An evaluation of CMIP5 precipitation simulations using ground observations over ten river basins in China
Xiuqin Yang, Bin Yong, Zhiguo Yu and Yuqing Zhang

ABSTRACT
Using the precipitation measurements obtained from 2,419 ground meteorological stations over China from 1960 to 2005 as benchmark, the performance of 21 single-mode precipitation data from the Coupled Model Intercomparison Project Phase 5 (CMIP5) were evaluated using Taylor diagrams and several statistical metrics. Based on statistical metrics, the models were ranked in terms of their ability to reproduce similar patterns in precipitation relative to the observations. Except in Southeast and Pearl river basins, research results show that all model ensemble means overestimate in the rest of the river basins, especially in Southwest and Northwest. The performance of CMIP5 models is quite different among each river basin; most models show significant overestimation in Northwest and Yellow and significant underestimations in Southeast and Pearl. The simulations are more reliable in Songhua, Liao, Yangtze, and Pearl than in other river basins according to spatial distribution and interannual variability. No individual model performs well in all the river basins both spatially and temporally. In Songhua, Liao, Yangtze, and Pearl, precipitation indices are more consistent with observations, and the spread among models is smaller. The multimodel ensemble selected from the most reasonable models indicates improved performance relative to all model ensembles.

Key words | China, CMIP5 precipitation, ground precipitation observations, ten river basins

HIGHLIGHTS
- The performance of CMIP5 models were evaluated using the precipitation measurements obtained from meteorological stations.
- The performance of CMIP5 models is quite different among each river basin.
- No individual model performs well in all the river basins both spatially and temporally.
- The multimodel ensemble selected from the most reasonable models indicates improved performance relative to all model ensembles.

INTRODUCTION
A direct consequence of global warming is changes in precipitation patterns because warmer temperatures accelerate evapotranspiration, change surface energy balance, and alter atmospheric circulation patterns (Durack et al. 2012; Berg et al. 2013; Guilbert et al. 2015). A 1.0 °C rise in global mean temperature is projected to increase global mean precipitation by 1–3% (IPCC 2013; Pendergrass & Hartmann 2014a, 2014b) and, as the warming continues to
accelerate, the rate of increase in global mean precipitation may also escalate (Hirabayashi et al. 2013; IPCC 2013; Chen et al. 2017). Despite the positive correlation between changes in global mean temperature and precipitation, the relationship remains unclear on regional scale because of the highly variable nature of precipitation on this scale (Guilbert et al. 2015). However, it is regional precipitation that has important effects on socioeconomics due to potentially severe impacts for water resources and agriculture (Cao et al. 2011; Wehner 2013; Wang et al. 2014; Li et al. 2015; Chen et al. 2017). Thus, it is crucial to know how well general circulation models (GCMs) can simulate current precipitation patterns on regional scale, as the primary tools for future climate projections under different greenhouse gas emission scenarios.

Phase 5 of the Coupled Model Intercomparison Project (CMIP5) (Taylor et al. 2012) combined the world’s leading GCMs to produce an ensemble of future climate projections that laid the scientific foundation for the assessment of climate change and its impact across the globe, by the world’s authority on this important global issue, the Intergovernmental Panel on Climate Change (IPCC). Since the debut of CMIP5 results in 2014, a series of studies have been conducted to assess the ability of CMIP5 models in simulating historical precipitation on global, regional, and basin scales (Liu et al. 2012; Chadwick et al. 2013; Joetzjer et al. 2013; Kharin et al. 2013; Mehran et al. 2014; Guilbert et al. 2015; Jiang et al. 2015; Li et al. 2016; Nguyen et al. 2017; Sun et al. 2018; Wu et al. 2018; Katiraei-Boroujerdy et al. 2019; Yang et al. 2019). By comparing the GCM simulated precipitation and observations in different parts of the world and during different periods, these studies revealed considerable variations in the skill of CMIP5 GCMs in capturing regional precipitation intensity and variability (Sillmann et al. 2013) and they attributed the large scatter to differences in model forcing, magnitude of the internal variability, and climatic sensitivity of individual models in different regions (Liu et al. 2012; Gaetani & Mohino 2013; Huang et al. 2013; Chen & Frauenfeld 2014; Mehran et al. 2014; Wang & Chen 2014; Jiang et al. 2015; Nguyen et al. 2017).

Several studies have evaluated the performance of CMIP5 GCMs in simulating precipitation in China (Chen & Frauenfeld 2014; Yue et al. 2016). Their results showed that the CMIP5 models can reproduce the spatial pattern and seasonal variability characteristics of the observed precipitation (Su et al. 2013; Chen & Frauenfeld 2014; Zhao et al. 2016). However, the models tend to overestimate seasonal and annual precipitation in western and northern China, but underestimate precipitation, especially for summer season over southeastern China, and the models generally are more skillful in simulating precipitation in eastern than western China (Chen & Frauenfeld 2014; Wang et al. 2016b). These CMIP5 precipitation evaluations were carried out for a specific region, such as northwestern and southern China (Chen et al. 2017; Yang et al. 2019), or areas such as the Tibetan, Loess, and Inner Mongolian Plateaus where precipitation is strongly influenced by topography (Su et al. 2013; Yue et al. 2016).

The current study expands the evaluation domain to the entirety of China with a focus on assessing CMIP5 GCMs’ ability to capture regional differences in precipitation pattern and variability. Although the study domain is the entirety of China, the evaluation is done on basin scale. Specifically, the Chinese mainland is divided into ten major river basins based on topographical and hydrological features and the skill of each CMIP5 model is quantitatively evaluated by comparison with in-situ precipitation observations within the basin. The best and worst performing models in the CMIP5 GCM suite are identified based on a set of comprehensive evaluation scores along with an estimate of uncertainty. In addition to the large domain and the use of multiple evaluation scores, the current study also differs from previous studies of similar nature by utilizing a very large precipitation observational network of more than 2,400 ground stations around China and over a period of more than four decades (1960–2005). Because basin-scale precipitation is directly linked to water resources and agriculture, the results from this study can be used to directly inform stakeholder and policymakers that rely on CMIP5 precipitation projections for future planning.

The remainder of this paper is organized as follows: the next section describes the datasets and methodology utilized. This is followed by a quantitative analysis of CMIP5 precipitation product errors, then a section presents a summary of our evaluation results. The final section provides a discussion and final conclusions.
DATASETS AND METHODOLOGY

The skill of 21 CMIP5 GCMs in simulating seasonal and annual precipitation and interannual variability in ten river basins in China is evaluated by comparison with observations from 2,419 meteorological stations during the historical period of 1960–2005 when long-term ground precipitation observations are available. Several quantitative evaluating metrics are employed to rank the models and identify those of superior performance and those that are least reliable as far as precipitation is concerned. Details about the models, observations and evaluation metrics are given below.

Data

The simulated monthly precipitation values, monthly zonal winds, and geopotential height from 21 CMIP5 models were obtained through data portals of the Earth System Grid Federation via the website http://www.ipcc-data.org/sim/gcm_monthly/AR5/Reference-Archive.html. Information about the models, including model names, resolution, as well as institutions that performed the simulations, are given in Table 1. The spatial resolution differs considerably among the models and to facilitate model intercomparison, the simulated monthly precipitation data from all models was interpolated onto a common grid

Table 1 | The 21 CMIP5 models included and used in this study (Samuels et al. 2018)

| No. | Model name | Modeling center | Historical period | Atmospheric resolution |
|-----|------------|-----------------|-------------------|-----------------------|
| 1   | ACCESS1-0  | Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology, Australia | 1850–2005 | 1.875° × 1.25° |
| 2   | BCC-CSM1-1 | Beijing Climate Center, China Meteorological Administration, China | 1850–2012 | 2.8125° × 2.8125° |
| 3   | BNU-ESM   | College of Global Change and Earth System Science, Beijing Normal University, China | 1850–2005 | 2.8125° × 2.8125° |
| 4   | CanESM2    | Canadian Centre for Climate Modelling and Analysis, Canada | 1850–2005 | 2.8125° × 2.8125° |
| 5   | CCSM4      | National Center for Atmospheric Research, USA | 1850–2005 | 1.875° × 0.625° |
| 6   | CESM1-BGC  | Community Earth System Model Contributors, USA | 1850–2005 | 1.875° × 0.625° |
| 7   | CNRM-CM5   | Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancée en Calcul Scientifique, France | 1850–2005 | 1.4118° × 1.4063° |
| 8   | CSIRO-MK3-6-0 | Commonwealth Scientific and Industrial Research Organization/Queensland Climate Change Centre of Excellence, Australia | 1850–2005 | 1.875° × 1.875° |
| 9   | GFDL-CM3   | Geophysical Fluid Dynamics Laboratory, USA | 1860–2005 | 2.5° × 2° |
| 10  | GFDL-ESM2G |                          | 1861–2005 | 2.5° × 2° |
| 11  | GFDL-ESM2M |                          | 1861–2005 | 2.5° × 2° |
| 12  | INMCM4     | Institute for Numerical Mathematics, Russia | 1850–2005 | 2° × 2.5° |
| 13  | IPSL-CM5A-L-R | Institute Pierre-Simon Laplace, France | 1850–2005 | 3.75° × 1.875° |
| 14  | IPSL-CM5A-M-R |                          | 1850–2005 | 2.5° × 1.2587° |
| 15  | MIROC5     | The University of Tokyo, National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology, Japan | 1850–2012 | 1.4063° × 1.4063° |
| 16  | MIROC-ESM-CHEM |                          | 1850–2005 | 2.8125° × 2.8125° |
| 17  | MIROC-ESM  |                          | 1850–2005 | 2.8125° × 2.8125° |
| 18  | MPI-ESM-LR | Max Planck Institute for Meteorology, Germany | 1850–2005 | 1.875° × 1.875° |
| 19  | MPI-ESM-MR |                          | 1850–2005 | 1.875° × 1.875° |
| 20  | MRI-CGCM3  | Meteorological Research Institute, Japan | 1850–2005 | 1.125° × 1.125° |
| 21  | NorESM1-M  | Norwegian Climate Centre, Norway | 1850–2005 | 2.5° × 1.875° |
with $1.5^\circ \times 1.5^\circ$ spatial spacing via the bilinear interpolation scheme.

Daily surface precipitation observations from 2,419 China Meteorological Administration (CMA) meteorological stations across China (Figure 1) (Wang et al. 2016a) were used as benchmark to evaluate the simulated precipitation from the 21 CMIP5 models. The homogeneity and reliability of the daily precipitation series were subjected to a series of strict quality control procedures at the National Meteorological Information Center (NMIC) of China that considered the changes in instrument type, station relocations, and trace biases (Ren et al. 2010).

To quantitatively assess performance of the model suite on basin scale and investigate basin-to-basin differences, the Chinese mainland was divided into ten major river basins according to the Ministry of Water Resources of China: Songhua, Liao, Hai, Yellow, Huai, Yangtze, Southeast, Pearl, Southwest, and Northwest river basins (Liu et al. 2013) (Figure 1 and Table 2). The boundaries of these ten river basins are defined based on topographic river basin divides (Yang et al. 2018).

Methods of evaluation

The comparison is carried out on the seasonal and annual time scale. The observed daily precipitation value at each ground station is first accumulated over seasonal and annual cycles to obtain seasonal and annual precipitation at that station. Similarly, the modeled monthly precipitation values at each grid cell are accumulated to seasonal and annual amounts. The model simulated precipitation is continuous and represents the rainfall rate within a defined area (grid cell), while observed precipitation is a point measurement. Similar to previous studies (Chokngamwong & Chiu 2008; Yong et al. 2010; Yang et al. 2016), the observed seasonal and annual precipitation values from all ground stations within each $1.5^\circ \times 1.5^\circ$ grid cell are first averaged to obtain the grid cell mean seasonal and annual precipitation before comparing with modeled seasonal and annual precipitation for that grid cell. Grid cells containing no ground stations are eliminated from the comparison.

Several statistics commonly used to compare model results against observations, including the mean bias (Bias), relative bias (RB), and mean absolute error (MAE) (Yang et al. 2016) are utilized to quantify the differences between simulated and observed precipitation for each model in each river basin. If $P_m^i$ and $P_o^i$ represent modeled and observed seasonal or annual precipitation averaged over all grid cells in a basin for the $i^{th}$ year between 1960 and 2005, then

$$\text{Mean bias (Bias)} = \frac{1}{n} \sum_{i=1}^{n} (P_m^i - P_o^i)$$  

$$\text{Relative bias (RB)} = \frac{\sum_{i=1}^{n} (P_m^i - P_o^i)}{\sum_{i=1}^{n} P_o^i} \times 100\%$$  

$$\text{Mean absolute error (MAE)} = \frac{1}{n} \sum_{i=1}^{n} |P_m^i - P_o^i|$$
where \( n = 46 \) is the total number of years in the 1960–2005 period.

In addition, a Taylor diagram (Taylor 2001) is utilized to provide a statistical summary of comparisons between modeled and observed interannual and interdecadal variability during the study period of 1960 to 2005. A Taylor diagram includes three statistical measures: correlation coefficient (CC), normalized standard deviation (NSTD), and normalized centralized root mean square error (NCRMSE). The statistics can be either spatial or temporal. In this case, the Taylor diagram is used to explore temporal relationship or, specifically, the relationship between modeled and observed interannual and interdecadal variability of annual mean precipitation.

\[
CC = \frac{\sum_{i=1}^{n} (P_{oi} - \overline{P_o})(P_{mi} - \overline{P_m})}{\sqrt{\sum_{i=1}^{n} (P_{oi} - \overline{P_o})^2 \cdot \sum_{i=1}^{n} (P_{mi} - \overline{P_m})^2}} \tag{4}
\]

where \( \overline{P_m} \) and \( \overline{P_o} \) are the modeled and observed precipitation averaged over the 46-year \((n = 46)\) study period.

The standard deviations (STD) of the observed and modeled precipitation are calculated, respectively, by

\[
\sigma_o = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_{oi} - \overline{P_o})^2} \tag{5}
\]

\[
\sigma_m = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_{mi} - \overline{P_m})^2} \tag{6}
\]

And the normalized STD, NSTD, is given by

\[
NSTD = \frac{\sigma_m}{\sigma_o} = \sqrt{\frac{\sum_{i=1}^{n} (P_{mi} - \overline{P_m})^2}{\sum_{i=1}^{n} (P_{oi} - \overline{P_o})^2}} \tag{7}
\]

The centralized RMSE (CRMSE) is defined as follows:

\[
CRMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_{mi} - \overline{P_m} - (P_{oi} - \overline{P_o}))^2} \tag{8}
\]

which, when normalized by the observed STD, yields NCRMSE

\[
NCRMSE = \frac{CRMSE}{\sigma_o} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} [(P_{mi} - \overline{P_m}) - (P_{oi} - \overline{P_o})]^2}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_{oi} - \overline{P_o})^2}} \tag{9}
\]

The statistics calculated at each grid cell are then averaged over all grid cells in a basin to obtain the basin mean values. CC can quantitatively measure the degree of similarity between two fields, which for the current study are two time series of annual precipitation during a 46-year period. The CRMSE and NCRMSE measure the degree of agreement in amplitude of the oscillation from the 46-year climatology. The STD and NSTD indicate the spread of the annual precipitation during the study period from the climatological values. Perfect model performance is indicated by \( CRMSE = NCRMSE = 0 \) and \( CC = NSTD = 1 \), and the model skill decreases as CRMSE and NCRMSE deviate from 0 and CC and NSTD move away from 1 (Jiang et al. 2015). The use of the Taylor diagram, which combines multiple statistics from multiple data sources into a single diagram (Taylor 2001; Tiwari et al. 2016; Zhang et al. 2018) enables a direct intercomparison of the 21 models in their ability to simulate the direction and amplitude of the interannual variability of precipitation during the study period.

| Order | River basin | Drainage area (10^4 km^2) | Number of cells |
|-------|-------------|---------------------------|-----------------|
| 1     | Songhua     | 93.5                      | 41              |
| 2     | Liao        | 31.4                      | 13              |
| 3     | Hai         | 32.0                      | 13              |
| 4     | Yellow      | 79.5                      | 31              |
| 5     | Huai        | 33.0                      | 11              |
| 6     | Yangtze     | 180.0                     | 67              |
| 7     | Southeast   | 24.5                      | 10              |
| 8     | Pearl       | 57.8                      | 8               |
| 9     | Southwest   | 84.4                      | 13              |
| 10    | Northwest   | 336.2                     | 57              |

Table 2 | Names, order, and drainage area of ten river basins over China
The overall skill of the models in simulating interannual and interdecadal variability is measured by a comprehensive evaluating indicator \( M_R \) that combines the ranking by multiple statistical variables (Schuenemann & Cassano 2010; Jiang et al. 2015; Zhang et al. 2018):

\[
M_R = 1 - \frac{1}{n \times m} \sum_{i=1}^{n} rank_i
\]

where \( m \) denotes the number of models that participated in evaluation and in this study \( m = 21 \); \( n \) is the number of variables used to evaluate model performance and here \( n = 4 \) for CC, NSTD, NCRMSE, and Bias; \( rank_i \) is the ranking of each model based on the \( i \)th variable. For each variable, the rank of the best-performing model is 1 and the worst is 21. If a model is ranked 1 for all four variables, then \( M_R = 1 - 1/(4 \times 21) \times 4 = 1 - 1/21 \). If a model is ranked last for all four variables, then \( M_R = 0 \). Thus, \( 0 \leq M_R \leq 1 \); the closer the \( M_R \) value is to 1, the closer the simulated precipitation is to the observation, as measured collectively by the four statistical variables.

**RESULTS**

**Spatial pattern**

The overall skill of all models in capturing the spatial distribution of precipitation across China and the seasonal variations are first examined by comparing the 21 model ensemble means to the observations from the 2,419 surface stations (Figure 2). The ground observations show that the spatial distribution of precipitation across China is dominated by an increasing tendency from the north and northwest towards the south and southeast. Under the influence of the East Asian summer monsoon (Ding & Chan 2005; Chen & Frauenfeld 2014), regions in southern China, including Southeast, eastern Pearl and the downstream of Yangtze, receive the highest amount of mean annual precipitation of >1,700 mm/yr. The lowest mean annual precipitation amount of <200 mm/yr is observed over northwestern China, especially in Northwest where the largest desert in China is found (Chen & Frauenfeld 2014). The 21 CMIP5 model ensemble means are able to capture this spatial pattern of annual precipitation across China. The amount of modeled precipitation, however, can differ significantly from the observed. Specifically, the ensemble means overestimate precipitation amount in the northern part of the Southwest and southwestern part of the Northwest river basins near the eastern edge of the Tibetan Plateau, but underestimate the precipitation amount in Pearl and the eastern part of the Yangtze River basins (Yue et al. 2016). The former may be attributed to the failure of coarse resolution models like the CMIP5 models to capture sharp terrain gradients and complex terrain processes (Mao & Robock 1998; Zhou & Li 2002; Bader et al. 2008; Chen et al. 2010b; Su et al. 2013) and the latter may be due to the inability of coarse resolution models in simulating convection precipitation during monsoon season (Feng et al. 2011; Sillmann et al. 2013).

The observed precipitation is marked by distinct seasonality with wet summers and dry winters across China except for the Northwest basin where annual cycle is small. The CMIP5 ensemble means successfully reproduce this seasonality. The pattern of departure of the simulated and the observed winter precipitation is small, and the pattern of departure in summer precipitation is similar to that of annual precipitation.

**Cross-basin comparison**

Next, the skill of each model in quantitatively simulating the observed basin-mean annual and seasonal precipitation is evaluated using the three statistics, Bias, RB, and MAE, for each of the river basins and the results are displayed via a ‘portrait diagram’ (Figures 3–5) to ease the comparison across the large number of models and basins. The values of all three statistics vary considerably among the ten basins for each model, and for each basin, the values also vary significantly across models. Comparing the magnitudes between the inter-model and inter-basin variations, the inter-basin variations are much larger, suggesting a strong dependency of model performance on geographic region and thus the necessity to document model bias and errors on regional scale.

For the annual precipitation, most models have positive Bias, or an overestimation, in all but the Southeast, Pearl, and Southwest river basins. The opposite occurs in the Southeast, Pearl, and Southwest basins where Bias is...
overwhelmingly negative, indicating an underestimation of precipitation in these regions where annual rainfall amounts are among the highest in the country. The consistency among the 21 models in the nature of bias is in agreement with the tendency of models to overestimate (underestimate) precipitation in dry (wet) regions (Chen & Frauenfeld 2014; Yue et al. 2016). An exception to this is the MIROC5 model, which produce a large positive Bias in the Southeast and Pearl river basins. Nearly all models have the largest RB in the Northwest basin, followed by the Yellow river basin, due probably to the small observed annual precipitation in the Northwest basin and a combination of the small observed annual precipitation and the large Bias in the Yellow river basin. The largest overestimations occur for

Figure 2 | Annual, summer, and winter average precipitation (mm/year) for the ground observations (left) and for the 21 CMIP5 models’ ensemble mean (right) over ten river basins for the period from 1960 to 2005.
Figure 3 | Portrait diagram for annual precipitation of (a) Bias (mm/year), (b) RB (%), and (c) MAE (mm/year) between the models and raingauge observational data during the 1960–2005 period over China.
Figure 4  Portrait diagram for summer precipitation of (a) Bias (mm/year), (b) RB (%), and (c) MAE (mm/year) between the models and raingauge observational data during the 1960–2005 period over China.
Figure 5 | Portrait diagram for winter precipitation of (a) Bias (mm/year), (b) RB (%), and (c) MAE (mm/year) between the models and raingauge observational data during the 1960–2005 period over China.
the Yellow and the Northwest basins, with RB values up to 150.3% for the Yellow and 979.3% for the Northwest basins. A significant overestimation also exists in most models for the Hai basin with RB values up to 97.4% (see Figure 3). In comparison, the underestimation is substantially smaller with the largest percentage of −61.1% in the Southwest basin. Models are quite consistent in the MAE and the values appear to be positively related to the amount of observed annual mean precipitation with larger MAE values (924.5–1,474.9 mm/year) in basins of higher annual rainfall amount (Southeast, Pearl, and Southwest) and smaller values (291.1–770.2 mm/year) in basins of lower annual rainfall amount (Northwest).

A similar model behavior occurs for summer precipitation (Figure 4). The largest overestimations of summer precipitation also occur for the Yellow and the Northwest basins, with RB values up to 171.6% for the Yellow and 424.9% for the Northwest basin, and the largest underestimation of 53% occurs for the Huai basin. The larger MAE values (174.7–394.9 mm/year) occur in basins of higher annual rainfall amount (Southeast and Pearl) and smaller values (71.6–211.8 mm/year) in basins of lower annual rainfall amount (Northwest).

For winter precipitation (Figure 5), the dependency of biases on the models and the basins is similar to that of summer except that the magnitudes of the biases for winter precipitation are considerably smaller. However, the values of RB are an order of magnitude larger compared to summer. For example, for the Northwest basin where the RB is generally the largest among the ten basins, the RB values range from around 200 to over 3,000% in winter compared to a maximum of just over 400% for summer. The large relative percent deviation across models and basins in winter suggests that the winter precipitation amount can be quite sensitive to model biases and that caution needs to be taken in interpreting the CMIP5 model simulated winter precipitation in China. For summer, the biases themselves can be large, but because of the large rainfall amount in summer, the relative biases are generally small, from a few percent to less than 50%, for most models and basins except for the Yellow river basin and the Northwest basin where the relative bias can be more than 100 and even 400%.

### Within-basin comparison

Given the large size of the basins, a small mean Bias over a basin does not necessarily mean better performance because large Bias with different sign at different parts of the basin may cancel each other out. The spatial variations of the three statistics within each basin are examined (Figure 6) in the form of a box-whisker plot showing the median (line in the box), the interquartile range (box) spanned by the 25th and 75th quartiles, and the minimum and maximum values (whiskers) during the period 1960–2005 for each of the 21 models and each of the ten basins. Despite the variations among the model suite, the heterogeneity of model performance, as measured by the interquartile range, is relatively large for four river basins (Yellow, Yangtze, Southwest, and Northwest) and relatively small for the other six basins (Songhua, Liao, Huai, Southeast and, to some degree, Hai and Pearl). The four basins with larger interquartile range also have larger range of values and contain more outliers. The extreme RB values for the Northwest river basin, which is an order of magnitude larger than those in the other basins, are a result of the small observed annual precipitation in this basin that contains the largest deserts in China. Except for one or two models, most models have an interquartile range containing zero within all basins except for the Northwest basin where the first quartile is above zero indicating consistent overestimation in this basin. Southeast and Pearl have the third quartile below zero for most models indicating a consistent underestimation across these basins.

### Interannual variability

The skill of the 21 models in capturing the observed interannual variability is evaluated and inter-compared via Taylor diagrams (Figure 7). For each basin, the CC between the observed and the modeled basin-mean annual precipitation time series, the standard deviation of the modeled time series normalized by that of the observations (NSTD), and the centralized root-mean square error of the modeled time series normalized by that of the observation (NCRMSE) calculated by Equations (4), (7), and (9), respectively, are shown in a single Taylor diagram for all 21 models. In most basins, the points are cluttered together, indicating
Figure 6 | Bias (mm/year) boxplots (left panel), RB (%) boxplots (middle panel), and MAE (mm/year) boxplots (right panel) of CMIP5 simulations compared with raingauge observations (1966–2005) for annual precipitation in ten river basins over China.
Figure 7 | Taylor diagram (Taylor 2001) comparing between the annual precipitation for the 21 CMIP5 models and raingauge observations during the 1960–2005 period in ten river basins over China. Please refer to the online version of this paper to see this figure in color: http://dx.doi.org/10.2166/nh.2021.151.

OBS. Observation;
1. ACCESS1-0; 2. BCC-CSM1-1; 3. BNU-ESM; 4. CanESM2; 5. CCSM4;
6. CESM1-BGC; 7. CNRM-CM5; 8. CSIRO-MK3-6-0; 9. GFDL-CM3;
10. GFDL-ESM2G; 11. GFDL-ESM2M; 12. INMCM4; 13. IPSL-CM5A-LR;
14. IPSL-CM5A-M-R; 15. MIROC5; 16. MIROC-ESM-CHEM; 17. MIROC-ESM;
18. MPI-ESM-LR; 19. MPI-ESM-MR; 20. MRI-CGCM3; 21. NorESM1-M
similar performance of the models. However, large scatters among models are found in the Yellow and especially the Northwest river basins. The large scatter in the Northwest basin might be a result of the spottiness of the ground stations and the highly complex terrain characterized by two plateaus (Qinbai-Tibet Plateau, Inner Mongolian Plateau), three basins (Jungar Basin, Tarim Basin, Qaidam Basin), and two major mountain ranges (Tianshan Mountains, Qilian Mountains). Previous studies have shown that coarse-resolution GCMs, such as those in CMIP5, have difficulty in resolving terrain-induced meso- and local-scale processes and their interactions with synoptic-scale processes (Zhou & Li 2002; Räisänen 2007; Chen et al. 2010b; Su et al. 2013; Yang et al. 2019), which often lead to systematic biases and large errors in the model simulations in regions of rapid elevation change (Bader et al. 2008). The magnitudes of the biases and errors can vary significantly across models depending on the sensitivity of model parameterizations to the underlying terrain variations and the models’ ability to capture the variations.

For most basins, the values of CC (dotted blue radial lines in Figure 7) generally fall between 0.7 and 0.9 for the majority of the models, suggesting that the modeled and observed basin-mean annual precipitation tend to go up and down together on interannual scale. The correlation is relatively low (CC < 0.6) between observations and model simulations for the Huai and Southeast basins, due possibly to inadequate simulations of the interannual variability in monsoonal circulations that dominate summer precipitation in these basins (Chen & Frauenfeld 2014). Biasutti (2013) reported similar discrepancies for Sahel rainfall trends in CMIP5 model outputs and suggested that the mismatch could be due to the decadal variability of sea surface temperatures, which was not captured well by the multi-model ensemble, and to the underestimation of aerosol effects in climate models. Over eastern China, the monsoon precipitation is mainly influenced by interdecadal variability over the central-eastern Pacific and the western tropical Indian Ocean (Zhou et al. 2008; Li et al. 2010), with aerosol forcing playing a complementary role (Song et al. 2014). The influences by modes of internal interdecadal variability, such as the Pacific Decadal Oscillation, are not captured well in CMIP5 models (Zhou et al. 2013; Song et al. 2014) and, as a result, the observed decadal variability in precipitation is not simulated well by the multi-model ensemble means (Chen & Frauenfeld 2014).

For the majority of the basins, the NSTD values mostly fall between 0.8 and 1.2, indicating the models are more or less adequate in simulating the observed amplitude of the interannual oscillation. However, over half of the models have NSTD values greater than 1.2 for the Yellow river basin and greater than 1.5 for the Northwest basins, indicating the simulated amplitude of interannual variability is 20–50% larger than those of the observed. This may be attributed to the influence of the terrain (Giorgi & Marinucci 1996) and, in the case of the Northwest basin, also to the sparseness of the stations (Chen et al. 2010a). Gauge locations tend to lie at low elevations relative to the surrounding terrain and in the Northwest river basin stations are mostly located in the northern part of the basin with a very limited number of stations in the southern part where the Tibetan Plateau is located. Adam et al. (2006) showed that a correction for orographic effects can result in a net precipitation increase of 20% in orographically influenced regions. In contrast, lower than 0.8 NSTD values or a significant underestimate of interannual variability are generally found in the Southeast and Pearl river basins. This underestimation of the magnitude of the year-to-year variations is related to the underestimation of the annual mean precipitation which has been attributed to the convective and microphysical parameterization schemes and the coarse resolutions of the GCMs that lead to large uncertainties in simulating the intensity and extent of the East Asian monsoon circulations (Huang et al. 2012). The NCRMSE values, which range between 0.4 and 0.8, show a similar dependency on models and basins.

**Ranking of the models**

The overall ranking of the 21 models in their performance of simulating temporal variability, as measured by four statistics, CC, NSTD, NCRMSE, and Bias, is determined by the $M_R$ score calculated by Equation (10) where $0 < M_R < 1$, with higher score indicating better performance (Figure 8). Figure 8 indicates that the performance of the models varies significantly from basin to basin with no model emerging as either the best for most basins or the worst.
For example, the BCC-CSM1-1 model performs well in northeastern China (Songhua), but worse in the rest of the river basins, and the BNU-ESM and MPI-ESM-MR models only show better performance in Southeast river basin. The CNRM-CM5, CSIRO-MK3-6-0, and IPSL-CM5A-M-R models do a better job in central China (Hai, Yellow, and Huai). GFDL-ESM2M and INMCM4 models have higher MR score in southeastern China (Southeast and Pearl).

Seven models (ACCESS1-0, BNU-ESM, CanESM2, MIROC5, MIROC-ESM-CHEM, MIROC-ESM, NorESM1-M) have MR score ≤ 0.5 in at least three or more basins, indicating relatively poor performance. Only six models (CNRM-CM5, CSIRO-MK3-6-0, INMCM4, IPSL-CM5A-L-R, IPSL-CM5A-M-R, MRI-CGCM3) have MR score ≥0.65 for at least three or more basins.

Considering all ten basins, the top ranking models are CNRM-CM5 and IPSL-CM5A-M-R with relatively high MR score in most basins, and the bottom ranking models are BNU-ESM, MIROC-ESM-CHEM, and MIROC-ESM. The varying ability to capture the interannual variation of precipitation is likely to be a combination of factors, such as model forcing, the magnitude of the internal variability, and the climatic sensitivity of individual models and model parameterizations. Most models have a relatively low MR score in Hai, Yellow, and Northwest river basins.

We choose the best models and worst models for each river basin based on the MR values, as shown in Table 3. In addition, the corresponding values of the RB for the top-ranked models (best) based on the MR scores are compared with the bottom-ranked models (worst) and with all models (all) (Figure 9). Except the Southwest river basin, the Songhua, Liao, Hai, Yellow, Huai, Yangtze, Southeast, Pearl, and the Northwest basins show significant improvement in RB distribution for the top-ranked models, with reduced mean and median and narrow interquartile ranges, compared to those of the bottom-ranked or of all models. The most noticeable improvement of the top MR scored models compared to all model ensembles is in the Northwest basin where the RB of the top ranking models is about 150% smaller and the 25th and 75th percentile errors are 40 and 350% smaller, respectively. The Southwest basin shows similar performance of the top MR scored models with regard to RB.
DISCUSSION AND CONCLUSIONS

Numerous previous studies have highlighted that climate simulations are subject to large biases and uncertainties (Brekke & Barsugli 2006) depending on phenomena, periods, and location. Due to structural errors, parameterization uncertainty, and coarse resolution, the outputs of climate models tend to exhibit large uncertainties and substantial biases relative to observational data in the amount and spatial-temporal distribution of precipitation that vary on local and regional scales (Sillmann et al. 2013; Steinschneider & Lall 2015). Consequently, the performances of CMIP5 models in simulating historical precipitation need to be evaluated thoroughly in order to better interpret the projections on future precipitation based on these models (Park et al. 2016).

The current study evaluated and ranked the performance of 21 CMIP5 models in simulating annual and seasonal precipitation in each of the ten river basins across China by using ground observations at 2,419 stations during the period 1960–2005 and comprehensive evaluation metrics. The results show that the 21 CMIP5 model ensembles adequately reproduce the spatial distribution of seasonal precipitation. There is a widespread overestimation among the models and basins especially in the basins, such as the Northwest basin, where annual precipitation is low. Several models, however, significantly underestimate annual and summer precipitation in the Liao, Southeast, Pearl, and Southwest river basins. The overestimation can be extremely large, with relative biases over 1,000% for summer precipitation and, in contrast, the underestimation is much smaller. Considerable performance heterogeneity is found within the Yellow, Yangtze, Southwest, and Southeast river basins.

| Model       | Songhua | Liao | Hai | Yellow | Huai | Yangtze | Southeast | Pearl | Southwest | Northwest |
|-------------|---------|------|-----|--------|------|---------|-----------|-------|------------|-----------|
| ACCESS1-0   | ✓       |      |     |        |      |         |           |       |            |           |
| BCC-CSM1-1  | ✓       |      |     |        |      |         |           |       |            |           |
| BNU-ESM     | ×       | ×    | ×   | ×      | ×    | ✓       |           |       |            |           |
| CanESM2     | ✓       | ×    | ×   | ×      |      |         |           |       |            |           |
| CCSM4       |         |      |     |        |      |         |           |       | ✓          |           |
| CESM1-BGC   |         |      |     |        |      |         |           |       | ✓          |           |
| CNRM-CM5    | ✓       | ✓    | ✓   | ✓      | ✓    | ✓       |           |       |            |           |
| CSIRO-MK3-6-0 |         | ✓    | ✓   | ✓      |      |         |           |       | ✓          |           |
| GFDL-CM3    | ✓       |      |     |        |      |         |           |       |            | ✓         |
| GFDL-ESM2G  |         |      |     |        |      |         |           |       | ✓          |           |
| GFDL-ESM2M  |         |      |     |        |      |         |           |       | ✓          |           |
| INMCM4      | ✓       | ✓    | ✓   | ✓      | ✓    | ✓       |           |       |            | ✓         |
| IPSL-CM5A-L-R | ⎚    | ✓    | ✓   |        |      |         |           |       |            | ✓         |
| IPSL-CM5A-M-R | ✓   | ✓    | ✓   | ✓      |      | ✓       |           |       |            | ✓         |
| MIROC5      | ×       |      |     |        |      |         |           |       | ✓          |           |
| MIROC-ESM-CHEM | ×   | ×    | ×   | ×      | ✓    | ✓       |           |       |            | ✓         |
| MIROC-ESM   | ×       | ×    | ×   | ×      | ✓    | ✓       |           |       | ✓          |           |
| MPI-ESM-LR  |         | ×    |     |        |      |         |           |       | ✓          |           |
| MPI-ESM-MR  |         |      |     |        |      |         |           |       | ✓          |           |
| MRI-CGCM3   | ✓       | ✓    | ✓   | ✓      | ✓    | ✓       |           |       | ✓          |           |
| NorESM1-M   |         |      |     |        |      |         |           |       | ✓          |           |

Note: ✓ indicates the chosen best models based on $M_R$ score $\geq 0.65$; × shows the chosen worst models based on $M_R$ score $\leq 0.3$. 
especially the Northwest basins where the evaluation statistics calculated at each 1.5° × 1.5° grid cell with the basin show large scatter.

Uncertainties on simulating large-scale circulations, e.g., East Asian subtropical westerly jet (EASWJ), western Pacific subtropical high (WPSH), and East Asian summer monsoon, are possible reasons for large model spread on simulating precipitation in both climatology and interannual variation (Huang et al. 2016). As is well-known, the anomalous location of EASWJ will have an important impact on precipitation across China, especially summer rainfall (Zhang et al. 2006). We especially examine summer zonal wind at 200 hPa for 21 CMIP5 models, along with the model spread during the period 1960–2005, as shown in Figure 10. In summer, there is an EASWJ core, namely, Tibetan pattern (85°–100°E), which anomaly can lead to the increase of summer precipitation over the downstream of Yangtze river basin and decrease over northern and southern China (Huang et al. 2013). Figure 10(a)–10(u) show large differences on simulating EASWJ in both intensity and location among 21 CMIP5 models. Compared to All-21-CMIP5-models multimodel ensemble mean (AMME), as shown in Figure 10(v), several models, such as CCSM4, CESM1-BGC, GFDL-CM3, IPSL-CM5A-M-R, and NorESM1-M, overestimate the intensity of Tibetan pattern, which is significantly related to overestimated precipitation simulations across downstream of Yangtze river basin. Compared to the AMME, although EASWJ simulations of several models, e.g., BCC-CSM1-1, BNU-ESM, GFDL-ESM2G, GFDL-ESM2M, MPI-ESM-LR, and MPI-ESM-MR, are much stronger, their maximum centers extend from Tibetan Plateau to the coastal ocean, which lead to a high probability of precipitation increasing (decreasing) over coastal ocean (eastern China). Moreover, Figure 10(w) shows that the large spread in simulating intensity and location of EASWJ among CMIP5 models mainly occurs in the principal subtropical jets’ area together with the coastal region, which may affect the precipitation uncertainties over each river basin in China. Similar findings have been reported by a previous study from Huang et al. (2015).

In addition, we also examine the geopotential height at 500 hPa in CMIP5 models in summer, along with the model spread for 1960–2005 (Figure 11). In summer, compared to AMME (Figure 11(v)), WPSH of several models, e.g., BNU-ESM,
CCSM4, CESM1-BGC, and MIROC-ESM, strengthens and moves westward, which may lead to the increase of summer precipitation over northern China. However, for CNRM-CM5, GFDL-CM3, GFDL-ESM2M, INMCM4, IPSL-CM5A-L-R, and MRI-CGCM3 models, WPSH retreats eastward, which relates to underestimate precipitation. In comparison with AMME, the simulated WPSH of MPI-ESM-LR and NorESM1-M has a similar pattern, which exhibits a high agreement with the observational precipitation. Figure 11(w) shows an obvious large center for model spread across the
northern Pacific, which indicates that most models have the ability to capture WPSH spatial distribution.

Overall, the present existing uncertainties of CMIP5 models in simulating summer precipitation in climatology across China are closely associated with the uncertainties on simulating large-scale atmospheric circulations, such as EASWJ and WPSH.

All models tend to track the interannual variability in all the basins very well with correlation coefficients generally greater than 0.8. The model simulated magnitudes of the
interannual oscillation tend to be within ±20% of the observed values except for the Yellow river basin where the magnitude is larger than 20% and the Northwest river basin where it is greater than 50% of the observed values. Based on a comprehensive score calculated using four evaluation metrics, models are ranked for their performance in each basin. For all but the Northwest basin, the relative biases of the best model subgroup ensemble are much smaller compared to all model ensemble and much smaller compared to the worse model subgroup ensemble. For the Northwest basin, although the best model group ensemble reduces the biases from the all model and the worse model ensembles, the biases and errors are still too large to warrant the use of the CMIP5 model predictions.

Individual models show different simulating skills to capture climatological variables because of different forcing, the magnitude of the internal variability, and the climatic sensitivity of individual models (Huang et al. 2013; Chen & Frauenfeld 2014; Wang & Chen 2014). The performance evaluations of the 21 CMIP5 models in reproducing annual and seasonal precipitation over each of the ten river basins in China can be used help policymakers and stakeholders to make informed decisions when choosing precipitation products generated by a large set of GCMs for future planning. The large uncertainties in some models and some regions identified from the study can also inform model developers in their model improvement efforts.

CMIP5 models can provide a good overview of both current and future climate globally, but it is difficult to evaluate climate change impacts at regional and local scales due to the coarse spatial resolution (Xue et al. 2007). Many researchers have previously indicated that CMIP5 models are limited in their coarse resolution and systematic bias (Christensen et al. 2008; Park et al. 2016). The coarse resolution prevents the models from sufficiently representing regional climatic processes. Räisänen (2007) concluded that CMIP5 models could not simulate many small-scale processes explicitly. The bias correction should be applied for each individual model (Christensen et al. 2008). The reproducibility of precipitation, as well as the recommended CMIP5 models, has reference value in future studies related to climate change and their impacts on ecohydrological simulation. This study indicates that some models exhibit much better agreement with the ground observations in terms of MR scores of precipitation than other models. However, this study also finds that there is no evident relationship between model performance and their horizontal resolution. For instance, the models with least biases for mean annual precipitation in Liao, Hai, and Yellow, e.g., CNRM-CM5 and CSIRO-MK3-6-0, have resolutions that are neither the highest nor lowest among the 21 models, therefore, the performance of the top models has no significant relationship with their resolution in this study, which shows that spatial resolution may not be a decisive factor influencing the ability of CMIP5 models to reasonably simulate precipitation variability, consistent with the conclusions of McMahon et al. (2015), Masson & Knutti (2011), Su et al. (2013), and Song & Zhou (2014).

It is important to note that different statistical measures used in the evaluation show different aspects of precipitation and their separate application may lead to different conclusions. This study chose some commonly used statistical methods to evaluate the performance of CMIP5 models. The best models may change depending on the statistical indicators used. The physics and the temporal and spatial characteristics of precipitation are complex. Improving the ability of models to simulate precipitation should be a priority for climate modelers (Su et al. 2013).

Large positive biases and poor model performance are found in the Northwest and Southwest basins at the eastern edge of the Tibetan Plateau, which is consistent with previous studies (Chen & Frauenfeld 2014). Jiang et al. (2015) inferred that the CMIP5 models’ overestimation over western and northern (Southwest and Northwest) China are closely related to the overestimation of southwestmonsoon along the east coast of the Arabian Peninsula and overestimation of the WPSH, inducing a stronger East Asian summer monsoon. The precipitation simulating skill in the basins is generally higher than that in the mountains and on the plateaus where the topographic gradients are usually sharper than basins (Yang et al. 2019). In basins, precipitation is mainly produced by large-scale precipitation-bearing systems instead of convectional precipitation induced by complex orography. A weather system in a basin is relatively simple, with less local convectional weather systems at a small scale, compared to mountains and plateaus. Chen & Frauenfeld (2014) also showed that
the models overestimated the precipitation along the eastern edge of the Tibetan Plateau to a large extent, particularly during summer (Yang et al. 2019). It is hoped that further enhancements to model resolution and improvements in microphysics and convective parameterizations will lead to improved skill in precipitation simulation for these regions.

ACKNOWLEDGEMENTS

The authors thank the modeling groups listed in Table 2 of this paper for making their simulations available for analysis. The authors thank the editor and anonymous reviewers, who helped to improve the earlier version of this article and provide their constructive comments. We thank Professor Lejiang Yu, who helped us to improve this article and provide many constructive comments. The authors thank the editor and anonymous reviewers, who helped to improve the earlier version of this paper for making their simulations available for analysis. The authors thank the editor and anonymous reviewers, who helped to improve the earlier version of this paper for making their simulations available for analysis.

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Adam, J. C., Clark, E. A., Lettenmaier, D. P. & Wood, E. F. 2006 Correction of global precipitation products for orographic effects. J. Clim. 19, 15–38.

Bader, D. C., Covey, C., Gutowski, W. J., Held, I. M., Kunkel, K. E., Miller, R. L., Tokmakian, R. T. & Zhang, M. H. 2008 Climate Models: an Assessment of Strengths and Limitations. A Report by the U.S. Climate Change Science Program and the Subcommittee on Global Change Research. Department of Energy, Office of Biological and Environmental Research, Washington, DC, USA, p. 124.

Berg, P., Moseley, C. & Haerter, J. O. 2013 Strong increase in convective precipitation in response to higher temperatures. Nat. Geosci. 6, 181.

Biasutti, M. 2013 Forced Sahel rainfall trends in the CMIP5 archive. J. Geophys. Res. Atmos. 118, 1613–1623.

Brekke, L. D. & Barsugli, J. J. 2013 Uncertainties in projections of future changes in extremes. In: Extremes in A Changing Climate: Detection, Analysis and Uncertainty (A. AghaKouchak, D. Easterling, K. Hsu, S. Schubert & S. Sorooshian, eds). Springer Netherlands, Dordrecht, The Netherlands, pp. 309–346.

Cao, L., Zhang, Y. & Shi, Y. 2011 Climate change effect on hydrological processes over the Yangtze River basin. Quat. Int. 244, 202–210.

Chadwick, R., Boutle, I. & Martin, G. 2013 Spatial patterns of precipitation change in CMIP5: why the rich do not get richer in the tropics. J. Clim. 26, 3803–3822.

Chen, L. & Frauenfeld, O. W. 2014 A comprehensive evaluation of precipitation simulations over China based on CMIP5 multimodel ensemble projections. J. Geophys. Res. Atmos. 119, 5767–5786.

Chen, D., Ou, T., Gong, L., Xu, C.-Y., Li, W., Ho, C.-H. & Qian, W. 2004 Spatial interpolation of daily precipitation in China: 1951–2005. Adv. Atmos. Sci. 27, 1221–1232.

Chen, H., Zhou, T., Neale, R. B., Wu, X. & Zhang, G. J. 2005 Performance of the new NCAR CAM3.5 in East Asian summer monsoon simulations: sensitivity to modifications of the convection scheme. J. Clim. 23, 3657–3675.

Chen, J., Gao, C., Zeng, X., Xiong, M., Wang, Y., Jing, C., Krysanova, V., Huang, J., Zhao, N. & Su, B. 2017 Assessing changes of river discharge under global warming of 1.5 °C and 2 °C in the upper reaches of the Yangtze River basin: approach by using multiple-GCMs and hydrological models. Quat. Int. 453, 63–73.

Choknamwong, R. & Chiu, L. S. 2008 Thailand daily rainfall and comparison with TRMM products. J. Hydrometeorol. 9, 256–266.

Christensen, J. H., Boberg, F., Christensen, O. B. & Lucas-Picher, P. 2008 On the need for bias correction of regional climate change projections of temperature and precipitation. Geophys. Res. Lett. 35, L20709.

Ding, Y. & Chan, J. C. L. 2005 The East Asian summer monsoon: an overview. MAP 89, 117–142.

Durack, P. J., Wijffels, S. E. & Matear, R. J. 2012 Ocean salinities reveal strong global water cycle intensification during 1950 to 2000. Science 336, 455–458.

Feng, L., Zhou, T., Wu, B., Li, T. & Luo, J.-J. 2011 Projection of future precipitation change over China with a high-resolution global atmospheric model. Adv. Atmos. Sci. 28, 464–476.

Gaetani, M. & Mohino, E. 2015 Decadal prediction of the Sahelian precipitation in CMIP5 simulations. J. Clim. 26, 7708–7719.

Giorgi, F. & Marinucci, M. R. 1996 An investigation of the sensitivity of simulated precipitation to model resolution and its implications for climate studies. Mon. Weather Rev. 124, 148–166.

Guilbert, J., Betts, A. K., Rizzo, D. M., Beckage, B. & Bomblies, A. 2015 Characterization of increased persistence and intensity of precipitation in the northeastern United States. Geophys. Res. Lett. 42, 1888–1893.
Mehran, A., AghaKouchak, A. & Phillips, T. J. 2014 Evaluation of CMIP5 continental precipitation simulations relative to satellite-based gauge-adjusted observations. J. Geophys. Res. Atmos. 119, 1695–1707.

Nguyen, P., Thorstensen, A., Sorooshian, S., Zhu, Q., Tran, H., Ashouri, H., Miao, C., Hsu, K. & Gao, X. 2017 Evaluation of CMIP5 model precipitation using PERSIANN-CDR. J. Hydrometeorol. 18, 2313–2330.

Park, C., Min, S.-K., Lee, D., Cha, D.-H., Suh, M.-S., Kang, H.-S., Hong, S.-Y., Lee, D.-K., Baek, H.-J., Boo, K.-O. & Kwon, W.-T. 2016 Evaluation of multiple regional climate models for summer climate extremes over East Asia. Clim. Dyn. 46, 2469–2486.

Pendergrass, A. G. & Hartmann, D. L. 2014a The atmospheric energy constraint on global-mean precipitation change. J. Clim. 27, 757–768.

Pendergrass, A. G. & Hartmann, D. L. 2014b Changes in the distribution of rain frequency and intensity in response to global warming. J. Clim. 27, 8372–8383.

Räisänen, J. 2007 How reliable are climate models? Tellus A 59, 2–29.

Ren, Z. H., Zhao, P., Zhang, Q., Zhang, Z. F., Cao, L. J., Yang, Y. R., Zou, F. L., Zhao, Y. F., Zhao, H. M. & Chen, Z. 2010 Quality control procedures for hourly precipitation data from automatic weather stations in China. Meteor. Mon. 36, 123–132.

Samuels, R., Hochman, A., Baharad, A., Givati, A., Levi, Y., Yosef, Y., Saaroni, H., Ziv, B., Harpaz, T. & Alpert, P. 2018 Evaluation and projection of extreme precipitation indices in the Eastern Mediterranean based on CMIP5 multi-model ensemble. Int. J. Climatol. 38, 2280–2297.

Schuenemann, K. C. & Cassano, J. J. 2020 Changes in synoptic weather patterns and Greenland precipitation in the 20th and 21st centuries: 2. Analysis of 21st century atmospheric changes using self-organizing maps. J. Geophys. Res. Atmos. 115, D20115.

Sillmann, J., Kharin, V. V., Zhang, X., Zwiers, F. W. & Bronaugh, R. 2015 Climate extremes indices in the CMIP5 multimodel ensemble: part 1. Model evaluation in the present climate. J. Geophys. Res. Atmos. 118, 1716–1733.

Song, F. & Zhou, T. 2014 Interannual variability of East Asian summer monsoon simulated by CMIP3 and CMIP5 AGCMs: skill dependence on Indian Ocean–Western Pacific anticyclone teleconnection. J. Clim. 27, 1679–1697.

Song, F., Zhou, T. & Qian, Y. 2014 Responses of East Asian summer monsoon to natural and anthropogenic forcings in the 17 latest CMIP5 models. Geophys. Res. Lett. 41, 596–603.

Steinschneider, S. & Lall, U. 2015 A hierarchical Bayesian regional model for nonstationary precipitation extremes in Northern California conditioned on tropical moisture exports. Water Resour. Res. 51, 1472–1492.

Su, F., Duan, X., Chen, D., Hao, Z. & Cuo, L. 2013 Evaluation of the global climate models in the CMIP5 over the Tibetan Plateau. J. Clim. 26, 3187–3208.

Sun, H., Wang, A., Zhai, J., Huang, J., Wang, Y., Wen, S., Zeng, X. & Su, B. 2018 Impacts of global warming of 1.5 °C and 2.0 °C on precipitation patterns in China by regional climate model (COSMO-CLM). Atmos. Res. 203, 83–94.
Taylor, K. E. 2001 Summarizing multiple aspects of model performance in a single diagram. J. Geophys. Res. Atmos. 106, 7183–7192.

Taylor, K. E., Stouffer, R. J. & Meehl, G. A. 2012 An overview of CMIP5 and the experiment design. B. Am. Meteorol. Soc. 93, 485–498.

Tiwari, P. R., Kar, S. C., Mohanty, U. C., Dey, S., Kumari, S. & Sinha, P. 2016 Seasonal prediction skill of winter temperature over North India. Theor. Appl. Climatol. 124, 15–29.

Wang, L. & Chen, W. 2014 A CMIP5 multimodel projection of future temperature, precipitation, and climatological drought in China. Int. J. Climatol. 34, 2059–2078.

Wang, Q., Wu, J., Lei, T., He, B., Wu, Z., Liu, M., Mo, X., Geng, G., Li, X., Zhou, H. & Liu, D. 2014 Temporal-spatial characteristics of severe drought events and their impact on agriculture on a global scale. Quat. Int. 349, 10–21.

Wang, L., Yuan, X., Xie, Z., Wu, P. & Li, Y. 2016a Increasing flash droughts over China during the recent global warming hiatus. Sci. Rep.-UK 6, 30571.

Wang, Y., Xu, Y., Lei, C., Li, G., Han, L., Song, S., Yang, L. & Deng, X. 2016b Spatio-temporal characteristics of precipitation and dryness/wetness in Yangtze River Delta, eastern China, during 1960–2012. Atmos. Res. 172–173, 196–205.

Webner, M. 2013 Methods of projecting future changes in extremes. In: Extremes in A Changing Climate: Detection, Analysis and Uncertainty (A. AghaKouchak, D. Easterling, K. Hsu, S. Schubert & S. Sorooshian, eds). Springer Netherlands, Dordrecht, The Netherlands, pp. 223–237.

Wu, Y., Zhong, P.-a., Xu, B., Zhu, F. & Fu, J. 2018 Evaluation of global climate model on performances of precipitation simulation and prediction in the Huaihe River basin. Theor. Appl. Climatol. 153, 191–204.

Xue, Y., Vasic, R., Janjic, Z., Mesinger, F. & Mitchell, K. E. 2007 Assessment of dynamic downscaling of the continental U.S. regional climate using the Eta/SSiB regional climate model. J. Clim. 20, 4172–4193.

Yang, X., Yong, B., Hong, Y., Chen, S. & Zhang, X. 2016 Error analysis of multi-satellite precipitation estimates with an independent raingauge observation network over a medium-sized humid basin. Hydrol. Sci. J. 61, 1813–1830.

Yang, X., Yong, B., Yin, Y. & Zhang, Y. 2018 Spatio-temporal changes in evapotranspiration over China using GLEAM_V3.0a products (1980–2014). Hydrol. Res. 49, 1330–1348.

Yang, Y., Bai, L., Wang, B., Wu, J. & Fu, S. 2019 Reliability of the global climate models during 1961–1999 in arid and semiarid regions of China. Sci. Total Environ. 667, 271–286.

Yong, B., Ren, L.-L., Hong, Y., Wang, J.-H., Gourley, J. J., Jiang, S.-H., Chen, X. & Wang, W. 2010 Hydrologic evaluation of multisatellite precipitation analysis standard precipitation products in basins beyond its inclined latitude band: a case study in Laohahe basin, China. Water Resour. Res. 46, W07542.

Yue, T., Zhao, N., Fan, Z., Li, J., Chen, C., Lu, Y., Wang, C., Xu, B. & Wilson, J. 2016 CMIP5 downscaling and its uncertainty in China. Global Planet. Change 146, 30–37.

Zhang, Y., Kuang, X., Guo, W. & Zhou, T. 2006 Seasonal evolution of the upper-tropospheric westerly jet core over East Asia. Geophys. Res. Lett. 33, (11).

Zhang, Y., You, Q., Chen, C., Ge, J. & Adnan, M. 2018 Evaluation of downscaled CMIP5 coupled with VIC model for flash drought simulation in a humid subtropical basin, China. J. Clim. 31, 1075–1090.

Zhao, T., Li, C. & Zuo, Z. 2016 Contributions of anthropogenic and external natural forcings to climate changes over China based on CMIP5 model simulations. Sci. China Earth Sci. 59, 503–517.

Zhou, T. & Li, Z. 2002 Simulation of the East Asian summer monsoon using a variable resolution atmospheric GCM. Clim. Dyn. 19, 167–180.

Zhou, T., Yu, R., Li, H. & Wang, B. 2008 Ocean forcing to changes in global monsoon precipitation over the recent half-century. J. Clim. 21, 3833–3852.

Zhou, T., Song, F., Lin, R. & Chen, X. 2013 The 2012 North China floods: Explaining an extreme rainfall event in the context of a longer-term drying tendency. 94, S49–S51.

First received 18 October 2020; accepted in revised form 12 March 2021. Available online 19 April 2021.