In this section we compare the five methods and determine the best one by three aspects, evaluation of historical simulation, future climate projection uncertainty and robustness of results. The model spreading (based on the 25–75th and 10–90th percentiles) and the inter-model standard deviation (SD) are used to assess the uncertainty (inter-model spread) of climate changes. The signal-to-noise ratio (SNR) is used to quantitatively measure the robustness of future projection (Hawkins and Sutton, 2011; Li et al., 2016; IPCC, 2021).

3.1.1 Simulation evaluation

Taylor diagrams and TSS are used to evaluate the overall skills of multi-model ensemble methods in reproducing the spatial pattern of climate indices during the reference period (1995–2014) (Figure 2). The performance with MME (gray markers) is generally low. The PCC of annual mean temperature (Tav) deduced from MME is 0.97, the centered RMSE is 0.25, the RSD is 1.05, and the TSS is 0.97. Total precipitation (Prcptot) from MME shows a mediocre performance, with PCC and TSS less than 0.86, the centered RMSE larger than 0.6. All other ensemble-processing methods show improvements compared to MME, with higher PCCs and TSSs, as well as smaller centered RMSEs. BMA shows the best performance, followed by ClimWIP. The best-performing BMA gives PCCs and TSSs superior to 0.99 for Tav and Prcptot. PCCs and TSSs for climate indices of ClimWIP are above 0.93. Overall, the results show that weighting schemes effectively enhance the capability of reproducing the spatial climate characteristics in China, and BMA has the best performance, with TSSs of Tav and Prcptot higher than 0.99, followed by ClimWIP, with their TSSs higher than 0.93.

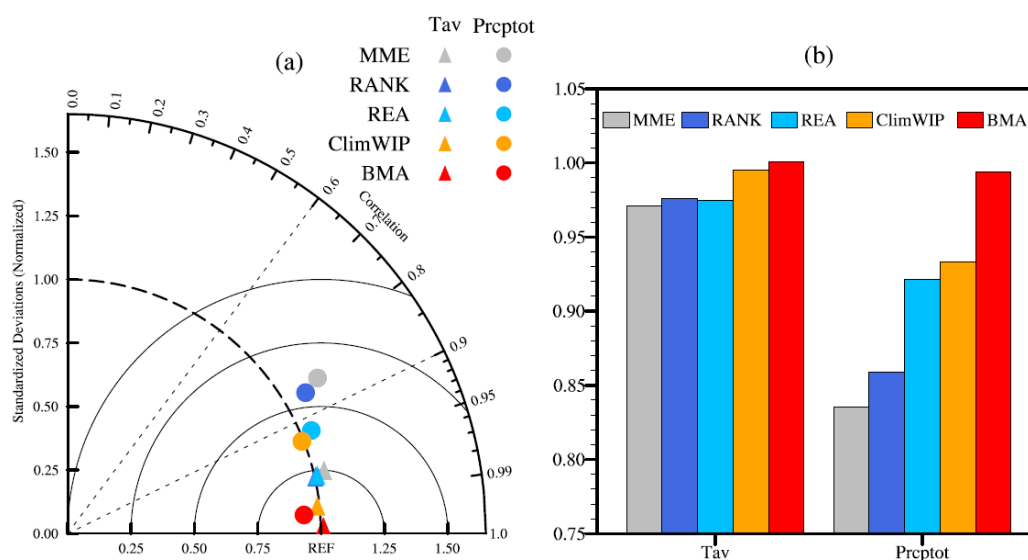


Figure 2 Taylor diagram and Taylor skill scores of climate indices (annual surface air temperature Tav and annual total amount of precipitation Prcptot) in China during 1995–2014.

3.1.2 Projection uncertainty and robustness

Climate projection is inevitably accompanied by differences due to diverse model pools and ensemble methods, but the reliability of the projection can be evaluated by the level of consistency of the estimates for future climate change. Changes in areal-mean temperature and relative precipitation over China between the reference (1995–2014) and future mean states based on different methods are shown in this section. The projection of future climate for 1.5 and 2 °C global warming targets (relative to preindustrial) under the SSP5-8.5 emission scenario are analyzed. The median changes and model spreading (based on the 25–75th and 10–90th percentiles) for each method are given.

The projected changes in annual mean temperature over China under the 1.5 and 2 °C warming targets are shown in Figure 3 and Figure S1. A notable feature is that the temperature is projected to increase across China, and the median estimates agree better among methods. When regionally averaged over China, relative to 1995–2014, the five methods project an increase between 1.1/1.8 and 1.2/1.9 °C under 1.5/2 °C warming, respectively. And the full range of the 10th and 90th percentiles for different methods are 0.6–1.5/1.3–2.3 °C (ClimWIP), 0.5–1.6/1.3–2.5 °C (BMA), 0.5–1.7/1.2–2.5 °C (RANK), 0.5–1.6/1.3–2.5 °C (REA), and 0.4–1.7/1.1–2.6 °C (MME), respectively. There is a reduction in the 10th–90th percentile ranges from about 10% (RANK) to 30% (ClimWIP), compared to MME. The greatest reduction is found in ClimWIP, exceeding 30%. ClimWIP has the smallest standard deviation (SD), only 0.37/0.41 under 1.5/2 °C warming. However, compared with MME, most other methods show slightly weaker warming in a few regions (e.g., NW and TB). In TB, MME has an increase of 1.9 °C, while other constrained methods show a range from 1.7 °C (ClimWIP) to 1.8 °C (REA) under 2 °C global warming (Figure S1). Compared with MME, the spread (25th–75th and 10th–90th percentile ranges) of the other four ensemble-processing methods is generally reduced in most subregions, as well as the regionally averaged standard deviation. The largest reduction is found in ClimWIP, particularly in SW and SE, exceeding 50%. It indicates that ClimWIP performs best in the reduction of uncertainty (inter-model spread).

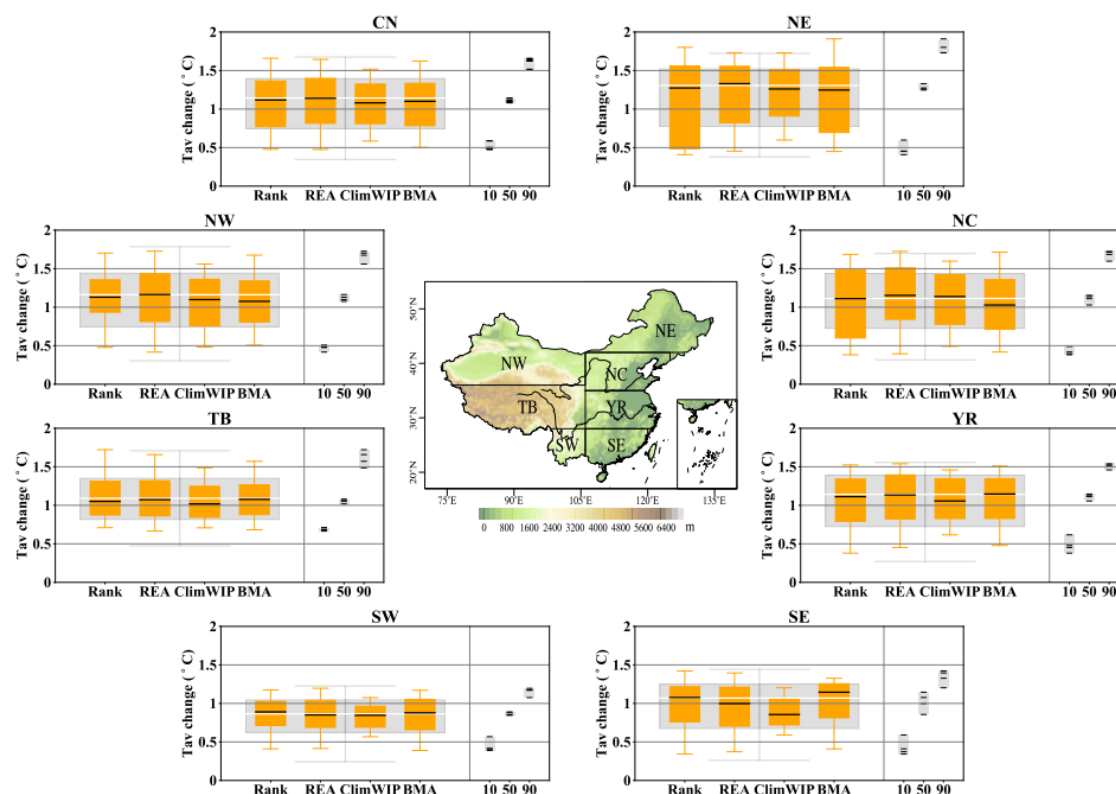


Figure 3 Projected changes (relative to 1995–2014) in Tav over China and the seven subregions under 1.5 °C global warming under SSP5-8.5 scenario. The lighter boxes give the unconstrained distributions, the orange boxes give the constrained distributions. Boxes indicate the

interquartile model spread (25th and 75th quantiles) with the horizontal line indicating the ensemble median and the whiskers showing the 10th and 90th percentile values. A synthesis of all constrained multi-model distributions is shown on the rightmost side. The bars represent the 10th, 50th, and 90th percentiles of the methods, and the shading indicates the full spread.

To quantitatively evaluate the uncertainty (inter-modal deviation) and robustness of the temperature change from different methods. Figure 4 and Figure S2 show the regionally averaged standard deviation (SD) and signal-to-noise ratio (SNR) of China and its seven subregions based on different methods under 1.5 and 2 °C warming targets. Results show that the SNR (SD) from the four more elaborated ensemble-processing methods is generally larger (smaller) than that from simple MME, indicating that the other four ensemble-processing methods are more credible, compared to the simple method of MME. The largest SNR is found in ClimWIP for Tav across China, with the SNR of CN up to 3.32 under 1.5 °C warming, and 5.02 under 2 °C warming. It indicates that the projection of Tav in ClimWIP is the most credible.

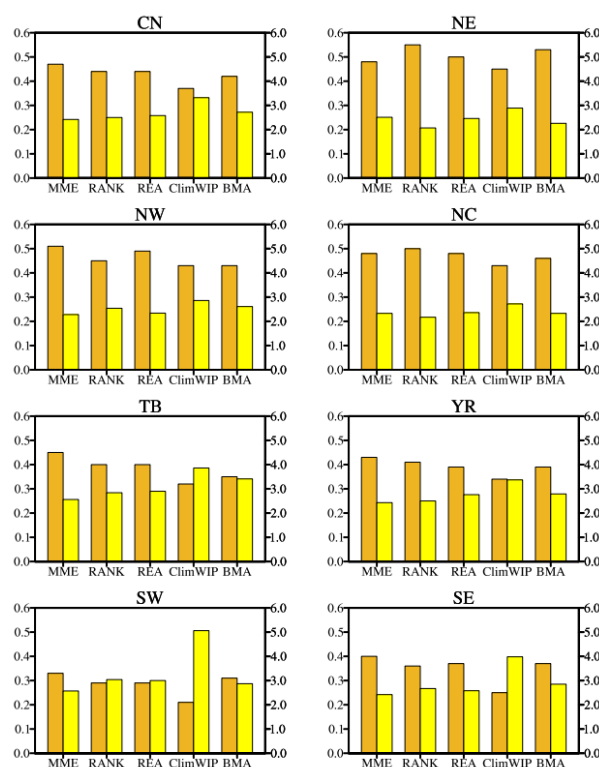


Figure 4 The areal-mean standard deviation (SD, golden) and signal-to-noise ratio (SNR, yellow) of Tav over China and its seven subregions by multi-model ensemble methods under 1.5 °C warming target. The scale of SD is on the left y-axis, and that of SNR is on the right y-axis.

The projected changes in total precipitation over China and its subregions under the 1.5 and 2 °C warming targets are shown in Figure 5 and Figure S3. Precipitation is projected to increase in China with a larger magnitude for higher warming conditions, as well as the inter-model spread. The median changes agree better among the five

ensemble-processing methods. When regionally averaged over China, relative to 1995–2014, Preptot increases from 4.5/8.5 to 5.4/11.2% under 1.5/2 °C warming, respectively. The full range of the 10th and 90th percentiles for different methods are –1.8–13.0/1.3–20.6% (BMA), –1.8–15.4/0.8–24.3% (ClimWIP), –2.7–16.2/–1.1–24.8% (RANK), –2.7–16.8/–0.4–25.4% (REA), and –5.0–15.8/–3.5–22.6% (MME), respectively. There is a reduction in the 10th–90th percentile ranges for BMA (28%) and ClimWIP (15%), in comparison to MME, but with comparable ranges for RANK and REA. The greatest reduction is found in BMA, by about 28%. BMA has the smallest SD, with about 5.8/7.7 under 1.5/2 °C warming, while that is 7.6/9.8 in MME. Furthermore, compared with MME, most methods show much heavier precipitation increases in several regions (e.g., NW and TB), especially in ClimWIP. For example, when regionally averaged over NW, ClimWIP projects an increase of 9.1% and 17.9% under 1.5 and 2 °C global warming, while the increase in MME is only 7.1% and 11.6%. In TB, the projected changes in ClimWIP for the total precipitation (exceeding 12.4% under the 2 °C warming) are almost twice higher than in MME (6.9%). A reduction in the spread for Preptot is found for all methods (compared to MME) in most subregions, except in NW and TB. The upper (90th) percentile are not constrained by most methods over these regions. A remarkable reduction is found in BMA, particularly in NW, exceeding 45% under 1.5 and 2 °C global warming. The reduction, in ClimWIP, in NE and SW is also obvious, up to 35%. The areal-mean SDs also can reproduce the above features. The SDs from ClimWIP and BMA are smaller than those from other methods, which indicates that ClimWIP and BMA perform better in the reduction of uncertainty (inter-model spread) for Preptot.

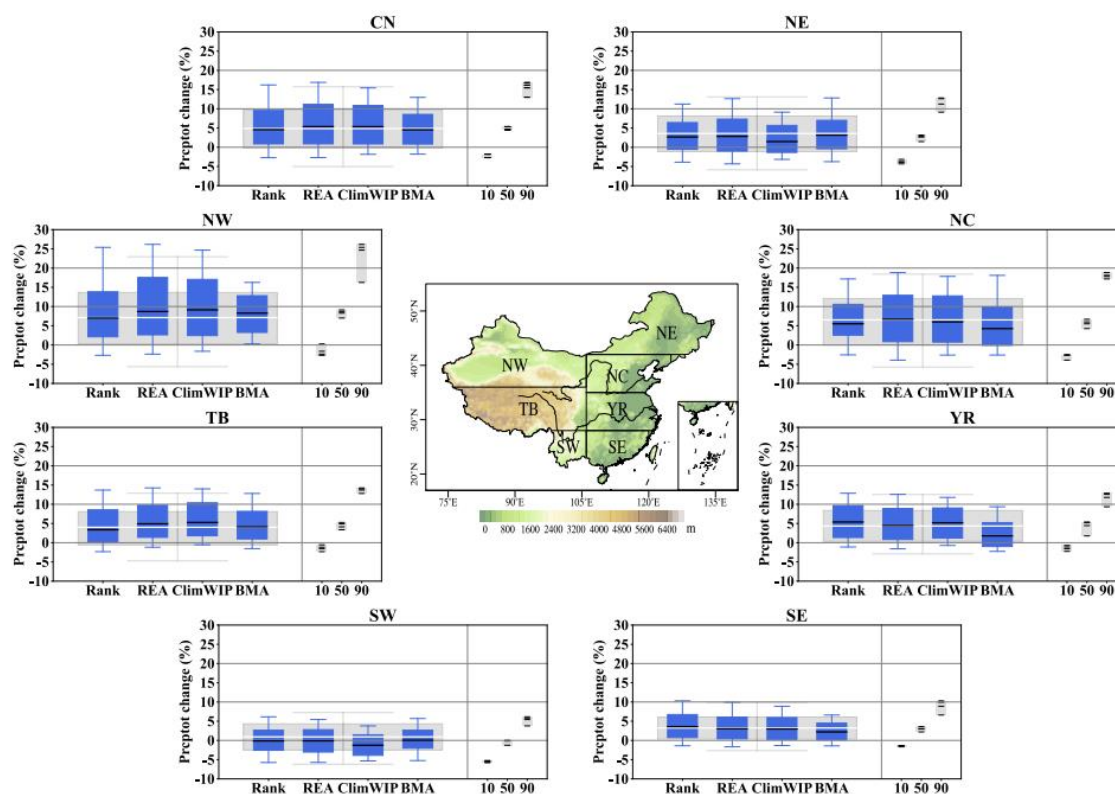


Figure 5 Same as in **Figure 3** but for Prcptot.

The SDs and SNRs for Prcptot by the ensemble methods under 1.5 and 2 °C warming targets are shown in Figure 6 and Figure S4. Similar to Tav, changes for Prcptot from the four more-elaborated ensemble-processing methods are generally more credible than those from MME, with larger SNRs and smaller SDs. The largest (smallest) SNR (SD) is in BMA for CN, up to 1.02 (5.8) under 1.5 °C warming, and 1.49 (7.73) under 2 °C warming. BMA also has the largest SNR in NW. However, in other subregions, the largest SNR is generally found in ClimWIP. It is clear that the projection of Prcptot in BMA and ClimWIP are more credible than in other methods.

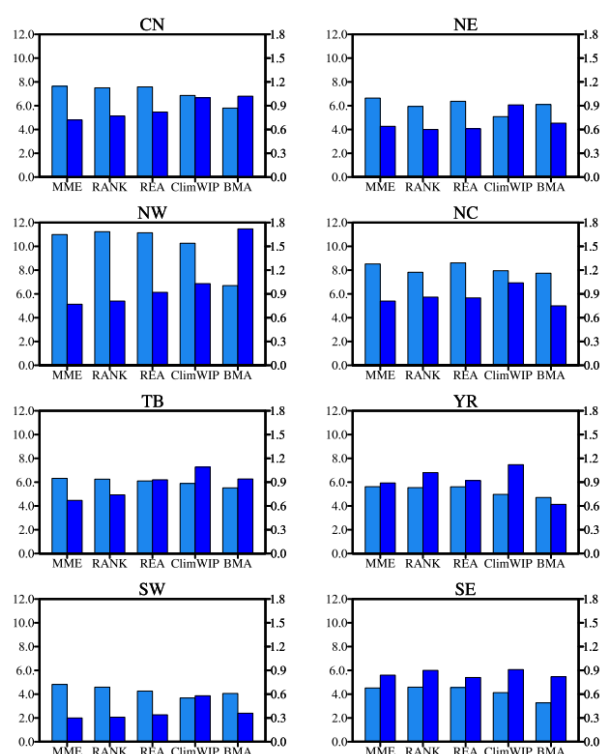


Figure 6 The areal-mean standard deviation (SD, light blue) and signal-to-noise ratio (SNR, blue) of Prcptot over China and its seven subregions by multi-model ensemble methods under 1.5 °C warming target. The scale of SD is on the left y-axis, and that of SNR is on the right y-axis.

The ensemble-processing methods provide reasonably robust median temperature and precipitation projections over China. Compared with MME, other four methods show less warming and enhanced wet conditions. They also reduce the uncertainty (inter-model spread) and increase the robustness of projections, with smaller SDs and larger SNRs. Further assessment, through a balanced consideration of both temperature and precipitation, allows to reach the conclusion that ClimWIP is the most optimal method to present credible projections in China. It shows good performance in simulating current climate, and has a reduction of about 30% for Tav, and 15% for Prcptot (relative to MME) for the uncertainty (10th–90th percentile ranges) of

projection. In addition, ClimWIP provides the most robust projection for Tav across China, and for Preptot in most subregions, as it has the largest SNRs over there.

In summary, RANK and REA show similar skills in presenting credible temperature and precipitation projections in China, but with inferior performance compared with ClimWIP and BMA. Let us now examine the underlying assumptions and their implementation in each ensemble-processing strategy to search a physical explanation for the obtained results. Actually, ClimWIP seems more judicious for both its concept and implementation, since its weighting scheme takes into account, simultaneously, model performance and independence. The consideration of model's independence in ClimWIP mainly consists of reducing the weights for models with high interdependence. While Rank and REA process their weighting scheme by only considering model performance, though they share the same performance metrics (RMSE and IVS), with assessment on mean state and interannual variability. BMA updates a prior probability distribution in light of new information provided by observations, and considers the seasonal cycle of indices to quantify model capabilities.

The final weights obtained for the five weighting schemes over China, sorted by descending order of each method are shown in Figure 7. For both Tav and Preptot, the distribution of weights among models is generally similar for the four ensemble-processing methods. Most weight is given to a small number of models, especially in RANK and ClimWIP. For the latter, almost no weight is given to those models in the tails in Figure 7. The final results are basically dominated by those models having larger weights. A small number of effective models in ensemble-processing methods contribute to the decrease of uncertainty (spread). This feature is even more pronounced at regional scales (figure not shown). Excessive concentration of weights might lead to overfitting because the ensemble subset size is small. But this issue does not seem to arise in our work because the number of effective models is still large enough.

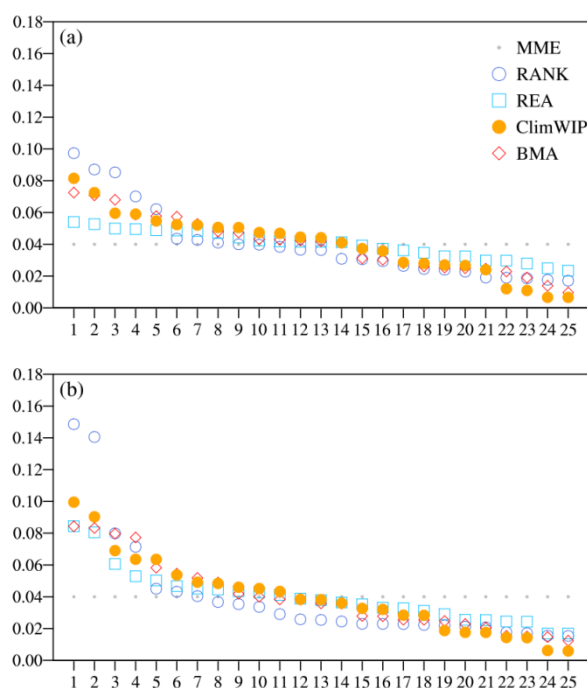


Figure 7 The models' weights obtained by the five ensemble-processing methods for (a) Tav and (b) Prcptot over China, sorted by descending order of each method.

3.2 Difference of projection reliability among regions

China has a vast territory with complex terrain and diverse climate conditions (Guo et al., 2020). There exist large differences for future projections of climate change and related uncertainties over subregions in China. We divided China into seven subregions according to their administrative boundaries, societal and geographical conditions. This section further analyses the projection differences among the seven subregions. Results show that ensemble-processing methods agreement is generally good for median temperature changes in most subregions, with a range less than 0.1 °C, except in SE. The increase for ClimWIP and REA is smaller than 1 °C, and that for BMA and RANK is larger than 1 °C in SE under 1.5 °C warming (Figure 3). Larger warming occurs at higher latitudes of China such as in NE, NW, and NC, which is consistent with previous studies (Shi et al., 2018; Wu et al., 2020; Zhu et al., 2021). The largest warming is found in NE with the median values exceeding 1.3/2.0 °C under the global warming of 1.5/2 °C. Followed by NC and NW, where the median increase is about 1.1/1.9 °C. The smallest warming is in SW with values exceeding only 0.9/1.5 °C. The warming is comparable across other regions (e.g., TB, YR), with an increase of approximately 1.0/1.8 °C under the global warming of 1.5/2 °C.

Regions with stronger warming have larger uncertainty, and smaller SNRs (Figure 4). NE has the largest spread among the seven regions, and the full range of absolute values of the 10th–90th percentiles can exceed 1.5 °C (0.4 to 1.9 °C) under 1.5 °C warming, and 1.5 °C (1.2 to 2.7 °C) under 2 °C warming. Even in ClimWIP (the best method), the values in NE range from 0.6 to 1.7 °C under 1.5 °C warming, and from 1.4 to 2.6 °C under 2 °C warming. The spread is also large in NC, ranging from 0.4 to 1.7 °C under 1.5 °C warming, and from 1.1 to 2.6 °C under 2 °C warming. This partly reflects the large projection uncertainty in north China. The largest SD is in NE (0.55), and the smallest SD is in SW (0.21) under 1.5 °C warming. The SNR is contrary to the SD. The largest SNR is in SW (5.06), and the smallest is in NE (2.07). The S is large in NE, but the N (spread among models) is much larger, which leads to smaller SNR over there. It is clear that temperature changes in SW have the highest robustness, but lower robustness in NE.

The ensemble methods also provide reasonably robust median precipitation projections over most subregions (Figure 5 and Figure S3). For example, in SE, they have an increase from 2.2/4.5% to 3.6/5.9% under 1.5/2 °C global warming. In NE, the increase ranges from 1.5/7.6% to 3.1/9.1% under 1.5/2 °C global warming. The full range is smaller than 2% over these places. But they agree less on a few regions, for instance, in YR (ranging from 1.8% to 5.4%) under 1.5 °C warming. The subregion in Eastern China is strongly influenced by the East Asian monsoon (Guo et al., 2020). Changes in the monsoon circulation could affect the precipitation (Chen et al., 2020). There are still many disagreements about future East Asian monsoon changes due to different choices of monsoon indices and uncertainties among models (You et al., 2022),

which might lead to the uncertainty in precipitation projections over there. In SW, the median change under 1.5 °C warming is approximately zero, except for ClimWIP, which points toward a decrease (−1.3%). The methods disagree on the sign of the change in SW under 1.5 °C warming. However, all methods agree on a median projection that points to a weak increase in rainfall under 2 °C warming. The signal of regional climate changes will amplify as the higher levels of global warming (IPCC, 2021). The precipitation in SW is largely influenced by the South Asian summer monsoon (SAM). The weakening of the SAM circulation could lead to the decrease in precipitation (Wu et al., 2019). However, a warmer Indian Ocean would provide increased source of moisture transport, leading to a larger moisture flux convergence and increasing precipitation there (Sharmila et al., 2015). Additionally, SW has complex geomorphic and topographic characteristics, which could lead to prominent regional climate. The Tibetan Plateau could significantly impact water vapor flux and atmospheric circulation via land surface processes, and consequently SW precipitation (Sun et al., 2017). All above could lead to the uncertain precipitation changes in SW.

The largest precipitation enhancement is found in NW, with the median increase up to 9.1/17.9% based on ClimWIP under the global warming of 1.5/2 °C. Followed by TB and NC, where the median increase is 12.4% and 11.8% under the global warming of 2 °C. The projected relatively precipitation increase is large in western China, due to the drier regional climate, which is consistent with the previous projections from CMIP5 (Zhou et al., 2014) and CMIP6 (Yang et al., 2021; Zhu et al., 2021). The increase of moisture convergence (Zhang et al., 2019), which mainly comes from the thermodynamic effect due to increasing atmospheric moisture under warming, and the enhanced transport of water vapor caused by the strengthening of the westerly wind circulation (Wu et al., 2019, Zhang et al., 2021) could contribute to the large increase of precipitation in western China. In addition, NW has the largest spread among the seven subregions, with the full range (the 10th–90th percentiles) exceeding 29% (−2.7–26.1%) under 1.5 °C warming, and more than 40% (−2.3–38.1%) under 2 °C warming among the methods. The spread in NC (from −3.9/−0.3 to 18.8/26.3% under 1.5/2 °C warming) is also large, which shows the notable projection uncertainty (inter-model spread) over these places.

There are also significant regional differences in the reliability of precipitation projection. The largest SD is in NW (11.23), followed by NC (7.95) under 1.5 °C warming. But for SNR, the largest value is in NW (1.72), and the smallest value is in SW (0.31) under 1.5 °C warming. It indicates that the projections for Prcptot in NW (SW) have the highest (lowest) robustness. The influence of S (projected change) on SNR is higher than that of N (inter-model spread) for Prcptot. The increases of precipitation will amplify at higher levels of global warming (IPCC, 2021). The SNRs of Prcptot would also increase with the warming targets. For example, in NW, the SNR for Prcptot under 1.5 °C warming is 1.72, while that under 2 °C warming is 2.14 in BMA. The SNRs are larger than 1 for most subregions (except SW) under 2 °C warming, but smaller than 1 under 1.5 °C warming. In addition, changes for prcptot have a lower confidence level than for Tav. SNR for Prcptot is smaller than that for Tav among all

methods and regions. The projections are credible ($\text{SNR}>2$) for Tav over all regions, while only in a few regions (e.g. NW in BMA) for Prcptot under 2 °C warming. Even SNR is lower than 1 for Prcptot over most regions under 1.5 °C warming.

Overall, the ensemble methods provide reasonably robust median temperature and precipitation projections over most subregions, but they agree less on a few regions (e.g., SE for Tav, SW and YR for Prcptot). The most robust projection is in SW for Tav, with the largest SNR, though the warming is weak. It would increase by 0.9 (0.6–1.1)/1.5 (1.1–1.7) °C under 1.5/2 °C warming based on ClimWIP. The highest warming is noted in NE, combined with the largest projection spread, with increase of 1.3 (0.6–1.7)/2.0 (1.4–2.6) °C under 1.5/2 °C warming, followed by NC, with increase of 1.1 (0.5–1.6)/1.9 (1.3–2.4) °C. For Prcptot, the largest increase and robustness occur in NW, it would increase up to 9.1 (–1.6–24.7)/17.9 (0.5–36.4)% under the global warming of 1.5/2 °C, based on ClimWIP. There is also significant and robust increase for precipitation in NC with values by 6.0 (–2.6–17.8)/ 11.8 (2.4–25.1)% under 1.5/2 °C warming. The smallest projection reliability is found in SW, its SNR is lower than 0.6 under warming climates. Actually, different methods disagree from each other, even on the sign of the change under 1.5 °C warming.

The increase in precipitation under warming is primarily attributed to the increase of moisture convergence, which comes mainly from the thermodynamic effect due to increasing atmospheric moisture (Zhang et al., 2019; Chen et al., 2020). The dynamic effect, such as monsoon circulation changes, dominates regional differences in the projected precipitation changes (Chen et al., 2020). Besides global warming, many other factors may also contribute to the precise changes of climate at the regional scale, including local feedbacks (Lorenz et al., 2016) and local-scale external forcings, such as aerosol concentrations (Dong et al., 2016; Zhou et al., 2020), land use and land cover changes (Findell et al., 2017). Moreover, internal variability plays an important role in the projection of climate changes at regional scales (Thompson et al., 2015; Li et al., 2022). A better understanding of complex drivers controlling regional climate changes, including climate system internal feedbacks and external forcings, is an urgent issue for future researches.

3.3 Future projection based on ClimWIP

Given its good performance in simulating current climate and providing credible projections for both temperature and precipitation, ClimWIP (the most optimal method) is now used for the projection of future climate for 1.5 and 2 °C global warming targets (relative to preindustrial), and the differences between the two warming levels in this section. Interpret the climate projections in a probabilistic way can address the issue of uncertainties due to inter-model differences to some extent (Chen et al., 2011, Li et al., 2016). So, we further discuss the likelihood of changes in climate indices exceeding certain thresholds. Combined with the increase values for Tav and Prcptot in section 3.2, we have chosen the thresholds of 1.5 °C for Tav, and 5% for Prcptot under 1.5 and 2 °C warming.

Figure 8 illustrates the spatial distributions of the projected changes and

probabilistic projections for Tav under 1.5 and 2 °C global warming, and the differences between the two warming levels, under SSP5-8.5 scenario. The annual temperature is projected to increase across China, with stronger warming under larger warming target. Relative to 1995–2014, Tav would increase by approximately 1.1 and 1.8 °C averaged over China for 1.5 and 2 °C global warming. Regions with large warming are mainly located over Northeast and Northwest China. Under an additional 0.5 °C global warming, the Tav will increase more than 0.5 °C across China. Some regions display substantial increases. For example, it will warm more than 1 °C in north part of NE, which is about 2 times higher than that of 0.5 °C global warming. The probability of warming is higher in the north than in the south of China. Under 1.5 °C global warming, the probability of warming greater than 1.5 °C threshold is greater than 50% only in a few parts of northern China. However, Tav is likely (>50%) to have an increase of more than 1.5 °C across China under 2 °C global warming, except for a small portion of South China. The probability for an increase of more than 1.5 °C even exceed 80% over parts of the Yangtze river basin and western part of northeastern China. An additional 0.5 °C warming dramatically increases the likelihood of warming exceeding the 1.5 °C threshold, especially in northern China.

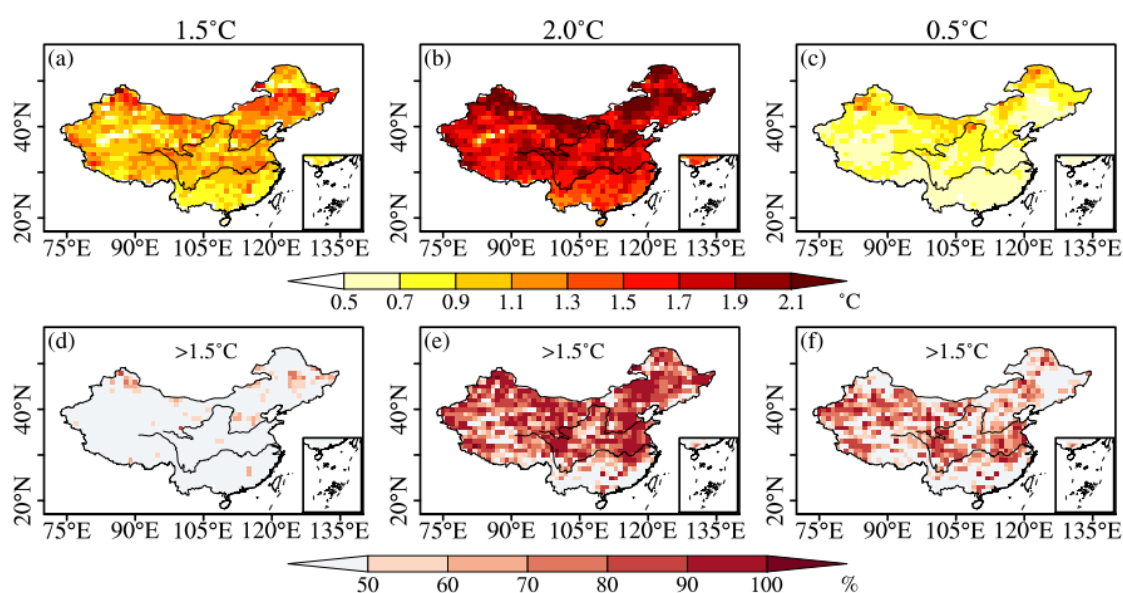


Figure 8 Spatial distributions of projected changes (a–c) for Tav over China under 1.5 and 2 °C global warming and the differences between the two warming levels under SSP5-8.5 scenario. Spatial patterns of probability (d–e) for Tav change exceeding the threshold of 1.5 °C over China under 1.5 and 2.0 °C global warming levels and their difference.

Figure 9 shows the spatial distribution of the projected changes and probabilistic projection for Prcptot. It is projected to increase in China with a larger magnitude for higher warming conditions. Relative to 1995–2014, Prcptot would increase by approximately 5.4 and 11.2% averaged over China for 1.5 and 2 °C global warming. Almost all regions witness an increase of Prcptot under the warming climates. However, it decreases in most parts of SW under 1.5 and 2 °C warming climates. Decreasing

precipitation combined with robust increasing temperature indicates that the drought risk in southwestern China would increase in the future, which is consistent with previous studies (Chen and Sun, 2015; Zhang et al., 2021). The region with a large increase is mainly located in western China. This feature becomes even more pronounced as GMST further rises. The additional half-degree warming leads to increases in Prcptot across China. Northwestern China has been shown to be more significantly associated to the additional warming, the increase can exceed 10% over there. It is more likely that precipitation increases significantly in north and western China. The probability of the precipitation change exceeding 5% is greater than 50% in the middle and lower reaches of the Yellow River area and northwestern China under 1.5 °C warming, and even a higher probability (>80%) in parts of northwestern China. Under 2 °C global warming, areas with the probability (>50%) of Prcptot change exceeding the 5% threshold expand to almost the whole country, except for western part of Northeast and South China. The additional half-degree warming increases the likelihood of Prcptot increase exceeding the 5% threshold, especially in northeastern China.

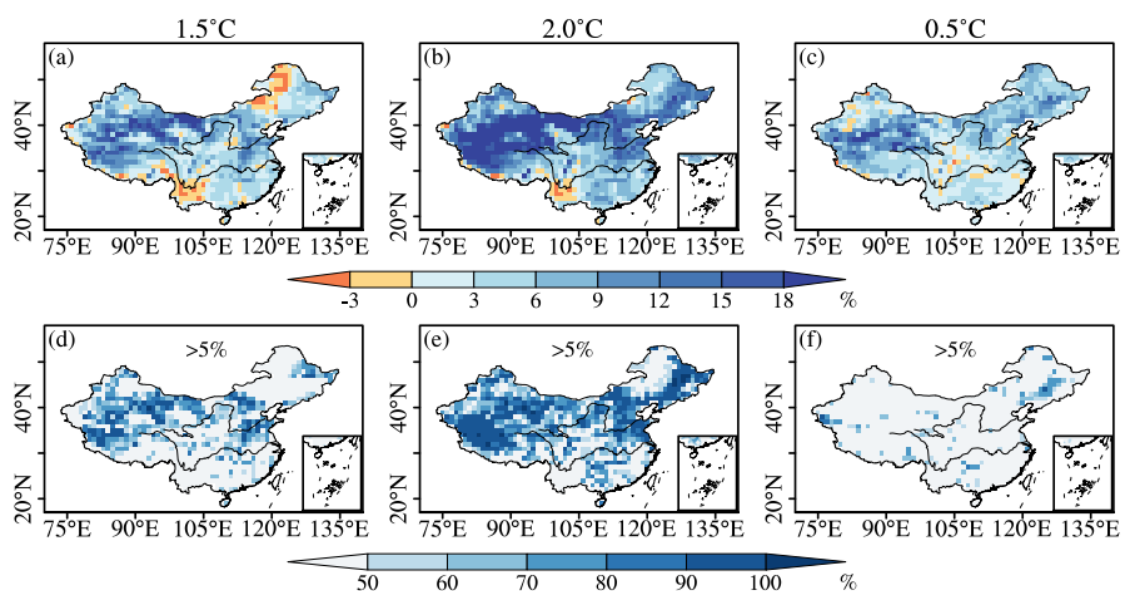


Figure 9 Spatial distributions of projected changes (a-c) for Prcptot over China under 1.5 and 2 °C global warming and the differences between the two warming levels under SSP5-8.5 scenario. Spatial patterns of probability (d-f) for Prcptot change exceeding the threshold of 5% over China under 1.5 and 2.0°C global warming levels and their difference.

Overall, there are increases in annual mean temperature and general humidification of climate for total precipitation over China under 1.5 and 2 °C global warming. The greatest warming occurs in northeastern China, and the largest increase in Prcptot occurs in western China. Under 2 °C global warming, the probability of T_{av} increasing by 1.5 °C is greater than 50% almost across whole China, except for South China. The probability of Prcptot increasing by 5% is greater than 50% in the north of the Yangtze river, except for western part of Northeast China. An additional 0.5 °C

global warming is projected to have substantial impacts on climate over China. The T_{av} increases more than $0.5\text{ }^{\circ}\text{C}$ across China. Almost all regions witness an increase of P_{cprtot} under the additional half-degree warming, and northwestern China suffers the strongest response.

4. Conclusion and discussion

In this study, using a common framework, we quantitatively compared five ensemble-processing methods in multi-model climate simulations, which helps to achieve a more reliable climate projection over China. Besides the simple methodology of MME, the other four methods use physically-valid and objective criteria to introduce constraints in their ensemble processing scheme. Beyond the methodological aspects, we also analyzed the projection in CMIP6 under high-emission pathways (SSP5-8.5) for two warming levels (1.5 and $2\text{ }^{\circ}\text{C}$). Results generally show increases in temperature, and an intensification of precipitation in the future warmer world. They are in agreement with our previous studies (Zhu et al., 2021). Compared with earlier studies, this work was designed to be more comprehensive, and made an objective and fair evaluation of projection agreement and uncertainty across the methods and regions by a common framework. It also provided more detailed information on the future changes in climate for China and its seven subregions. Main findings are summarized as follows:

(1) Compared with MME, the other four ensemble-processing methods effectively enhance the capability of capturing spatial climate characteristics over China. In addition, they provide reasonably consistent median temperature and precipitation projections over China, with reduction of uncertainty (inter-model spread) and increase of robustness (larger SNRs). ClimWIP is the most optimal method to present good performance in simulating current states and credible future projections for both temperature and precipitation. The uncertainty (10th–90th percentile ranges) reduction (compared to MME) is about 30% for T_{av} , and 15% for P_{cprtot} in China, with a certain variation among different regions.

(2) Reliability of projections is found dependent on investigated regions and indices. The projection of T_{av} is more credible than P_{cprtot} , as the SNR of T_{av} is larger than 2 among all regions, while lower than 1 for P_{cprtot} over most regions under $1.5\text{ }^{\circ}\text{C}$ warming. The most robust projection is in SW for T_{av} , with the largest SNR, though the warming is weak. The largest inter-model spread is noted in NE, followed by NC. For P_{cprtot} , the largest increase and robustness occur in NW, followed by NC, and the inter-model spread is large there. The smallest SNR and worst projection agreement are noted in SW, the methods even disagree on the sign of the change for P_{cprtot} over there under $1.5\text{ }^{\circ}\text{C}$ warming.

(3) Based on ClimWIP, T_{av} increases by about $1.1/1.8\text{ }^{\circ}\text{C}$ (with reference to 1995–2014) across China under $1.5/2\text{ }^{\circ}\text{C}$ global warming levels (regarding pre-industrial), while P_{cprtot} increases by about 5.4/11.2%, respectively. The most remarkable warming occurs in NE, with increase of 1.3 (0.6 – 1.7)/ 2.0 (1.4 – 2.6) $^{\circ}\text{C}$ (ensemble's median and range of the 10th–90th percentiles) under $1.5/2\text{ }^{\circ}\text{C}$ warming, followed by NC and NW. The smallest but most robust warming is in SW, with values exceeding 0.9 (0.6 – 1.1)/ 1.5 (1.1 – 1.7) $^{\circ}\text{C}$. The significant and robust increases in wetness occur in

NW and NC, with increase up to 9.1 (−1.6–24.7)/17.9 (0.5–36.4)% and 6.0 (−2.6–17.8)/11.8 (2.4–25.1)% under 1.5/2 °C warming, respectively. Under 2 °C global warming, the probability of T_{av} increasing by 1.5 °C is greater than 50% almost across whole China. The probability of P_{rcptot} increasing by 5% is greater than 50% in the north of the Yangtze river. For the additional half-degree warming, T_{av} increases more than 0.5 °C across China. Almost all regions witness an increase of P_{rcptot} , with northwestern China showing the largest increase.

Our results show that the median estimate is generally consistent across the ensemble-processing methods, while the spread is highly dependent on the method used. The projection information has to be combined with the end-user's needs. Our results are confident in the constrained median projection, so users with a relatively low level of risk aversion who care for the most likely climate outcome could use results from any of these methods. Those users with a higher level of risk aversion may need to consider carefully which method to rely on. Ultimately all methods, including the unconstrained model-equality strategy, are based on implicit and explicit assumptions that can be challenged. Therefore, a single recommendation for selecting or combining methods for various situations would be impossible and not very meaningful. The meaning of “optimal” is not general and can vary depending on the specific application, e.g., the specific variable, geographic region, time period, metrics, and so on. Meanwhile, special attentions should be paid to the risk of overfitting in a small-size ensemble subset, when selecting the “optimal” methods (Herger et al., 2018a, b). One promising way to resolve this is that of a coordinated perfect model test (Brunner et al., 2020). Many researches have already used such a test with some success (Brunner et al., 2019; Ribes et al., 2021).

This study uses individual member of the CMIP6 model to project future changes in regional climate, mainly considering the model uncertainty. Internal variability typically plays an important role in the nearer term, at local spatial scales, and in spatially and temporally heterogeneous variables such as precipitation (Hawkins and Sutton, 2009). It can be isolated from model uncertainty using large ensembles (Brunner et al., 2020). More attention should be paid to abundant climate variables (e.g., extremes and compound events) using large ensembles in the future.

Acknowledgements

We would like to acknowledge the World Climate Research Programme's Working Group on Coupled Modeling, which is responsible for CMIP6. We thank the climate modeling groups for producing simulations and making their model outputs available. This work is supported by the National Key Research and Development Program of China (Grant No. 2017YFA0603804), the National Natural Science Foundation of China (Grant No. 42275184), and the Postgraduate Research and Practice Innovation Program of Government of Jiangsu Province (Grant No. KYCX22_1135). Laurent Li acknowledges French GENCI for allocation of computing resources.

References

- Brunner L, Lorenz R, Zumwald M, Knutti R. 2019. Quantifying uncertainty in European climate projections using combined performance-independence weighting. *Environ Res Lett*, 14: 124010.
- Brunner L, McSweeney C, Ballinger A P, Befort D J, Benassi M, Booth B, Coppola E, Vries H D, Harris G, Hegerl G C, Knutti R, Lenderink G, Lowe J, Nogherotto R, O'Reilly C, Qasmi S, Ribes A, Stocchi P, Undorf S. 2020. Comparing methods to constrain future European climate projections using a consistent framework. *J Clim*, 33(20): 8671–8692.
- Chen H P, Sun J Q. 2015. Changes in drought characteristics over china using the standardized precipitation evapotranspiration index. *J Clim*, 28: 5430–5447.
- Chen W L, Jiang Z H, Li L. 2011. Probabilistic projections of climate change over China under the SRES A1B scenario using 28 AOGCMs. *J Clim*, 24: 4741–4756.
- Chen Z M, Zhou T J, Zhang L X, Chen X L, Zhang W X, Jiang J. 2020. Global Land Monsoon Precipitation Changes in CMIP6 Projections. *Geophys Res Lett*, 47(14): e2019GL086902.
- CMA Climate Change Centre. 2021. Blue Book on Climate Change in China (2021). Beijing: Science Press.
- Dong B W, Sutton R T, Chen W, Liu X D, Lu R Y, Sun Y. 2016. Abrupt summer warming and changes in temperature extremes over Northeast Asia since the mid-1990s: drivers and physical processes. *Adv Atmos Sci*, 33: 1005–23.
- Findell K L, Berg A, Gentine P, Krasting J P, Lintner B R, Malyshev S, Santanello Jr J A, Shevliakova E. 2017. The impact of anthropogenic land use and land cover change on regional climate extremes. *Nat Commun*, 8: 989.
- Giorgi F, Mearns L O. 2022. Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the “reliability ensemble averaging” (REA) method. *J Clim*, 15: 1141–1158.
- Giorgi F, Mearns L O. 2003. Probability of regional climate change based on the Reliability Ensemble Averaging (REA) method. *Geophys Res Lett*, 30: 1629.
- Guo L Y, Jiang Z H, Chen D L, Treut H L, Li L. 2020. Projected precipitation changes over China for global warming levels at 1.5 °C and 2 °C in an ensemble of regional climate simulations: impact of bias correction methods. *Clim Change*, 162(2): 623–643.
- Hawkins E, Sutton R. 2009. The potential to narrow uncertainty in regional climate predictions. *Bull Am Meteorol Soc*, 90(8): 1095–1107.
- Hawkins E, Sutton R. 2011. The potential to narrow uncertainty in projections of regional precipitation change. *Clim Dyn*, 37(1): 407–418.
- Herger N, Abramowitz G, Knutti R, Angélil O, Lehmann K, Sanderson B M. 2018a. Selecting a climate model subset to optimise key ensemble properties. *Earth Syst Dyn*, 9: 135–151.
- Herger N, Angélil O, Abramowitz G, Donat M, Stone D, Lehmann K. 2018b. Calibrating climate model ensembles for assessing extremes in a changing climate. *JGR-Atmos*, 123: 5988–6004.
- Hidalgo H G, Alfaro E J. 2015. Skill of CMIP5 climate models in reproducing 20th

- century basic climate features in central America. *Int J Climatol*, 35: 3397–421.
- Hu T, Sun Y, Zhang X B. 2017. Temperature and precipitation projection at 1.5 and 2 °C increase in global mean temperature. *Chinese Sci Bull*, 62: 3098–3111. (in Chinese)
- IPCC. 2021. Summary for Policymakers: Climate Change 2021. The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.
- Jiang Z H, Li W, Xu J J, Li L. 2015. Extreme precipitation indices over China in CMIP5 models. Part I: Model evaluation. *J Clim*, 28: 8603–8619.
- Knutti R, Furrer R, Tebaldi C, Cermak J, Meehl G A. 2010. Challenges in combining projections from multiple climate models. *J Clim*, 23: 2739–58.
- Knutti R, Masson D, Gettelman A. 2013. Climate model genealogy: generation CMIP5 and how we got there. *Geophys Res Lett*, 40: 1194–9.
- Knutti R, Sedláček J, Sanderson B M, Lorenz R, Fischer E M, Eyring V. 2017. A climate model projection weighting scheme accounting for performance and interdependence. *Geophys Res Lett*, 44: 1909–1918.
- Li T, Jiang Z H, Treut H, Li L, Zhao L L, Ge L L. 2021a. Machine learning to optimize climate projection over China with multi-model ensemble simulations. *Environ Res Lett*, 16: 094028.
- Li T, Jiang Z H, Zhao L L, Li L. 2021b. 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. *J Meteorol Res*, 35: 184–97.
- Li W, Jiang Z H, Xu J J, Li L. 2016. Extreme precipitation indices over China in CMIP5 models. Part II: Probabilistic projection. *J Climate*, 29: 8989–9004.
- Li W, Jiang Z H, Zhang X B, Li L, Sun Y. 2018. Additional risk in extreme precipitation in China from 1.5 °C to 2.0 °C global warming levels. *Sci Bull*, 63(4): 228–234.
- Li W, Wang Q R, Zhu H H, Tang Z F. 2022. Influence of internal variability on future changes in extreme wind speed over North America. *Glob Planet Change*, 218: 103968.
- Lorenz R, Argüeso D, Donat M G, Pitman A J, Hurk B V D, Berg A, Lawrence D M, Chéruf F, Ducharne A, Hagemann S, Meier A, Milly P C D, Seneviratne S I. 2016. Influence of land-atmosphere feedbacks on temperature and precipitation extremes in the GLACE-CMIP5 ensemble. *J Geophys Res*, 121:607–23.
- Murphy J M, Sexton D M, Barnett D N, Jones G S, Webb M J, Collins M, Stainforth D A. 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature*, 430: 768–772.
- O'Neill B C, Tebaldi C, van Vuuren D P, Eyring V, Friedlingstein P, Hurtt G, Knutti R, Kriegler E, Lamarque J F, Lowe J, Meehl G A, Moss R, Riahi K, Sanderson B M. 2016. The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geosci Model Dev*, 9: 3461–3482.
- Raftery A E, Gneiting T, Balabdaoui F, Polakowski M. 2005. Using Bayesian model

- averaging to calibrate forecast ensembles. *Mon Weather Rev*, 133 (5): 1155–1174.
- Ribes A, Qasmi S, Gillett N P. 2021. Making climate projections conditional on historical observations. *Sci Adv*, 7(4): 1–10.
- Sanderson B M, Knutti R, Caldwell P. 2015. Addressing interdependency in a multimodel ensemble by interpolation of model properties. *J Clim*, 28: 5150–70.
- Schmidt G A, Shindell D T, Tsigaridis K. 2014. Reconciling warming trends. *Nat Geosci*, 7 (3): 158e160.
- Sharmila S, Joseph S, Sahai A K, Abhilash S, Chattopadhyay R. 2015. Future projection of Indian summer monsoon variability under climate change scenario: an assessment from CMIP5 climate models. *Glob Planet Change*, 124: 62–78.
- Shi C, Jiang Z H, Chen W L, Li L. 2018. Changes in temperature extremes over China under 1.5 °C and 2 °C global warming targets. *Adv Climate Change Res*, 9: 120–129.
- Smith I, Chandler E. 2010. Refining rainfall projections for the Murray Darling Basin of south-east Australia—the effect of sampling model results based on performance. *Clim Change*, 102 (3–4): 377–393.
- Sun S L, Chen H S, Ju W M, Wang G J, Sun G, Huang J, Ma H D, Gao C J, Hua W J, Yan G X. 2017. On the coupling between precipitation and potential evapotranspiration: contributions to decadal drought anomalies in the Southwest China. *Clim Dyn*, 48(11): 3779–3797.
- Tan J L, Jiang Z H, Ma T T. 2016. Projections of future surface air temperature change and uncertainty over China based on the Bayesian Model Averaging. *Acta Meteor Sinica*, 74: 583–597. (in Chinese)
- Taylor K E. 2001. Summarizing multiple aspects of model performance in a single diagram. *JGR-Atmos*, 106: 7183–7192.
- Thompson D W, Barnes E A, Deser C, Foust W E, Phillips A S. 2015. Quantifying the role of internal climate variability in future climate trends. *J Clim*, 28 (16): 6443–6456.
- United Nations Framework Convention on Climate Change (UNFCCC). 2015. Adoption of the Paris Agreement. Preprints. United Nations Office at Geneva, Switzerland (FCCC/CP/2015/L.2019/Rev.2011), 1–32.
- Wu J, Gao X J. 2013. A gridded daily observation dataset over China region and comparison with the other datasets. *Chinese J Geophy*, 56: 1102–1111, (in Chinese with English abstract)
- Wu J, Han Z Y, Xu Y, Zhou B T, Gao X J. 2020. Changes in extreme climate events in China under 1.5 °C–4 °C global warming targets: Projections using an ensemble of regional climate model simulations. *JGR-Atmos*, 125(2): e2019JD031057.
- Wu S Y, Wu Y J, Wen J H. 2019. Future changes in precipitation characteristics in China. *Int J Climatol*, 39(8): 3558–3573.
- Xu Y, Gao X J, Giorgi F. 2002. Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections. *Clim Res*, 41: 61–81.
- Yang X L, Zhou B T, Xu Y, Han Z Y. 2021. CMIP6 evaluation and projection of temperature and precipitation over China. *Adv Atmos Sci*, 38(5): 817–830.

- You Q L, Jiang Z H, Yue X, Guo W D, Liu Y G, Cao J, Li W, Wu F Y, Cai Z Y, Zhu H H, Li T, Liu Z Y, He J H, Chen D L, Pepin N, Zhai P M. 2022. Recent frontiers of climate changes in East Asia at global warming of 1.5 °C and 2 °C. *NPJ Clim*, 5(1): 1–17.
- Zhang L X, Chen Z M, Zhou T J. 2021. Human influence on the increasing drought risk over Southeast Asian monsoon region. *Geophys Res Lett*, 48: e2021GL093777.
- Zhang Q, Yang J H, Wang W, Ma P L, Lu G Y, Liu X Y, Yu H P, Fang F. 2021. Climatic warming and humidification in the arid region of Northwest China: Multi-scale characteristics and impacts on ecological vegetation. *J Meteor Res*, 35(1): 113–127.
- Zhang W X, Zhou T J, Zhang L X, Zou L W. 2019. Future Intensification of the Water Cycle with an Enhanced Annual Cycle over Global Land Monsoon Regions. *J Clim*, 32(17): 5437–5452.
- Zhou B T, Wen Q H, Xu Y, Song L C, Zhang X B. 2014. Projected changes in temperature and precipitation extremes in China by the CMIP5 multimodel ensembles. *J Clim*, 27: 6591–6611.
- Zhou T J, Zou L W, Chen X L. 2019. Commentary on the Coupled Model Intercomparison Project Phase 6 (CMIP6). *Clim Change Res*, 15: 445–456. (in Chinese with English abstract)
- Zhou T J, Zhang W X, Zhang L X, Zhang X B, Qian Y, Peng D D, Ma S M, Dong B W. 2020. The dynamic and thermodynamic processes dominating the reduction of global land monsoon precipitation driven by anthropogenic aerosols emission. *Sci China Earth Sci*, 63: 919–933.
- Zhu H H, Jiang Z H, Li J, Li W, Sun C X, Li L. 2020. Does CMIP6 inspire more confidence in simulating climate extremes over China? *Adv Atmos Sci*, 37(10): 1119–1132.
- Zhu H H, Jiang S, Jiang Z H. 2022. Projection of Climate Extremes over China in Response to 1.5/2 °C Global Warming Based on the Reliability Ensemble Averaging. *Adv Earth Sci*, 37, 612–626. (in Chinese)
- Zhu H H, Jiang Z H, Li L. 2021. Projection of climate extremes in China, an incremental exercise from CMIP5 to CMIP6. *Sci Bull*, 66(24): 2528–2537.

Supplementary Materials

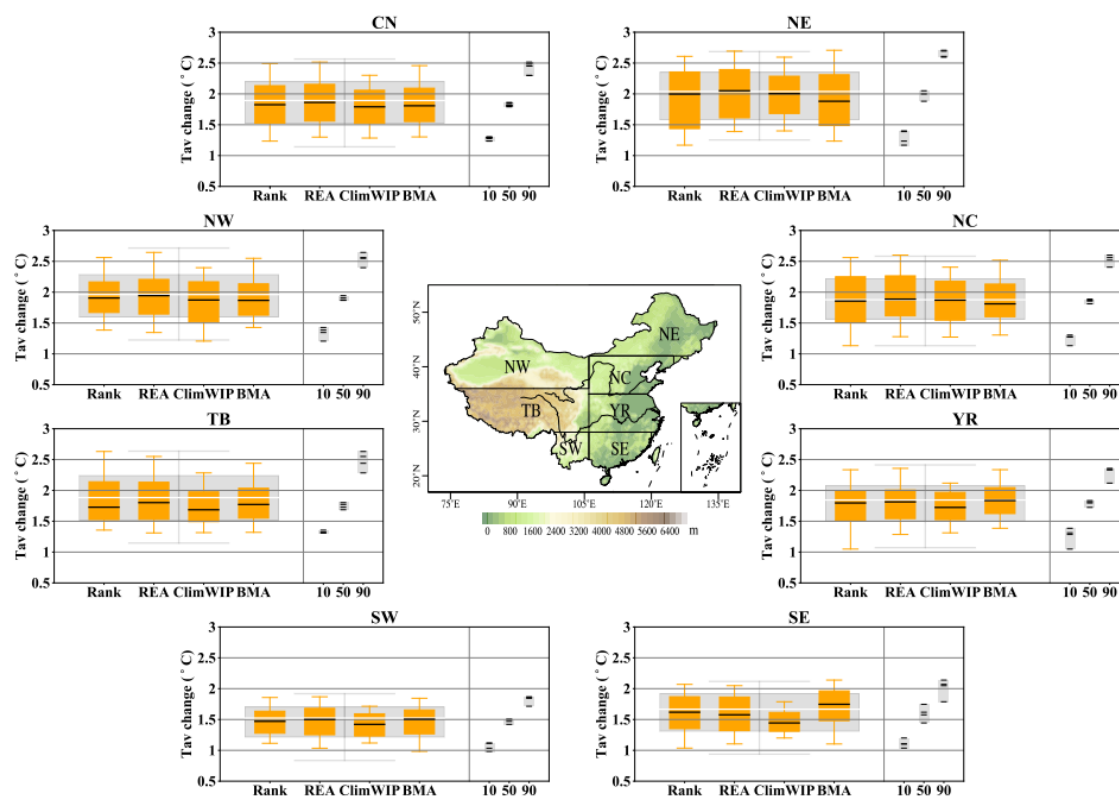


Fig. S1. As in Fig. 3., but for the changes under 2 °C global warming target.

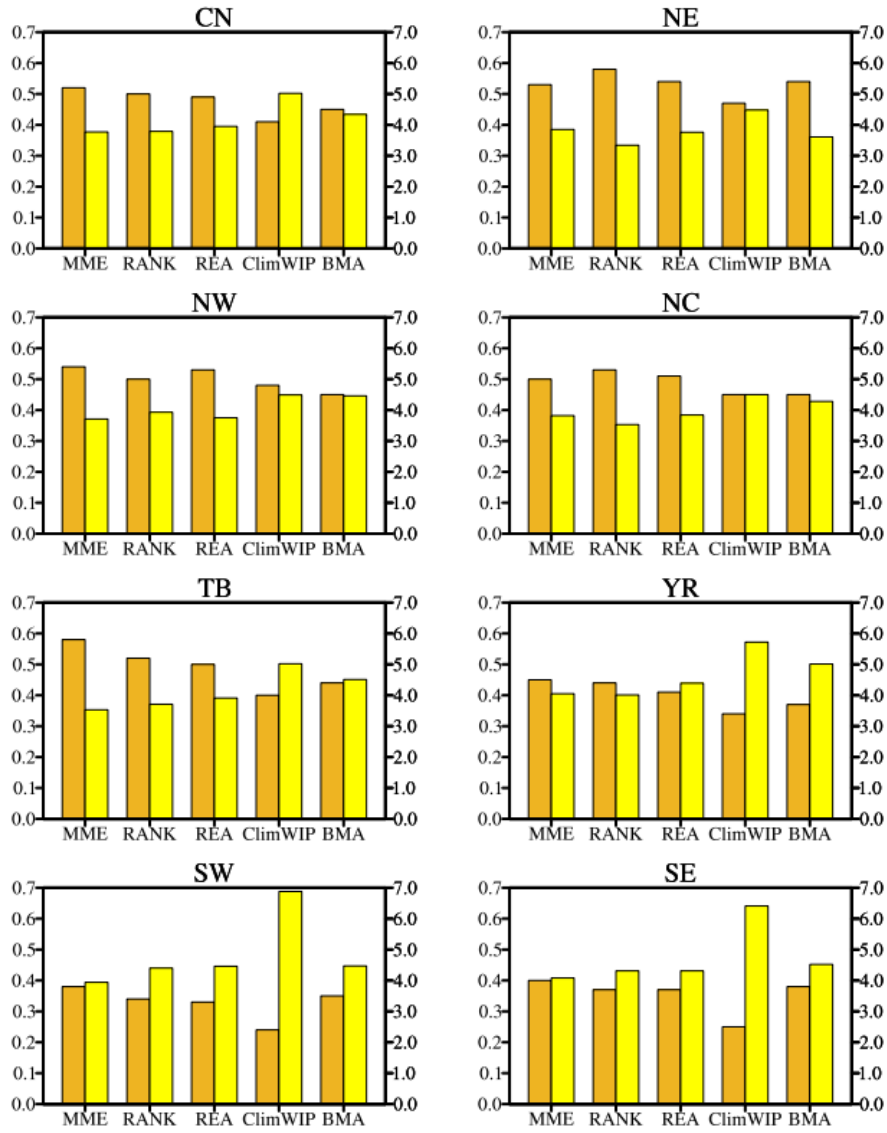


Fig. S2. Same as in Fig. 4. but for the SD and SNR under 2 °C global warming target.

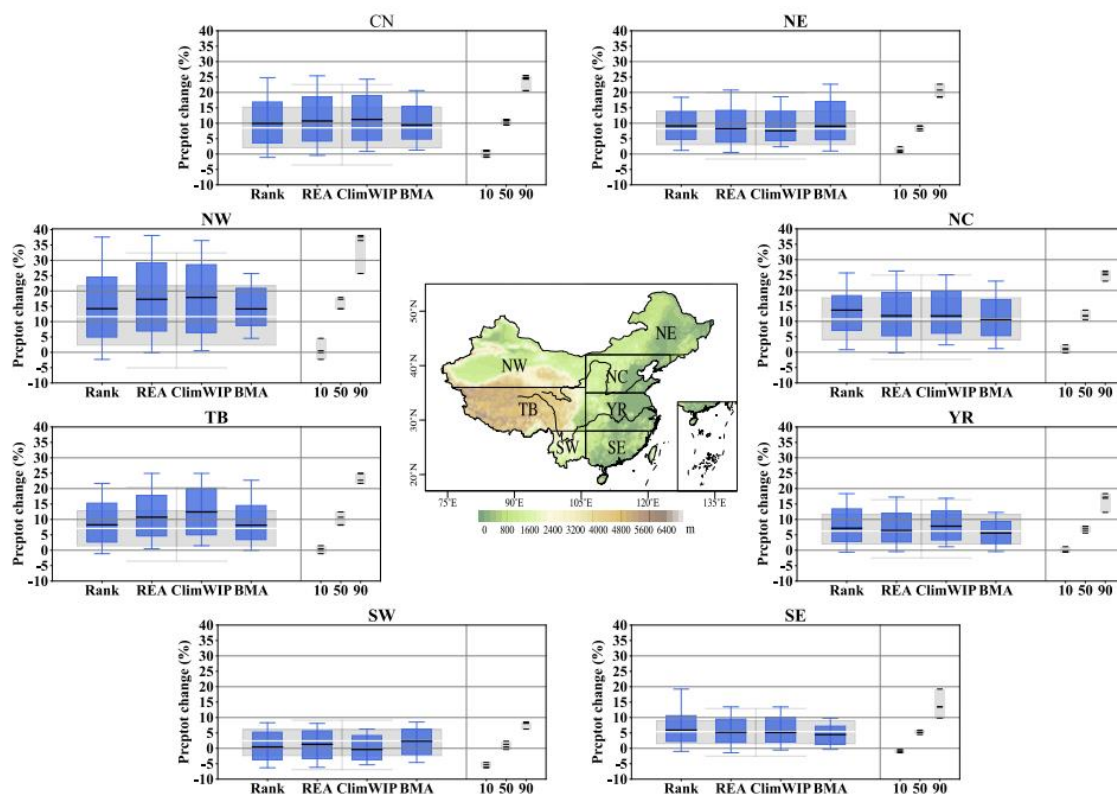


Fig S3. Same as in Fig. S1. but for Preptot.

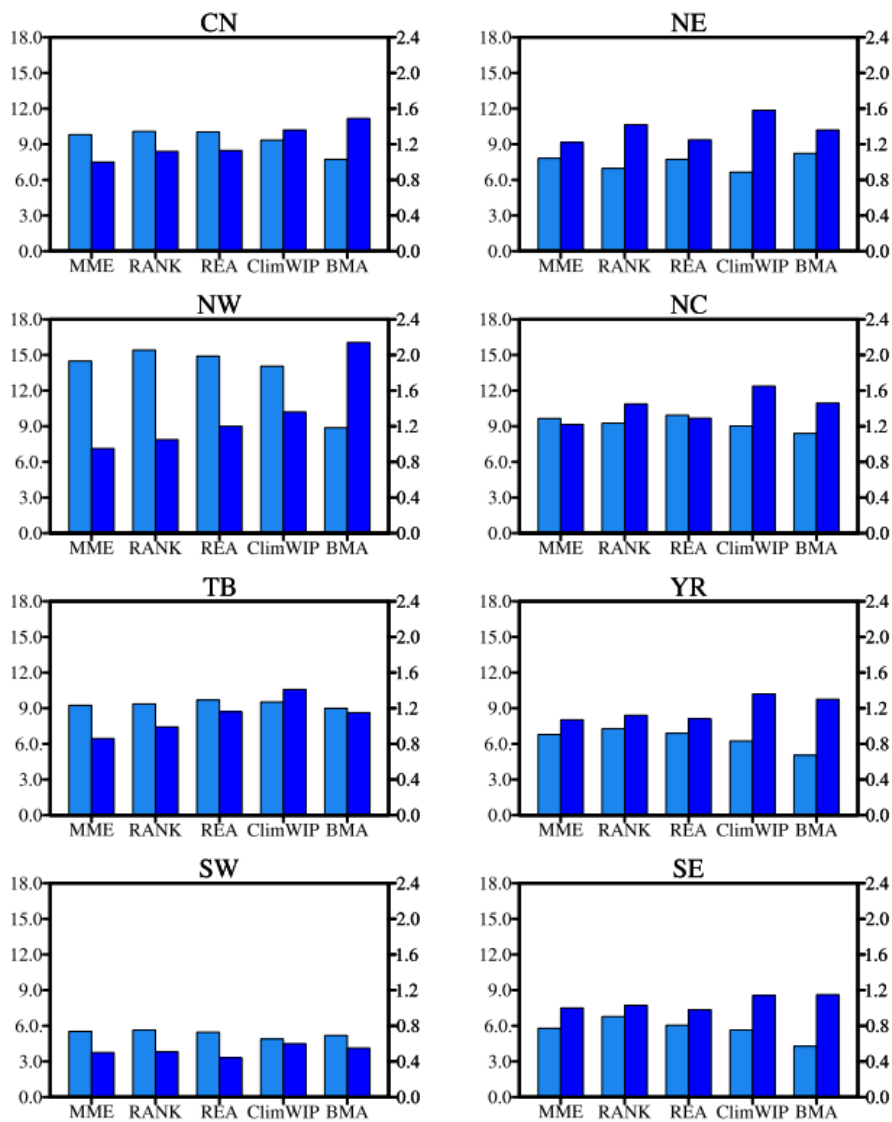


Fig. S4. Same as in Fig. S2. but for Precptot.

Table S1. Model acronyms, modeling centers and countries, as well as the atmospheric resolutions, of 25 CMIP6 global climate models.

| Model acronym | Modeling center and country | Atmospheric resolution (lat × lon) |
|-----------------|--|---------------------------------------|
| ACCESS-CM2 | Commonwealth Scientific and Industrial Research | 1.25°×1.875° |
| ACCESS-ESM1-5 | Organization and Bureau of Meteorology (Australia) | 1.25°×1.875° |
| BCC-CSM2-MR | Beijing Climate Center, China Meteorological Administration (China) | 1.125°×1.125° |
| CanESM5 | Canadian Centre for Climate Modelling and Analysis (Canada) | 2.8°×2.8° |
| CMCC-ESM2 | Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (Italy) | 0.9375°×1.25° |
| CNRM-CM6-1 | Centre National de Recherches Météorologiques–Centre | 1.4°×1.4° |
| CNRM-ESM2-1 | Européen de Recherche et de Formation Avancée en Calcul Scientifique (France) | 1.4°×1.4° |
| EC-Earth3 | EC-EARTH consortium | 0.7°×0.7° |
| EC-Earth3-Veg | | 0.7°×0.7° |
| FGOALS-g3 | LASG, Institute of Atmospheric Physics, Chinese Academy of Sciences and Center for Earth System Science, Tsinghua University (China) | 2.25°×2° |
| GFDL-CM4 | NOAA Geophysical Fluid Dynamics Laboratory (USA) | 1°×1.25° |
| GFDL-ESM4 | | 1°×1.25° |
| HadGEM3-GC31-LL | Met Office Hadley Centre (UK) | 1.25°×1.875° |
| INM-CM4-8 | Institute for Numerical Mathematics, Russian Academy | 1.5°×2° |
| INM-CM5-0 | of Science (Russia) | 1.5°×2° |
| IPSL-CM6A-LR | L’Institut Pierre-Simon Laplace (France) | 1.26°×2.5° |
| MIROC6 | National Institute for Environmental Studies, | 1.4°×1.4° |
| MIROC-ES2L | The University of Tokyo (Japan) | 2.8°×2.8° |
| MPI-ESM-1-2-HR | Max Planck Institute for Meteorology (Germany) | 0.9375°×0.9375° |
| MPI-ESM-1-2-LR | | 1.875°×1.875° |
| MRI-ESM2-0 | Meteorological Research Institute (Japan) | 1.125°×1.125° |
| NESM3 | Nanjing University of Information Science and Technology (China) | 1.875°×1.875° |
| NorESM2-LM | Norwegian Climate Centre (Norway) | 1.875°×2.5° |
| NorESM2-MM | | 0.94°×1.25° |
| UKESM1-0-LL | Met Office Hadley Centre (UK) | 1.25°×1.875° |

Table S2. The timing of 1.5 and 2 °C warming for individual models

| Model acronym | SSP585 | |
|-----------------|--------|------|
| | 1.5 °C | 2 °C |
| ACCESS-CM2 | 2025 | 2038 |
| ACCESS-ESM1-5 | 2028 | 2039 |
| BCC-CSM2-MR | 2032 | 2045 |
| CanESM5 | 2012 | 2023 |
| CMCC-ESM2 | 2028 | 2038 |
| CNRM-CM6-1 | 2028 | 2041 |
| CNRM-ESM2-1 | 2032 | 2045 |
| EC-Earth3 | 2025 | 2036 |
| EC-Earth3-Veg | 2012 | 2028 |
| FGOALS-g3 | 2028 | 2046 |
| GFDL-CM4 | 2029 | 2041 |
| GFDL-ESM4 | 2040 | 2053 |
| HadGEM3-GC31-LL | 2020 | 2030 |
| INM-CM4-8 | 2030 | 2046 |
| INM-CM5-0 | 2030 | 2046 |
| IPSL-CM6A-LR | 2018 | 2034 |
| MIROC6 | 2040 | 2053 |
| MIROC-ES2L | 2034 | 2047 |
| MPI-ESM1-2-HR | 2033 | 2049 |
| MPI-ESM1-2-LR | 2034 | 2049 |
| MRI-ESM2-0 | 2026 | 2038 |
| NESM3 | 2021 | 2034 |
| NorESM2-LM | 2043 | 2057 |
| NorESM2-MM | 2039 | 2054 |
| UKESM1-0-LL | 2023 | 2032 |