Comparison of Statistical and Dynamic Downscaling Techniques in Generating High-Resolution Temperatures in China from CMIP5 GCMs[✉]

LEI ZHANG, YINLONG XU, CHUNCHUN MENG, XINHUA LI, AND HUAN LIU

*Institute of Environment and Sustainable Development in Agriculture,
Chinese Academy of Agricultural Sciences, Beijing, China*

CHANGGUI WANG

Met Office Hadley Centre, Exeter, United Kingdom

(Manuscript received 22 February 2019, in final form 2 October 2019)

ABSTRACT

In aiming for better access to climate change information and for providing climate service, it is important to obtain reliable high-resolution temperature simulations. Systematic comparisons are still deficient between statistical and dynamic downscaling techniques because of their inherent unavoidable uncertainties. In this paper, 20 global climate models (GCMs) and one regional climate model [Providing Regional Climates to Impact Studies (PRECIS)] are employed to evaluate their capabilities in reproducing average trends of mean temperature (T_m), maximum temperature (T_{max}), minimum temperature (T_{min}), diurnal temperature range (DTR), and extreme events represented by frost days (FD) and heat-wave days (HD) across China. It is shown generally that bias of temperatures from GCMs relative to observations is over $\pm 1^\circ\text{C}$ across more than one-half of mainland China. PRECIS demonstrates better representation of temperatures (except for HD) relative to GCMs. There is relatively better performance in Huanghuai, Jianghuai, Jiangnan, south Yangzi River, and South China, whereas estimation is not as good in Xinjiang, the eastern part of northwest China, and the Tibetan Plateau. Bias-correction spatial disaggregation is used to downscale GCMs outputs, and bias correction is applied for PRECIS outputs, which demonstrate better improvement to a bias within $\pm 0.2^\circ\text{C}$ for T_m , T_{max} , T_{min} , and DTR and ± 2 days for FD and HD. Furthermore, such improvement is also verified by the evidence of increased spatial correlation coefficient and symmetrical uncertainty, decreased root-mean-square error, and lower standard deviation for reproductions. It is seen from comprehensive ranking metrics that different downscaled models show the most improvement across different climatic regions, implying that optional ensembles of models should be adopted to provide sufficient high-quality climate information.

1. Introduction

It is well known that climate is changing continuously under the intricate influences of natural and artificial factors at global and regional scales. As emphasized in the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, global surface mean temperature has an average increment of 0.85°C during the period of 1880–2012 and is likely to increase by 0.3°C – 4.8°C by the end of the twenty-first century (IPCC 2013). China experiences a significant temperature increase concurrent

with global warming. According to China Ministry of Science and Technology (2015), a warming trend of 0.9°C – 1.5°C during 1909–2011 is detected in China, which is higher than the global level. Mean temperature is projected to rise more than 3.4°C and 5.7°C under the RCP4.5 and RCP8.5 scenarios, respectively, in China by the end of the twenty-first century (Zhu et al. 2017). With regard to local features specifically, the warming magnitude varies within regions across China. General knowledge identifies warmer trends in northern regions relative to southern regions (Zhang et al. 2011), but scientific research should be further improved for investigating the vulnerability of climate change and future climate risk at regional or local scales and then developing the corresponding adaptation strategies.

The provision of robust climate information with a multimodel, multimethod, and multiscale (M^5S) method

[✉] Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JAMC-D-19-0048.s1>.

Corresponding author: Yinlong Xu, xuyinlong@caas.cn

can assist decision-making responding to climate change as emphasized by the World Climate Research Programme and Global Framework for Climate Services (Hewitt et al. 2012). The global Coupled Model Intercomparison Project (CMIP) already provides multimodel data resources. Global climate models (GCMs) have proven to be unable to resolve the details of regional climate change features because of the limitation of their coarse resolution and the intermodels' uncertainties (Moise et al. 2015; Ning and Bradley 2015; Marotzke et al. 2017). To bridge these gaps, downscaling methods, that is, statistical and dynamical downscaling, are multimethod ways to get fine-resolution projections to compensate for the deficiencies in the raw resolution of GCMs (Rummukainen 2010; Guo and Wang 2016; Shi et al. 2018; Chen et al. 2018).

Statistical downscaling is conducted to build the quantitative statistical relationships between coarse-resolution variables in GCMs and fine-resolution or local observations during the recorded period. We then apply these relationships to project future climate variables (Werner and Cannon 2016; Yu et al. 2015; Cannon 2018). Statistical downscaling methods cited in most related works can be grouped into four categories: scaling methods, regression-based approaches, weather-pattern-based approaches, and weather generators (Schoof 2013). Typical methods referring to these categories include physical scaling (Gaur and Simonovic 2018), bias-correction spatial downscaling (Ning et al. 2015; Werner and Cannon 2016), stepwise clustered downscaling (Zhai et al. 2019), and artificial neural networks (Chadwick et al. 2011), and these are widely adopted to handle downscaling issues around countries and local regions. Bias-correction spatial downscaling can be considered appropriate for its direct dealing on daily outputs by capturing biased statistical characteristics, as compared with other methods (Bürger et al. 2012). Dynamical downscaling is implemented to nest a fine-resolution regional climate model (RCM) driven by boundary conditions from the GCM (Guo and Wang 2016; Wang et al. 2017; Liang et al. 2019). Nowadays, there are more than 60 RCMs developed to provide climate simulations for regions around the world, and more than 12 RCMs are applied for regional climate modeling over China or East Asia (Sangelanton et al. 2019). Representative RCMs—for example, Providing Regional Climates to Impact Studies (PRECIS), Regional Climate Model (RegCM), and Weather Research and Forecasting (WRF) Model—are gradually being developed with the crucial improved understanding of sophisticated-systematic interactions throughout process modules in China (Yu et al. 2015; Shi et al. 2018; Zhu et al. 2017; Guo et al. 2018; Zhou et al. 2018). With these advances, the resolution of RCMs is getting finer and finer, and the main features of the climate including

average trend and extreme events are becoming more reliable and available to assess the impacts of climate change in local areas.

Statistical downscaling is preferred by users for its relatively low computational requirements and fast calculations, whereas dynamic downscaling is appreciated by researchers for its superiority of embracing more systematic characteristics in relation to topography, climatic dynamical process, and so on. Statistical and dynamic downscaling outputs have been widely used in the fields of climate change, natural hazards, agriculture, and hydrology across all of China and different local regions (Guo et al. 2018; Huang et al. 2018; Liang et al. 2019; Zhai et al. 2019). For example, climate simulations from statistical downscaling models are assessed through their reproduction of long-term trends of temperature and precipitation, with more attention on local regions rather than on China as a whole (Fan et al. 2011, 2013, 2015). Comparatively, with the help of commonly used RCMs, for example, PRECIS and RegCM, a number of works are being done to assess and project climate change and their subsequent impacts, covering main features of the mean climatologies and extreme events (Xu et al. 2010; Yang et al. 2010; Gao et al. 2012, 2013, 2017; Fan et al. 2014; Zhang et al. 2017; Li et al. 2018). Even though these works have been proven to enhance multiscale information, there is still a shortage of systematic analysis on high-resolution scenarios, and few comparisons have been conducted between statistical and dynamic downscaling methods, which need to be enhanced for future climate change assessments. Previous efforts (Kotlarski et al. 2014; Trzaska and Schnarr 2014; Sangelanton et al. 2019) have discovered that no single approach is best, and a method's performance is dependent on the desired spatial-temporal resolution of outputs and the climate characteristics in a specific region. It is alternatively accepted that ensemble statistical and dynamic downscaling methods are better to quantify climate change impacts (Kreienkamp et al. 2019). Moreover, it is of great concern to correct systematic biases between simulations and observations before making use of the simulations from models. The process of bias correction has been proved to be important in climate impact studies not only for GCM outputs but also for RCM outputs (Ahmed et al. 2013). The confusing usage of contrasting datasets from these models can offer widely different indications of the direction of past and future climate change. However, although preprocessing techniques for bias correction have been adopted in some statistical downscaling methods (Werner and Cannon 2016; Rahman et al. 2019), to date there is still a lack of postprocessing in RCMs and

some statistical downscaling methods without preprocessing for GCMs (Zhu et al. 2017; Guo et al. 2018; Zhai et al. 2019). Therefore, an optimization is to make better use of corrected outputs from GCMs and RCMs to have a combined understanding of plausible changes for regions and to further apply them to impacts studies.

Given the growing requirements for local or site-specific projection information, ongoing efforts are being made. Serving as an attempt to meet the aim of M²S, we explore to what extent current cognition and techniques can promote regional climate change information and bias-corrected information to improve the understanding of weather–climate change in average, extreme, and related stability. The objective of this study is to provide systematic comparisons in the performance of statistical and dynamic downscaling techniques in generating temperatures across different climatic regions in China at three levels: that is, the difference between raw GCMs and RCM outputs, the difference between statistical downscaling and dynamic downscaling outputs, and the difference between raw outputs and bias-corrected outputs. This paper is structured as follows: section 2 is devoted to a brief description of the study region, simulated and observational datasets, method of statistical downscaling for 20 GCMs, and bias correction for PRECIS; section 3 is for the evaluation in terms of temperature simulations as compared with observations; in section 4 we discuss and reach conclusions about the reliability of associated methods and findings; and the main conclusion is summarized in section 5.

2. Data and methods

a. Study region and data sources

Since there are multiple land types and atmospheric circulation systems across a wide geographical scope in China, local regions witness different climatic conditions and have inevitable different responses to climate change. The whole of mainland China is divided into 12 climatic regions in this paper (Fig. 1): Northeast China (NE), North China (NC), Inner Mongolia (IM), Huanghuai (HH), Xinjiang (XJ), the eastern part of northwest China (WE), Jianghuai (JH), Jiangnan (JHa), the south Yangzi River (SY), South China (SC), Southwestern China (SW), and the Tibetan Plateau (TP), according to the definition from China's National Weather Service (Huang et al. 2018). The elevation increases from east to west, with a span from -150 to 8434 m according to data from the digital elevation model (DEM) at a resolution of $1\text{ km} \times 1\text{ km}$ according to the Shuttle Radar Topography Mission.

In this study, 20 GCMs in the CMIP5 experiment are collected from the Program for Climate Model Diagnosis and Intercomparison (<https://pcmdi.llnl.gov/index.html>),

FIG. 1. Study regions, with DEM.

with descriptions listed in Table 1. It is obviously illustrated that the resolution of GCMs is different and mostly larger than 1.0° . The data involved in GCMs, that is, daily mean near-surface air temperature T_m , maximum near-surface air temperature T_{max} , and minimum near-surface air temperature T_{min} , cover the time period 1961–2005 and a future period (2006–2100) under RCP4.5 and RCP8.5 scenarios.

The RCM utilized in this study is version 2.1 of PRECIS, with a horizontal resolution of 25 km with 19 vertical levels (Wang et al. 2017). PRECIS is an atmospheric and land surface model of limited area that can be run over any part of the globe based on the Met Office HadRM3P (Massey et al. 2015), which had been used widely for climate change impact studies in various countries (Van Khiem et al. 2014; Rao et al. 2014). Similar to GCMs, the time slices for T_m , T_{max} , and T_{min} are 1961–2005 and 2006–2100 under RCP4.5 and RCP8.5 scenarios.

As references, observations including T_m , T_{max} , and T_{min} are extracted at a resolution of $0.25^\circ \times 0.25^\circ$, interpolated as an extension of dataset CN05 (Xu et al. 2009), over the period of 1961–2005. Because the resolution of PRECIS is close to that of the observations, PRECIS is bilinearly interpolated to the outputs as the resolution of $0.25^\circ \times 0.25^\circ$.

b. BCSD for GCMs outputs

Because of the coarse resolution and locally biased statistical characteristics of GCM outputs, one statistical downscaling method, bias-correction spatial disaggregation (BCSD), is proposed for daily data in this study, since it has

TABLE 1. General information for GCMs in CMIP5. Expansions for the model acronyms can be found online (<https://www.ametsoc.org/PubsAcronymList>).

Model	Resolution	Institution	Country
ACCESS 1.0	$1.875^\circ \times 1.25^\circ$	Commonwealth Scientific and Industrial Research Organization and Bureau of Meteorology	Australia
BNU-ESM	$2.8^\circ \times 2.8^\circ$	Beijing Normal University	China
CanESM2	$2.8^\circ \times 2.8^\circ$	Canadian Centre for Climate Modeling and Analysis	Canada
CCSM4	$0.9^\circ \times 1.25^\circ$	National Center for Atmospheric Research	United States
CNRM-CM5	$1.4^\circ \times 1.4^\circ$	Centre National de Recherches Météorologiques	France
CSIRO Mk3.6.0	$1.875^\circ \times 1.875^\circ$	Commonwealth Scientific and Industrial Research Organization	Australia
GFDL CM3	$2.5^\circ \times 2.0^\circ$	Geophysical Fluid Dynamics Laboratory	United States
GFDL-ESM2G	$2.5^\circ \times 2.0^\circ$	Geophysical Fluid Dynamics Laboratory	United States
GFDL-ESM2M	$2.5^\circ \times 2.0^\circ$	Geophysical Fluid Dynamics Laboratory	United States
HadGEM2-ES	$1.875^\circ \times 1.25^\circ$	Hadley Centre	United Kingdom
INM-CM4.0	$1.5^\circ \times 2^\circ$	Institute for Numerical Mathematics	Russia
IPSL-CM5A-LR	$3.75^\circ \times 1.895^\circ$	Institute Pierre-Simon Laplace	France
IPSL-CM5A-MR	$2.5^\circ \times 1.27^\circ$	Institute Pierre-Simon Laplace	France
MIROC5	$1.4^\circ \times 1.4^\circ$	Atmosphere and Ocean Research Institute	Japan
MIROC-ESM	$2.8^\circ \times 2.8^\circ$	Atmosphere and Ocean Research Institute	Japan
MIROCESM-CHEM	$2.8^\circ \times 2.8^\circ$	Atmosphere and Ocean Research Institute	Japan
MPI-ESM-LR	$1.875^\circ \times 1.875^\circ$	Max Planck Institute for Meteorology	Germany
MPI-ESM-MR	$1.875^\circ \times 1.875^\circ$	Max Planck Institute for Meteorology	Germany
MRI-CGCM3	$1.125^\circ \times 1.125^\circ$	Meteorological Research Institute	Japan
NorESM1-M	$1.895^\circ \times 2.5^\circ$	Norwegian Climate Centre	Norway

been widely theorized by previous research and international institutions (Ning et al. 2015; Werner and Cannon 2016) to be useful. Under the assumption that climatological intramonthly variability does not change, BCSD is used to correct large-scale bias by comparing coarse-resolution GCM outputs with observations and then interpolating the bias-corrected anomalies into a fine-scale panel of observations to get fine-resolution GCM outputs. The detailed algorithm is as follows.

First, the long-term monthly trend for each variable (T_m , T_{max} , and T_{min}) is extracted from the GCM data:

$$x'_{ij} = x_{ij} - x_{i,tre}, \quad (1)$$

where x'_{ij} and x_{ij} are the detrended value and raw value of GCM output on yearday j for month of the year i , respectively. $x_{i,tre}$ is the long-term monthly trend for month i (e.g., the series of the value in all Januaries during the period is taken to extract the trend of January), calculated as a 9-yr running average the same way as for the detrend process in generating NASA Earth Exchange Global Daily Downscaled Projections (NEX-GDDP) (Thrasher et al. 2012). This is calculated for the retrospective period (1961–2005) and the prospective period (2006–2100 under RCP4.5 and RCP8.5).

Preliminarily, finescale observation is bilinearly interpolated to coarse-resolution observation corresponding to GCM. The cumulative distribution functions (CDF) for the coarse-resolution observation and for the retrospective GCM simulation are generated by pooling and sorting

the corresponding source values (within 30 days around day j of year) over the period 1961–2005. Then, the CDF of the GCM simulation is corrected by the CDF of coarse-resolution observation. To be specific, the value of a GCM in any CDF quantile is translated to the corresponding value of coarse-resolution observation in the same CDF quantile. Assuming that the CDF of the GCM simulation is stable across the retrospective period and the prospective period, the CDF of the simulation during the prospective period is corrected by a corresponding CDF of coarse-resolution observation. Therefore, the possible bias in the statistical structure of the original GCM output is removed after this correcting procedure. After this step, the previously extracted climate trend $x_{i,tre}$ is added back to the adjusted GCM climate values.

The third step, named spatial disaggregation, spatially interpolates the adjusted GCM data to the finer-resolution data. The multidecade daily climatologies (i.e., the average for each day of the year over 1961–2005) of the observation are generated at both fine and coarse resolutions. Scaling factors are calculated as the differences between the bias-corrected GCM data and the coarse-resolution observation. The scaling factors are then bilinearly interpolated to the fine-resolution ($0.25^\circ \times 0.25^\circ$) grid. Last, the scaling factors are applied by addition or shifting on the fine-resolution observation climatologies to obtain the desired downscaled climate outputs. As such, final fine-resolution results (daily T_m , T_{max} , and T_{min} during 1961–2005 and 2006–2100 under RCP4.5 and

RCP8.5) are reproduced for the 20 GCMs that can be used to project the potential change of temperatures.

c. Bias correction for RCM outputs

As a result of systematic errors caused by physics and parameterization schemes in the climate model, the bias correction for daily PRECIS output is carried out here, following the same algorithm used in correcting the biases in the GCM. Each variable (T_m , T_{max} , and T_{min}) is detrended through Eq. (1). The CDF of the PRECIS simulation during 1961–2005 is corrected by the CDF of observation during 1961–2005, and same correction process is also applied for projections during 2006–2100 under RCP4.5 and RCP8.5. Bias-corrected PRECIS output is generated by adding a preextracted climate trend back to the above corrected values.

d. Validation of GCM and RCM simulations

The performance of GCM and RCM simulations, including T_m , T_{max} , T_{min} , and diurnal temperature range (DTR, quantified as $T_{max} - T_{min}$), are generally evaluated by some metrics (McSweeney et al. 2015; Ahmed et al. 2019), namely, the difference between simulation and observation (bias), the spatial correlation coefficient (SCC), the root-mean-square error (RMSE), and standard deviations (STD), as follow:

$$\text{bias}_i = \text{sim}_i - \text{obs}_i, \quad (2)$$

$$\text{SCC} = \frac{\sum_{i=1}^n \sum_{j=1}^m (\text{sim}_{i,j} - \overline{\text{sim}_{i,j}})(\text{obs}_{i,j} - \overline{\text{obs}_{i,j}})}{\sqrt{\sum_{i=1}^n \sum_{j=1}^m (\text{sim}_{i,j} - \overline{\text{sim}_{i,j}})^2 \sum_{i=1}^n \sum_{j=1}^m (\text{obs}_{i,j} - \overline{\text{obs}_{i,j}})^2}}, \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\text{sim}_i - \text{obs}_i)^2}{n}}, \quad \text{and} \quad (4)$$

$$\text{STD} = \sqrt{\frac{\sum_{i=1}^n [(\text{sim}_i - \text{obs}_i) - (\overline{\text{sim}} - \overline{\text{obs}})]^2}{n}}, \quad (5)$$

where sim_i and obs_i are simulation and observation at grid i and $\overline{\text{sim}}$ and $\overline{\text{obs}}$ are the mean values of the simulation and observation of all grids.

Among many extreme climate events, heat wave and frost are two typical extreme events of particular interest for their significant effects on human communities, agroecosystems, and the socioeconomic development (Jones et al. 2015; Kukal and Irmak 2018; Liu et al. 2018). Thus, the number of heat-wave days (HD, represented by

the days with T_{max} greater than 35°C) and frost days (FD, calculated by the days with T_{min} less than 0°C) are selected to assess the performance of statistical and dynamic downscaling outputs on extreme events.

Symmetrical uncertainty (SU), based on the concept of information entropy, is employed additionally to evaluate the ability of GCMs and RCM in simulating temperatures, with the advantage of measuring the overall character of the distributions and conditional dependencies between the variables (Ahmed et al. 2019). Prior to obtaining SU, the entropy of observation [$H(\text{obs})$] and the entropy of observation given the value of simulation [$H(\text{obs}|\text{sim})$] are respectively calculated by Eqs. (6) and (7):

$$H(\text{obs}) = - \int p(\text{obs}) \log_2[p(\text{obs})] \quad \text{and} \quad (6)$$

$$H(\text{obs}|\text{sim}) = - \int p(\text{sim}) \int p(\text{obs}|\text{sim}) \log_2[p(\text{obs}|\text{sim})], \quad (7)$$

where $p(\text{obs})$ and $p(\text{obs}|\text{sim})$ are the prior probability of all values of observation and conditional probability of observation under the condition of simulation. SU is thus expressed as follows:

$$\text{SU} = 2 \left[\frac{H(\text{obs}) - H(\text{obs}|\text{sim})}{H(\text{obs}) + H(\text{sim})} \right], \quad (8)$$

where SU ranges from 0 to 1, implying from no agreement to perfect agreement of simulation relative to observation.

The systematic comparisons of simulations are performed between GCMs and PRECIS, between statistical and dynamic downscaling, as well as raw outputs and bias-corrected outputs, at the spatial and temporal dimensions in the baseline period of 1961–90.

An overall performance of one model can be evaluated by a comprehensive ranking metric (MR) (Jiang et al. 2015; Ahmed et al. 2019), which is defined as

$$\text{MR} = 1 - \frac{\sum_{i=1}^n \text{rank}_i}{mn}, \quad (9)$$

where m is the number of models; n is the number of indices, that is, 4, referring to SCC, RMSE, STD, and SU, respectively; and rank_i is the rank of index i .

3. Results

a. Comparison of bias between raw GCMs/PRECIS simulations and observation

There are obvious differences in T_m generated by raw GCMs, as seen by the spatial bias between

FIG. 2. Bias for Tm (°C) between raw GCMs simulations and observation (period: 1961–90).

simulations and observations illustrated in Fig. 2. A better agreement evaluated as a bias within $\pm 1^\circ\text{C}$ is recorded over HH, JH, SY, and SC (Table 2). Another general consistency is captured over TP where Tm bias is less than -3°C . For other regions, no uniform simulations are obtained from all GCMs. Specifically, the central and northern regions of mainland China (i.e., NE, IM, NC, Hh, JH, and Jha) show a negative bias in BNU-ESM, CNRM-CM5, INM-CM4.0, IPSL-CM5A-LR, and NorESM1-M, while they show a positive bias in CanESM2, MIROC5, MIROC-ESM, and MIROC-ESM-CHEM. The Tm over XJ and WE is generally simulated with

a bias over $\pm 1^\circ\text{C}$. The “statistic of grids” (Table 3) shows that there are 46.7% and 45.1% of all grids within $\pm 1^\circ\text{C}$ in CCSM4 and IPSL-CM5A-MR, respectively, which is better than other models. The higher percentage of grids with a bias over 1°C is nearly 61.8% in MIROC5, followed by 43.1% in MIROC-ESM and 40.7% in MIROC-ESM-CHEM. However, INM-CM4.0, CNRM-CM5, IPSL-CM5A-LR, and NorESM1-M are highlighted in generating a bias less than -1°C in 84.4%, 78.0%, 75.2%, and 74.9% of all grids, respectively; 59.2% of all grids in INM-CM4.0 have a bias of less than -3°C , which is much higher than others. For PRECIS

TABLE 2. Selection of raw models with good performance in simulating variables (with more than 50% grids showing better performance). Numeric symbols 1–21 represent as follows—1: ACCESS1.0, 2:BNU-ESM, 3:CCSM4, 4:CanESM2, 5:CNRM-CM5, 6:CSIRO Mk3.6.0, 7:GFDL CM3, 8:GFDL-ESM2G, 9:GFDL-ESM2M, 10:INM-CM4.0, 11:HadGEM2-ES, 12:IPSL-CM5A-LR, 13:IPSL-CM5A-MR, 14:MIROC5, 15:MIROC-ESM, 16:MIROC-ESM-CHEM, 17:MPI-ESM-LR, 18:MPI-ESM-MR, 19: MRI-CGCM3, 20:NorESM1-M, and 21:PRECIS.

Region	Tm (within $\pm 1^\circ\text{C}$)	Tmax (within $\pm 1^\circ\text{C}$)	Tmin (within $\pm 1^\circ\text{C}$)	DTR (within $\pm 1^\circ\text{C}$)	FD (within ± 5 days)	HD (within ± 5 days)
NE	3, 7, 8, 9, 17, 18, 21	3, 4, 14, 15, 16, 21	1, 4, 7, 13, 19, 20		1, 7, 20, 21	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
NC	3, 13, 19, 21	1, 5, 21	1, 4, 13, 18, 19, 21	4, 10, 19		4, 7, 8, 9, 12, 13, 15, 17, 18, 20
IM	3, 8, 9, 19, 21	3, 17, 21	2, 4, 6, 7, 13, 19, 20			1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
HH	1, 3, 5, 6, 10, 13, 16, 17, 18, 19, 21	1, 3, 4, 5, 6, 10, 17, 18, 19, 21	1, 6, 13, 18, 19, 21	1, 2, 3, 4, 6, 13, 14, 15, 16, 19	1, 9, 21	4, 7, 8, 13, 17, 18
XJ						1, 2, 4, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
WE	10	5	21	21		8, 12, 13, 17, 18, 19
JH	3, 4, 10, 13, 17, 18, 19, 21	3, 4, 13, 17, 18, 21	3, 13, 17, 18, 21	3, 5, 6, 7, 10, 14, 17, 18, 19	1, 17	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
Jha	6, 13, 19, 21	3, 4, 6, 17, 18, 19, 20, 21	13, 15, 21	5, 14, 17, 18, 19, 21	1	8, 12, 13, 17, 18
SY	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, 14, 17, 18, 19, 21	2, 3, 4, 6, 7, 8, 9, 14, 17, 18, 20, 21	1, 3, 6, 8, 9, 13, 14, 17, 18, 21	5, 6, 7, 17, 18, 19, 21	1, 8, 9, 14, 15, 16	2, 3, 4, 7, 13, 14, 17, 18
SC	1, 2, 3, 5, 6, 7, 8, 9, 10, 12, 13, 14, 17, 19	2, 4, 5, 6, 7, 13, 17, 18	1, 2, 3, 5, 6, 8, 10, 12, 13, 14, 17, 18, 21	5, 6, 8, 9, 17, 18, 19, 21	1, 2, 4, 5, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21	1, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 17, 18, 19, 20
SW	1, 7, 8, 13, 19, 21	6, 17, 21	1, 17, 18, 21	5, 8, 9, 20, 21	14, 15, 17	1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21
TP						1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21

TABLE 3. Percent statistic of grids for bias between raw GCMs/PRECI-S simulations and observations. Model numeric symbols 1–21 are as in Table 2. The six values in each cell are for the ranges ≤ -3 , from -3 to -1 , from -1 to 0, from 0 to 1, 1–3, and $>3^\circ\text{C}$ for Tm, Tmax, Tmin, and DTR and ≤ -5 , from -5 to -3 , from -3 to -1 , from -1 to 0, 0–5, 5–15, and >15 days for FD and HD.

Model	Tm	Tmax	Tmin	DTR	FD	HD
1	17.3, 27.1, 21.2, 16.1, 12.7, 5.6	36.8, 39.3, 13.2, 5.4, 3.4, 2.0	10.5, 21.9, 19.5, 19.0, 17.8, 11.2	36.1, 47.1, 11.2, 4.9, 0.7, 0.0	27.8, 16.3, 14.2, 11.9, 14.4, 15.4	5.8, 5.8, 45.6, 29.5, 11.0, 2.2
2	26.8, 23.6, 13.8, 12.2, 10.6, 13.0	40.6, 26.0, 10.6, 4.9, 10.6, 7.3	20.3, 26.8, 9.8, 13.8, 13.8, 15.4	32.5, 38.2, 13.0, 6.5, 8.9, 0.8	25.2, 6.5, 8.9, 4.9, 11.4, 43.1	4.1, 4.1, 35.8, 37.4, 10.6, 8.1
3	12.3, 26.6, 26.6, 20.1, 10.4, 4.0	26.1, 30.3, 16.6, 14.2, 10.9, 1.9	16.3, 28.5, 14.6, 16.0, 19.1, 5.5	19.7, 37.7, 11.0, 9.0, 17.0, 5.5	22.9, 18.1, 8.8, 7.9, 14.4, 27.9	1.7, 5.9, 47.4, 35.8, 7.0, 2.2
4	18.7, 18.7, 9.8, 15.4, 26.0, 11.4	15.4, 19.5, 19.5, 21.9, 17.9, 5.7	30.1, 12.2, 14.6, 11.4, 15.4, 16.3	14.6, 19.5, 8.9, 13.0, 23.6, 20.3	22.8, 14.6, 5.7, 3.3, 14.6, 39.0	0.0, 2.4, 30.1, 30.1, 17.1, 20.3
5	34.1, 43.6, 14.5, 5.0, 2.3, 0.6	20.9, 32.2, 21.5, 7.6, 10.7, 7.0	50.8, 36.6, 5.2, 4.5, 2.1, 0.8	2.3, 9.1, 10.1, 15.5, 21.5, 41.5	1.6, 2.1, 2.1, 8.3, 19.6, 66.3	1.6, 3.3, 36.8, 24.8, 15.5, 18.0
6	18.3, 22.9, 20.4, 17.9, 14.3, 6.1	35.8, 29.0, 16.9, 10.8, 5.7, 1.8	18.3, 24.0, 20.4, 13.6, 11.8, 11.8	30.5, 26.9, 16.5, 13.6, 11.8, 0.7	17.9, 7.2, 6.1, 7.2, 14.7, 47.0	2.1, 3.9, 35.8, 31.9, 18.6, 7.5
7	22.4, 31.8, 26.0, 11.5, 5.7, 2.6	64.6, 22.4, 6.3, 4.7, 1.6, 0.5	18.2, 28.6, 17.2, 14.6, 16.7, 4.7	49.5, 26.6, 9.9, 7.3, 6.8, 0.0	5.2, 9.4, 5.7, 12.5, 21.4, 45.8	5.7, 9.9, 56.8, 26.0, 1.6, 0.0
8	14.6, 24.0, 21.4, 19.3, 14.6, 6.3	50.0, 31.3, 6.8, 5.2, 2.1, 4.7	8.8, 16.1, 15.6, 13.5, 30.7, 15.1	68.2, 19.8, 6.3, 5.7, 0.0, 0.0	22.4, 13.0, 7.8, 16.7, 17.7, 22.4	6.3, 7.3, 47.9, 33.9, 4.7, 0.0
9	14.1, 26.6, 20.3, 18.2, 14.1, 6.8	53.6, 29.7, 4.7, 4.7, 2.6, 4.7	8.8, 16.7, 12.0, 16.1, 31.3, 15.1	67.7, 20.8, 6.8, 4.7, 0.0, 0.0	19.8, 15.6, 8.8, 15.1, 19.8, 20.8	6.3, 9.9, 53.6, 25.5, 4.2, 0.5
10	17.8, 32.2, 20.2, 14.4, 11.7, 3.7	34.1, 42.2, 11.2, 5.6, 4.9, 2.0	11.9, 34.1, 19.5, 12.4, 14.6, 7.3	24.9, 41.5, 20.2, 9.8, 3.7, 0.0	26.3, 11.5, 6.3, 10.2, 17.8, 27.8	6.1, 5.4, 42.7, 28.8, 13.4, 3.7
11	59.2, 25.2, 8.3, 3.2, 3.8, 0.3	47.8, 34.1, 8.9, 4.5, 2.9, 1.9	89.2, 7.3, 1.3, 0.6, 1.6, 0.0	2.9, 2.2, 3.2, 9.6, 23.6, 58.6	0.6, 1.6, 1.3, 6.5, 1.89, 8.3	4.1, 9.9, 42.0, 34.4, 8.3, 1.3
12	30.1, 45.1, 12.0, 9.0, 3.8, 0.0	68.4, 15.8, 5.3, 2.3, 8.3, 0.0	21.1, 39.9, 19.6, 8.7, 5.2, 2.3	43.6, 25.6, 12.0, 5.3, 11.3, 2.3	6.0, 6.8, 3.8, 9.8, 23.3, 50.4	6.8, 12.0, 75.2, 3.8, 2.3, 0.0
13	14.8, 29.3, 24.0, 21.1, 10.2, 0.7	38.2, 32.6, 8.9, 9.5, 8.9, 2.0	12.8, 20.1, 30.6, 14.1, 17.4, 4.9	37.8, 29.9, 7.9, 9.5, 10.2, 4.6	9.2, 7.9, 4.9, 12.8, 23.7, 41.5	3.0, 7.2, 72.0, 11.5, 4.3, 2.0
14	7.8, 5.8, 9.5, 15.1, 41.1, 20.7	16.7, 17.6, 15.9, 15.1, 25.8, 8.9	5.6, 5.0, 5.4, 11.8, 36.6, 35.7	43.2, 31.2, 13.2, 7.8, 4.3, 0.2	43.4, 22.1, 12.4, 5.0, 6.4, 10.7	0.8, 2.7, 38.8, 25.0, 11.2, 21.5
15	17.9, 15.4, 9.8, 13.8, 27.6, 15.4	31.7, 18.7, 15.4, 7.3, 14.6, 12.2	14.6, 11.4, 6.5, 14.6, 26.0, 26.8	52.0, 21.1, 7.3, 7.3, 9.8, 2.4	33.3, 13.8, 11.4, 10.6, 8.9, 21.9	1.6, 5.7, 32.5, 27.6, 8.9, 23.6
16	17.9, 16.3, 8.9, 16.3, 26.8, 13.8	32.5, 20.3, 13.0, 7.3, 15.4, 11.4	14.6, 10.6, 9.8, 12.2, 27.6, 25.2	50.4, 23.6, 8.9, 4.1, 9.8, 3.3	32.5, 16.3, 6.5, 11.4, 9.8, 23.6	1.6, 4.1, 34.1, 22.8, 10.6, 26.8
17	9.3, 14.3, 13.6, 25.8, 26.5, 10.4	19.7, 29.4, 19.4, 14.0, 10.0, 7.5	7.2, 8.2, 8.2, 19.4, 34.4, 22.6	36.6, 41.6, 12.2, 7.2, 2.5, 0.0	45.2, 19.0, 9.0, 8.6, 7.9, 10.4	2.5, 9.0, 66.0, 15.4, 6.8, 0.4
18	10.0, 14.0, 14.3, 22.6, 28.3, 10.8	19.7, 30.8, 19.4, 12.9, 10.0, 7.2	7.9, 7.9, 9.7, 21.5, 30.5, 22.6	36.6, 41.2, 11.5, 7.5, 2.5, 0.7	45.5, 17.6, 7.9, 9.0, 8.2, 11.8	2.5, 10.4, 64.5, 16.5, 5.7, 0.4
19	21.8, 30.8, 26.1, 14.6, 5.0, 1.7	34.9, 41.5, 15.9, 4.2, 2.4, 1.2	19.7, 27.8, 19.4, 17.5, 12.0, 3.5	20.5, 41.3, 20.1, 12.8, 5.2, 0.1	8.1, 14.9, 11.9, 12.7, 16.4, 36.0	0.9, 3.4, 42.4, 25.9, 10.7, 16.7
20	26.6, 48.3, 12.3, 5.4, 4.4, 3.0	53.2, 30.1, 9.4, 3.5, 1.5, 2.5	22.2, 35.0, 21.2, 9.4, 5.9, 6.4	38.4, 27.6, 13.3, 5.4, 13.8, 1.5	14.3, 13.8, 8.4, 8.4, 13.8, 41.4	5.4, 13.8, 66.0, 14.3, 0.5, 0.0
21	10.9, 12.0, 17.3, 25.3, 27.9, 6.6	14.1, 15.5, 18.2, 21.4, 24.4, 6.3	14.4, 16.6, 22.0, 19.8, 19.7, 7.4	9.8, 18.1, 17.7, 21.7, 25.0, 7.7	7.9, 13.3, 15.8, 16.0, 19.9, 27.2	0.3, 0.3, 32.5, 23.4, 16.9, 26.7

FIG. 3. Bias for variables (Tm, Tmax, Tmin, and DTR: °C; FD and HD: days) between raw PRECIS simulation and observation (period: 1961–90).

(Fig. 3), Tm is underestimated with a bias less than -1°C over the TP, similar to GCMs. Tm has a bias within $\pm 1^{\circ}\text{C}$ over the central and northern regions of mainland China and is overestimated over the southern and most northwestern regions. There are 42.6% of all grids with a bias within $\pm 1^{\circ}\text{C}$, followed by 34.5% with a bias over 1°C and 22.9% with a bias of less than -1°C (Table 3). Comparisons between raw PRECIS and GCMs indicate better performance in

PRECIS than HadGEM2-ES, and similar performance to MPI-ESM-LR.

Bias of Tmax seems to follow a general underestimation in most GCMs (Fig. S1 in the online supplemental material), relative to that of Tm. Tmax is well simulated over NE, HH, JH, Jha, SY, and SC (Table 2), while it is underestimated in XJ, WE, and TP. With regard to SY and SC, better simulations with a bias within $\pm 1^{\circ}\text{C}$ are drawn from BNU-ESM, CCSM4, CSIRO Mk3.6.0,

MIROC5, MPI-ESM-LR and MPI-ESM-MR, while underestimation with a bias less than -1°C is drawn from ACCESS1.0, CNRM-CM5, GFDL CM3, GFDL-ESM2G, GFDL-ESM2M, HadGEM2-ES, INM-CM4.0, and MRI-CGCM3. Underestimation with a bias over 1°C occurs with IPSL-CM5A-MR, MIROC-ESM, and MIROC-ESM-CHEM. Tmax across central regions (including NC, HH, JH, and Jha) are estimated well in CCSM4, CNRM-CM5, HadGEM2-ES, and MRI-CGCM3, while they are overestimated in MIROC5 and underestimated in other GCMs. Cold bias is shown in IM and NE for most GCMs except CanESM2 and MIROC5. As shown in Table 3, statistical comparisons demonstrate a prominent detection of cold bias less than -1°C in GFDL CM3, IPSL-CM5A-LR, GFDL-ESM2M, NorESM1-M, INM-CM4.0, GFDL-ESM2G, MRI-CGCM3, HadGEM2-ES, ACCESS1.0, IPSL-CM5A-MR, BNU-ESM, and CSIRO Mk3.6.0, with over 65.0% of total grids. MIROC5 is highlighted, having 34.7% of total grids in a major warm bias over 1°C , and CanESM2 is also prominent, producing 41.5% of total grids in a bias within $\pm 1^{\circ}\text{C}$. Similar to Tm, Tmax generated by PRECIS is overestimated over southern and northwestern regions, underestimated over TP, and better estimated over central and most northern regions (Table 2 and Fig. 3). The percentages of grids with a bias more negative than -1°C , within $\pm 1^{\circ}\text{C}$, and more positive than 1°C are 29.7%, 39.6%, and 30.8% (Table 3), respectively. Relative to GCMs, much more agreement of Tmax is detected in PRECIS.

Raw GCMs' outputs illustrate inconsistent simulations for Tmin (Fig. S2 in the online supplemental material), in comparison with some consistent simulations for Tm and Tmax. Tmin in NE, NC, IM, HH, JH, SY, and SC are closer to observation, derived from at least four GCMs (Table 2). Conspicuously, CNRM-CM5 and INM-CM4.0 point out a significant national cold bias. Besides, similar underestimation from CCSM4, GFDL CM3, IPSL-CM5A-LR, MRI-CGCM3, and NorESM1-M is detected around southern HH, JH, Jha, SY, SC, SW, and TP. Over NE, IM, NC, WE, and XJ, Tmin is overestimated in GFDL-ESM2G, GFDL-ESM2M, MIROC5, MIROC-ESM, MIROC-ESM-CHEM, MPI-ESM-LR, and MPI-ESM-MR. Comparing the maps of percent statistics of grids in Table 3, it is clear that INM-CM4.0 and CNRM-CM5 are the two models that most underestimate Tmin with a percentage of grids with a bias less than -1°C of 96.5% and 87.4%, respectively. It is accepted that only IPSL-CM5A-MR produces more grids with better simulation. Drawing from Fig. 3 and Table 3, it is evident that Tmin in PRECIS is underestimated over TP and overestimated over northern NE, eastern IM, eastern XJ, and central SC. A good estimate is recorded in 41.8% of all grids, higher than overestimated ratio (31.1%)

and underestimated ratio (27.1%). The performance of Tmin in PRECIS is better than most of GCMs, except IPSL-CM5A-MR (well estimated in 44.7% of all grids).

Integrating the simulations of Tmax and Tmin, the performance of DTR is obviously distinct among GCMs (Fig. S3 in the online supplemental material). Only HH, JH, Jha, SY, SC, and SW are well estimated in some GCMs (Table 2). There are major findings that DTR over XJ, WE, and TP are underestimated in GCMs except CanESM2, CNRM-CM5, and INM-CM4.0. A widespread underestimation of DTR over NE, IM, NC, and HH is derived from most GCMs excluding CNRM-CM5 and INM-CM4.0. Most differences of DTR bias are recorded over JH, Jha, SY, SC and SW, where DTR is well estimated from CNRM-CM5, MIROC5, MPI-ESM-MR, and MRI-CGCM3; overestimated from BNU-ESM, CanESM2, CCSM4, INM-CM4.0, IPSL-CM5A-MR, MIROC-ESM, MIROC-ESM-CHEM, and NorESM1-M; and underestimated from ACCESS1.0 and HadGEM2-ES. Notably, the percentage of grids with a bias over 1°C is 82.2% and 63.0% in INM-CM4.0 and CNRM-CM5, respectively, which is much higher than other GCMs (Table 3). The percent of grids with a bias less than -1°C is more than 50.0% in most GCMs except CanESM2, CNRM-CM5, and INM-CM4.0. Scattered bias of DTR in PRECIS is exhibited across China as shown in Fig. 3, as a result of a difference between Tmax and Tmin. In the percent statistic of grids for DTR bias, it is deduced that 39.4% of total grids produce better estimation, followed by 32.7% and 27.9% referring to underestimation and overestimation, respectively. It is agreed that PRECIS is better than GCMs in simulating DTR. Overall, the simulations from both GCMs and PRECIS are nevertheless characterized by some biases, suggesting the inevitability of bias correction for raw outputs. Of course, temperatures are well estimated from PRECIS due to its higher resolution relative to GCMs.

When we explore the performance of models in simulations on extreme events, it is highlighted that well-estimated FD in a bias within ± 5 days is located in SY and SC from GCMs (Table 2 and online supplemental Fig. S4), and well-estimated HD is in most regions except central parts of mainland China (Table 2 and online supplemental Fig. S5). Generally, overestimation in FD is shown in NC, HH, JH, Jha, SY, and SW, for most GCMs. FD is underestimated in NE and IM from CanESM2, CCSM4, MIROC-ESM, MIROC, MIROC-ESM-CHEM, MPI-ESM-MR, and MPI-ESM-LR. For PRECIS, FD is well estimated in eastern regions and overestimated in XJ, TP, and northern SW, suggesting better performance than most GCMs. Percent statistics of grids (Table 3) also show better estimation in PRECIS (31.8%) is larger than GCMs (2.9%–26.1%). HD across

FIG. 4. Bias for Tm (°C) between SD-GCMs simulations and observation (period: 1961–90).

NC, HH, JH, Jha, SY, and SW is either underestimated or overestimated in most GCMs. Compared to GCMs, more overestimation of HD is drawn from PRECIS, which produces higher HD bias across central-eastern regions and XJ. The percent of overestimation is 43.6%, larger than GCMs. Overall, the performance of FD in PRECIS is better than GCMs, but HD does not perform as well as CGMs.

b. Comparison of bias between SD-GCMs/corrected PRECIS simulations and observation

The performance of reproduced Tm from statistical-downscaling GCMs (SD-GCMs) is also evaluated to

quantify the extent of improvement from the BCSD method. Extracting from Fig. 4, bias for Tm between SD-GCMs simulations and observation is mostly within $\pm 0.2^{\circ}\text{C}$, much less than that between raw GCMs' simulations and observation. It is clearly indicated that 60.0% of all grids have a bias within $\pm 0.2^{\circ}\text{C}$ for most SD-GCMs, among which the percentage is more than 90.0% for NorESM1-M, CanESM2, IPSL-CM5A-LR, BNU-ESM, GFDL-ESM2M, MPI-ESM-MR, and MPI-ESM-LR (Table 4). It is also noticed that there are still some biases with values exceeding 0.2°C in some SD-GCMs. A warm bias existed over NE, IM, NC, and

TABLE 4. Percent statistic of grids for bias between SD-GCMs/corrected PRECIS simulations and observation. The numeric symbols 1–21 are as in Table 2. The six values in each cell are for the ranges ≤ -0.5 , from -0.5 to -0.2 , from -0.2 to 0 , 0.2 – 0.5 , and $>0.5^\circ\text{C}$ for Tm, Tmax, Tmin, and DTR and ≤ -5 , from -5 to -2 , from -2 to 0 , 0 – 2 , 2 – 5 , and >5 days for FD and HD.

Model	Tm	Tmax	Tmin	DTR	FD	HD
1	0.0, 3.6, 34.6, 54.4, 7.3, 0.0	0.0, 10.0, 58.4, 30.9, 0.8, 0.0	0.0, 7.9, 40.6, 42.9, 8.6, 0.0	0.0, 39.0, 51.4, 9.6, 0.0, 0.0	2.2, 34.0, 43.0, 17.9, 2.9, 0.0	0.0, 2.1, 49.0, 43.7, 3.7, 1.6
2	0.0, 0.5, 47.9, 49.6, 2.0, 0.0	0.0, 12.6, 58.4, 29.0, 0.0, 0.0	0.0, 0.4, 29.1, 53.8, 16.7, 0.0	0.0, 30.1, 58.5, 11.4, 0.0, 0.0	10.2, 39.3, 31.8, 15.2, 3.3, 0.1	0.0, 2.2, 47.0, 44.3, 5.7, 0.7
3	0.0, 0.5, 22.5, 66.0, 10.9, 0.0	0.0, 16.1, 38.3, 34.3, 9.6, 1.8	0.0, 9.4, 50.1, 27.9, 12.5, 0.0	0.0, 7.7, 41.7, 44.1, 6.5, 0.0	15.6, 11.1, 32.3, 15.9, 6.5, 18.7	2.9, 5.8, 48.9, 31.5, 6.7, 4.3
4	0.0, 0.0, 48.8, 50.3, 0.8, 0.0	0.0, 0.4, 66.8, 32.4, 0.4, 0.0	0.0, 1.0, 39.5, 49.9, 9.6, 0.0	0.0, 16.2, 50.9, 30.4, 2.5, 0.0	2.8, 17.9, 42.8, 28.6, 7.3, 0.7	0.3, 4.8, 48.8, 39.8, 5.7, 0.5
5	0.0, 0.0, 5.6, 80.5, 13.9, 0.0	3.0, 36.9, 56.3, 3.8, 0.0, 0.0	0.0, 0.0, 4.2, 76.0, 19.8, 0.0	6.7, 47.9, 44.2, 1.1, 0.0, 0.0	10.2, 34.7, 27.7, 20.8, 5.4, 1.1	0.3, 2.4, 45.5, 43.1, 7.8, 0.9
6	0.0, 0.0, 6.3, 80.6, 13.2, 0.0	0.0, 0.0, 18.6, 74.1, 7.2, 0.0	0.0, 0.0, 3.2, 57.1, 39.6, 0.0	0.0, 17.5, 64.6, 17.6, 0.3, 0.0	9.4, 41.2, 34.8, 13.0, 1.5, 0.1	0.0, 0.3, 41.4, 45.3, 12.1, 0.9
7	0.0, 0.6, 26.1, 60.0, 13.3, 0.0	0.0, 5.6, 47.3, 42.4, 4.8, 0.0	0.0, 0.3, 10.5, 55.6, 33.2, 0.4	0.2, 41.4, 50.7, 7.7, 0.0, 0.0	7.4, 28.7, 37.2, 23.2, 3.1, 0.4	0.5, 4.6, 56.5, 34.8, 2.8, 0.8
8	0.0, 0.1, 19.3, 68.0, 12.6, 0.0	0.0, 6.1, 37.4, 54.0, 2.5, 0.0	0.0, 0.0, 7.5, 51.5, 41.0, 0.0	0.0, 41.7, 50.9, 7.4, 0.0, 0.0	6.0, 26.8, 47.4, 19.0, 0.7, 0.0	0.2, 5.3, 45.6, 41.7, 6.4, 0.8
9	0.0, 0.1, 24.0, 73.0, 3.0, 0.0	0.0, 8.4, 61.3, 29.7, 0.5, 0.0	0.0, 0.0, 7.5, 62.6, 29.9, 0.0	0.1, 53.1, 44.0, 2.9, 0.0, 0.0	2.4, 23.1, 46.3, 25.7, 2.6, 0.0	0.1, 2.3, 58.0, 30.6, 7.2, 1.8
10	0.0, 0.0, 0.3, 48.8, 49.0, 2.0	0.0, 0.0, 6.3, 47.2, 46.5, 0.0	0.0, 0.0, 0.8, 39.7, 52.7, 6.9	0.0, 14.0, 64.9, 18.1, 3.0, 0.0	63.3, 21.3, 11.1, 4.1, 0.1, 0.0	0.4, 4.6, 46.3, 41.9, 5.6, 1.1
11	0.0, 0.1, 17.0, 47.2, 35.4, 0.3	0.0, 0.7, 25.2, 37.8, 35.6, 0.7	0.0, 0.8, 18.9, 45.3, 30.5, 4.5	0.0, 14.1, 45.4, 30.7, 9.7, 0.0	21.1, 36.0, 15.2, 10.5, 9.8, 7.5	0.6, 4.0, 46.2, 36.1, 10.6, 2.5
12	0.0, 1.6, 41.1, 56.9, 0.5, 0.0	0.0, 9.8, 56.7, 33.5, 0.0, 0.0	0.0, 0.2, 24.9, 65.1, 9.9, 0.0	0.0, 23.1, 67.5, 9.4, 0.0, 0.0	3.1, 32.2, 38.3, 19.8, 5.1, 1.4	1.4, 10.2, 70.6, 17.1, 0.7, 0.0
13	0.0, 0.0, 27.6, 60.5, 11.9, 0.0	0.0, 0.4, 39.6, 53.7, 6.4, 0.0	0.0, 0.5, 19.1, 52.2, 28.1, 0.0	0.0, 13.1, 68.6, 17.9, 0.4, 0.0	2.2, 31.3, 44.7, 18.6, 2.6, 0.6	0.4, 5.4, 68.7, 25.1, 0.4, 0.0
14	0.0, 0.0, 5.5, 64.5, 29.1, 1.0	0.0, 0.0, 13.5, 65.0, 21.4, 0.1	0.0, 0.0, 0.9, 49.0, 47.8, 2.3	0.4, 22.2, 55.3, 21.9, 0.2, 0.0	60.2, 27.5, 42.3, 22.4, 1.8, 0.0	0.0, 0.1, 49.3, 41.4, 8.9, 0.3
15	0.0, 4.0, 19.6, 42.6, 32.4, 1.5	0.0, 7.1, 27.3, 38.8, 25.4, 1.4	0.0, 1.8, 12.1, 38.2, 45.1, 2.9	0.0, 20.8, 56.8, 20.4, 2.1, 0.0	5.8, 47.9, 29.1, 11.4, 4.3, 1.5	0.3, 3.1, 44.9, 36.2, 9.1, 6.3
16	0.0, 0.6, 16.5, 54.4, 28.3, 0.3	0.0, 0.8, 27.5, 44.9, 25.5, 1.2	0.0, 0.3, 7.9, 44.1, 46.0, 1.7	0.0, 21.8, 50.8, 24.5, 2.9, 0.0	8.6, 34.0, 39.4, 14.4, 3.0, 0.5	0.1, 2.6, 50.5, 33.1, 7.8, 6.0
17	0.0, 0.0, 21.8, 69.4, 8.8, 0.0	0.0, 0.3, 42.3, 54.9, 2.5, 0.0	0.0, 0.0, 10.0, 57.1, 32.9, 0.0	0.0, 25.6, 62.0, 12.3, 0.0, 0.0	4.6, 40.0, 42.0, 12.6, 0.8, 0.0	0.0, 2.5, 58.0, 37.5, 1.8, 0.2
18	0.0, 2.4, 47.8, 46.9, 3.0, 0.0	0.0, 8.1, 57.4, 33.6, 0.8, 0.0	0.0, 1.5, 34.0, 44.3, 20.2, 0.0	0.0, 22.0, 65.5, 12.2, 0.3, 0.0	1.6, 34.7, 36.0, 23.0, 4.6, 0.1	0.1, 4.4, 61.6, 31.5, 2.1, 0.3
19	0.0, 0.0, 4.4, 47.3, 48.2, 0.0	0.0, 0.1, 6.4, 74.2, 19.3, 0.0	0.0, 1.3, 6.7, 34.7, 57.1, 0.3	0.0, 22.0, 52.7, 20.6, 4.5, 0.3	15.0, 49.8, 28.5, 5.2, 1.5, 0.1	0.0, 0.5, 48.6, 40.7, 9.4, 0.8
20	0.0, 0.0, 31.4, 68.3, 0.3, 0.0	0.0, 2.3, 47.8, 48.4, 1.5, 0.0	0.0, 0.0, 17.8, 73.6, 8.7, 0.0	0.0, 15.6, 58.9, 22.9, 2.5, 0.0	3.3, 28.4, 39.7, 25.1, 3.5, 0.1	0.0, 4.8, 65.1, 29.3, 0.7, 0.0
21	0.0, 3.8, 43.6, 52.2, 0.4, 0.0	0.0, 7.0, 63.4, 29.5, 0.0, 0.0	0.5, 3.9, 28.1, 60.7, 6.7, 0.0	0.0, 13.1, 73.2, 10.3, 2.8, 0.6	1.9, 12.1, 39.0, 37.6, 8.4, 1.0	0.1, 1.6, 45.5, 42.6, 9.8, 0.4

FIG. 5. Bias for variables (Tm, Tmax, Tmin, and DTR: °C; FD and HD: days) between corrected PRECIS simulation and observation (period: 1961–90).

eastern WE that was mainly derived from HadGEM2-ES, MRI-CGCM3, INM-CM4.0, MIROC-ESM, MIROC, and MIROC-ESM-CHEM. No more than 4.0% of all grids show a bias between -0.5°C and -0.2°C , and a cold bias less than -0.5°C is not tracked. With regard to PRECIS (Fig. 5), similar improvement is shown in a bias within $\pm 0.2^{\circ}\text{C}$ from corrected outputs, and a cold bias is tracked only in a few parts of XJ. Considering

percent statistics in Table 4, there are 95.8%, 3.8%, and 0.4% for Tm bias within $\pm 0.2^{\circ}\text{C}$, less than -0.2°C , and over 0.2°C , respectively. Maps of comparison of probability distribution of Tm (Figs. 6 and 7) indicate good consistency between corrected models and observation that is much better than that between raw outputs and observation.

Definite improvement of Tmax is ascertained from SD-GCMs, characterized by a bias within $\pm 0.2^{\circ}\text{C}$ across

FIG. 6. Probability distribution of T_m from raw GCMs, SD-GCMs, and observations (the red line represents the probability distribution from observations; period: 1961–90).

major regions (Fig. 8). A bias less than -0.2°C is located in some regions, for example, XJ and WE in CNRM-CM5, NE in BNU-ESM, and TP in CCSM4 and MIROC-ESM. Besides, a bias above 0.2°C is also present in some regions, for example, NE, IM, and NC in HadGEM2-ES, INM-CM4.0, MIROC-ESM, and MIROC-ESM-CHEM;

SW in MIROC-ESM-CHEM; and TP in HadGEM2-ES. It is clear that over 53.5% of all grids are estimated in a bias within $\pm 0.2^\circ\text{C}$ from all SD-GCMs, among which the percentage is over 90.0% for CanESM2, MPI-ESM-LR, NorESM1-M, IPSL-CM5A-MR, CSIRO Mk3.6.0, GFDL-ESM2G, GFDL-ESM2M, MPI-ESM-MR,

FIG. 7. Probability distribution of T_m , $T_{\max}$, $T_{\min}$, and DTR from raw PRECIS, corrected PRECIS, and observations (The red line indicates the probability distribution from observations; period: 1961–90).

FIG. 8. Bias for $T_{\max}$ ($^{\circ}\text{C}$) between SD-GCMs simulations and observation (period: 1961–90).

and IPSL-CM5A-LR (Table 4). A bias above 0.2°C is found in 46.5%, 36.4%, 26.8%, 26.8%, and 21.5% of grids from HadGEM2-ES, INM-CM4.0, MIROC-ESM, MIROC-ESM-CHEM, and MIROC, respectively. A bias less than -0.2°C is present in 39.9% from CNRM-CM5. It is shown from Fig. 5 and Table 4 that 93.0% of all grids for PRECIS are corrected in a bias within $\pm 0.2^{\circ}\text{C}$, and only 7.0% are still in a bias less than -0.2°C .

Relative to the raw output of GCMs, it is sure that $T_{\min}$ for SD-GCMs is improved. As presented in Fig. 9, a bias within $\pm 0.2^{\circ}\text{C}$ is seen for the corrected GCMs in most regions, and a bias over 0.2°C is detected in some

northern regions (i.e., NE, IM, and NC). Although the percent of grids for $T_{\min}$ in a bias within $\pm 0.2^{\circ}\text{C}$ is relatively lower than for T_m and $T_{\max}$, there are still over 50.0% grids in most SD-GCMs among which the highest percent is 91.4% from NorESM1-M, as shown in Table 4. There are over 40.0% grids in a bias over 0.2°C for HadGEM2-ES, MRI-CGCM3, MIROC5, MIROC-ESM, MIROC-ESM-CHEM, and GFDL-ESM2G. For PRECIS in Fig. 5 and Table 4, $T_{\min}$ is also corrected in a bias within $\pm 0.2^{\circ}\text{C}$, which is obviously improved over most regions accounting for 88.5% across mainland China.

FIG. 9. Bias for T_{min} (°C) between SD-GCMs simulations and observation (period: 1961–90).

Apparent evaluations for DTR between SD-GCMs simulations and observation state that DTR over most regions are statistically downscaled to reduce the bias within $\pm 0.2^\circ\text{C}$ (Fig. 10). There are other exceptions; for example, a bias less than 0.2°C is illustrated over some northern regions in most SD-GCMs and only some central and southern regions in INM-CM4.0, MIROC-ESM, MIROC-ESM-CHEM, MRI-CGCM3, and NorESM1-M. Statistical results for grids from Table 4 show that the percent of a bias within $\pm 0.2^\circ\text{C}$ ranges from 45.3% for CNRM-CM5 to 86.5% for IPSL-CM5A-MR, and the percent of a bias less than -0.2°C varies from 7.7% for

CCSM4 to 54.7% for CNRM-CM5, while the percent of a bias over 0.2°C is below 9.8% for all SD-GCMs. As shown in Fig. 5 and Table 4, a bias within $\pm 0.2^\circ\text{C}$ is shown in 83.5% grids, while a bias less than -0.2°C is shown in 13.1% grids, and a bias over 0.2°C is shown in only 3.4% grids. From comparisons of probability distributions of T_{max}, T_{min}, and DTR for raw outputs, corrected models, and observations (Figs. S6–S8 in the online supplemental material), the agreeable improvement is clear in the relation between corrected models and observations, much better than that between raw models and observations.

FIG. 10. Bias for DTR ($^{\circ}\text{C}$) between SD-GCMs simulations and observation (period: 1961–90).

The improvements of FD and HD are identified from SD-GCMs in comparison with raw GCMs (Figs. 11 and 12). Except HadGEM2-ES, INM-CM4.0, MIROC-ESM, and MRI-CGCM3, FD with a bias within ± 2 days is observed in NE, NC, SY and SW. FD is statistically downscaled to a bias within ± 2 days in most regions from corrected PRECIS (Fig. 5). As shown in Table 4, the percent statistic for FD announces that 76.6% grids for PRECIS, higher than SD-GCMs (15.2%–71.9%), implying PRECIS performs better than SD-GCMs in reproducing FD. When GCMs are downscaled, HD derived from most SD-GCMs is reproduced in a bias

within ± 2 days across most regions. It is also noted that HD in SC is estimated with a bias over 5 days from GFDL-ESM2M, MIROC-ESM, and MIROC-ESM-CHEM, while the bias of HD in SC is less than -5 days from CCSM4 and INM-CM4.0. HD from corrected PRECIS with a bias within ± 2 days is demonstrated across most regions (Fig. 5). From the percent statistic (Table 4), more than 90.0% grids are extracted in a bias within ± 2 days for HD from ACCESS1.0, BNU-ESM, GFDL CM3, IPSL-CM5A-MR, MIROC5, MPI-ESM-LR, MPI-ESM-MR, and NorESM1-M, higher than PRECIS (88.1%).

FIG. 11. Bias for FD (days) between SD-GCMs simulations and observation (period: 1961–90).

To explore the temporal characteristics with the help of models, it is clear that the yearly simulations vary in a wide range among raw GCMs/PRECIS outputs (Fig. 13). Most GCMs produce underestimated T_m , T_{min} , and DTR but overestimated FD and HD. Among models, the yearly simulated T_m is largest for MIROC5 (6.43° – 7.81° C) but smallest in INM-CM4.0 (1.32° – 2.76° C); T_{max} is largest for PRECIS (1.61° – 3.07° C) but smallest for GFDL CM3 (7.21° – 9.24° C); T_{min} is largest for MIROC5 (11.50° – 13.29° C) but smallest for IPSL-CM5A-LR (-2.48° to -0.51° C); DTR is largest for CanESM2 (12.29° – 12.88° C) but smallest for GFDL-ESM2G (7.70° – 8.57° C)

and GFDL-ESM2M (7.96° – 8.48° C); FD is largest for INM-CM4.0 (222–246 days), but relatively smaller for MIROC5, MPI-ESM-LR, and MPI-ESM-MR; HD is relatively higher from PRECIS, CanESM2, and MIROC-ESM-CHEM but relatively smaller from GFDL CM3, GFDL-ESM2G, and GFDL-ESM2M. Overall, it is also agreed that the yearly simulations from PRECIS (except HD) is better than most GCMs. Comparatively, yearly reproductions from SD-GCMs/corrected PRECIS are much closer to observation, fluctuating within a small differential magnitude, as drawn from Fig. 13.

FIG. 12. Bias for HD (days) between SD-GCMs simulations and observation (period: 1961–90).

c. Comparison of metrics between GCMs/PRECIS simulations and observation

Yearly SCC for variables between raw GCMs/PRECIS simulations and observation vary during 1961–90 (Fig. 14). Generally, a relatively lower SCC for Tm is detected in BNU-ESM, MIROC-ESM, and MIROC-ESM-CHEM in which the ranges of SCC are 0.86–0.89, 0.86–0.90, and 0.85–0.90, respectively, while SCC is nearly above 0.90 in other raw GCMs and PRECIS. Similarly, SCC for Tmax from BNU-ESM, GFDL-ESM2G, GFDL-ESM2M, MIROC-ESM, and MIROC-ESM-CHEM is 0.82–0.86,

0.86–0.89, 0.86–0.89, 0.82–0.87, and 0.81–0.87, respectively, while SCC for other raw GCMs and PRECIS is more than 0.90. For Tmin, SCC is relatively lower for MIROC-ESM (0.87–0.91) and MIROC-ESM-CHEM (0.87–0.91), while SCC is generally 0.90–0.97 for other raw GCMs/PRECIS outputs. SCC for DTR is less than 0.80 for raw GCMs/PRECIS outputs, revealing bad performance compared to Tmax and Tmin. SCC for FD is relatively lower for BNU-ESM (0.89–0.92) and INM-CM4.0 (0.88–0.93). SCC for HD is distinctly different among GCMs, ranging from BNU-ESM (0.02–0.61) to MRI-CGCM3 (0.44–0.82). Comparatively, SCC for

FIG. 13. Yearly variation of raw GCMs/PRECIS simulations, SD-GCMs/corrected PRECIS simulations, and observations (T_m , T_{max} , T_{min} , and DTR: $^{\circ}\text{C}$; FD and HD: days).

variables between SD-GCMs/corrected PRECIS simulations and observation is significantly higher than that between raw GCMs/PRECIS outputs and observation.

Maps of variation of yearly RMSE for T_m (Fig. 15) indicate that RMSE is 2.1° – 5.5°C between raw GCMs/PRECIS outputs and observation, and the relatively higher value is 4.0° – 5.5°C in INM-CM4.0. Similar to T_m , RMSE for T_{min} is found between 2.2° and 8.4°C for models, in which INM-CM4.0 and CNRM-CM5 are the top two, with respectively higher RMSE being 6.6° – 8.4°C and 5.0° – 6.3°C . As for T_{max} , multiple models (i.e., MIROC-ESM, MIROC-ESM-CHEM, IPSL-CM5A-LR, GFDL CM3, GFDL-ESM2G, and GFDL-ESM2M) are detected in RMSE with a variation of 4.3° – 6.0°C . Higher RMSE for DTR is recorded in CNRM-CM5, GFDL CM3, GFDL-ESM2G, and GFDL-ESM2M, with respective values of

2.0° – 3.6°C , 2.3° – 3.0°C , 3.6° – 3.3°C , and 2.6° – 3.1°C , while RMSE in other models generally varies within 1.0° – 2.3°C . RMSE for FD varies within 21.32–87.41 days from GCMs/PRECIS outputs, and higher values are shown in INM-CM4.0 (64.71–87.41 days). For HD, RMSE ranges from 4.38 to 32.57 days, and higher values are shown for MIROC-ESM (12.21–32.57 days) and MIROC-ESM-CHEM (12.87–31.86 days). Evidence from RMSE for variables between SD-GCMs/corrected PRECIS simulations and observation indicate lower values relative to raw GCMs/PRECIS, with general values of 0.3° – 1.4°C , 0.3° – 1.8°C , 0.3° – 1.6°C , 0.2° – 1.3°C , 6.93–18.18 days, and 1.97–15.14 days, respectively.

Similar variations are seen for STD as to what extent differences between GCMs/PRECIS simulations and observation are consistent for T_m , T_{max} , T_{min} , and DTR,

FIG. 14. Variation of yearly SCC for variables between GCMs/PRECIS simulations and observations (the boxes indicate the upper and lower quartile of the value during 1961–90; the whiskers indicate the extreme value; the pairs of boxes of the same color indicate SCC for the same model, with the left one being for raw GCMs or PRECIS and the right one being for SD-GCMs or corrected PRECIS).

as demonstrated in Fig. 16. General STD greater than 1.7°C , 2.1° , and 2.0°C is obviously demonstrated for Tm, Tmax, and Tmin, respectively. Higher STD is shown for MIROC-ESM and MIROC-ESM-CHEM, with values over 4.0° , 4.5° , and 3.9°C for Tm, Tmax, and Tmin, respectively. For DTR, STD is 1.9° – 2.7°C for CNRM-CM5, higher than those of other models. The range of STD is 21.02–54.88 days for FD and 4.28–28.14 days for HD, varying among GCMs/PRECIS. The improvement of SD-GCMs/corrected PRECIS is also evidenced by decreased STD relative to that of raw GCMs/PRECIS outputs.

Figure 17 indicates that SU for GCMs/PRECIS simulations is generally smaller than 0.78, 0.69, 0.65, 0.62, 0.77, and 0.66 for Tm, Tmax, Tmin, DTR, FD, and HD, though SU varies among years in models. A relatively lower SU for Tmax, Tmin, and DTR is identified from CCSM4, with respective magnitudes of 0.31–0.55, 0.32–0.49, and 0.17–0.48. For FD and HD, the mean value of yearly SU is smaller but the range of SU is larger relative to Tm, Tmax, Tmin, and DTR. SU tends to be much larger, with respective values greater than 0.80, 0.70, 0.77, 0.71, 0.73, and 0.45 for Tm, Tmax, Tmin, DTR, FD, and HD,

FIG. 15. As in Fig. 14, but for RMSE.

when it is measured from SD-GCMs/corrected PRECIS simulations.

The comparisons for Tm, Tmax, Tmin, and DTR are additionally conducted between simulations and observation during seasons (i.e., MAM, JJA, SON, and DJF), as shown in Fig. S9 in the online supplemental material. The capability of simulations from SD-GCMs/corrected PRECIS is obviously improved compared to that from raw GCMs/PRECIS outputs, which is proved by the evidence of increased SCC and decreased RMSE and STD during seasons. Furthermore, seasonal comparisons indicate that the simulated performance of variables in

summer and autumn are typically better than those in spring and winter.

Derived from the four metrics, that is, SCC, RMSE, STD, and SU, the MR index shows the general performance of SD-GCMs/corrected PRECIS outputs in reproducing the features of temperatures. From the top five ranks of models (Table 5), it can be deduced that different models are better at generating Tm, Tmax, Tmin, DTR, FD, and HD in different regions. Overall, IPSL-CM5A-LR, CSIRO Mk3.6.0, PRECIS, NorESM1-M, and IPSL-CM5A-MR are relatively better at reproducing these variables.

FIG. 16. As in Fig. 14, but for STD.

4. Discussion

In the field of climate change assessment, to some extent GCMs are generally unable to capture climate characteristics at the regional scale due to their coarse-resolution and higher-resolution-associated uncertainties. The outputs of GCMs are evaluated in this study to certify the performance of GCMs in simulating temperatures. For Tm, Tmax, Tmin, DTR, FD, and HD, biases between simulations and observation are significantly different across different regions in China. This qualitative

estimation is also addressed in [Bannister et al. \(2017\)](#) who diagnosed that the majority of GCMs could not capture the basic spatial pattern in the metrics of mean value and probability distribution functions of temperatures, and particular detections revealed that temperatures were underestimated even in a cold bias exceeding -3°C across some regions. Similar previous works ([Chen and Frauenfeld 2014](#); [Guo and Wang 2016](#)) were also conducted across China, in relation to temperature simulations from several GCMs. Several studies ([Chen and Frauenfeld 2014](#); [Guo and Wang 2016](#))

FIG. 17. As in Fig. 14, but for SU.

agree that considerable discrepancies are produced between coarse-resolution GCMs and observations when simulating the climate across China, such as significantly lower temperature estimation over the TP. Regarding individual GCMs, MIROC5 does not perform well and misses many details at the regional scale, such as a large bias across the TP (Yu et al. 2015). These results from multimodels and individual models agree with the evaluations presented in this work, which shows obvious underestimation across TP in particular. To our knowledge, these discrepancies can be attributed to local features such as unconsolidated topography, vegetation, etc. in the GCM simulations (Deb et al. 2018).

The reference for cold bias across the Tibetan Plateau is widely inherited from insufficient high-quality observations, which result in errors since they are subjectively extrapolated from measurements at a very limited number of stations (Liang et al. 2019). Simulations at the season level also assist to evaluate the performance of GCMs and PRECIS. It is consistent that better simulations are considered in summer and autumn from most GCMs and PRECIS. Similar seasonal simulated performance is also diagnosed from GCMs and RCMs in other studies (Bucchignani et al. 2014; Chen and Fraunfeld 2014; Yu et al. 2015; Liang et al. 2019).

TABLE 5. The top five ranked models for simulating the variables. The numeric symbols 1–21 are as in Table 2.

Region	Tm	Tmax	Tmin	DTR	FD	HD
NE	12, 20, 2, 4, 13	12, 20, 13, 18, 9	21, 12, 20, 13, 2	5, 11, 21, 12, 13	12, 18, 21, 9, 1	18, 12, 13, 20, 10
NC	12, 20, 17, 5, 8	12, 7, 1, 17, 8	12, 13, 6, 5, 20	21, 14, 6, 12, 18	21, 17, 14, 18, 10	13, 12, 18, 20, 17
IM	12, 20, 6, 9, 3	9, 12, 21, 20, 6	12, 20, 21, 6, 8	18, 13, 20, 12, 7	21, 8, 12, 4, 9	12, 13, 20, 7, 17
HH	2, 12, 13, 11, 5	4, 13, 12, 20, 18	9, 10, 11, 12, 6	15, 13, 21, 12, 14	21, 6, 18, 14, 13	17, 13, 18, 21, 4
XJ	4, 20, 17, 13, 6	4, 17, 13, 6, 19	4, 20, 17, 5, 13	6, 18, 17, 13, 19	17, 9, 8, 21, 7	13, 14, 20, 17, 5
WE	17, 3, 14, 20, 12	9, 17, 21, 12, 6	17, 20, 13, 2, 6	2, 6, 8, 17, 9	17, 1, 18, 6, 14	20, 6, 18, 13, 12
JH	12, 2, 13, 9, 11	13, 12, 7, 12, 21	12, 18, 2, 21, 9	10, 7, 21, 6, 12	21, 6, 20, 1, 13	19, 13, 21, 11, 2
Jha	13, 2, 4, 18, 17	13, 8, 18, 4, 12	13, 4, 2, 21, 6	13, 21, 14, 7, 6	21, 12, 14, 11, 6	17, 18, 7, 12, 21
SY	12, 21, 1, 11, 2	12, 11, 1, 13, 6	18, 6, 12, 21, 11	6, 14, 12, 13, 21	21, 14, 13, 18, 20	12, 19, 21, 17, 20
SC	13, 12, 5, 3, 6	20, 13, 4, 12, 19	12, 18, 8, 14, 6	6, 4, 10, 21, 2	14, 21, 7, 8, 9	10, 13, 19, 20, 12
SW	17, 20, 6, 1, 10	12, 11, 7, 20, 6	20, 5, 6, 12, 17	20, 21, 14, 6, 12	18, 13, 17, 20, 6	14, 12, 11, 17, 13
TP	20, 21, 6, 9, 21	21, 20, 2, 18, 12	20, 14, 13, 12, 9	20, 19, 15, 13, 16	18, 21, 8, 1, 9	13, 7, 4, 10, 11

To assess the impacts of climate change, it is necessary to use bias-corrected fine-resolution projections at the local scale using the two ways, namely, statistical downscaling and dynamic downscaling, as a replacement of raw outputs. Here, BCSD is implemented for its advantage of capturing biased statistical characteristics of daily outputs from GCMs (Ning et al. 2015; Werner and Cannon 2016). Previous comparisons of represented climatic extremes from five downscaling methods have indicated that BCSD performs better on temperature-related indices (Bürger et al. 2012). The techniques referring to bias correction are generally summarized in two ways (Durman et al. 2001; Cannon 2018; Sangelanton et al. 2019), that is, mean correction method (correction factors based on the difference or ratio between simulation and observation) and quantile mapping method (transfer functions based on cumulative distribution functions of simulation and observation). The quantile mapping method, widely adopted by previous researchers (Teutschbein and Seibert 2012; Chen et al. 2013; Crimp et al. 2019) to produce climate change scenarios for impact studies, has been proven to be better than the mean correction method. Therefore, the correction process is consistent not only in the process of BCSD for GCMs but also in bias correction for PRECIS outputs. The evaluation of the corrected outputs shows significant improvement for temperature simulations compared to raw GCMs and PRECIS simulations. A new-generation statistically downscaled dataset, NEX-GDDP, in high spatial (25 km) and temporal (daily) resolutions, has recently been publicized for climate change research, including Tmax and Tmin during the present and future periods under RCP4.5 and RCP8.5. Since it is free access, the dataset has been used to assess climate change over Southeast Asia and China (Bao and Wen 2017; Raghavan et al. 2018). To quantify the improvement of statistical downscaling for GCMs, comparative results (Table S1 in the online supplemental material) from the

metrics for Tmax and Tmin indicate that larger SCC and smaller RMSE and STD come from SD-GCMs relative to NEX-GDDP. Because of same statistical downscaling method in them, the differential performance is highly attributed to the quality of corresponding observation datasets that establish math-statistics mapping between simulation and observation. For NEX-GDDP, Global Meteorological Forcing Dataset (GMFD) is employed as the observational data, which are biased for some reasons referring to generating GMFD (Raghavan et al. 2018). However, for the historical simulations, persistent bias is still inevitable between SD-GCMs/corrected PRECIS and observation, which is in agreement with previous findings (Watanabe et al. 2012; Chen et al. 2018) about negligible difference in bias correction for model outputs. Of course, it will remain an ongoing concern and be possible to optimize the results with the adoption of a high-precision observational dataset.

When comparisons are characterized between raw PRECIS and GCMs outputs, PRECIS generally yields better simulations in temperatures. This means that RCM can reproduce more reliable characteristics in higher resolution, with more realistic topographic forcing and comprehensive physical processes. As stated in previous comparisons obtained from GCM and RCM simulations (Guo and Wang 2016), RCM is a method offering better improvement to simulations, such as a significant decrease in bias and better reproduction of regional details. However, relative to SD-GCMs, there is no significant superiority in the outputs from corrected PRECIS. Moreover, the MR index highlights different models that better reproduce temperatures across each region. These findings suggest that it would be effective to use comprehensive application of SD-GCMs and corrected PRECIS outputs. Thus, it is inevitable to export potential misleading conclusions, if the individual model with overestimation or underestimation is used for future projections, which has been proved by previous

researchers (Wu et al. 2015; Wu and Huang 2016) who highlighted that the individual model might not generate satisfying projections, especially for extreme events. Therefore, it is generally accepted that the performance of multimodel ensemble outperform than individual model (Miao et al. 2014; Akinsanola et al. 2018). Besides, ensemble results can improve spatial simulation to some extent, superior to most models (Sun et al. 2015). An optional multimodel ensemble of better regional models in high resolution is an important way for capturing and reducing uncertainties in climate simulations and projections, which generate more reliable outputs and will be implemented later.

We point out that some uncertainties are unavoidable in this work. GCM structure itself, the bias correction and spatial disaggregation during statistical downscaling for GCMs, and the bias correction for PRECIS are the main causes of differences in the process of climate change assessments and projections (Chen et al. 2011a,b; Liu et al. 2013; Aryal et al. 2019). For instance, different patterns of underestimation and overestimation are still extracted from the downscaled GCMs outputs or corrected PRECIS, albeit with small absolute bias value. This can be explained by uncertainties in models that are constructed based on different parameterizations, physics formulations, and numeric implementations (Zhai et al. 2019). Moreover, the observation dataset is affected by a number of potential uncertainties, such as interpolated method and spatial density of observational stations, since it is obtained through an interpolation of values at stations.

5. Conclusions

This study evaluates the performance of simulated temperatures (Tm, Tmax, Tmin, DTR, FD, and HD) for raw GCMs/PRECIS and SD-GCMs/corrected PRECIS, in the spatial-temporal dimension. It is demonstrated that raw GCMs and PRECIS outputs can reproduce Tm, Tmax, Tmin, and DTR with a bias within $\pm 1^\circ\text{C}$ and FD and HD are reproduced with a bias within ± 5 days only in few regions, which implies the necessity of bias correction for raw outputs. Relative to GCMs, better outputs for temperatures except HD can be gotten from PRECIS, owing to its finer resolution. Regional evaluations of Tm, Tmax, Tmin, DTR, and FD from raw GCMs/PRECIS perform better in HH, JH, JHa, SY, and SC but not as well in XJ, WE, and TP. For HD, better simulation lies in most regions excluding central parts. These comparatively distinct raw simulations agree with most previous evaluations produced from coarse-resolution GCMs and can further provide optional usage of models for targeted assessments. Comparatively, reproduced temperatures are significantly improved in

SD-GCMs/corrected PRECIS outputs, characterized by a bias within $\pm 0.2^\circ\text{C}$ (for Tm, Tmax, Tmin, DTR) and ± 2 days (for FD and HD) in majority regions. Besides, SCC, RMSE, and STD for raw outputs and corrected outputs are additionally conducted to give evidence of the importance of using SD-GCMs/corrected PRECIS in the measure of yearly and seasonal values. When we comprehensively evaluate the performance, it is obvious that SD-GCMs/corrected PRECIS outputs have different capabilities of reproducing temperatures across different climatic regions. In general, IPSL-CM5A-LR, CSIRO Mk3.6.0, PRECIS, NorESM1-M, and IPSL-CM5A-MR are relatively better. Systematic comparisons in this study supply a better knowledge of statistical and dynamic downscaling techniques. These noticeable results with respect to models' performance in regions are a prerequisite or basis for the selection of models when we further project climate change and related impacts at the national or regional scales.

Acknowledgments. This work is financially supported from the Key Projects of China's national twelfth 5-year Science and Technology Pillar Program (2013BAC09B04) and the Key 948 Project (2011-G9). It is an honor to get help from group members of regional climate modeling at the Hadley Centre.

REFERENCES

- Ahmed, K. F., G. L. Wang, J. Silander, A. M. Wilson, J. M. Allen, R. Horton, and R. Anyah, 2013: Statistical downscaling and bias correction of climate model outputs for climate change impact assessment in the U.S. northeast. *Global Planet. Change*, **100**, 320–332, <https://doi.org/10.1016/j.gloplacha.2012.11.003>.
- Ahmed, K., S. Shahida, D. A. Sachindrac, N. Nawazb, and E. S. Chung, 2019: Fidelity assessment of general circulation model simulated precipitation and temperature over Pakistan using a feature selection method. *J. Hydrol.*, **573**, 281–298, <https://doi.org/10.1016/j.jhydrol.2019.03.092>.
- Akinsanola, A. A., V. O. Ajayi, A. T. Adejare, O. E. Adeyeri, I. E. Gbode, K. O. Ogunjobi, G. Nikulin, and A. T. Abolude, 2018: Evaluation of rainfall simulations over West Africa in dynamically downscaled CMIP5 global circulation models. *Theor. Appl. Climatol.*, **132**, 437–450, <https://doi.org/10.1007/s00704-017-2087-8>.
- Aryal, A., S. Shrestha, and M. S. Babel, 2019: Quantifying the sources of uncertainty in an ensemble of hydrological climate-impact projections. *Theor. Appl. Climatol.*, **135**, 193–209, <https://doi.org/10.1007/s00704-017-2359-3>.
- Bannister, D., M. Herzog, H. F. Graf, J. S. Hosking, and C. A. Short, 2017: An assessment of recent and future temperature change over the Sichuan Basin, China, using CMIP5 climate models. *J. Climate*, **30**, 6701–6722, <https://doi.org/10.1175/JCLI-D-16-0536.1>.
- Bao, Y., and X. Y. Wen, 2017: Projection of China's near- and long-term climate in a new high-resolution daily downscaled dataset NEX-GDDP. *J. Meteor. Res.*, **31**, 236–249, <https://doi.org/10.1007/s13351-017-6106-6>.

- Bucchignani, E., M. Montesarchio, L. Cattaneo, M. P. Manzi, and P. Mercogliano, 2014: Regional climate modeling over China with COSMO-CLM: Performance assessment and climate projections. *J. Geophys. Res. Atmos.*, **119**, 12 151–12 170. <https://doi.org/10.1002/2014JD022219>.
- Bürger, G., T. Q. Murdock, A. T. Werner, S. R. Sobie, and A. J. Cannon, 2012: Downscaling extremes—An intercomparison of multiple statistical methods for present climate. *J. Climate*, **25**, 4366–4388. <https://doi.org/10.1175/JCLI-D-11-00408.1>.
- Cannon, A. J., 2018: Multivariate quantile mapping bias correction: An N-dimensional probability density function transform for climate model simulations of multiple variables. *Climate Dyn.*, **50**, 31–49. <https://doi.org/10.1007/s00382-017-3580-6>.
- Chadwick, R., E. Coppola, and F. Giorgi, 2011: An artificial neural network technique for downscaling GCM outputs to RCM spatial scale. *Nonlinear Processes Geophys.*, **18**, 1013–1028. <https://doi.org/10.5194/npg-18-1013-2011>.
- Chen, J., F. P. Brissette, and R. Leconte, 2011a: Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *J. Hydrol.*, **401**, 190–202. <https://doi.org/10.1016/j.jhydrol.2011.02.020>.
- , —, A. Poulin, and R. Leconte, 2011b: Overall uncertainty study of the hydrological impacts of climate change for a Canadian watershed. *Water Resour. Res.*, **47**, W12509. <https://doi.org/10.1029/2011WR010602>.
- , —, D. Chaumont, and M. Braun, 2013: Finding appropriate bias correction methods in downscaling precipitation for hydrologic impact studies over North America. *Water Resour. Res.*, **49**, 4187–4205. <https://doi.org/10.1002/wrcr.20331>.
- , H. Chen, and S. L. Guo, 2018: Multi-site precipitation downscaling using a stochastic weather generator. *Climate Dyn.*, **50**, 1975–1992. <https://doi.org/10.1007/s00382-017-3731-9>.
- Chen, L., and O. W. Frauenfeld, 2014: Surface air temperature changes over the twentieth and twenty-first centuries in China simulated by 20 CMIP5 models. *J. Climate*, **27**, 3920–3937. <https://doi.org/10.1175/JCLI-D-13-00465.1>.
- China Ministry of Science and Technology, 2015: *China's Third National Assessment Report on Climate Change*. Science Press, 900 pp., <http://book.sciencereading.cn/shop/book/Booksimple/onlineRead.do?id=B84F220E8E0184E0883F02750A7CB6C3F000&bookPageNum=1>.
- Crimp, S., H. D. Jin, P. Kokic, S. Bakar, and N. Nicholls, 2019: Possible future changes in South East Australian frost frequency: An inter-comparison of statistical downscaling approaches. *Climate Dyn.*, **52**, 1247–126. <https://doi.org/10.1007/s00382-018-4188-1>.
- Deb, P., M. S. Babel, and A. F. Denis, 2018: Multi-GCMs approach for assessing climate change impact on water resources in Thailand. *Model. Earth Syst. Environ.*, **4**, 825–839. <https://doi.org/10.1007/s40808-018-0428-y>.
- Durman, C. F., J. M. Gregory, D. C. Hassel, R. G. Jones, and J. M. Murphy, 2001: A comparison of extreme European daily precipitation simulated by a global and a regional climate model for present and future climates. *Quart. J. Roy. Meteor. Soc.*, **127**, 1005–1015. <https://doi.org/10.1002/qj.49712757316>.
- Fan, F. X., R. S. Bradley, and M. A. Rawlins, 2014: Climate change in the northeastern US: Regional climate model validation and climate change projections. *Climate Dyn.*, **43**, 145–161. <https://doi.org/10.1007/s00382-014-2198-1>.
- Fan, L. J., C. B. Fu, and D. L. Chen, 2011: Long-term trend of temperature derived by statistical downscaling based on EOF analysis. *J. Meteor. Res.*, **25**, 327–339. <https://doi.org/10.1007/s13351-011-0308-0>.
- , D. L. Chen, C. B. Fu, and Z. W. Yan, 2013: Statistical downscaling of summer temperature extremes in northern China. *Adv. Atmos. Sci.*, **30**, 1085–1095. <https://doi.org/10.1007/s00376-012-2057-0>.
- , Z. W. Yan, D. L. Chen, and C. B. Fu, 2015: Comparison between two statistical downscaling methods for summer daily rainfall in Chongqing, China. *Int. J. Climatol.*, **35**, 3781–3797. <https://doi.org/10.1002/joc.4246>.
- Gao, X. J., Y. Shi, D. F. Zhang, J. Wu, F. Giorgi, and Z. M. Ji, 2012: Uncertainties of monsoon precipitation projections over China: Results from two high-resolution RCM simulations. *Climate Res.*, **52**, 213–226. <https://doi.org/10.3354/cr01084>.
- , M. L. Wang, and F. Giorgi, 2013: Climate change over China in the 21st century as simulated by BCCCSM1.1-RegCM4.0. *Atmos. Ocean. Sci. Lett.*, **6**, 381–386. <https://doi.org/10.1080/16742834.2013.11447112>.
- , Y. Shi, Z. Y. Han, M. L. Wang, J. Wu, and D. F. Zhang, 2017: Performance of RegCM4 over major river basins in China. *Adv. Atmos. Sci.*, **34**, 441–455. <https://doi.org/10.1007/s00376-016-6179-7>.
- Gaur, A., and S. P. Simonovic, 2018: Application of physical scaling towards downscaling climate model precipitation data. *Theor. Appl. Climatol.*, **132**, 287–300. <https://doi.org/10.1007/s00704-017-2088-7>.
- Guo, D. L., and H. J. Wang, 2016: Comparison of a very-fine-resolution GCM with RCM dynamical downscaling in simulating climate in China. *Adv. Atmos. Sci.*, **33**, 559–570. <https://doi.org/10.1007/s00376-015-5147-y>.
- Guo, J. H., G. H. Huang, and X. Q. Wang, 2018: Dynamically-downscaled projections of changes in temperature extremes over China. *Climate Dyn.*, **50**, 1045–1066. <https://doi.org/10.1007/s00382-017-3660-7>.
- Hewitt, C., S. Mason, and D. Walland, 2012: The Global Framework for Climate Services. *Nat. Climate Change*, **2**, 831–832. <https://doi.org/10.1038/nclimate1745>.
- Huang, D. P., L. Zhang, G. Gao, and S. Sun, 2018: Projected changes in population exposure to extreme heat in China under a RCP8.5 scenario. *J. Geogr. Sci.*, **28**, 1371–1384. <https://doi.org/10.1007/s11442-018-1550-5>.
- IPCC, 2013: *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, 1535 pp., <https://doi.org/10.1017/CBO9781107415324>.
- Jiang, Z., W. Li, J. Xu, and L. Li, 2015: Extreme precipitation indices over China in CMIP5 models. Part I: Model evaluation. *J. Climate*, **28**, 8603–8619. <https://doi.org/10.1175/JCLI-D-15-0099.1>.
- Jones, B., B. C. O'Neill, L. McDaniel, S. Mcginnis, L. O. Mearns, and C. Tebaldi, 2015: Future population exposure to US heat extremes. *Nat. Climate Change*, **5**, 652–655. <https://doi.org/10.1038/nclimate2631>.
- Kotlarski, S., and Coauthors, 2014: Regional climate modeling on European scales: A joint standard evaluation of the EURO-CORDEX RCM ensemble. *Geosci. Model Dev.*, **7**, 1297–1333. <https://doi.org/10.5194/gmd-7-1297-2014>.
- Kreienkamp, F., A. Paxian, B. Früh, P. Lorenz, and C. Matulla, 2019: Evaluation of the empirical-statistical downscaling method EPISODES. *Climate Dyn.*, **52**, 991–1026. <https://doi.org/10.1007/s00382-018-4276-2>.
- Kukal, M. S., and S. Irmak, 2018: U.S. agro-climate in 20th century: Growing degree days, first and last frost, growing season length, and impacts on crop yields. *Sci. Rep.*, **8**, 6977. <https://doi.org/10.1038/s41598-018-25212-2>.
- Li, X. H., Y. L. Xu, C. C. Meng, L. Zhang, and C. G. Wang, 2018: Analysis on the changes of agro-meteorological thermal indices in Northeast China under RCP4.5 scenario using the

- PRECIS2.1. *Atmosphere*, **9**, 323, <https://doi.org/10.3390/atmos9080323>.
- Liang, X. Z., and Coauthors, 2019: CWRf performance at downscaling China climate characteristics. *Climate Dyn.*, **52**, 2159–2184, <https://doi.org/10.1007/s00382-018-4257-5>.
- Liu, L. L., T. Fischer, T. Jiang, and Y. Luo, 2013: Comparison of uncertainties in projected flood frequency of the Zhujiang River, South China. *Quat. Int.*, **304**, 51–61, <https://doi.org/10.1016/j.quaint.2013.02.039>.
- Liu, Q., and Coauthors, 2018: Extension of the growing season increases vegetation exposure to frost. *Nat. Commun.*, **9**, 426, <https://doi.org/10.1038/s41467-017-02690-y>.
- Marotzke, J., and Coauthors, 2017: Climate research must sharpen its view. *Nat. Climate Change*, **7**, 89–91, <https://doi.org/10.1038/nclimate3206>.
- Massey, N., and Coauthors, 2015: Weather@home—Development and validation of a very large ensemble modelling system for probabilistic event attribution. *Quart. J. Roy. Meteor. Soc.*, **141**, 1528–1545, <https://doi.org/10.1002/qj.2455>.
- McSweeney, C. F., R. G. Jones, R. W. Lee, and D. P. Rowell, 2015: Selecting CMIP5 GCMs for downscaling over multiple regions. *Climate Dyn.*, **44**, 3237–3260, <https://doi.org/10.1007/s00382-014-2418-8>.
- Miao, C., and Coauthors, 2014: Assessment of CMIP5 climate models and its multi-model ensemble prediction in temperature changes over northern Eurasia. *Environ. Res. Lett.*, **9**, 055007, <https://doi.org/10.1088/1748-9326/9/5/055007>.
- Moise, A., L. Wilson, M. Grose, P. Whetton, I. Watterson, J. Bhend, and J. Bathols, 2015: Evaluation of CMIP3 and CMIP5 models over the Australian region to inform confidence in projections. *Aust. Meteor. Oceanogr. J.*, **65** (1), 19–53, <https://doi.org/10.22499/2.6501.004>.
- Ning, L., and R. S. Bradley, 2015: NAO and PNA influences on winter temperature and precipitation over the eastern United States in CMIP5 GCMs. *Climate Dyn.*, **46**, 1257–1276, <https://doi.org/10.1007/s00382-015-2643-9>.
- , E. E. Riddle, and R. S. Bradley, 2015: Projected changes in climate extremes over the northeastern United States. *J. Climate*, **28**, 3289–3310, <https://doi.org/10.1175/JCLI-D-14-00150.1>.
- Raghavan, S. V., J. Hur, and S. Y. Liong, 2018: Evaluations of NASA NEX-GDDP data over Southeast Asia: Present and future climates. *Climatic Change*, **148**, 503–518, <https://doi.org/10.1007/s10584-018-2213-3>.
- Rahman, M. A., Y. S. Lou, S. Nahid, and O. Victor, 2019: Analysis of reference evapotranspiration (ET₀) trends under climate change in Bangladesh using observed and CMIP5 data sets. *Meteor. Atmos. Phys.*, **131**, 639–655, <https://doi.org/10.1007/s00703-018-0596-3>.
- Rao, K. K., S. K. Patwardhan, A. Kulkarni, K. Kamala, S. S. Sabade, and K. K. Kumar, 2014: Projected changes in mean and extreme precipitation indices over India using PRECIS. *Global Planet. Change*, **113**, 77–90, <https://doi.org/10.1016/j.gloplacha.2013.12.006>.
- Rummukainen, M., 2010: State-of-the-art with regional climate models. *Wiley Interdiscip. Rev.: Climate Change*, **1**, 82–96, <https://doi.org/10.1002/wcc.8>.
- Sangelanton, L., A. Russo, and F. Gennaretti, 2019: Impact of bias correction and downscaling through quantile mapping on simulated climate change signal: A case study over central Italy. *Theor. Appl. Climatol.*, **135**, 725–740, <https://doi.org/10.1007/s00704-018-2406-8>.
- Schoof, J. T., 2013: Statistical downscaling in climatology. *Geogr. Compass*, **7**, 249–265, <https://doi.org/10.1111/gec3.12036>.
- Sun, Q., C. Miao, and Q. Duan, 2015: Projected changes in temperature and precipitation in ten river basins over China in 21st century. *Int. J. Climatol.*, **35**, 1125–1141, <https://doi.org/10.1002/joc.4043>.
- Shi, Y., G. Wang, and X. J. Gao, 2018: Role of resolution in regional climate change projections over China. *Climate Dyn.*, **51**, 2375–2396, <https://doi.org/10.1007/s00382-017-4018-x>.
- Teutschbein, C., and J. Seibert, 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, <https://doi.org/10.1016/j.jhydrol.2012.05.052>.
- Thrasher, B., E. P. Maurer, C. McKellar, and P. B. Duffy, 2012: Technical Note: Bias correcting climate model simulated daily temperature extremes with quantile mapping. *Hydrol. Earth Syst. Sci.*, **16**, 3309–3314, <https://doi.org/10.5194/hess-16-3309-2012>.
- Trzaska, S., and E. Schnarr, 2014: A review of downscaling methods for climate change projections. CIESIN USAID Rep., 45 pp., http://www.ciesin.org/documents/Downscaling_CLEARED_000.pdf.
- Van Khien, M. V., G. Redmond, C. Mcsweeney, and T. Thuc, 2014: Evaluation of dynamically downscaled ensemble climate simulations for Vietnam. *Int. J. Climatol.*, **34**, 2450–2463, <https://doi.org/10.1002/joc.3851>.
- Wang, C. G., J. Liang, and K. I. Hodges, 2017: Projections of tropical cyclones affecting Vietnam under climate change: Downscaled HadGEM2-ES using PRECIS 2.1. *Quart. J. Roy. Meteor. Soc.*, **143**, 1844–1859, <https://doi.org/10.1002/qj.3046>.
- Watanabe, S., S. Kanae, S. Seto, P. J. F. Yeh, Y. Hirabayashi, and T. Oki, 2012: Intercomparison of bias-correction methods for monthly temperature and precipitation simulated by multiple climate models. *J. Geophys. Res.*, **117**, D23114, <https://doi.org/10.1029/2012JD018192>.
- Werner, A. T., and A. J. Cannon, 2016: Hydrologic extremes—An intercomparison of multiple gridded statistical downscaling methods. *Hydrol. Earth Syst. Sci.*, **20**, 1483–1508, <https://doi.org/10.5194/hess-20-1483-2016>.
- Wu, C. H., and G. R. Huang, 2016: Projection of climate extremes in the Zhujiang River basin using a regional climate model. *Int. J. Climatol.*, **36**, 1184–1196, <https://doi.org/10.1002/joc.4412>.
- , —, and H. J. Yu, 2015: Prediction of extreme floods based on CMIP5 climate models: A case study in the Beijiang River basin, South China. *Hydrol. Earth Syst. Sci.*, **19**, 1385–1399, <https://doi.org/10.5194/hess-19-1385-2015>.
- Xu, Y., X. J. Gao, Y. Shen, C. H. Xu, Y. Shi, and F. Giorgi, 2009: A daily temperature dataset over China and its application in validating a RCM simulation. *Adv. Atmos. Sci.*, **26**, 763–772, <https://doi.org/10.1007/s00376-009-9029-z>.
- , —, and F. Giorgi, 2010: Upgrades to the reliability ensemble averaging method for producing probabilistic climate-change projections. *Climate Res.*, **41**, 61–81, <https://doi.org/10.3354/cr00835>.
- Yang, H. L., Y. L. Xu, L. Zhang, J. Pan, and X. Li, 2010: Projected change in heat waves over China using the PRECIS climate model. *Climate Res.*, **42**, 79–88, <https://doi.org/10.3354/cr00860>.
- Yu, E. T., J. Q. Sun, H. P. Chen, and W. L. Xiang, 2015: Evaluation of a high-resolution historical simulation over China: Climatology and extremes. *Climate Dyn.*, **45**, 2013–2031, <https://doi.org/10.1007/s00382-014-2452-6>.
- Zhai, Y. Y., G. Huang, X. Q. Wang, X. Zhou, C. Lu, and Z. Li, 2019: Future projections of temperature changes in Ottawa, Canada through stepwise clustered downscaling of multiple

- GCMs under RCPs. *Climate Dyn.*, **52**, 3455–3470, <https://doi.org/10.1007/s00382-018-4340-y>.
- Zhang, Q., J. Li, C. Y. David, and X. Chen, 2011: Observed changes of temperature extremes during 1960–2005 in China: Natural or human-induced variations? *Theor. Appl. Climatol.*, **106**, 417–431, <https://doi.org/10.1007/s00704-011-0447-3>.
- Zhang, Y. J., L. Fu, J. Pan, and Y. L. Xu, 2017: Projected changes in temperature extremes in China using PRECIS. *Atmosphere*, **8**, 15, <https://doi.org/10.3390/atmos8010015>.
- Zhou, X., G. H. Huang, and X. Q. Wang, 2018: Dynamically-downscaled temperature and precipitation changes over Saskatchewan using the PRECIS model. *Climate Dyn.*, **50**, 1321–1334, <https://doi.org/10.1007/s00382-017-3687-9>.
- Zhu, J. X., G. Huang, X. Q. Wang, and G. H. Cheng, 2017: Investigation of changes in extreme temperature and humidity over China through a dynamical downscaling approach. *Earth's Future*, **5**, 1136–1155, <https://doi.org/10.1002/2017EF000678>.