



HAL
open science

Machine learning to optimize climate projection over China with multi-model ensemble simulations

Tong Li, Zhihong Jiang, Hervé Le Treut, Laurent Li, Lilong Zhao, Lingling Ge

► **To cite this version:**

Tong Li, Zhihong Jiang, Hervé Le Treut, Laurent Li, Lilong Zhao, et al.. Machine learning to optimize climate projection over China with multi-model ensemble simulations. *Environmental Research Letters*, 2021, 16 (9), pp.094028. 10.1088/1748-9326/AC1D0C . hal-04343423

HAL Id: hal-04343423

<https://hal.science/hal-04343423>

Submitted on 13 Dec 2023

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.

LETTER • **OPEN ACCESS**

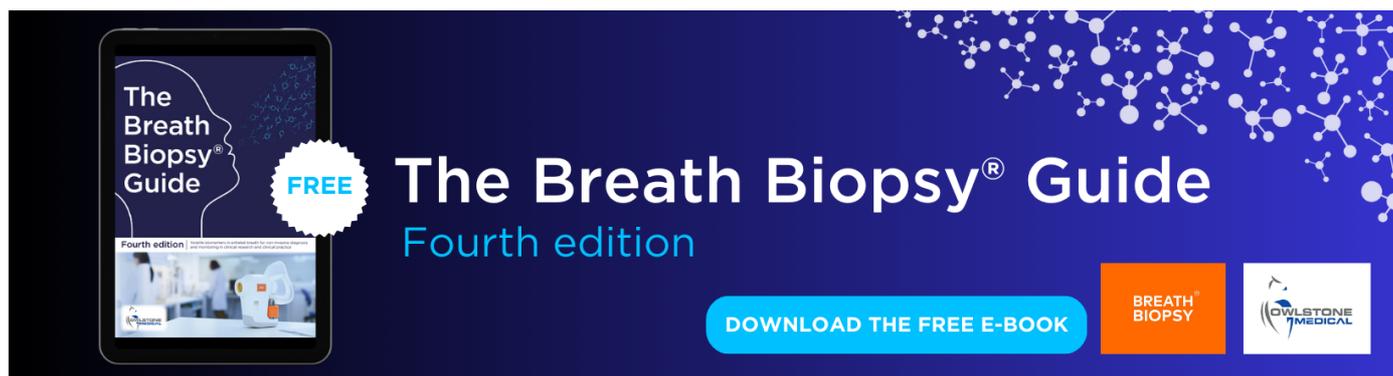
Machine learning to optimize climate projection over China with multi-model ensemble simulations

To cite this article: Tong Li *et al* 2021 *Environ. Res. Lett.* **16** 094028

View the [article online](#) for updates and enhancements.

You may also like

- [Future changes and seasonal variability of the directional wave spectra in the Mediterranean Sea for the 21st century](#)
Andrea Lira-Loarca and Giovanni Besio
- [Uncertainty in future agro-climate projections in the United States and benefits of greenhouse gas mitigation](#)
Erwan Monier, Liyi Xu and Richard Snyder
- [Evidence for increasing global wheat yield potential](#)
Jose Rafael Guarin, Pierre Martre, Frank Ewert *et al.*



The Breath Biopsy® Guide
Fourth edition

FREE

DOWNLOAD THE FREE E-BOOK

BREATH BIOPSY

OWLSTONE MEDICAL

ENVIRONMENTAL RESEARCH
LETTERS

LETTER

OPEN ACCESS

RECEIVED
21 June 2021REVISED
5 August 2021ACCEPTED FOR PUBLICATION
12 August 2021PUBLISHED
27 August 2021

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.

Machine learning to optimize climate projection over China with
multi-model ensemble simulationsTong Li¹, Zhihong Jiang^{2,*}, Hervé Le Treut³, Laurent Li³, Lilong Zhao¹ and Lingling Ge⁴

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, People's Republic of China

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, People's Republic of China

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

⁴ Jiangsu Key Laboratory of Big Data Analysis Technology, School of Information and Control, Nanjing University of Information Science and Technology, Nanjing 210044, People's Republic of China

* Author to whom any correspondence should be addressed.

E-mail: zhjiang@nuist.edu.cn**Keywords:** ensemble-processing strategy, climate extremes, observational constraint, random forest, China regional climateSupplementary material for this article is available [online](#)**Abstract**

The multi-model ensemble approach is generally considered as the best way to explore the advantage and to avoid the weakness of each individual model, and ultimately to achieve the best climate projection. But the design of an optimal strategy and its practical implementation still constitutes a challenge. Here we use the random forest (RF) algorithm (from the category of machine learning) to explore the information offered by the multi-model ensemble simulations within the Coupled Model Intercomparison Project Phase 6. Our objective is to achieve a more reliable climate projection (mean climate and extremes) over China. RF is furthermore compared to two other ensemble-processing strategies of different nature, one is the basic arithmetic mean (AM), and another is the linear regression across the ensemble members. Our results indicate that RF effectively enhances the capability in capturing spatial climate characteristics. Regions with complex topography, such as the Tibetan Plateau and its periphery, show the most significant improvements. RF projects less future warming but enhanced wet conditions across China. It also produces larger spatial variability and more small-scale features. The most obvious increase of precipitation is in the northern part and the periphery of the Tibetan Plateau. The projected changes in RF for strong precipitation are almost twice higher than in AM, while in the northwestern area, weaker increases of precipitation are projected by RF, which indicates larger spatial inhomogeneity of its projection.

1. Introduction

Global warming has altered the mean and extreme climate in many regions of the world, and this warming trend will undoubtedly continue (Hulme 2016). Global climate models (GCMs) play a crucial role in generating future projections to examine the potential impacts of climate change. The ability of GCMs to reproduce observed features of the past and current climate increases our confidence to correctly make future projections (Palmer *et al* 2005, Semenov and Stratonovitch 2010). Climate

projection is inevitably accompanied by uncertainties, with available physically-based models being imperfect (Knutti *et al* 2013, Hidalgo and Alfaro 2015). The multi-model ensemble approach is useful to explore the advantage and to avoid the weakness of individual models, and ultimately to achieve the best climate projection. But the design of an optimal ensemble-processing strategy and its practical implementation still constitute a challenge (Knutti *et al* 2010, 2013). The arithmetic mean (hereafter called AM) is the simplest and mostly-used method to deal with a multi-model

ensemble (Knutti *et al* 2010, Sanderson *et al* 2015). Subsequently, more complex statistical methods such as the Bayesian methods (Robertson *et al* 2004, Tan *et al* 2016) or weighted averages, which consider the simulation skills and model inter-dependence, have been developed (Xu *et al* 2010, Jiang *et al* 2015, Knutti *et al* 2017, Brunner *et al* 2020). These methods allow tuning particular parameters or weights and constraining uncertainties with historical observations. Most of these strategies or methods, however, rely on the concept of linear regression (LR) based on some specific relationships or indices, potentially neglecting useful information.

With observations as a target or a constraint, machine learning (ML) is a useful tool to extract more information from multi-model data. Significant advancements have been reported with application of heuristic ML for uses in weather forecast, climate prediction, and reconstruction of missing climate information (Ham *et al* 2019, Reichstein *et al* 2019, Kadow *et al* 2020). ML has considerable advantages in solving non-linear, high-dimensional, and hierarchical problems to retrieve implicit patterns in complex relationships (Alizamir *et al* 2018, Guo *et al* 2019, Li *et al* 2020). With such general properties, ML can better extract important dynamical and physical processes within climate models and fully explore useful information (Wang *et al* 2018, Reichstein *et al* 2019). This would lead to a hybrid approach for future climate projection, which combines the strengths of physical modeling and mathematical algorithms of ML (Reichstein *et al* 2019, Watson-Parris 2020).

Under the framework of the Coupled Model Intercomparison Project (CMIP), a large number of climate simulations have been performed and released publicly. CMIP is an unprecedented effort and has entered its 6th phase (CMIP6) (Eyring *et al* 2016), with more models and a larger ensemble of simulations compared to its predecessor (CMIP5) (Liang *et al* 2020, Zhu *et al* 2020a). It offers exciting new opportunities for expanding our knowledge of the Earth system through the exploration of big data with advanced ML concepts and algorithms. The present study uses the random forest (RF), a powerful ML algorithm that is based on the decision tree and able to extract non-linear relations and behaviors (Breiman *et al* 1984, Breiman 2001). For the purpose of demonstration, RF is contrasted to the AM, the simplest ensemble-processing strategy, as well as the basic LR applied to the ensemble members. We want to check whether RF can effectively enhance our skill to mimic observed properties and to make reliable future climate projections. This work is a part of our general efforts of climate change mitigation and adaptation in China. It focuses on the recommended targets of 1.5 °C, and 2 °C global warming levels, following the Paris Agreement (UNFCCC 2015). The geographic domain of our investigation is mainland

China where a reliable dataset of observed climate is available.

The rest of the paper is organized as follows. Section 2 describes the data, methodology, and the three algorithms involved in our study, together with the skill metrics for evaluation. Followed in section 3 are the main results of the methodological assessment in present-day and future climate projection. Finally, conclusions and a few discussions are provided in section 4.

2. Data and methods

2.1. Study area and data used

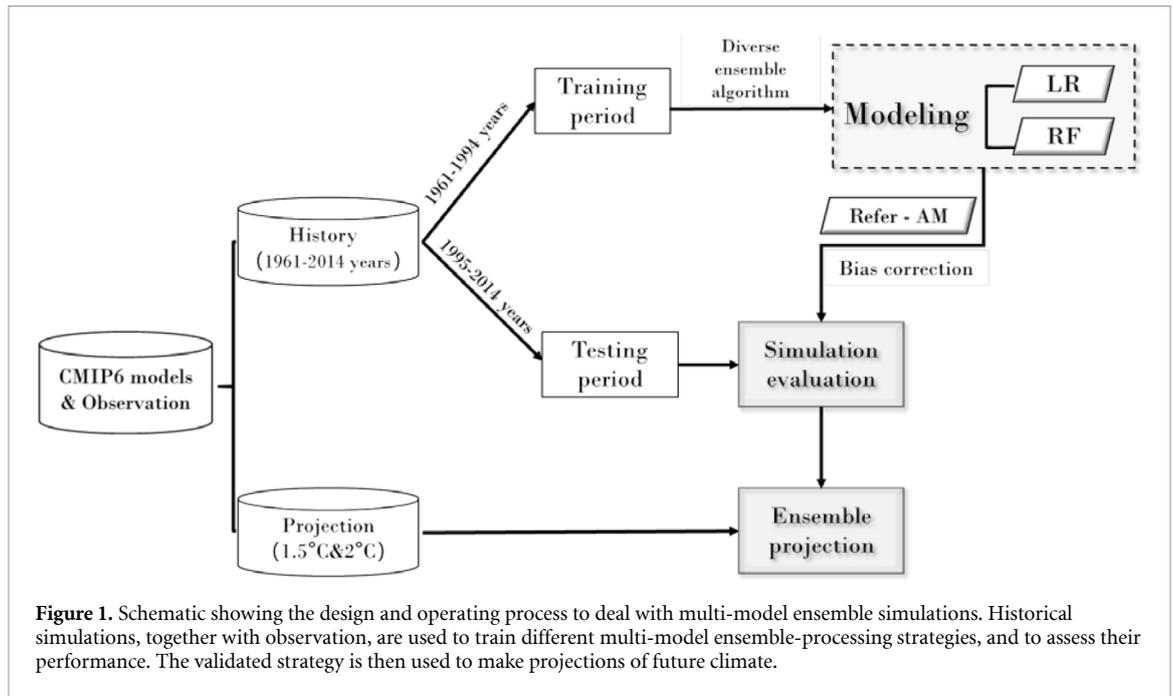
This work focuses on mainland China, a territory highly susceptible to climate change due to its complex topography and strongly-pronounced monsoonal characteristics (Fu *et al* 2008, Piao *et al* 2010). A high-quality *in situ* dataset (CN05.1), including conventional surface climatic variables, is employed for the calibration of all our approaches to develop a reliable multi-model ensemble-processing strategy. The daily gridded dataset covers 1961–2014, with a spatial resolution of $0.25^\circ \times 0.25^\circ$ over whole China. Wu and Gao (2013) provide detailed information about this dataset.

On the other hand, 24 CMIP6 models' historical simulations and future projections from shared socioeconomic pathway (SSP5-8.5) are used to construct the multi-model ensemble and to generate 1.5 °C, 2 °C and 3 °C warming projection. These models were selected on the sole criterion of data availability for our purpose of determining warming targets at 1.5 °C, 2 °C and 3 °C. All CMIP6 data were retrieved through the data portals of the Earth System Grid Federation, which can be obtained from <https://esgf-node.llnl.gov/search/cmip6/>. Some essential characteristics of the used models are listed in table S1 (available online at stacks.iop.org/ERL/16/094028/mmedia). Only their first realization (r1i1p1f1) was used in this work.

2.2. Methods

2.2.1. Climate indices

The present study employed six quantitative indices, including mean temperature (TAS), annual maximum (hottest daytime) temperature (TX_x), annual minimum (coldest nighttime) temperature (TN_n), total precipitation in wet days (PRCPTOT), annual maximum consecutive 5 d precipitation amount (RX5DAY) and annual total precipitation for events exceeding the 95th percentile (R95P, an indication of strong precipitation). These indices are useful in capturing climate change information and have been widely used to identify and monitor extreme climate (Zhang *et al* 2011, Zhu *et al* 2020a). They are derived from daily precipitation and temperature CMIP6 datasets following the recommendation by the Expert



Team on Climate Change Detection and Indices (<http://etccdi.pacificclimate.org/>). Indices from different models and observation were first calculated at their original grid and then interpolated, using bilinear interpolation, onto a common $1^\circ \times 1^\circ$ grid comprising 928 geographic locations across China. The three ensemble-processing strategies, AM, LR and RF, were then practiced on this common grid to ensure fairness and to facilitate their inter-comparison.

The study adopted the criteria used by Shi *et al* (2018) in defining the calendar year for models to reach 1.5°C and 2°C global warming thresholds. A time window of 21 years, including the 10 years before and after the nominative year, is used to deduce the climate statistics. A similar approach has been utilized in a few recent studies (e.g. Sun *et al* 2019, Guo *et al* 2020).

2.2.2. Strategies in processing multi-model ensemble

Figure 1 shows an overall flow chart of our designed processing. Historical simulations, together with observation, are divided into the training period 1961–1994 (34 years) and the testing period 1995–2014 (20 years). The testing period also serves as the historical reference for future warming projection. Our procedure is separately applied to each of the six climate indices with the general goal to explore, as much as possible, the properties of observation. The basic principle is to minimize the loss function (here the mean squared error) representing the deviation between the multi-model ensemble output and the observation. Once the training procedure is accomplished, the optimized multi-model ensemble-processing scheme can then be used to produce results for the testing period. Finally, future

projections under the 1.5°C , 2°C and 3°C global warming were conducted.

The AM is the simplest and widely-used ensemble-processing strategy. There is no parameter to optimize and it is incapable of learning from training data, which would constitute a biased reference to fairly evaluate other ensemble-processing strategies. To ensure a fair comparison with LR or RF, a linear scaling is used in AM to remove biases of climate models with their domain-mean deviation from observation (Lenderink *et al* 2007, Teutschbein and Seibert 2012). The temperature (T) is corrected with an additive term on original value and precipitation (P) with a multiplier:

$$T_{\text{cor}} = T_{\text{ori}} + \mu(T_{\text{obs}}) - \mu(T_{\text{ori}}) \quad (1)$$

$$P_{\text{cor}} = P_{\text{ori}} \times \frac{\mu(P_{\text{obs}})}{\mu(P_{\text{ori}})} \quad (2)$$

where subscripts denote corrected (cor), raw (ori), and observed (obs) values, and μ represents averaging over the domain.

The result of AM after bias correction is included here as a comparison baseline. It is worth mentioning that this linear scaling bias correction has no impact on the projections. This is due to the fact that methods measuring future changes are absolute change for temperature and relative change for precipitation.

LR is a basic linear algorithm, suitable for resolving regression problems across multiple models or members in an ensemble. It fits a linear model to minimize the sum of squared errors. Its general form can be written as: $Y = a_0 + A \cdot X$, where $X(i, k)$ is the input spatial field ($i = 1, \dots, 928$) from the 24 models ($k = 1, \dots, 24$) and $Y(i)$ is the output spatial field

($i = 1, \dots, 928$). The regression coefficients a_0 and A_k , ($k = 1, \dots, 24$) were fitted with data in the training period comprising 34 years from 1961 to 1994.

A linear model is not always inferior to non-linear models, depending on the nature of the problem to resolve (Choubin *et al* 2016, Xu *et al* 2020). The practical realization of LR used in this paper was done through the function ‘LR’ in the module ‘sklearn.linear_model’ in Python 3.8 (https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression).

RF solves regression problems by growing an ensemble of decision trees based on binary recursive partitioning (Breiman *et al* 1984, Breiman 2001). Although each individual regression (done at the level of leaves or terminal nodes) is still linear, but it is operated in a reduced range among the total samples. This is why RF can solve non-linear problems and reveal complex behaviors hidden in the data samples. Its randomness manifests in two particular points. Firstly, the samples used to construct each decision tree of the forest is a random subset of the total samples. They are generally drawn with replacement under the strategy of bootstrapping. Secondly, for each partitioning node, only a randomly-formed subset of features is used to split samples into binary branches. The size of this subset is generally around the square root of the number of total features. Under such conditions, RF is quite time consuming for its operation, but it has an excellent performance, with large tolerance to imperfections of samples, and good capacity to avoid overfitting. In our work, the function ‘RF Regressor’ from the Python package ‘sklearn.ensemble’ (Pedregosa *et al* 2012) was used (<https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor>). For the training procedure, we have data covering 34 years, from 1961 to 1994, and 928 spatial points. The total number of samples into our RF training is thus $34 \times 928 = 31\,552$. Each of the 24 climate models is treated as a feature in our RF implementation. After RF is trained, it is used in the testing period from 1995 to 2014 to validate its performance. Similarly, it is used to make the future projection under the specific warming thresholds.

The ‘Bayesian Optimization’ was used to find the best hyperparameters implemented in the RF algorithm (Shahriari *et al* 2016) (<http://rmcantin.github.io/bayesopt/html/bopttheory.html>). It has a higher efficiency than other methods, such as ‘grid search’ or ‘randomized search’. We thus optimized four important parameters of the RF algorithm, the number of trees in the forest (n_estimators), the maximum depth of the tree (max_depth), the minimum number of samples required to split an internal node (min_samples_split), and the number of features to consider when looking for the best split (max_features). Within 30 iterations, the Bayesian optimization process generally converges to optimal

parameters for a specific climate index. The optimal parameters for the six indices are shown in table S2.

As other statistical tools, ML methods do not inspire confidence if we cannot ensure an appropriate interpretation on their derived features, patterns, and rules. In the RF model which consists of establishing a set of decision trees with internal nodes and leaves, the importance of input features or variables (climate models in our case) can be measured by the variance reduction attributed to each feature (total variance before the splitting node minus the sum of the same variance in the two split groups). In our case of multiple decision trees, the final measure of importance is the sum from all trees in the forest. It is furthermore normalized among all features or variables to ensure that the total sum is unity. This ‘relative importance’ can help understanding the importance of each climate model in the ensemble-processing strategy. Relevant analysis and results are shown in supplementary materials text S1 and figure S1.

2.2.3. Skill evaluation metrics

Taylor diagram (Taylor 2001) and skill score are standard tools providing a concise statistical summary of spatial characteristics between the simulation and observation. The Taylor diagram can show three aspects of statistical information: pattern correlation coefficient, a ratio of the centered standard deviations, and root mean square error, any two of them being independent (Li *et al* 2021). A good simulation would be that both the pattern correlation coefficient and the ratio of standard deviations are close to one, and the root mean square error is close to zero (Taylor 2001, Jiang *et al* 2015).

Taylor skill score (TSS), calculated as in equation (3), is a numerical summary of the Taylor diagram to express a synthetic measure:

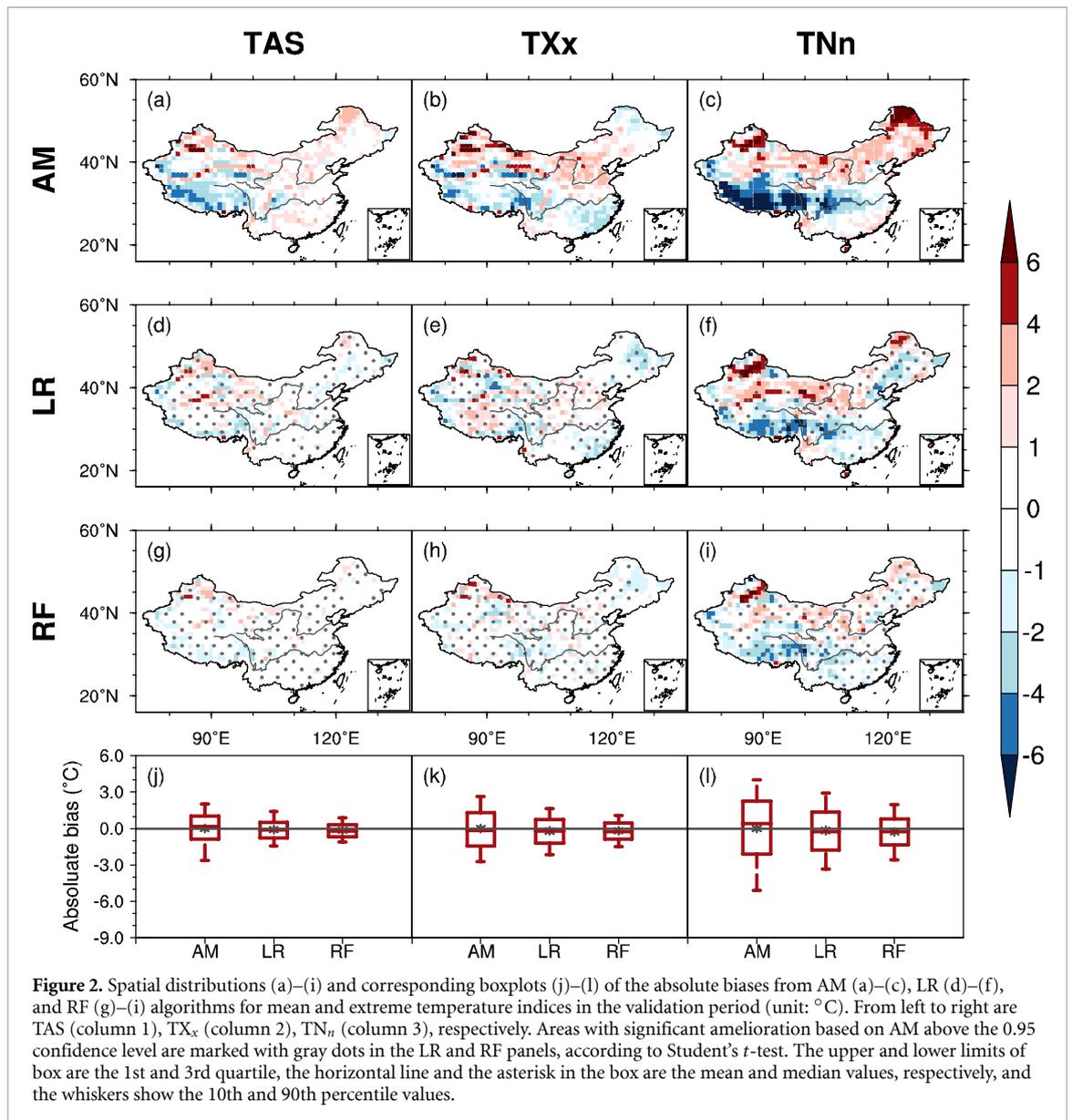
$$\text{TSS} = \frac{4(1 + R_m)^2}{\left(\frac{\sigma_m}{\sigma_0} + \frac{\sigma_0}{\sigma_m}\right)^2 (1 + R_0)^2} \quad (3)$$

where R_m is the spatial correlation coefficient of climatological mean between simulation and observation, R_0 is the maximum correlation coefficient attainable set here to 0.999, σ_m and σ_0 are the standard deviations of the simulated and observed spatial patterns in climatological means, respectively. The closer the value of TSS is to one, the better the agreement between the simulation and observation. This skill score has been generally used in many previous researches (Wang *et al* 2018, Zhu *et al* 2020a, Ngoma *et al* 2021).

3. Results

3.1. Performance evaluation

To assess the ability of our three schemes dealing with the multi-model ensemble simulations, the spatial patterns and corresponding distribution boxplots for



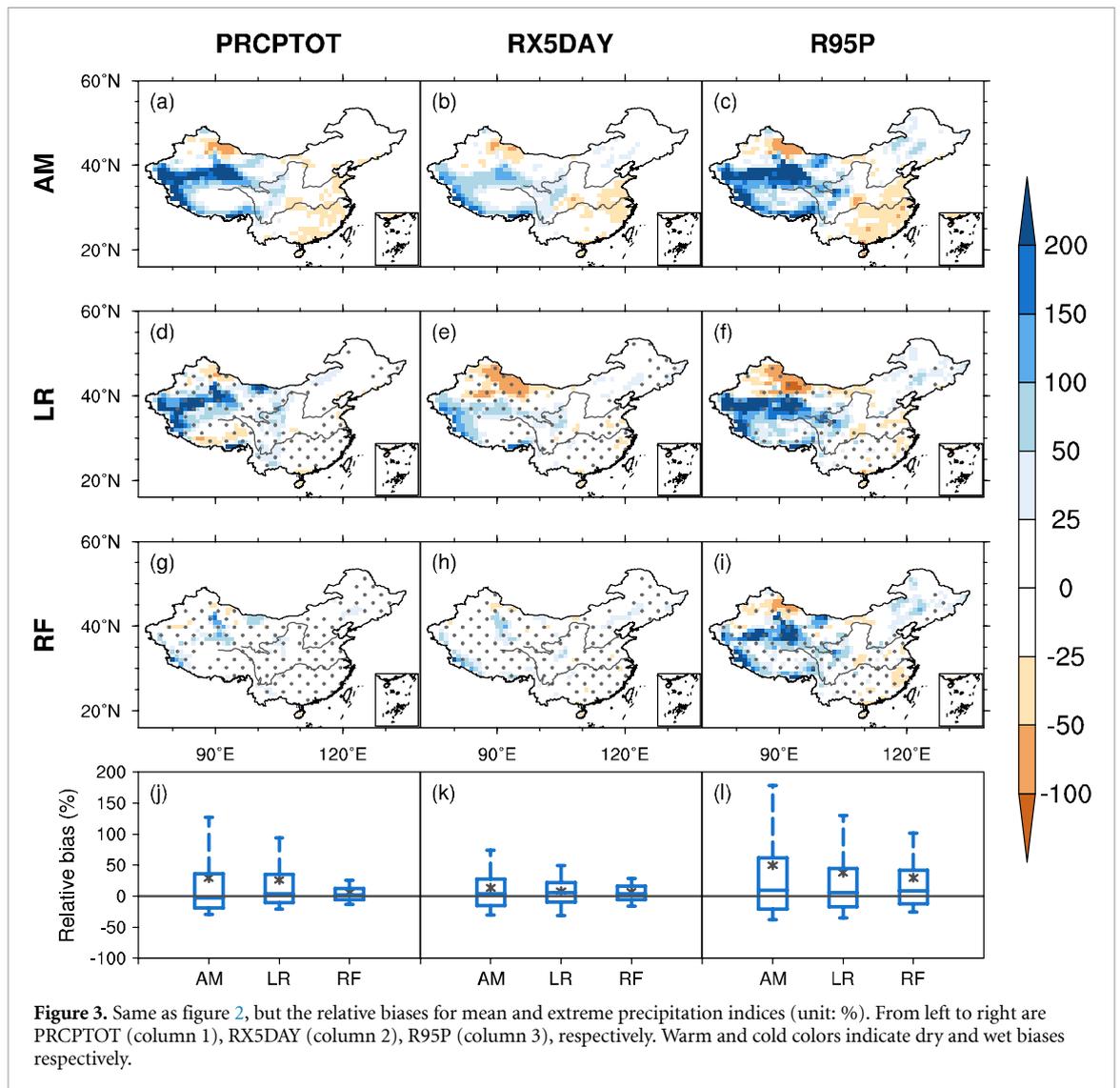
biases of all indices against observations across China during the validation period are examined (figures 2 and 3). Darker colors and far-from-zero bars represent higher deviations from observation. To facilitate visual inspection and interpretation of differential fields, the climatology from observation in the validation period from 1995 to 2014 is exhibited in figure S2. A general feature that can be observed in figures 2 and 3 is that the three schemes exhibit similar patterns of spatial bias distribution, and AM (with a bias correction included) shows the largest biases. Compared with AM, biases from LR and RF are reduced across almost the whole domain.

Cold biases ($\Delta T < -6$ °C) from AM are mainly concentrated in the Tibetan Plateau and the middle and upper reaches of the Yangtze River for all temperature indices. They are significantly reduced in LR and RF (figures 2(d)–(i) vs (a)–(c)). The amelioration of RF is especially remarkable, only some scattered areas exist with bias exceeding 2 °C. Higher cold

biases (with focus on the 10th percentile biases in boxplots) depict a decrease from AM to RF, with values -2.62 °C to -1.12 °C for TAS, -2.73 °C to -1.46 °C for TX_x , and -5.07 °C to -2.58 °C for TN_n .

Similar characteristic holds true for precipitation indices (figures 3(d)–(i) vs (a)–(c)). Areas with large biases in AM are reduced in LR and RF, especially in the Tibetan Plateau and its periphery where there are the largest wet biases. Higher wet biases (with focus on the 90th percentile in the boxplots) are reduced from 127% in AM to 25% in RF for PRCPTOT. Similarly, RX5DAY shows a reduction from 74% to 28%, and R95P from 178% to 101%. RF is the best performing, and the biases are lower than 50% over almost the whole territory of China for PRCPTOT and RX5DAY. Higher wet biases for R95P exist in the Tarim Basin and the Qilian Mountains with complex topography.

Taylor diagram and TSS are presented in figure 4 to show a concise statistical analysis of the three



ensemble-processing strategies in the evaluation period. There is a general weak performance with AM (gray markers). The correlation coefficients of all temperature indices deduced from AM are between 0.94 and 0.97, the standardized deviations vary between 0.96 and 1.05, and the TSSs are lower than 0.97. LR and RF schemes show an extra improvement, compared to AM. The best-performing RF gives correlation coefficient, and TSSs all superior to 0.98 and 0.99. Precipitation indices from AM show an unsatisfactory performance, with all spatial correlation coefficients less than 0.88, standardized deviations between 0.64 and 0.80, and the lowest value of TSSs reaching only 0.71. RF has the best efficiency, with precipitation indices comparable to temperature indices. In terms of TSSs, RF improves them from 0.82 in AM to 0.98 for PRCPTOT, from 0.79 to 0.95 for RX5DAY and from 0.71 to 0.89 for R95P.

Overall, the results provide clear evidence that LR and RF schemes effectively enhance the capability of reproducing the spatial climate characteristics in China, especially in western China where, with

complex topography, most significant biases manifest in AM. RF has the best performance, with TSSs of all temperature indices at the level of 0.98 and 0.99, and remarkably improves the skill scores of precipitation indices to a level higher than 0.89. Temperature indices generally have a better performance than precipitation, but the improvement for precipitation indices is more significant and substantial.

Beyond the mean state, it is also interesting to check how well our ensemble-processing schemes can produce their interannual variability. We now assess the temporal standard deviation during the validation period from 1995 to 2014, with the interannual standard deviation from observation shown in figure S3. The result of evaluation is shown in figure S4, and expressed as a ratio of standard deviations between the simulation and observation. Only the mean states of TAS and PRCPTOT are exhibited as illustration. This ratio is generally smaller than 1.0, reflecting the fact that the ensemble-processing strategies present a reduced interannual variability. Such a result is expected, since any ensemble-processing strategy, due to

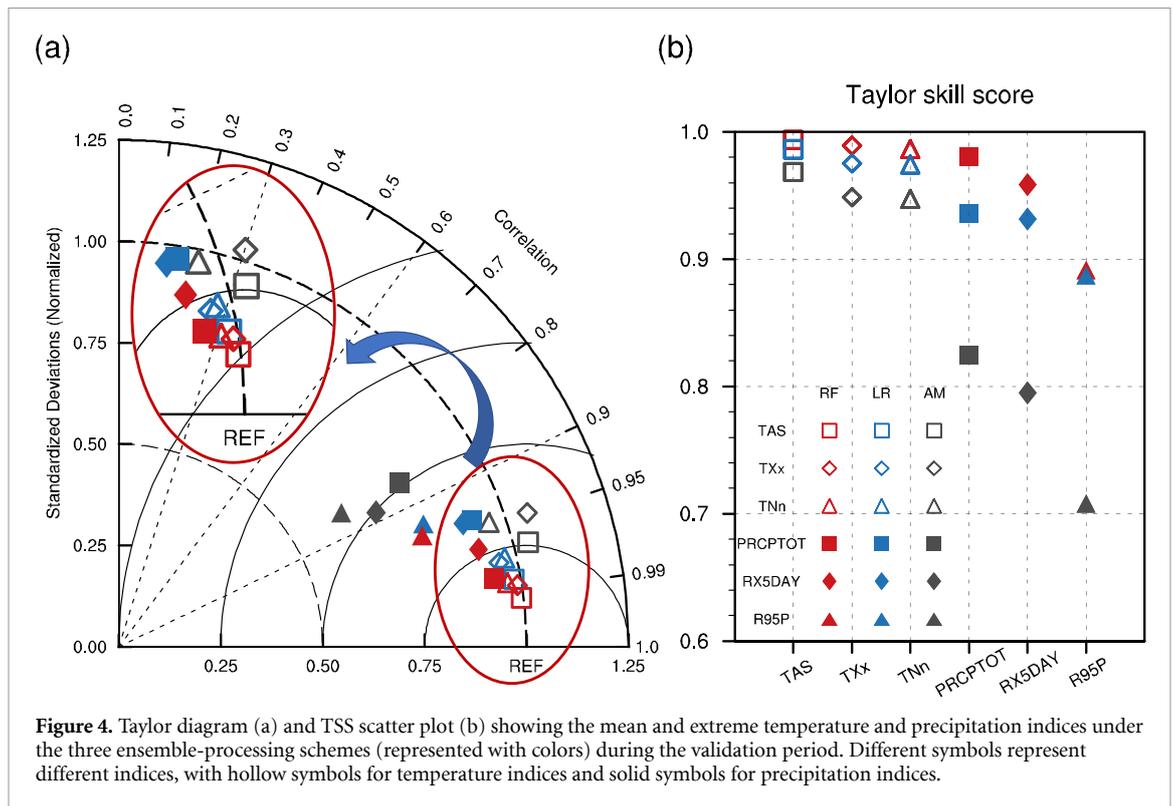


Figure 4. Taylor diagram (a) and TSS scatter plot (b) showing the mean and extreme temperature and precipitation indices under the three ensemble-processing schemes (represented with colors) during the validation period. Different symbols represent different indices, with hollow symbols for temperature indices and solid symbols for precipitation indices.

its nature of mixing different simulations, reduces the interannual variability. In the case of AM, if all members in the ensemble are sequentially independent and possess an identical standard deviation, then the ensemble average from N members would reduce the standard deviation by a factor of $1/\sqrt{N}$. In our configuration of 24 models, this factor is about 0.20. The actual ratio for the mean temperature indices is larger than this expected value, but its counterpart for precipitation indices is smaller (all indices are not shown). We believe that this behavior is due to the fact that temperature indices have a consistent warm trend among models, but precipitation indices do not. Let us now inspect the cases of RF and LR, since a regression relationship is used to combine the 24 models (or a subset), the reduction of interannual variability is less pronounced. It is necessary to point out that when the regression is ill-fitted (with large negative coefficients for certain members, for example), the interannual variability can even be augmented.

3.2. Projection of future climate

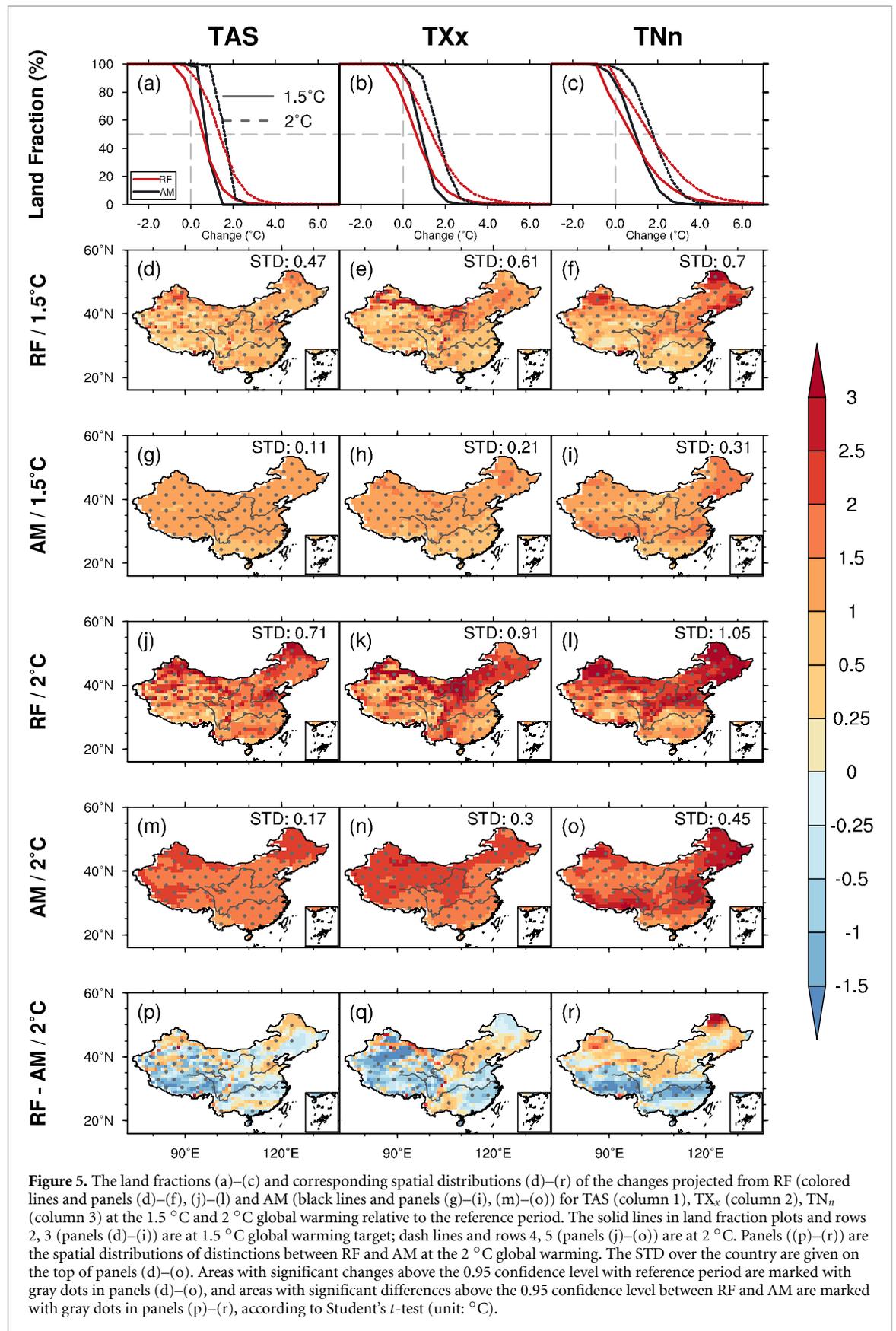
Given its good performance in dealing with multi-model ensemble simulations, RF is now used for the regional projection of future climate for 1.5 °C, 2 °C and 3 °C global warming targets (relative to preindustrial), under the SSP5-8.5 emission scenario. The widely-used AM scheme is also shown as a baseline and reference. As a conventional practice, the target warming levels are relative to pre-industrial

(1861–1900), while the projected changes are relative to 1995–2014. For the sake of conciseness, only temperature and precipitation indices under 1.5 °C and 2 °C warming targets are given in the main text, the results under 3 °C warming target being placed in supplementary materials.

3.2.1. Temperature indices

The land fraction from whole China territory with projected changes exceeding the abscissa's values is plotted in figures 5(a)–(c) in the form similar to a curve of the cumulative probability distribution function. Results are shown for both RF and AM, for all temperature indices, and for the 1.5 °C and 2 °C warming targets, respectively. Figures 5(d)–(o) show their corresponding spatial pattern of changes, while the difference between RF and AM under the 2 °C warming level is shown in figures 5(p)–(r).

Let us firstly examine the median value which is an emblematic figure since it separates the entire territory across China into two equal halves. Changes of mean and extreme temperature projected by RF are lower than those by AM (figure 5). Under the 2 °C warming target, but relative to nowadays, RF shows a median change of TAS, TX_x, and TN_n at 1.35 °C, 1.37 °C and 1.64 °C, which are lower than the counterpart in AM, by about 0.23 °C, 0.31 °C and 0.15 °C. Recent studies based on CMIP6 models show higher transient climate response and equilibrium climate sensitivity than what shown by previous versions of these models in CMIP5. Consequently, the projection of future



climate in CMIP6 is also stronger than in CMIP5 (Gettelman *et al* 2019, Nijse *et al* 2020, Zelinka *et al* 2020). However, with some observational constraints, the projected warming is reduced compared with

non-constrained projection (Brunner *et al* 2020, Liang *et al* 2020, Tokarska *et al* 2020). Our results of multi-model ensemble projection seem to agree with this conclusion. Observation plays important

role in the ML RF ensemble-processing scheme, similar to a role of observation-based constraining, which lowers the projected warming compared to unconstrained AM.

Obvious differences are detected between the land fraction curves of RF and AM. Extreme changes have higher probability of occurrence in more areas in RF as its curves have longer tails. Under the 2 °C global warming target, AM does not project warmer mean temperature exceeding 2.6 °C, but RF suggests a likelihood of 9% over China with such a warming level. In terms of geographic distribution of TAS, larger spatial variability is detected in RF, as the spatial standard deviations (STD) are almost twice larger than that in AM (figures 5(j) vs (m)). Large magnitudes of warming projected by RF are found in the western part of Northeastern China and the north part of Northwestern China under 1.5 °C warming. Under the 2 °C warming target, the warming in these areas would further expand and strengthen, the northern and eastern periphery of the Tibetan Plateau also shows significant warming, exceeding 2.5 °C. Meanwhile, AM projects a smoother distribution, with warming uniformly enhanced (exceeding 2.5 °C warming) in the area north of 45° N and part of the Tibetan Plateau. From the difference between RF and AM (figure 5(p)), it is clear that, except a few regions with drastic increases in RF, the warming projected from RF is generally lower about 0.25 °C–1 °C than that of AM in almost the whole country, especially in the Tibetan Plateau, where the difference is significant under the 0.95 confidence level.

Regarding the spatial pattern, a similar behavior holds for extreme temperature TX_x . Significantly enhanced warming over Northeastern China, the Tianshan Mountains, as well as the Loess Plateau is projected in RF, with a magnitude of 2.5 °C above current world under the 2 °C warming target. Areas with larger increases from AM are evenly distributed in Northeastern China and the entire Northwest region. The change of TX_x in different regions has large distinction in the projection of RF, the STD are more than three times larger than that in AM under both 1.5 °C and 2 °C warming targets (figures 5(e) vs (h) and figures 5(k) vs (n)).

For the minimum temperature TN_n , sensitive areas from RF are distributed in Northeastern China, in the Yellow River Basin and in sparse areas in Northwestern China, while the warming projected by AM is more widely distributed in the southeast, extending to the south of the Yangtze River Basin.

These results show broad similarities with those from GCMs (Shi *et al* 2018, Sui *et al* 2018, Yang *et al* 2018), i.e. Northwestern China, Northeastern China, and the Tibetan Plateau are particularly sensitive to global warming. Compared to AM, RF shows more detailed information and larger inhomogeneity, and it exhibits a closer correlation with topography. More pronounced hotspots can be observed in RF.

3.2.2. Precipitation indices

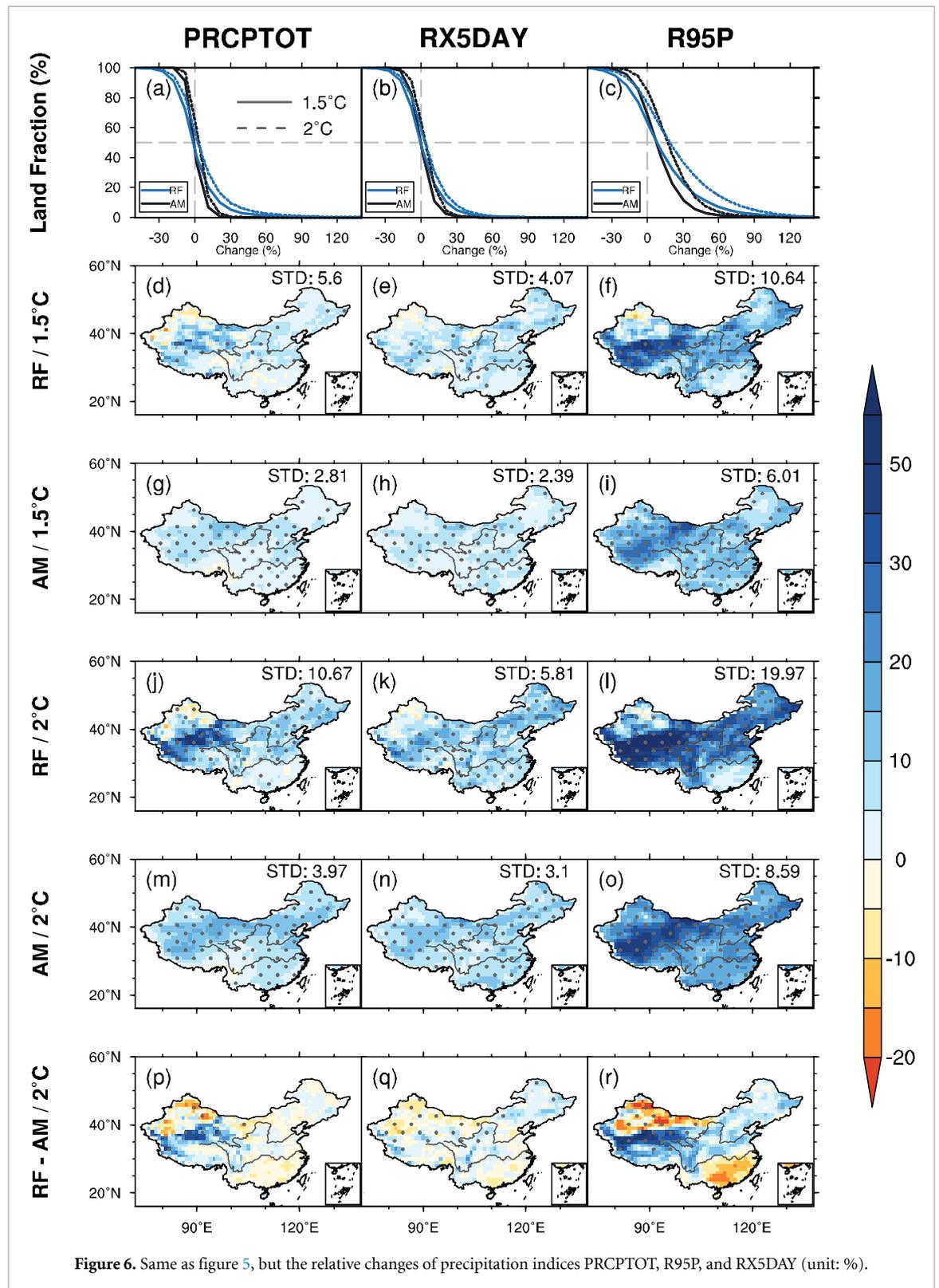
Mean and extreme precipitation projections are presented in figure 6, both RF and AM project increased precipitation over most of China in response to global warming. For the median value across China that separates the whole territory into two equal halves, RF shows an increase of 3%, 4%, and 19% for PRCPTOT, RX5DAY, and R95P under the 2 °C global warming, which is almost the same as the counterpart in AM (only about 0.6%, 0.1% and 1.5% higher). For the change of total precipitation (PRCPTOT) under the 1.5 °C and 2 °C global warming targets, the fraction of lands where increase exceeds 40% is projected to be almost non-existent over China in AM, while that fraction in RF is above 6%. That means a higher risk for intense precipitation in RF projection.

In terms of geographic distribution, RF and AM show good consistency, but there are substantial differences of magnitude (figures 6(d)–(r)). Small-scale features in RF are more significant, and the amplitude of increase is also higher. Large-increase areas of total precipitation (more than 30%) in RF are concentrated in the region of the Tsaidam Basin and Qilian Mountains. For the case of AM, enhanced precipitation (10%–20%) is more evenly distributed in the whole western area, extends from the Tibetan Plateau, northeastward, stretching to the Loess Plateau and its northern area. In the northwestern area, significant lower precipitation change is projected by RF compared with AM. In other areas, the changes of total precipitation projected by RF are generally more notable than that in AM.

Changes of RX5DAY show a close resemblance to total precipitation in terms of intensity and main geographic patterns. But precise areas of remarkable increase have some differences, especially in RF. Significant enhancement is found in most part of the Tibetan Plateau and patchy areas in Northeastern China, where the magnitudes exceed 20% under the 2 °C global warming. Contrasted with RF, AM suggests smaller increases of RX5DAY, but with a more homogeneous geographic distribution. Almost all the territory would see an increase within the 15% threshold. As shown in figure 5(q), significant differences between AM and RF are found in the northwestern region.

Changes of R95P exceeding 50% in RF concentrate in the Tibetan Plateau and the Yellow River Basin, where the changes are almost twice higher than in AM (higher about 20%–30%). Meanwhile, in the southeastern and northwestern regions, the projected increase in strong precipitation from RF are not noticeable, which is lower than that projected by AM.

Further comparison of our RF projections with previous studies using high-resolution regional climate models (RCMs) shows some similarities, especially in complex-terrain areas. Zhu *et al* (2020b), using WRF v3.7.1, showed that, for total precipitation



and extreme events, the Tibetan Plateau and regions outside China's northwestern boundaries are particularly sensitive to climate change, conclusion very consistent with our results. Similar patterns from our RF projection for RX5DAY were also present in Li *et al* (2018b) using five RCMs involved in the CORDEX-East Asia project. Our results are also comparable to Li *et al* (2018a) using FROALS as a

dynamical downscaling model, together with a statistical downscaling tool. It is worthy of note that our projected R95P pattern in RF is very close to what found with WRF v3.5.1 when it was applied to China (Bao *et al* 2015).

The ML RF algorithm uses the concept of multi-regression decision trees. It can efficiently solve non-LR problems and achieve good matching to

observation, in both temporal and spatial domains, as well demonstrated in Crawford *et al* (2019) and Pang *et al* (2017). Our results shown here are consistent with its intrinsic properties and with our expectation on it.

4. Conclusion and discussion

In this work, three different ensemble-processing strategies, AM, LR, and RF (ML decision tree algorithm), are used to explore information offered by the multi-model ensemble climate simulations of CMIP6. The main idea was to find the best way of processing the ensemble simulations to mimic observational climatic properties and to give a more reliable projection of future climate. AM is the simplest and most intuitive strategy. LR advocates the vision of a LR approach to establish the relationship between simulations and observations, but it cannot necessarily represent any physical rules governing the climate system. RF is one of the most advanced ML algorithms. It can extract non-linear and complex relations among climate models, instead of making a simple evaluation of models' apparent performance as in other ensemble-processing strategies. This leads to a hybrid approach that we advocate for climate change issues, which combines physical modeling and ML strengths, thus giving confidence in retrieving more valuable information.

The performance of the three schemes was assessed in the validation period (20 years, from 1995 to 2014). Compared with AM, LR and RF effectively enhance the capability of capturing spatial climate characteristics over China. Improvement in areas with complex terrain is the most significant such as in the periphery of the Tibetan Plateau. RF performs well, with the TSS of temperature indices being of 0.98 and 0.99, and that of precipitation indices higher than 0.89. It was also revealed that the internal variability, such as the interannual-scale standard deviation, can not be correctly reproduced by any of our ensemble-processing strategies which were designed, after all, to calculate the mean state of our expectation.

After an inter-comparison of performance, RF was selected as the optimal scheme and used to investigate climate changes in the 1.5 °C, 2 °C and 3 °C warmer worlds under the SSP5-8.5 emission scenario. Compared with AM, RF shows less warming and enhanced wet conditions at the national scale of China. In terms of median changes across China, mean temperature (TAS), annual maximum (hottest daytime) temperature (TX_x), and annual minimum (coldest nighttime) temperature (TN_n) show 1.35 °C, 1.37 °C and 1.64 °C warming relative to 1995–2014 period, respectively, under the 2 °C global warming level, when RF is used. They are lower than their counterpart in AM, especially for TX_x, lower about 0.31 °C. The median changes of total precipitation in wet days (PRCPTOT), annual maximum consecutive

5 d precipitation amount (RX5DAY), and annual total precipitation for events exceeding the 95th percentile (R95P) projected in RF are 3%, 4%, and 19%, respectively, similar with the counterpart in AM.

Regarding the geographic distribution, RF would see larger warming in Northeastern China and the northern part of Northwestern China. Tianshan Mountain, Loess Plateau area for TX_x, and the Yellow River Basin for TN_n are also regions of hotspots. Meanwhile except the regions with intensified warming, the warming projected from RF is generally lower than that of AM. That indicates a larger spatial variability and more pronounced local-scale characteristics of RF. For the projection of TX_x, the spatial standard deviation can be three times larger compared with that in AM.

RF also projects more intense precipitation in most part of China. For example, in the region of the Tsaidam Basin and the Qilian Mountains, the projected changes in RF for the strong precipitation (partly exceeding 50% under the 2 °C warming) are almost twice higher than in AM. Meanwhile in the northwestern area, for all precipitation indices, weaker increases of precipitation compared with AM are projected by RF. AM shows however much more homogeneous features.

It is interesting to point out that the geographic structure of climate projection in RF shows a resemblance to that from dynamical downscaling with high-resolution models or from statistical downscaling (Li *et al* 2018a, Zhu *et al* 2020b). This indicates that the ML algorithm RF could capture detailed information at local scale, certainly due to its ability to behave as do those dynamic models with higher spatial resolution. This is quite reasonable since the high-resolution observation seems to play its role in constraining the ensemble-processing strategy RF which is able to manipulate complex nonlinear processes across multiple models. We believe that using advanced ML techniques can provide a new perspective to retrieve more information from large amounts of data and make more reliable climate projections.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://esgf-node.llnl.gov/search/cmip6/>.

Acknowledgments

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 2017YFA0603804 and Grant 2018YFC1507704), and

the Postgraduate Research and Practice Innovation Program of Government of Jiangsu Province (KYCX21_0940).

References

- Alizamir M, Azhdary Moghadam M, Hashemi Monfared A and Shamsipour A 2018 Statistical downscaling of global climate model outputs to monthly precipitation via extreme learning machine: a case study *Environ. Prog. Sustain. Energy* **37** 1853–62
- Bao J, Feng J and Wang Y 2015 Dynamical downscaling simulation and future projection of precipitation over China *J. Geophys. Res.* **120** 8227–43
- Breiman L 2001 Random forests *Mach. Learn.* **45** 5–32
- Breiman L, Friedman J, Olshen R and Stone C J 1984 *Classification and Regression Trees* Mathematics, Computer Science, Medicine (New York: Wadsworth International Group) p 368
- Brunner L, Pendergrass A G, Lehner F, Merrifield A L, Lorenz R and Knutti R 2020 Reduced global warming from CMIP6 projections when weighting models by performance and independence *Earth Syst. Dyn.* **11** 995–1012
- Choubin B, Khalighi-Sigaroodi S, Malekian A and Kisi Ö 2016 Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals *Hydrol. Sci. J. Des Sci. Hydrol.* **61** 1001–9
- Crawford J, Venkataraman K and Booth J 2019 Developing climate model ensembles: a comparative case study *J. Hydrol.* **568** 160–73
- Eyring V, Bony S, Meehl G A, Senior C A, Stevens B, Stouffer R J and Taylor K E 2016 Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization *Geosci. Model Dev.* **9** 1937–58
- Fu C, Jiang Z, Guan Z, He J and Xu Z 2008 Impacts of climate change on water resources and agriculture in China *Regional Climate Studies of China* Regional Climate Studies (Berlin: Springer) pp 447–64
- Gettelman A, Hannay C, Bacmeister J T, Neale R, Pendergrass A G, Danabasoglu G, Lamarque J, Fasullo J T, Bailey D A and Lawrence D M 2019 High climate sensitivity in the community earth system model version 2 (cesm2) *Geophys. Res. Lett.* **46** 8329–37
- Guo L, Jiang Z, Chen D, Le Treut H and 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** 623–43
- Guo L, Jiang Z, Ding M, Chen W and Li L 2019 Downscaling and projection of summer rainfall in eastern China using a nonhomogeneous hidden Markov model *Int. J. Climatol.* **39** 1319–30
- Ham Y G, Kim J H and Luo J J 2019 Deep learning for multi-year enso forecasts *Nature* **573** 568–72
- Hidalgo H G and 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
- Hulme M 2016 1.5 °C and climate research after the Paris agreement *Nat. Clim. Change* **6** 222–4
- Jiang Z H, Li W, Xu J J and Li L 2015 Extreme precipitation indices over China in CMIP5 models. Part I: model evaluation *J. Clim.* **28** 8603–19
- Kadow C, Hall D M and Ulbrich U 2020 Artificial intelligence reconstructs missing climate information *Nat. Geosci.* **13** 408–13
- Knutti R, Furrer R, Tebaldi C, Cermak J and Meehl G A 2010 Challenges in combining projections from multiple climate models *J. Clim.* **23** 2739–58
- Knutti R, Masson D and 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 and Eyring V 2017 A climate model projection weighting scheme accounting for performance and interdependence *Geophys. Res. Lett.* **44** 1909–18
- Lenderink G, Buishand A and van Deursen W 2007 Estimates of future discharges of the river Rhine using two scenario methodologies: direct versus delta approach *Hydrol. Earth Syst. Sci.* **11** 1145–59
- Li D, Zou L and Zhou T 2018a Extreme climate event changes in China in the 1.5 °C and 2 °C warmer climates: results from statistical and dynamical downscaling *J. Geophys. Res.* **123** 10215–30
- Li H, Chen H, Wang H and Yu E 2018b Future precipitation changes over China under 1.5° C and 2.0° C global warming targets by using cordex regional climate models *Sci. Total Environ.* **640–641** 543–54
- Li T, Jiang Z, Zhao L and Li L 2021 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 X, Li Z, Huang W and Zhou P 2020 Performance of statistical and machine learning ensembles for daily temperature downscaling *Theor. Appl. Climatol.* **140** 571–88
- Liang Y, Gillett N P and Monahan A H 2020 Climate model projections of 21st century global warming constrained using the observed warming trend *Geophys. Res. Lett.* **47** e2019GL086757
- Ngoma H, Wen W, Ayugi B, Babaousmail H, Karim R and Ongoma V 2021 Evaluation of precipitation simulations in CMIP6 models over uganda *Int. J. Climatol.* **41** 4743–68
- Nijssen F J M M, Cox P M and Williamson M S 2020 An emergent constraint on transient climate response from simulated historical warming in CMIP6 models *Earth Syst. Dyn.* **2020** 1–14
- Palmer T N, Shutts G J, Hagedorn R, Doblasreyes F J, Jung T and Leutbecher M 2005 Representing model uncertainty in weather and climate prediction *Annu. Rev. Earth Planet. Sci.* **33** 163–93
- Pang B, Yue J, Zhao G and Xu Z 2017 Statistical downscaling of temperature with the random forest model *Adv. Meteorol.* **2017** 1–11
- Pedregosa F et al 2012 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30
- Piao S, Ciais P, Huang Y, Shen Z, Peng S, Li J, Zhou L, Liu H, Ma Y and Ding Y 2010 The impacts of climate change on water resources and agriculture in China *Nature* **467** 43–51
- Reichstein M, Camps-Valls G, Stevens B, Jung M, Denzler J, Carvalhais N and Prabhat 2019 Deep learning and process understanding for data-driven Earth system science *Nature* **566** 195–204
- Robertson A W, Lall U, Zebiak S E and Goddard L 2004 Improved combination of multiple atmospheric GCM ensembles for seasonal prediction *Mon. Wea. Rev.* **132** 2732–44
- Sanderson B M, Knutti R and Caldwell P 2015 Addressing interdependency in a multimodel ensemble by interpolation of model properties *J. Clim.* **28** 5150–70
- Semenov M A and Stratonovitch P 2010 Use of multi-model ensembles from global climate models for assessment of climate change impacts *Clim. Res.* **41** 1–14
- Shahriari B, Swersky K, Wang Z, Adams R P and de Freitas N 2016 Taking the human out of the loop: a review of Bayesian optimization *Proc. IEEE* **104** 148–75
- Shi C, Jiang Z H, Chen W L and Li L 2018 Changes in temperature extremes over China under 1.5 °C and 2 °C global warming targets *Adv. Clim. Change Res.* **9** 120–9
- Sui Y, Lang X and Jiang D 2018 Projected signals in climate extremes over China associated with a 2 °C global warming under two RCP scenarios *Int. J. Climatol.* **38** e678–e697

- Sun C, Jiang Z, Li W, Hou Q and Li L 2019 Changes in extreme temperature over China when global warming stabilized at 1.5 °C and 2.0 °C *Sci. Rep.* **9** 14982
- Tan J, Jiang Z and Ma T 2016 Projections of future surface temperature change and uncertainty over China based on Bayesian model averaging *Acta Meteor. Sin.* **74** 583–97
- Taylor K E 2001 Summarizing multiple aspects of model performance in a single diagram *J. Geophys. Res.* **106** 7183–92
- Teutschbein C and Seibert J 2012 Bias correction of regional climate model simulations for hydrological climate-change impact studies: review and evaluation of different methods *J. Hydrol.* **456–457** 12–29
- Tokarska K B, Stolpe M B, Sippel S, Fischer E M, Smith C J, Lehner F and Knutti R 2020 Past warming trend constrains future warming in CMIP6 models *Sci. Adv.* **6** eaaz9549
- UNFCCC: United Nations Framework Convention on Climate Change 2015 Decision 1/cp.21. The Paris agreement (available at: <http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf> (Accessed 18 August 2021))
- Wang B, Zheng L, Liu D L, Ji F, Clark A and Yu Q 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** 4891–902
- Watson-Parris D 2020 Machine learning for weather and climate are worlds apart *Phil. Trans. R. Soc. A* **379** 20200098
- Wu J and Gao X J 2013 A gridded daily observation dataset over China region and comparison with the other datasets *Chin. J. Geophys.* **56** 1102–11 (in Chinese)
- Xu L, Chen N, Zhang X and Chen Z 2020 A data-driven multi-model ensemble for deterministic and probabilistic precipitation forecasting at seasonal scale *Clim. Dyn.* **54** 3355–74
- Xu Y, Gao X and Giorgi F 2010 Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections *Clim. Res.* **41** 61–81
- Yang Y, Tang J, Wang S and Liu G 2018 Differential impacts of 1.5 °C and 2 °C warming on extreme events over China using statistically downscaled and bias-corrected CESM low-warming experiment *Geophys. Res. Lett.* **45** 9852–60
- Zelinka M D, Myers T A, Mccoy D T, Pochedley S, Caldwell P, Ceppi P, Klein S A and Taylor K E 2020 Causes of higher climate sensitivity in CMIP6 models *Geophys. Res. Lett.* **47** e2019GL085782
- Zhang X, Alexander L, Hegerl G C, Jones P, Tank A K, Peterson T C, Trewin B and Zwiers F W 2011 Indices for monitoring changes in extremes based on daily temperature and precipitation data *Wiley Interdiscip. Rev. Clim. Change* **2** 851–70
- Zhu H, Jiang Z, Li J, Li W, Sun C and Li L 2020a Does CMIP6 inspire more confidence in simulating climate extremes over China? *Adv. Atmos. Sci.* **37** 1119–32
- Zhu X, Wei Z, Dong W, Ji Z, Wen X, Zheng Z, Yan D and Chen D 2020b Dynamical downscaling simulation and projection for mean and extreme temperature and precipitation over central Asia *Clim. Dyn.* **54** 3279–306