INTERNATIONAL JOURNAL OF CLIMATOLOGY
*Int. J. Climatol.* **37**: 657–671 (2017)
Published online 21 April 2016 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/joc.4731

RMetS
Royal Meteorological Society

# Evaluation of global climate models for downscaling applications centred over the Tibetan Plateau

Jianwei Xu,[a†] Yanhong Gao,[a*] Deliang Chen,[b] Linhong Xiao[a] and Tinghai Ou[b]

[a] *Key Laboratory of Land Surface Process and Climate Change in Cold and Arid Regions, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences, Lanzhou, China*
[b] *Department of Earth Sciences, University of Gothenburg, Sweden*

**ABSTRACT:** Quality of a downscaling depends primarily on the quality of the driving global climate model (GCM). In this study, historical atmospheric conditions simulated by 14 GCMs in CMIP5 are evaluated for downscaling applications centred over the Tibetan Plateau (TP) with ERA-Interim reanalysis as reference. Another reanalysis NCEP-DOE is also used to estimate the uncertainty associated with the reanalyses. Performances of six frequently used GCM variables, involving atmospheric circulation, air temperature and humidity, are evaluated in terms of biases, spatial correlation coefficient, mean absolute error as well as distinct seasonal features. To detect distributional biases, the two-sample Kolmogorov–Smirnov test (KS test) is applied to both the original time series and their anomalies on the monthly scale. A spatial ranking scheme is finally applied to objectively quantify overall relative merits of the GCMs over this region. We found that differences between two reanalysis datasets are negligible over this region. Regarding the GCMs' performances, the biases of the simulated variables show remarkable differences among models. Sea level pressure and 500 hPa geopotential height are well simulated by all the GCMs, whereas specific humidity at 600 hPa has a significant dry bias and temperature at 500 hPa has a sizable cold bias. The spatial pattern of the upper-tropospheric circulation is relatively poorly simulated. The KS test suggests that the climatic mean and higher order moments play about an equal role in causing the errors. According to the ranking scores, CCSM4, CNRM-CM5, MPI-ESM-LR, NorESM1-M, MIROC4h, MPI_ESM_MR and CSIRO-MK are relatively superior to other GCMs for this region.

KEY WORDS global climate models; Tibetan Plateau; CMIP5; downscaling

*Received 17 February 2015; Revised 3 March 2016; Accepted 7 March 2016*

## 1. Introduction

The Tibetan Plateau (TP), termed as the roof of the world, is the highest plateau in the world with an area proximal 2.5 million km$^2$ and an average elevation above 4000 m. As the source of several major rivers in Asia, the TP is also called 'water tower of Asia' (Immerzeel *et al.*, 2010), supporting hundreds of millions of people in the downstream. In recent decades, the TP has been experiencing more pronounced warming than other regions at same latitude (Liu and Chen, 2000; Rangwala *et al.*, 2013), which has had a significant impact on permafrost degradation (Wang *et al.*, 2000), glacial retreat (Yao *et al.*, 2007) and desertification (Xue *et al.*, 2009). How climate over the TP will change in the future is of great significance and receives much attention (Chen *et al.*, 2015; Su *et al.*, 2016).

Global climate models (GCMs) have provided valuable information on climate change and long-term climate projections at global to sub-continental scale (IPCC, 2013). The World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 5 (CMIP5) provides a state-of-the-art multi-model dataset, which was used by the Intergovernmental Panel on Climate Change (IPCC) for its fifth assessment report on climate change. With improved models CMIP5 are expected to perform better than those in the former phase – CMIP3 (Taylor *et al.*, 2012). Indeed, CMIP5 models have been found to have smaller bias than CMIP3 models in reproducing atmospheric downward longwave radiation (Ma *et al.*, 2014) and precipitation over China (Chen and Frauenfeld, 2014). The same holds true for simulation of East Asia monsoon characteristics (Sperber *et al.*, 2013; Wei *et al.*, 2013) and El Niño-Southern Oscillation (ENSO) (Bellenger *et al.*, 2014). Although CMIP5 models represent an improvement compared with those of CMIP3, they still have remarkable biases in depicting regional climate information. Previous studies (Cattiaux *et al.*, 2013; Chen *et al.*, 2012; Su *et al.*, 2013; Chen and Frauenfeld, 2014) have identified large error in the magnitude and trend of precipitation, surface air temperature and 10-m wind speed over the TP and in other regions. Almost all CMIP5 models suffer from a common bias in the thermodynamic structure of boreal

* Correspondence to: Y. Gao, Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences (CAS), 320 Donggang West Rd, 730000, Lanzhou, Gansu, P. R. China. E-mail: gaoyh@lzb.ac.cn

†Present address: Jianwei Xu is now at Hunan University of Arts and Science, China.

summer monsoons which is caused by an overly smoothed representation of topography west of the TP (Boos and Hurley, 2013). Further, CMIP5 models still exhibit some biases in simulating both winter and summer East Asia monsoon characters (Sperber *et al.*, 2013; Wei *et al.*, 2013; Gong *et al.*, 2014).

The horizontal resolution of most present-day GCMs is in the order of a few hundred kilometres (Meehl *et al.*, 2007), which limits their ability to represent complex topography, land surface characteristics and other processes in the climate system. This prevents GCMs from generating realistic and reliable climate change information at fine scales, which is imperative for developing suitable adaptation and mitigation strategies at the regional-to-local scale (Giorgi *et al.*, 2009). Therefore, downscaling GCMs outputs using either a regional climate model (RCM) or statistical downscaling model is necessary and useful because they add value to the driving GCMs (Feser *et al.*, 2011; Gao *et al.*, 2012; Soares *et al.*, 2012; Fan *et al.*, 2013).

Downscaling are conducted based on the assumption that GCMs can realistically reproduce large-scale atmospheric characteristics, such as middle-tropospheric temperature and humidity which are essential for downscaling (Brands *et al.*, 2013). However, all GCMs surfer from systematic biases to some extent, and RCMs or statistical models could inherit these biases from their driving GCMs. By comparing GCM-driven CRCM5 simulation with reanalysis-driven CRCM5 simulation for Africa, Laprise *et al.* (2013) found that biases from GCMs have deleterious consequences on the skill of CRCM5 at reproducing specific regional climate features. Four GCM-driven RCM simulations were conducted for Africa and the results show that the geographical distribution of mean sea level pressure (SLP), surface temperature and seasonal precipitation are strongly affected by the driving GCMs (Dosio *et al.*, 2014). This means that bias in the driving GCM was to a large extent passed over to the RCM output.

As for spectral nudging, large scales in the interior of RCMs domain are nudged towards the coarse resolution driving GCMs, the large-scale part of RCMs simulation is essentially independent from the choice of the lateral boundary treatment, location of the domain, size of the domain or the resolution of the regional model. This simplifies many problematic aspects of the lateral boundary forcing and turns the complex problem of the traditional dynamical downscaling into a much simpler one (Hong and Kanamitsu, 2014). As a result, spectral nudging is increasingly used in dynamical downscaling (e.g. Alexandru *et al.*, 2009; Heikkila *et al.*, 2011; Xu and Yang, 2015), which gives rise to more attention in GCM evaluation in the interior of a RCMs domain. Statistical downscaling can use local (e.g. GCM grid) or field (e.g. whole domain) predictors (Benestad *et al.*, 2008). The former would be appropriate in regions with small biases, whereas the latter may be more appropriate in regions with large bias (Gutiérrez *et al.*, 2013). Given the important effect of GCM bias on downscaling and strong desire of improving quality of downscaling, bias of GCM should be identified and assessed.

To reduce the RCM bias caused by GCM, some studies (Bruyère *et al.*, 2014; Xu and Yang, 2015) have performed bias correction on the GCM boundary conditions. No matter which downscaling scheme is chosen, it is a key step to evaluate GCM and to select GCM with small bias before a downscaling.

Previous studies (Xu and Xu, 2012; Su *et al.*, 2013) have evaluated the ability of GCMs in reproducing near surface variables, such as near surface air temperature and precipitation over the TP. However, from the point view of downscaling, atmospheric variables at different levels, such as SLP, mid-tropospheric humidity, temperature and upper-tropospheric circulation, remain to be evaluated over the TP and a larger domain. Herein, these variables of the historical run in CMIP5 are assessed in comparison to ERA-Interim reanalysis (ERA-Int for short), which is considered as the best among the most popular reanalyses to express the water cycle climatology and climate change in the TP (Gao *et al.*, 2014). As outlined by Brands *et al.* (2012), reanalysis datasets also suffer from biases and the difference between two distinct reanalysis datasets is an effective and useful estimator of observational uncertainty which hinders reliable validation of GCMs. Hence, a comparison between ERA-Int and another reanalysis (NCEP-DOE) is also made to indicate reanalysis uncertainty over the study domain.

This work is arranged as follows: datasets and methodology used are introduced in Section 2; results are presented in Section 3; Section 4 summarizes and discusses the main findings.

## 2. Data and methodology

### 2.1. Datasets and study domain

CMIP5 provides a state-of-the-art multi-model dataset, which was applied in the Intergovernmental Panel on Climate Change (IPCC) Fifth Assessment Report (AR5) to address some of the scientific questions that arose during preparation of AR4 (Taylor *et al.*, 2012). We intend to evaluate a subset (14) of the GCMs in CMIP5 in this study (Table 1). Air temperature, specific humidity, geopotential height and wind are the main variables which are used to constitute driving conditions in dynamical downscaling or used as predictor in statistical downscaling. Six atmospheric variables (Table 2) at different vertical levels are evaluated, including the SLP, (the same as the other variables), the specific humidity at 600 hPa (Q600), the air temperature and the geopotential height at 500 hPa (T500 and Z500), the wind components at 200 hPa (U200 and V200), which are usually evaluated for dynamic downscaling (Brands *et al.*, 2013; Jury *et al.*, 2015). In fact, these variables were also frequently used in statistical downscaling (e.g. Benestad *et al.*, 2008).

SLP is a key variable for low-level wind, and the change of SLP differences between the continent and sea is the most evident indicator of evolution of East Asia winter

 

Table 1. Information of the 14 GCMs used for evaluation.

| Model name | Institute | Nation | Resolution |
|---|---|---|---|
| CanESM2 | CCCma | Canada | 2.8125° × 2.8125° L35 |
| CCSM4 | NCAR | USA | 0.9° × 1.25° L26 |
| CNRM-CM5 | CNRM | France | 1.4° × 1.4° L31 |
| CSIRO-MK | CSIRO | Australia | 1.875° × 1.875° L18 |
| GFDL-CM3 | GFDL | USA | 2° × 2.5° L48 |
| GFDL-ESM2M | GFDL | USA | 2° × 2.5° L24 |
| GISS-E2-H | NASA | USA | 2° × 2.5° L40 |
| IPSL-CM5A-LR | IPSL | France | 1.89° × 3.75° L39 |
| MIROC4h | MIROC | Japan | 0.5625° × 0.5625° L56 |
| MIROC-ESM | MIROC | Japan | 2.8125° × 2.8125° L80 |
| MPI-ESM-LR | MPI | Germany | 1.875° × 1.875° L47 |
| MPI-ESM-MR | MPI | Germany | 1.875° × 1.875° L95 |
| MRI-CGCM3 | MRI | Japan | 1.125° × 1.125° L48 |
| NorESM1-M | NCC | Norway | 1.875° × 2.5° L26 |

Table 2. Variables analyzed in this study.

| Acronyms | Description | Unit |
|---|---|---|
| SLP | Sea level pressure | hPa |
| Q600 | Specific humidity at 600 hPa | g kg$^{-1}$ |
| T500 | Temperature at 500 hPa | K |
| Z500 | Geopotential height at 500 hPa | m$^2$ s$^{-2}$ |
| U200 | U-wind at 200 hPa | m s$^{-1}$ |
| V200 | V-wind at 200 hPa | m s$^{-1}$ |

monsoon (Wei *et al.*, 2013). Precipitation over the TP is strongly associated with vapour transportation which usually appears a maximum at 600 hPa for the TP. Owing to the high altitude of the TP, variables at 500 hPa can strongly affect near-surface variables over the TP. In addition, the Western Pacific subtropical high, which is an important synoptic system for East Asia, is usually described using geopotential height at 500 hPa. Another key synoptic circulation for the TP is the South Asia high which usually hovers over the TP at 200 hPa (Yeh and Gao, 1979). The wind at that level is strongly affected by the subtropical westerly jet stream which plays an essential role in dynamical aspects of East Asia Summer Monsoon (Feng *et al.*, 2014).

The GCMs' outputs of the historical runs are obtained from the Earth System Grid Federation Portal (data are available online at http://pcmdi9.llnl.gov/esgf-web-fe/). GCMs datasets at monthly timescale were chosen in this study, which may lead to some limitations of the results. However, the major conclusion concerning the climatology can be expected to remain valid, because GCMs performance on the monthly timescale is correlated to model performance on shorter temporal scales (Jury *et al.*, 2015).

ERA-Int from the European Centre for Medium Range Weather Forecast (ECMWF) (Dee *et al.*, 2011) is used as a reference for the evaluation. It is proved to be the best among the widely available reanalyses to describe the surface air temperature and water cycle in the TP (Wang and Zeng, 2012; Gao *et al.*, 2014). NCEP-DOE AMIP-II (NCEP-DOE; Kanamitsu *et al.*, 2002) is compared with ERA-Int to estimate the

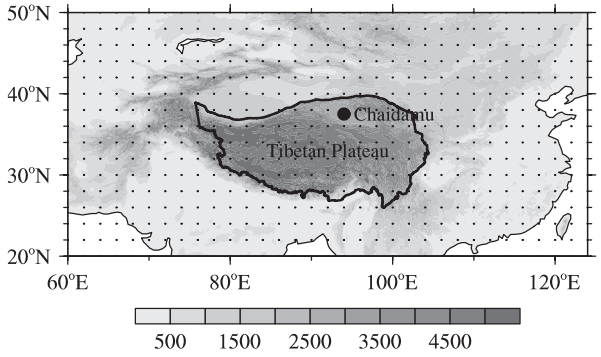


Figure 1. Study domain and topography (unit: m). The Tibetan Plateau (TP) is circled in solid line. Black dots refer to the 2° × 2° grids.

uncertainty associated with the reanalyses. Datasets of the two reanalyses are available online at http://www.ecmwf.int/research/era/do/get/ERA-Interim and http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html, respectively.

The study domain in this paper is limited to 20–50°N and 60–124°E, mainly covering the East Asia (Figure 1). This area is chosen to be much larger than the area occupied by the TP (the boundary is indicated by the bold line in Figure 1), which makes the study valuable for a variety downscaling applications including those focused on the TP (e.g. Xue *et al.*, 2014). Given the distinct seasonal climatology in the TP and different large-scale circulation control over the south and north of the TP (Yao *et al.*, 2013), seasonal assessments for three areas [whole TP (TP), to the north of 35°N (North) and to the south of the 35°N (South)] in the study domain are considered in the evaluation.

### 2.2. Methodology

As horizontal resolutions of the GCMs and reanalysis products vary widely, ranging from 0.5625° to 3.75°, all datasets are regridded to a regular 2° × 2° grid by bilinear interpolation – a commonly used regridding approach used in model intercomparison and evaluation (e.g. Brands *et al.*, 2013; Laprise *et al.*, 2013) to facilitate the evaluation and comparisons. Annual and seasonal climatology are evaluated by two statistics. The spatial correlation coefficient (SCC) is used to quantitatively evaluate model's ability to capture spatial pattern, while mean absolute error (MAE) is employed to indicate mean errors of the simulation which are averaged over the whole domain. These statistics are first calculated at monthly scale during the period from 1979 to 2005, and then multi-yearly averaged. Following the annual and seasonal climatology assessment, the two-sample Kolmogorov–Smirnov test (KS test; Wilks and Haman, 2006) is used to compare distributional features of the two samples under the null hypothesis that the samples are drawn from the same underlying theoretical probability distribution. Herein, this test is conducted between each GCM or NCEP-DOE and ERA-Int monthly time series at each grid. To detect biases in high-order moments, KS test is also applied to the anomalies of the

 

two paired samples. Anomalies are obtained from original series by subtracting monthly climatic mean at each grid. The statistic is defined following Brands *et al.* (2013), denoting that:

$$ks = \max_{i=1}^{2n} |E(z_i) - I(z_i)| \quad (1)$$

where $n$ is 324 in this study, representing the number of months during 1979–2005. The value of $z_i$ denotes the *ith* data value of the sorted joined sample and $E(z_i)$ and $I(z_i)$ are the empirical cumulative frequencies from two samples. Commonly, the probability ($p$ value) of the test statistic according to null hypothesis is used to check the distributional similarity. As usual, the 5% significance level is chosen. Therefore, if $p$ value is larger than 0.05, the null hypothesis is accepted, which means that the distributional difference is not significant at the 5% confidence level. To better demonstrate the result, $p$ value is transformed to 10 based logarithms (Brands *et al.*, 2013). As a result, $p$ value larger than −1.301 indicates that there is a significant distributional similarity.

To quantify overall relative performance and give an objective ranking of the 14 GCMs, a ranking method focusing on spatial aspects are utilized for each variable. The spatial ranking scheme is defined following Gettelman *et al.* (2010) to evaluate spatial pattern, considering SCC, spatial mean error (ME), and standard deviation (SD). This method has been used in earlier studies to quantify relative performance of multi-models (Douglass *et al.*, 1999; Waugh and Eyring, 2008). A score based on monthly mean data is defined as follows;

$$g_m = \max\left(0, 1 - \frac{1}{n}\sum_{i=1}^{n}\frac{|\mu_{i\text{obs}} - \mu_{i\text{mod}}|}{n_g\sigma_{i\text{obs}}}\right) \quad (2)$$

where $\mu$ is the value of monthly mean from either a model or observation, and $n$, representing the number of months, is 324 in the study. $\sigma$ is the SD calculated for each month, $n_g$ is a scaling factor for observational SD, commonly set to 3 (Douglass *et al.*, 1999; Waugh and Eyring, 2008). The value of $g_m$ ranges from zero to unity and if the model difference from observation is greater than 3 times of the SD, $g_m$ is set to zero. When model is definitely consistent with observation, $g_m$ is unity. According to the above description, the spatial ranking method is also applied to another two statistics: $g_c$ for SCC and $g_v$ for SD. The mathematical formulas following Gettelman *et al.* (2010) are given as follows:

$$g_c = \left(\frac{1}{n}\sum_{i=1}^{n} C\left(V_{i\text{obs}}, V_{i\text{mod}}\right) + 1\right)/2 \quad (3)$$

where $C$ means the spatial correlation coefficient between modelled and observed climatologies.

$$g_v = \max\left(0, 1 - \frac{1}{n}\sum_{i=1}^{n}\frac{|\sigma_{i\text{obs}} - \sigma_{i\text{mod}}|}{n_g\sigma_{i\text{obs}}}\right) \quad (4)$$

A composited mean score is then calculated as the linear combination of the three scores: $g_{sum} = \left(g_m + g_c + g_v\right)/3$. The mean score representing uncertainty and inter-model discrepancy (Gettelman *et al.*, 2010) is finally used to rank the models evaluated.

## 3. Results

### 3.1. Annual and seasonal biases

#### *3.1.1. Annual performance*

Figures 2–7, respectively show annual biases in six variables of SLP, Q600, T500, Z500, U200 and V200 of the 14 GCMs compared with ERA-Int. Differences between NCEP-DOE and ERA-Int are also shown in Figures 2(o)–7(o) to indicate the uncertainty of the reanalyses. Apparently, NCEP-DOE is in close agreement with ERA-Int for all the six variables, with a minimum MAE and maximum SCC compared with all the GCMs. For four variables in the mid-upper levels including T500, Z500, U200 and V200, SCC between NCEP-DOE and ERA-Int are higher than 0.99. For most variables, the differences between the two reanalyses within the study domain are generally very small. But for Q600, NCEP-DOE shows larger values over the TP and smaller values at the south edge of the TP than those of ERA-Int. Thus, we conclude that the uncertainty associated with the reanalyses is quite small and practically negligible over the study domain except for the specific humidity over the TP. In fact, Gao *et al.* (2014) established that ERA-Int presents the most reliable information about the moisture over the TP among several other reanalyses evaluated. Hence, ERA-Int is taken as the reference with confidence to evaluate performances of the 6 variables in the 14 GCMs in the following.

Figures 2(a)–(o) show the biases of SLP in the 14 GCMs against ERA-Int (Figure 2(p)). ERA-Int exhibits a pronounced pressure gradient with high SLP in the north and low values in the south of the study domain. Biases of the SLP in the 14 GCMs range from −8.0 to 4.0 hPa. Most GCMs display larger biases in the TP than the surroundings. CNRM-CM5 outperforms other GCMs with the minimal MAE of 2.3 hPa and the highest SCC of 0.89. SLP is underestimated by five GCMs (CanESM2, CSIRO-MK, GFDL-CM3, MPI-ESM-LR and MPI-ESM-MR) and overestimated by six (CCSM4, IPSL-CM5A-LR, MIROC4h, MIROC-ESM, MRI-CGCM3, and NorESM1-M) in the TP. IPSL-CM5A-LR possesses the maximum MAE of 6.3 hPa and minimum SCC of 0.47.

Figure 3(a)–(o) display distribution of Q600 biases of the GCMs and NCEP-DOE compared with ERA-Int as well as Q600 climatology of ERA-Int (Figure 3(p)). A wet centre is located to the southeastern TP in contrast to the arid northwestern TP in ERA-Int, most probably because of the water vapour transport by the South Asian Summer Monsoon and East Asian Summer Monsoon. Most GCMs underestimate Q600, especially to the south of the TP. Half of all the GCMs (CCSM4, CNRM-CM5, GFDL-CM3, MIROC4h, MPI-ESM-LR, MPI-ESM-MR and MRI-CGCM3) follow similar spatial

 

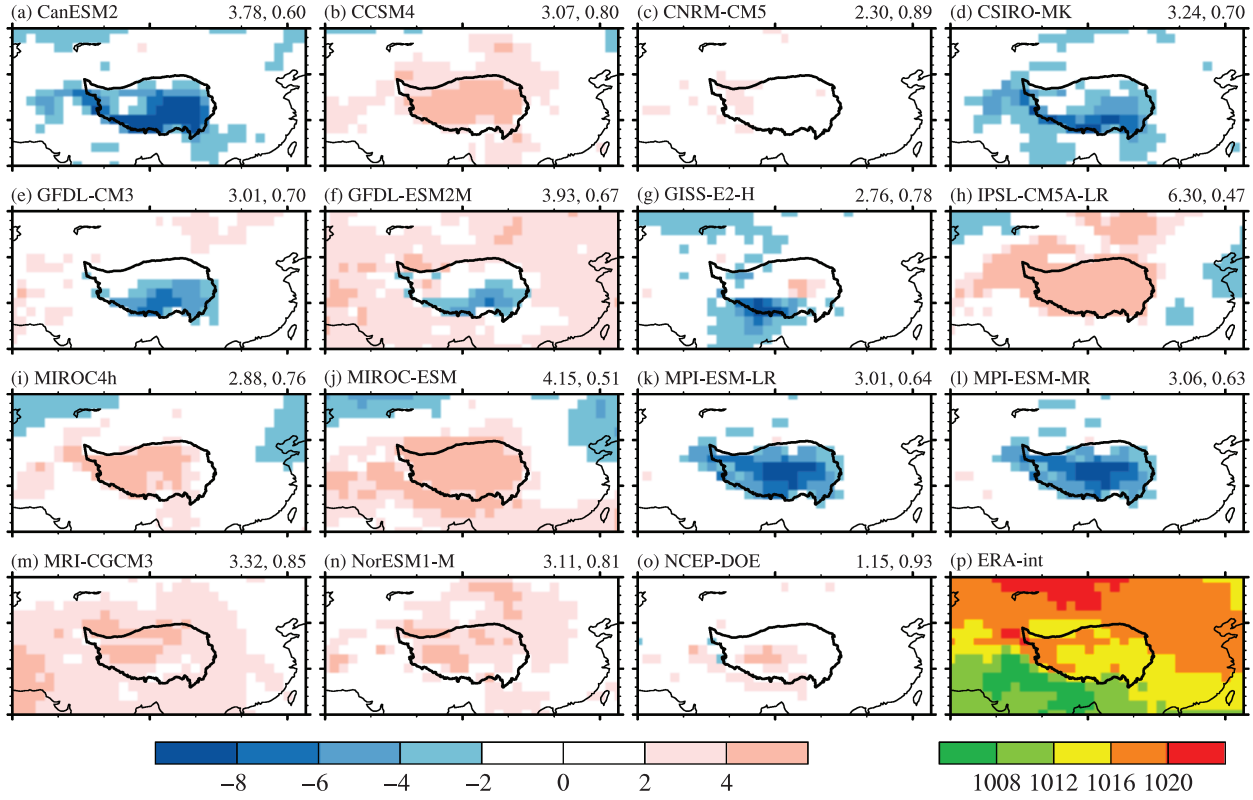


Figure 2. Distribution of biases for surface level pressure SLP (unit: hPa) simulated by the 14 GCMs during 1979–2005 (a–n) and NCEP-DOE (o) compared with ERA-int. The annual mean SLP climatology in ERA-Int is shown in (p). Numbers at top-right of each plot are monthly mean absolute error (MAE) and spatial correlation coefficient (SCC), respectively.

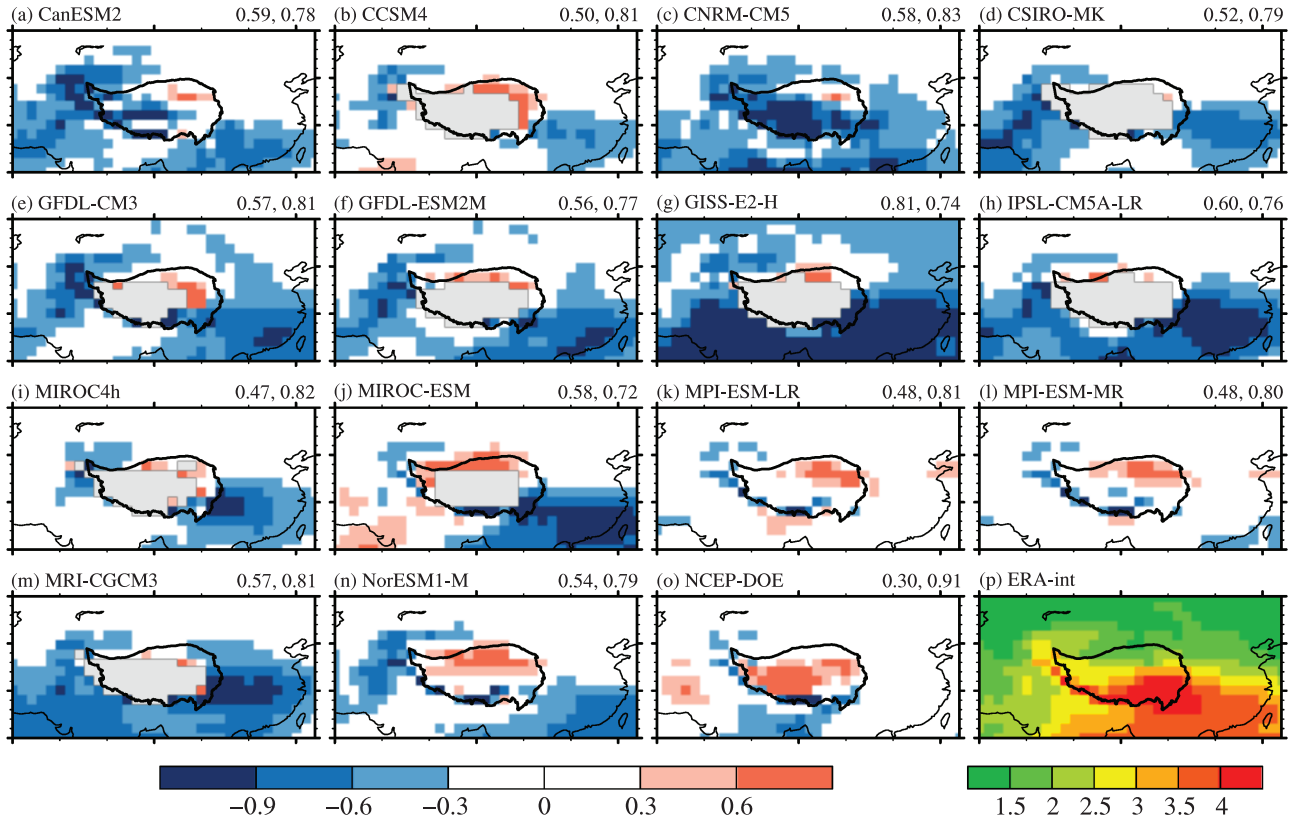


Figure 3. The same as Figure 2, but for specific humidity at the 600 hPa Q600 (unit: g $kg^{-1}$). Grey area represents missing data which is not available when convert from the sigma coordinate to pressure coordinate because the model elevation is above the height of 600 hPa.

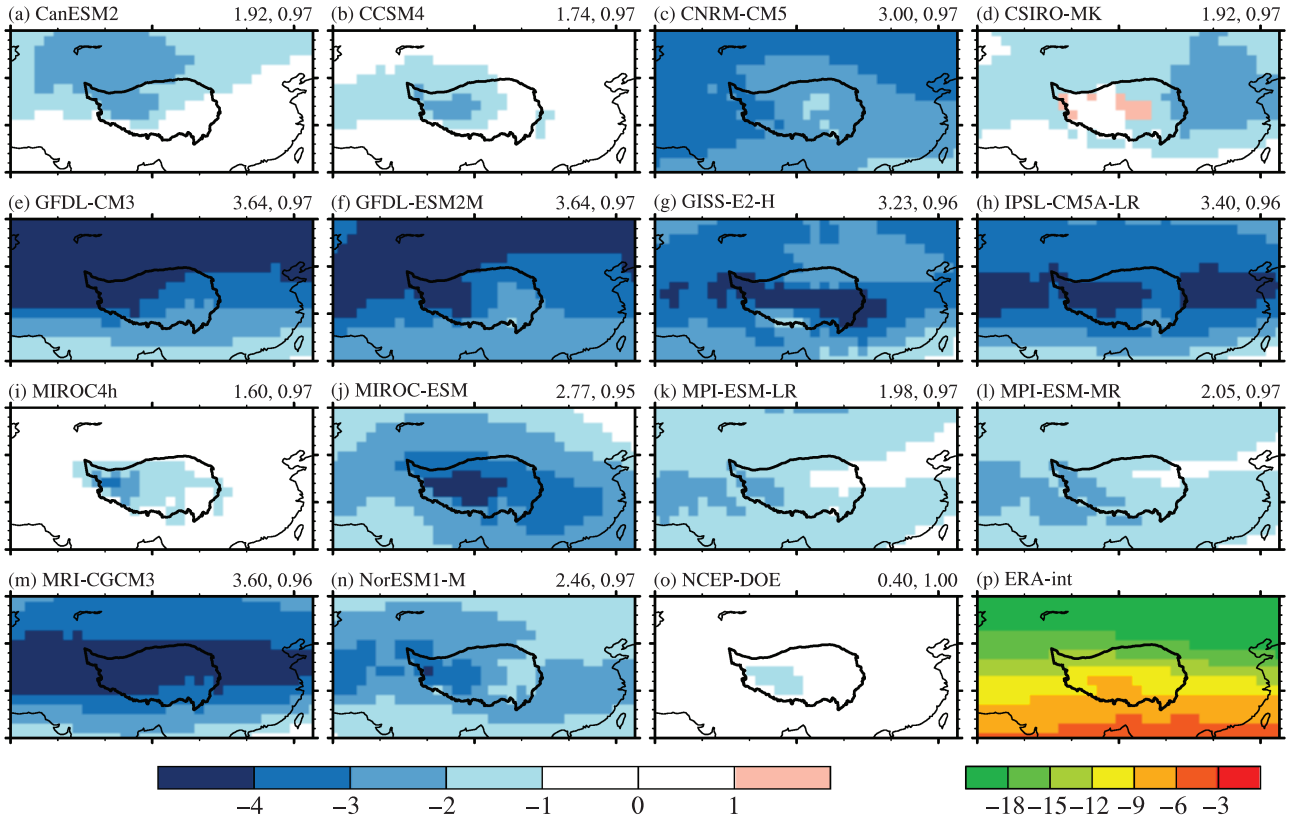


Figure 4. The same as Figure 2, but for air temperature at 500 hPa T500 (unit: °C). Grey area represents missing data.

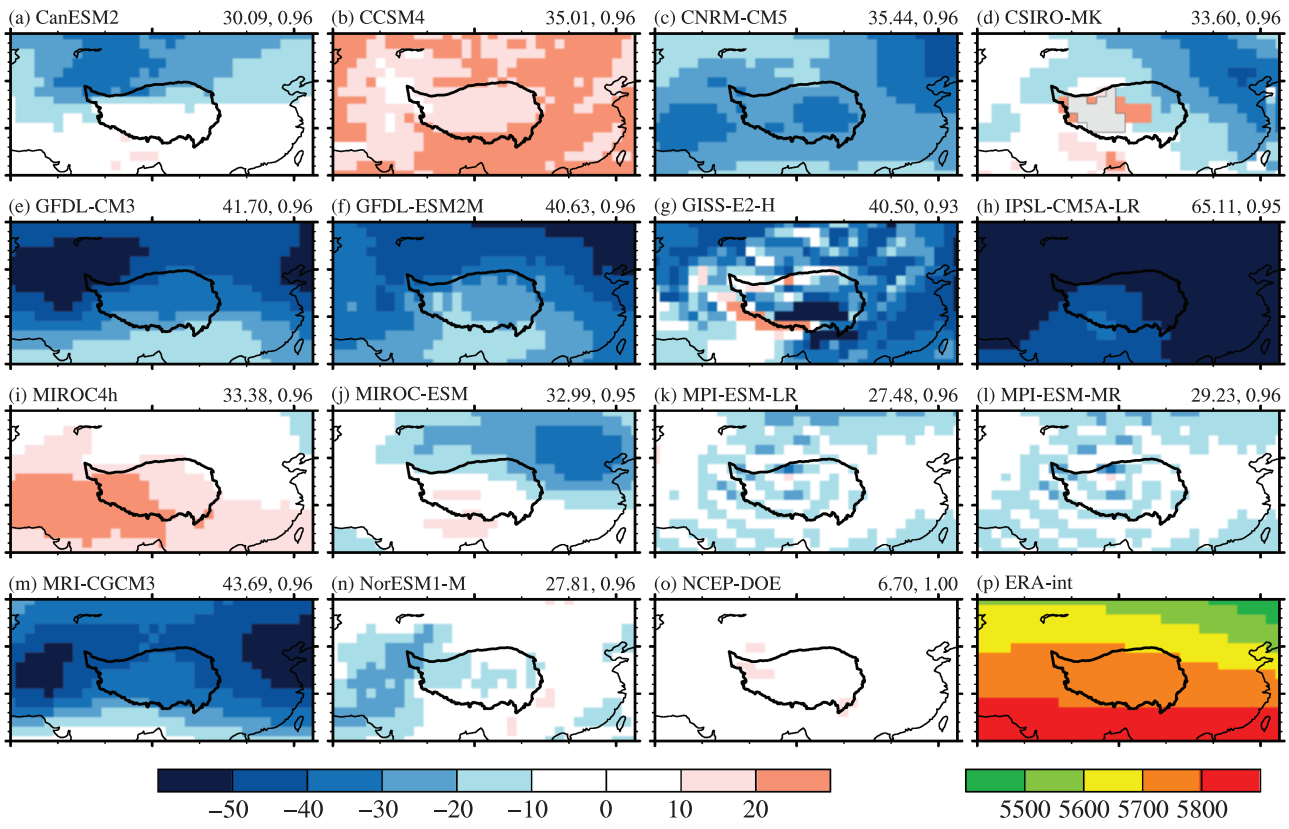


Figure 5. The same as Figure 2, but for geopotential height at 500 hPa Z500 (unit: $m^2\ s^{-2}$). Grey area represents missing data.

 

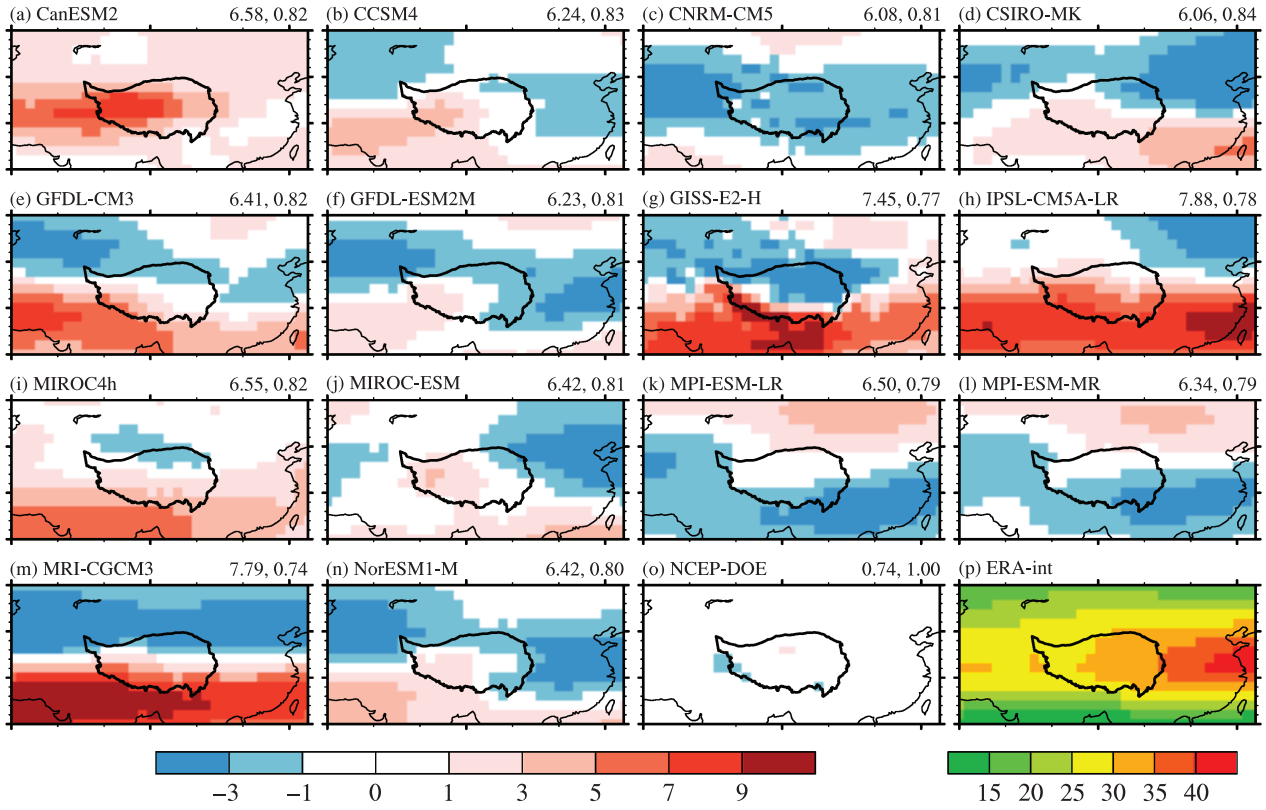


Figure 6. The same as Figure 2, but for zonal wind at 200 hPa U200 (unit: m $s^{-1}$).

pattern as ERA-Int with SCC larger than 0.8. Four of them (MIROC4h, MPI-ESM-LR, MPI-ESM-MR and CCSM4) possess MAE less than 0.5 g $kg^{-1}$. GISS-E2-H exhibits substantial negative biases. The underestimation in the annual Q600 to a large extent comes from the wet season (Figure 9(b) and (h)).

Similarly, Figures 4 and 5 present biases distribution of T500 and Z500 of the GCMs. GCMs show predominant cold biases in T500 (Figure 4), which may be caused by penetration of dry and cold air from the deserts of western Asia due to an overly smoothed representation of topography west of the TP (Boos and Hurley, 2013). Apart from these cold biases, all the GCMs successfully reproduce the spatial pattern with SCC larger than 0.95. With a relatively small cold biases over the TP, MIROC4h and CCSM4 outperform other GCMs in terms of T500, with a MAE of 1.6 and 1.7 °C, respectively. Remarkable cold biases are shown in GFDL-CM3, GFDL-ESM2M, MRI-CGCM3, IPSL-CM5A-LR and GISS-E2-H with MAE bigger than 3.0 °C. In the case of Z500, CCSM4 and MIROC4h overestimate Z500 but the remaining GCMs underestimate it. The five GCMs with relatively poor performance in T500 also show underestimation in Z500. Overwhelming negative biases appear in IPSL-CM5A-LR. Similar to T500, the underestimation in annual Z500 to a large extent comes from the cold season, while the slightly overestimation in CCSM4 and MIROC4h mainly comes from warm season (Figure 9).

Figures 6 and 7 present biases distributions of the wind components at 200 hPa (U200 and V200). Wind components in the GCMs are not simulated so well as SLP, Z500 and T500. Large discrepancy exists in terms of spatial pattern in the wind components, especially in V200 with SCC less than 0.5 (Figure 7). CSIRO-MK and CCSM4 possess higher SCC as 0.84 and 0.83 in U200, and smaller MAE than other GCMs compared with ERA-Int. GISS-E2-H, IPSL-CM5A-LR and MRI-CGCM3 show poor performance in U200 with pronounced overestimation to the south of the TP and underestimation to the north (Figure 6). For V200, CSIRO-MK, IPSL-CM5A-LR and MIROC-ESM present stronger northerly component than ERA-Int in the TP, which comes from biases in winter (Figure 9).

Owing to the complex topography in the TP, some of the differences discussed so far may be because of the differences in the height representation in the models. Figure 8 shows the differences between the heights in the 14 models plus the NCEP-DOE and those used in ERA-Int. It is interesting to note that most differences appear along the boundaries of the TP where dramatic elevation changes occur. Owing to the fairly big difference in the resolutions, this type of differences is unavoidable. Another notable feature is that the differences between the two reanalysis models are comparable to those between the GCMs and the ERA-Int model. Given that the differences of the six variables between the two reanalyses are generally negligible, and there is no strong signal of the direct topographic influence on the simulated variables, the different topography representations are not considered a major factor in the differences among the models.

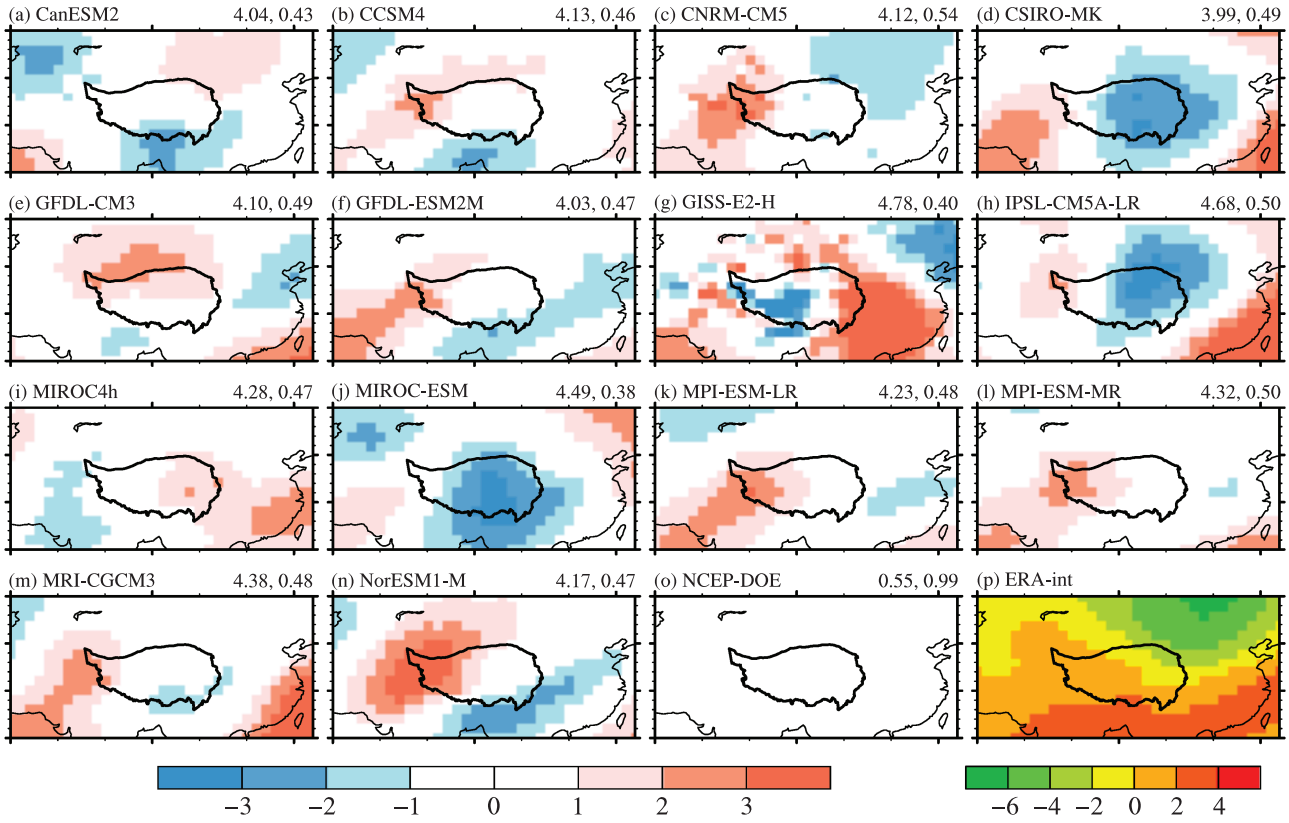


Figure 7. The same as Figure 2, but for meridional wind at 200 hPa V200 (unit: m $s^{-1}$).

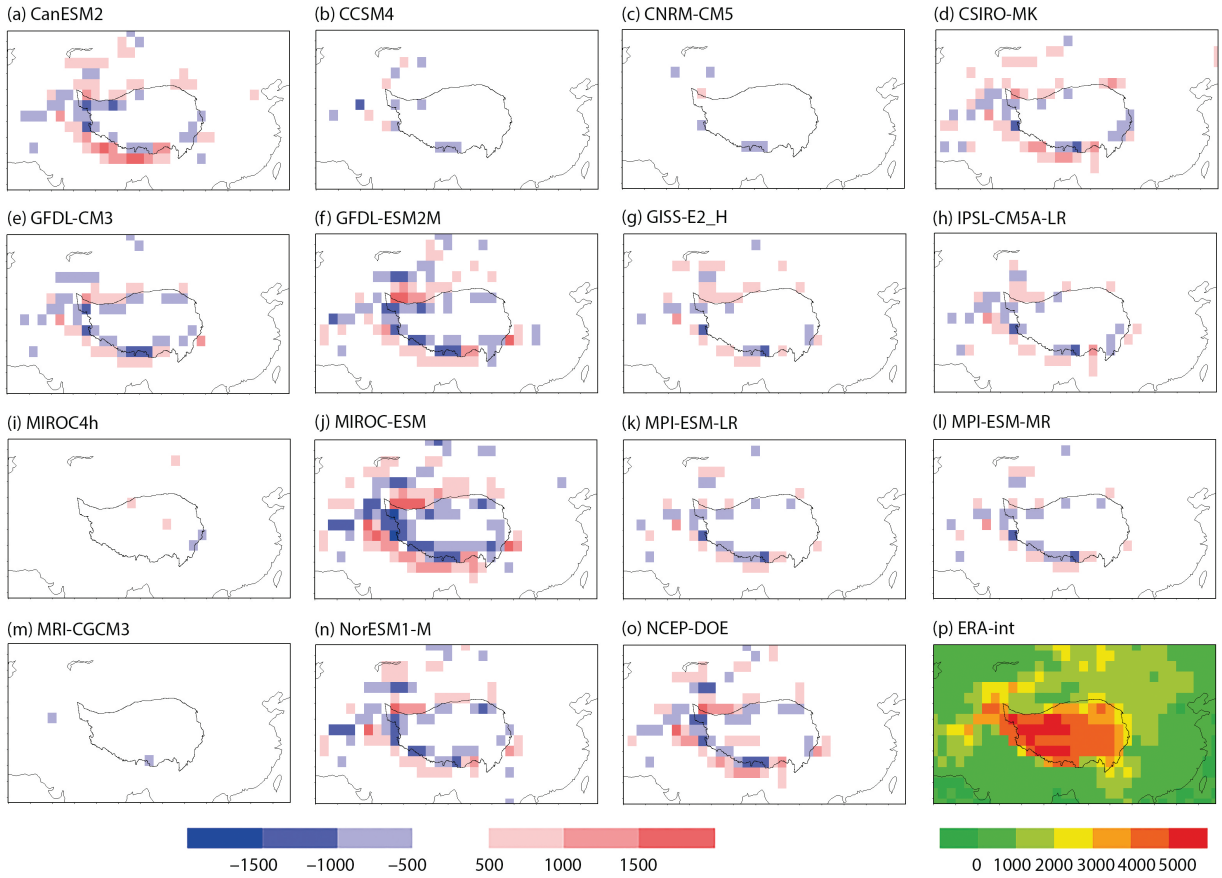


Figure 8. The same as Figure 2, but for topography (unit: m).

 

Table 3. Seasonal (JJA and DJF) mean errors (ME) of SLP (unit: hPa), Q600 (unit: g kg$^{-1}$), T500 (unit: °C), Z500 (unit: m$^2$ s$^{-2}$), U200 and V200 (unit: m s$^{-1}$) compared with ERA-Int for the 14 GCMs, GCM ensemble and NCEP-DOE in the TP during 1979–2005.

| | SLP | | Q600 | | T500 | | Z500 | | U200 | | V200 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | JJA | DJF | JJA | DJF | JJA | DJF | JJA | DJF | JJA | DJF | JJA | DJF |
| CanESM2 | −5.08 | 0.79 | −0.59 | −0.16 | −0.14 | −1.26 | 10.91 | −27.29 | 1.16 | 2.89 | 0.68 | −0.84 |
| CCSM4 | 0.34 | 2.81 | −0.20 | −0.04 | 0.48 | −0.95 | 36.17 | 11.44 | −0.51 | 0.28 | 0.88 | −0.75 |
| CNRM-CM5 | −0.43 | 1.49 | −0.76 | −0.13 | −1.54 | −3.95 | −6.47 | −43.14 | −2.22 | −0.47 | 0.41 | −0.52 |
| CSIRO-MK | −3.94 | 0.04 | −0.36 | −0.26 | −0.33 | −2.35 | 16.16 | −37.77 | −0.80 | −1.03 | 0.68 | −0.41 |
| GFDL-CM3 | −0.92 | 1.57 | −0.64 | −0.27 | −3.04 | −4.52 | −18.65 | −64.16 | 3.18 | −0.21 | 0.60 | 0.46 |
| GFDL-ESM2M | 0.52 | 3.37 | −0.64 | −0.22 | −2.93 | −4.16 | −20.04 | −49.85 | 0.91 | −1.64 | 0.51 | −0.42 |
| GISS-E2-H | −0.47 | −1.78 | −1.53 | −0.18 | −3.65 | −2.63 | −34.20 | −21.86 | 2.82 | 0.69 | 1.47 | −0.11 |
| IPSL-CM5A-LR | 2.18 | 5.33 | −0.69 | −0.14 | −2.39 | −3.53 | −41.85 | −78.93 | 1.31 | 3.37 | 1.82 | −0.92 |
| MIROC4h | 0.79 | 1.05 | −0.22 | −0.21 | 0.33 | −1.49 | 32.65 | −8.18 | 1.95 | 2.42 | 0.40 | 0.29 |
| MIROC-ESM | 1.16 | 3.38 | −0.13 | −0.05 | −1.11 | −3.12 | 9.66 | −18.42 | −2.43 | −0.01 | −0.07 | −1.00 |
| MPI-ESM-LR | −1.50 | −0.07 | −0.03 | 0.00 | −1.34 | −2.09 | −3.82 | −20.87 | 0.25 | −0.05 | 0.57 | 0.30 |
| MPI-ESM-MR | −1.53 | −0.44 | −0.04 | −0.01 | −1.30 | −2.27 | −3.43 | −23.80 | 0.64 | −0.15 | 0.53 | 0.19 |
| MRI-CGCM3 | 1.72 | 3.11 | −0.78 | −0.16 | −2.67 | −3.63 | −12.23 | −58.20 | 1.58 | 1.83 | 1.23 | −0.28 |
| NorESM1-M | 0.04 | 3.71 | −0.45 | −0.05 | −1.11 | −2.79 | 2.42 | −14.18 | −0.93 | −0.88 | 0.56 | 0.10 |
| NCEP-DOE | 0.57 | 0.45 | −0.14 | 0.14 | −0.36 | −0.22 | 11.42 | 1.56 | −0.20 | −0.28 | −0.16 | −0.24 |
| ENS | −0.51 | 1.74 | −0.50 | −0.13 | −1.48 | −2.77 | −2.34 | −32.52 | 0.49 | 0.50 | 0.73 | −0.28 |

### *3.1.2. Seasonal variation*

Most part of the study domain is dominated by different monsoon systems in summer and winter, which results in distinct circulations and climates in different seasons. Thus it is interesting to evaluate GCMs' performances in simulating seasonal characters. Summer (JJA) and winter (DJF) performance are separately evaluated. To distinguish the biases with different sign between DJF and JJA, the mean error (ME), rather than MAE, is used in Table 3 and Figure 9.

*3.1.2.1. Sea level pressure:* Half of the GCMs underestimate (half slightly overestimate) SLP in the study domain, causing a slight ensemble underestimation in the JJA ME averaged in the domain (Table 3). Larger biases exist over the TP than in the surroundings as seven GCMs underestimate and IPSL-CM5A-LR overestimates (Figure 9(a)). CNRM-CM5, CCSM4 and NorESM1-M possess smaller biases than others (Table 3). However, the top three in terms of the spatial correlation coefficient are CNRM-CM5, CCSM4 and MRI-CGCM3 in JJA, following NCEP-DOE (Table 4).

In winter, SLP is characterized by a strong land-sea pressure gradient as the Siberian high located in the Eurasian continent and a low pressure, the Aleutian low, over the sea. Change of SLP differences between the continent and sea is the most evident indicator of evolution of East Asia winter monsoon (Wei *et al.*, 2013). In contrast to the summer, most GCMs overestimate SLP in winter, leading to an ensemble overestimation (Table 3). As the same as in JJA, biases in the TP are larger than the surroundings. In fact, IPSL-CM5A-LR overestimates SLP in the TP by about 40 hPa (Figure 9(g)). The enhanced land-sea pressure gradient in GCMs results in a stronger low-level northerly winds along the coast of East Asia than the observations, which confirm what Gong *et al.* (2014) have found. CSIRO-MK, MPI-ESM-LR and MPI-ESM-MR outperform others in terms of the ME; while CNRM-CM5, MPI-ESM-LR and MRI-CGCM3 perform better than others in terms of SCC.

*3.1.2.2. Specific humidity at 600 hPa (Q600):* In summer, as Asia summer monsoon prevails in the East Asia and South Asia, moisture is transported from the ocean to the land, which results in higher Q600 in summer than in winter, especially in the South domain. All the GCMs underestimate Q600 in JJA (Table 3), particularly to the south of the TP (Figure 9(b)), which is likely to account for precipitation underestimation in the East Asia (Sperber *et al.*, 2013) and South Asia (Boos and Hurley, 2013). All the model simulations show large spread in ME from −0.03 to −0.78 g kg$^{-1}$ (Table 3) and SCC from 0.77 to 0.89 (Table 4) in the study domain. MPI-ESM-LR, MPI-ESM-MR, MIROC-ESM and CCSM4 relatively outperform others in terms of ME and SCC. For Q600, the whole TP is above the 600 hPa for eight GCMs, which may reduce the Q600 mean absolute errors of these GCMs because of the missing values set in the TP. However, it will not strongly affect the inter-comparison results because the area with missing values is relatively small and Q600 biases mostly occur to the south of 35°N.

The values of Q600 and biases are smaller in winter than in summer. NCEP-DOE overestimates Q600 over the TP, which points to an important uncertainty associated with reanalysis. Similar to the ensemble dry bias in JJA, an ensemble dry bias also exists in DJF but with much smaller magnitude than that in JJA (Figure 9(h) and Table 3). DJF possesses smaller ME and SCC than JJA, which means that most GCMs have serious limitation in reproducing spatial pattern of Q600 in winter when the moisture content is low. MPI-ESM-LR, MPI-ESM-MR and CCSM4 outperform others in terms of ME.

*3.1.2.3. T500 and Z500:* Spatial distributions of T500 and Z500 for all the 14 GCMs are highly correlated with

 

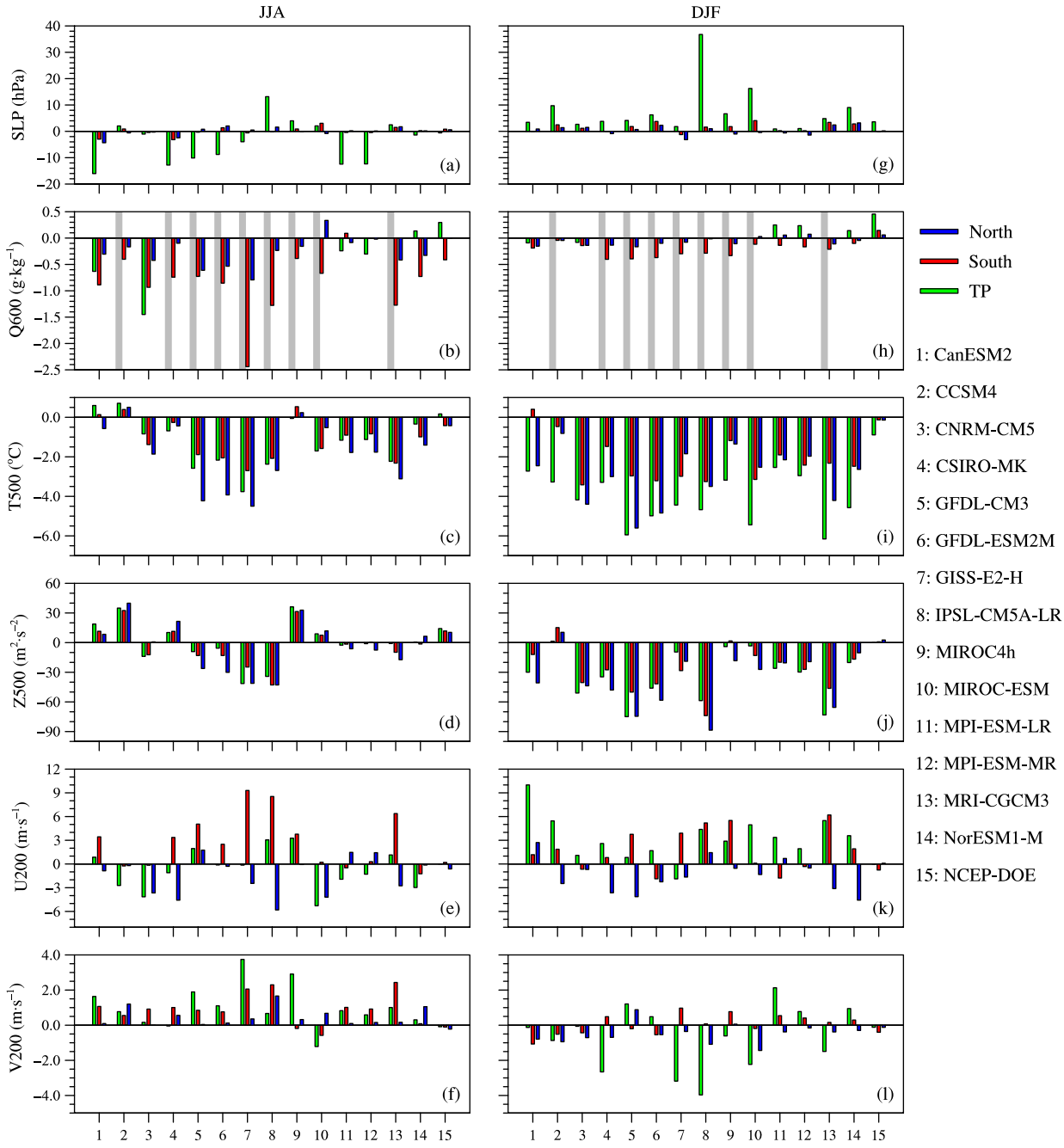


Figure 9. Seasonal (JJA and DJF) biases for SLP (unit: hPa) (a, g), Q600 (unit: g kg$^{-1}$) (b, h), T500 (unit: °C ) (c, i), Z500 (unit: m$^2$ s$^{-2}$) (d, j), U200 (unit: m s$^{-1}$) (e, k) and V200 (unit: m s$^{-1}$) (f, l) averaged over the TP (TP), to the south of 35 °N (South) and to the north of 35 °N (North) in the study domain. Grey bars represent missing value over the TP.

those of the ERA-Int. All the SCCs are higher than 0.9 (Table 4). The ensemble means of all the models for both variables are underestimated in JJA and DJF (Table 3). The biases in DJF are larger than those in JJA. CanESM2, MIROC4h, CSIRO-MK and CCSM4 have smaller ME for T500 in the study domain (Table 3). CCSM4 performs better than the others as it exhibits small biases for T500 outside of the TP. MIROC4h and CCSM4 have better performance in DJF for Z500 than the others.

In summer, the western Pacific subtropical high hovers overhead the East Asia. Meanwhile, due to the heating of the TP, a warm centre appears in the southern TP. It constitutes a strong south-northward gradient in geopotential height at 500 hPa. The majority of the GCMs have cold biases in JJA except for CCSM4 and MIROC4h. The underestimation of T500 in JJA in the south (Figure 9(c)) may be partly related to the reduced precipitation in the South Asia (Boos and Hurley, 2013) and the resultant reduction of latent heat from condensation, while invasion of cold air from the high latitude may result in cold biases in the north in DJF (Gong *et al.*, 2014). Most GCMs show larger cold biases in the north than south in DJF (Figure 9(i)). All the GCMs possess large negative biases in Z500 except for CCSM4 in DJF (Figure 9(j)

 

Table 4. Seasonal (JJA and DJF) spatial correlation coefficient (SCC) of SLP, Q600, T500, Z500, U200 and V200 with ERA-Int for the 14 GCMs, GCM ensemble and NCEP-DOE in the TP during 1979–2005.

| | SLP | | Q600 | | T500 | | Z500 | | U200 | | V200 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | JJA | DJF | JJA | DJF | JJA | DJF | JJA | DJF | JJA | DJF | JJA | DJF |
| CanESM2 | 0.16 | 0.83 | 0.81 | 0.70 | 0.95 | 0.98 | 0.93 | 0.98 | 0.90 | 0.84 | 0.39 | 0.59 |
| CCSM4 | 0.86 | 0.73 | 0.88 | 0.68 | 0.96 | 0.97 | 0.92 | 0.98 | 0.90 | 0.87 | 0.40 | 0.59 |
| CNRM-CM5 | 0.89 | 0.90 | 0.86 | 0.78 | 0.96 | 0.98 | 0.92 | 0.98 | 0.90 | 0.87 | 0.42 | 0.65 |
| CSIRO-MK | 0.37 | 0.85 | 0.84 | 0.66 | 0.96 | 0.98 | 0.92 | 0.97 | 0.89 | 0.88 | 0.44 | 0.68 |
| GFDL-CM3 | 0.43 | 0.84 | 0.87 | 0.72 | 0.96 | 0.98 | 0.94 | 0.97 | 0.88 | 0.87 | 0.35 | 0.65 |
| GFDL-ESM2M | 0.38 | 0.80 | 0.87 | 0.63 | 0.96 | 0.98 | 0.94 | 0.98 | 0.90 | 0.84 | 0.37 | 0.60 |
| GISS-E2-H | 0.71 | 0.77 | 0.68 | 0.70 | 0.92 | 0.97 | 0.86 | 0.97 | 0.77 | 0.85 | 0.14 | 0.63 |
| IPSL-CM5A-LR | 0.76 | 0.16 | 0.80 | 0.67 | 0.92 | 0.98 | 0.90 | 0.97 | 0.74 | 0.85 | 0.38 | 0.60 |
| MIROC4h | 0.82 | 0.71 | 0.89 | 0.71 | 0.96 | 0.98 | 0.93 | 0.98 | 0.90 | 0.87 | 0.34 | 0.66 |
| MIROC-ESM | 0.55 | 0.34 | 0.77 | 0.57 | 0.91 | 0.97 | 0.91 | 0.97 | 0.86 | 0.87 | 0.24 | 0.57 |
| MPI-ESM-LR | 0.29 | 0.88 | 0.90 | 0.69 | 0.96 | 0.98 | 0.93 | 0.98 | 0.90 | 0.86 | 0.39 | 0.65 |
| MPI-ESM-MR | 0.30 | 0.85 | 0.89 | 0.70 | 0.96 | 0.98 | 0.94 | 0.97 | 0.90 | 0.86 | 0.41 | 0.59 |
| MRI-CGCM3 | 0.86 | 0.86 | 0.80 | 0.76 | 0.95 | 0.97 | 0.93 | 0.97 | 0.86 | 0.87 | 0.41 | 0.62 |
| NorESM1-M | 0.81 | 0.80 | 0.87 | 0.69 | 0.96 | 0.97 | 0.93 | 0.97 | 0.91 | 0.87 | 0.41 | 0.60 |
| NCEP-DOE | 0.91 | 0.93 | 0.92 | 0.87 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.99 |
| ENS | 0.59 | 0.74 | 0.84 | 0.69 | 0.95 | 0.98 | 0.92 | 0.97 | 0.87 | 0.86 | 0.36 | 0.62 |

and Table 3). Underestimation in Z500 is highly related to the underestimation in T500. The underestimation is much pronounced in the North than the South, which means a weak western Pacific subtropical high for most GCMs. It is responsible for underestimation of summer rainfall intensity as indicated by Feng *et al.* (2014). The increased thermal gradient and the north-south geopotential height gradient enhance the westerly in the North (Figure 9(k)) and result in a slightly stronger westerly jet stream in the winter (Gong *et al.*, 2014). There are also strong underestimations over the climatological East Asian trough region, which means a deeper East Asian trough in the GCMs than in ERA-Int (Gong *et al.*, 2014).

*3.1.2.4. Wind components:* U200 in summer and winter shows contrasting spatial climatology patterns in the Asia monsoon region, especially over the TP. Summer is characterized by relatively weak westerly in the North and easterly in the South. Contrary to summer, winter is characterized by strong westerly covering the whole domain, with a centre on the south of TP. And there exists a strong westerly jet stream over the East Asia, called the East Asia westerly jet. In summer, the southerly prevails in the west of the study domain and northerly in the east because the South Asia high located in the upper troposphere dominants the Asia summer monsoon region. In winter, the southerly prevails to the south of the TP due to the Hadley circulation. The GCMs capture the spatial pattern of the zonal component better than the meridional one in JJA and DJF (Table 4), which indicates that there is more room for improvement of the monsoon dynamics over Asia.

Nine GCMs overestimate and five underestimate U200 in JJA (Table 3). Overestimation in the South and underestimation in the North in JJA (Figure 9e) indicate that the easterly in the South and the westerly in the North are both underestimated, which may be related to the underestimation of the South Asia High in the southern TP (Duan *et al.*, 2013). MPI-ESM-LR, CCSM4 and MPI-ESM-MR outperform other GCMs with smaller ME and higher SCC. In DJF, most GCMs overestimate U200 in the TP and the South and slightly underestimate U200 in the North (Figure 9(k)).

Pattern of the meridional component (V200) shows consistence between NCEP-DOE and ERA-Int (Figures 8(f) and (l)). However, it is not well reproduced by the GCMs in JJA with SCC ranging from 0.14 to 0.44 (Table 4 and Figure 9(f)). All the GCMs overestimate the southerly in the study domain, practically in the South of the TP except MIROC-ESM. It is noted that GISS-E2-H, IPSL-CM5A-LR and MRI-CGCM3 suffer from large biases (higher than 1 m $s^{-1}$). In DJF, nine GCMs underestimate the southerly (Table 3). The overestimation in JJA and underestimation in DJF suggest that the strength of the monsoon circulation is overstimulated in the majority of the GCMs.

### 3.2. KS test

Following the spatial characteristics of climatology evaluation for the six variables, we assess the distributional similarity by the KS test.

Table 5 shows NCEP-DOE presents similar distribution with ERA-Int in most of the domain as the grid number with significant distributional difference is less than 200 for most variables. At mid-troposphere, reanalysis uncertainty is very small for T500. Regarding Z500, significant distributional differences are detected over the southern region, but these differences are significantly reduced when the anomalies were used to the same analysis. At the upper troposphere, reanalysis uncertainties for U200 and V200 are negligible with almost no significant difference in the whole domain. In spite of the general similarity, considerable differences are found for SLP and Q600 over the TP and adjacent southwestern edge (figure not shown). In general, uncertainty in reanalyses for all the six variables outside the TP is generally negligible. The

 

Table 5. The number of grids at which P value in logarithm of the KS test is less than the threshold −1.301 for original time series (ORG) and anomalies (ANO) of 6 variables in 14 GCMs and NCEP-DOE, indicating the distributional difference is significant at the 5% confidence level (The total grid number is 528).

| | SLP | | Q600 | | T500 | | Z500 | | U200 | | V200 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ORG | ANO | ORG | ANO | ORG | ANO | ORG | ANO | ORG | ANO | ORG | ANO |
| CanESM2 | 486 | 322 | 445 | 305 | 459 | 203 | 451 | 346 | 449 | 167 | 362 | 129 |
| CCSM4 | 416 | 180 | 242 | 163 | 369 | 314 | 527 | 269 | 284 | 142 | 320 | 40 |
| CNRM-CM5 | 245 | 112 | 449 | 373 | 528 | 351 | 508 | 363 | 238 | 37 | 305 | 3 |
| CSIRO-MK | 442 | 296 | 336 | 213 | 447 | 232 | 443 | 469 | 365 | 45 | 363 | 96 |
| GFDL-CM3 | 342 | 194 | 437 | 296 | 528 | 222 | 528 | 405 | 396 | 72 | 341 | 43 |
| GFDL-ESM2M | 515 | 244 | 400 | 269 | 528 | 246 | 528 | 247 | 277 | 37 | 294 | 62 |
| GISS-E2-H | 428 | 267 | 464 | 449 | 528 | 280 | 494 | 224 | 453 | 157 | 433 | 223 |
| IPSL-CM5A-LR | 417 | 244 | 306 | 312 | 528 | 233 | 528 | 473 | 436 | 64 | 387 | 182 |
| MIROC4h | 375 | 103 | 276 | 134 | 314 | 234 | 471 | 369 | 312 | 30 | 308 | 117 |
| MIROC-ESM | 456 | 259 | 398 | 348 | 510 | 313 | 396 | 342 | 328 | 190 | 411 | 118 |
| MPI-ESM-LR | 279 | 158 | 301 | 218 | 480 | 114 | 225 | 140 | 327 | 69 | 289 | 17 |
| MPI-ESM-MR | 308 | 182 | 364 | 184 | 492 | 181 | 200 | 195 | 329 | 49 | 240 | 22 |
| MRI-CGCM3 | 432 | 157 | 380 | 355 | 528 | 243 | 518 | 384 | 495 | 73 | 297 | 98 |
| NorESM1-M | 425 | 280 | 348 | 322 | 499 | 325 | 232 | 170 | 384 | 156 | 362 | 46 |
| NCEP-DOE | 198 | 87 | 241 | 197 | 53 | 14 | 253 | 41 | 2 | 0 | 11 | 0 |

GCMs mainly exhibit pronounced differences at low levels. At the upper troposphere, much smaller differences than low levels are detected in the GCMs compared with ERA-Int (Table 5). Compared with other variables, U200 and V200 suffer from smaller distributional biases, particularly in high order moments. Uncertainty associated with reanalyses in SLP and Q600 over the TP will leave adverse impacts on GCMs evaluation and eventual downscaling. In contrast to the TP, reanalysis uncertainty outside the TP is generally negligible. GCMs' performance can be reliably assessed against ERA-Int with high confidence. Subsequently, we analyze the performances of the GCMs to detect which region suffers from distributional biases and whether these differences originate from the mean or higher order moment biases.

In the case of SLP, five GCMs, namely CNRM-CM5, GFDL-CM3, MIROC4h, MPI-ESM-LR and MPI-ESM-MR, suffer from smaller distributional biases for original series, especially in the eastern Asia (figure not shown). Pronounced differences in high-order moments calculated from anomalies are still revealed in the TP for most GCMs, which is consistent with remarkable biases found in Figure 2.

Compared with SLP, biases for Q600 are generally more evident and widespread, and the spatial patterns are considerably diversified among models (Table 5). While GFDL and MPI models share similar pattern for anomalies, MIROC4h and CCSM4 outperform other models as they do not have significant distributional biases outside the TP (figure not shown). Significant distributional biases exist in the TP for original series in all the GCMs, but they are largely removed in anomalies for nine GCMs with exception of small area in the north TP, which is also in accordance with the biases depicted in Figure 3.

In the mid-troposphere, notable distributional differences are found for T500 in all GCMs and cover almost the whole domain (figure not shown), whereas CCSM4 and MIROC4h outperform the remaining models as they do not suffer from pronounced biases in the north. Biases in high-order moments are largely reduced and mainly located in the south, especially for MPI-ESM-LR and MPI-ESM-MR, which suggests that distributional biases mainly result from climatic cold biases. With regard to Z500, biases appear over most of the domain except for Nor-ESM1-M, MPI-ESM-LR and MPI-ESM-MR as they only generate biases in the southwest (figure not shown). Most GCMs exhibit striking distributional biases over India for Z500 anomalies.

In general, for all the variables evaluated, considerable distributional differences are detected but the differences are reduced by about half in almost all the GCMs when the anomalies are considered, which suggests that the errors are partly caused by biases which in turn are associated with a shift of climatic mean and partly by higher order moments. This is different compared twith the regions in Europe and Africa (Brands *et al.*, 2013), which may be attributed to the highly heterogeneous land surface over the TP.

### 3.3. Ranking scores

The spatial ranking scores are calculated based on the spatial correlation efficient (SCC), ME and SD of the six variables, and the results are presented in Figure 10. As shown in Figure 10(a), all the models have comparably high-spatial SCC for Z500 and T500 but lower values for V200. As shown in Figure 10(a)–(c), scores of SCC and SD for V200 are smaller than those of ME, which indicates that all the GCMs have large errors in reproducing spatial pattern of V200, whereas they are able to reproduce mean value which is mainly due to offsets between negative and positive errors.

To further examine the difference in model performance, we compare the mean scores averaged over the three quantities for each variable (Figure 10(d)). Performance of most models is reasonable for Z500 and T500, followed by U200, Q600 and SLP, but relatively poor for

 

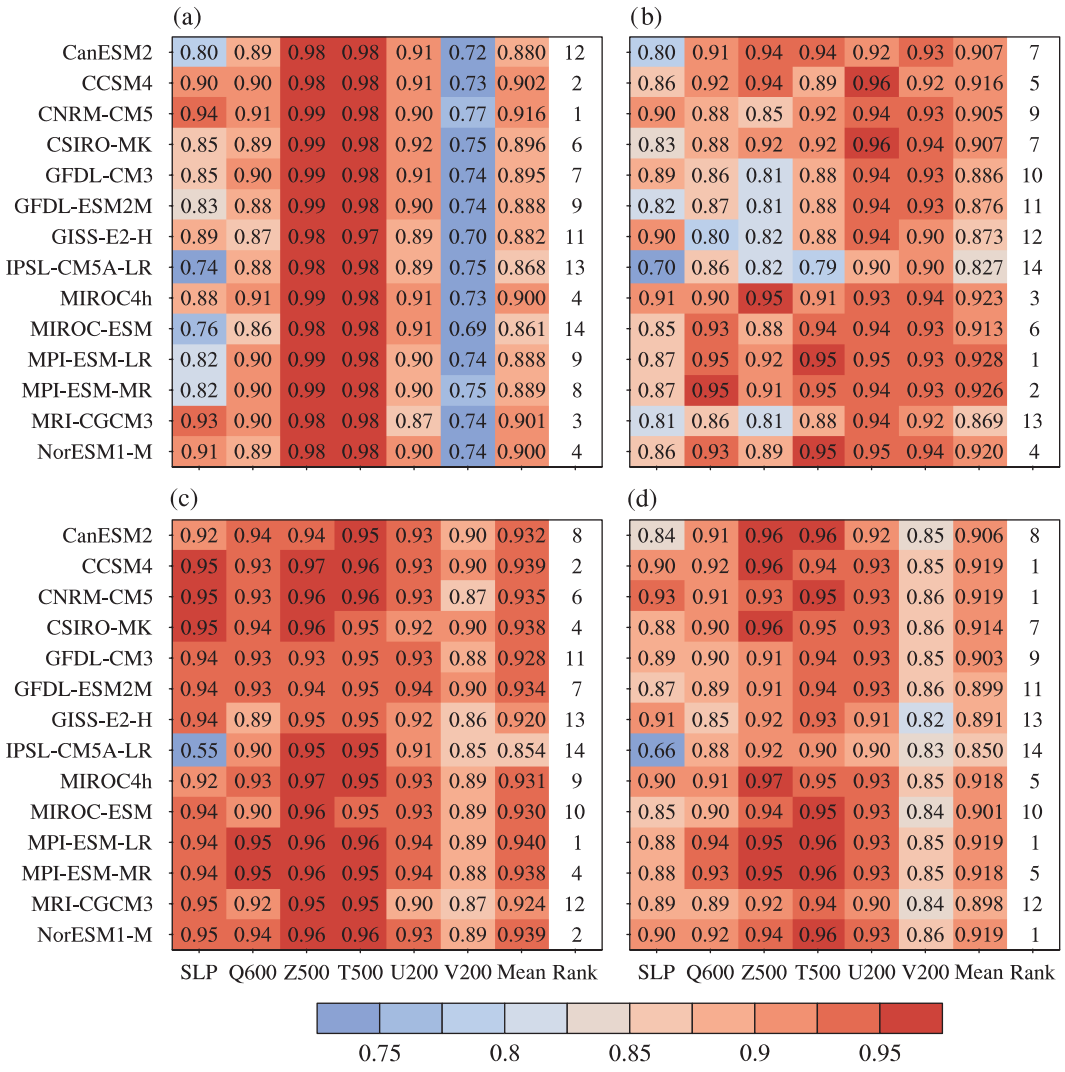


Figure 10. Spatial ranking scores for (a) spatial correlation efficient (SCC), (b) spatial mean error (ME), (c) standard deviation (SD) and (d) mean scores of six variables from 14 GCMs. The penultimate column represents mean ranking scores averaged over six variables and the last column is the ranking according to the mean ranking scores in each panel.

V200. Although no model performs best or worst at all aspects, we can determine overall skills of the models for the study domain with help of the averaged ranking scores. In general, seven top GCMs including CCSM4, CNRM-CM5, MPI-ESM-LR, NorESM1-M, MIROC4h, MPI_ESM_MR and CSIRO-MK are identified, with a composited mean score of 0.92. The four relatively poor models are GFDL-ESM2M, MRI-CGCM3, GISS-E2-H and IPSL-CM5A-LR.

## 4. Summary and conclusions

Downscaling is a frequently used approach to get regional or local scale climate information. However, GCMs could pass over their errors to the downscaling results through the large-scale forcing. Therefore, identifying and understanding GCM's error for large-scale variables used in downscaling is critical to a successful downscaling. In this study, five upper level and one low level atmospheric variables including sea level pressure (SLP), Q600, air temperature and geopotential height at 500 hPa (Z500 and T500), wind components at 200 hPa (U200 and V200), in 14 GCMs of CMIP5 are evaluated over the TP and its surroundings. ERA-Int, which is proven to outperform several other reanalyses in the study region, is used as the reference for the evaluation. Differences between NCEP-DOE and ERA-Int are analyzed to indicate the uncertainty in the reanalyses. The evaluation focuses on spatial characteristics of annual and seasonal biases and distributional features which are checked by the KS test. Finally, a spatial ranking scheme is used to identify relatively more skillful GCMs for downscaling in this region.

Uncertainty associated with reanalysis as expressed by the differences between the two reanalyzes is quite small except for the specific humidity at low level in the study region. ERA-Int possesses the best performance in water cycle climatology and changes in the TP compared with several other reanalysis products (Gao *et al.*, 2014) because it utilizes four-dimensional variational data assimilation (Dee and Uppala, 2009). Moreover, more observations of temperature and humidity are used in ERA-Int than in NCEP-DOE. These advantages may have

 

made ERA-Int superior to NCEP-DOE for the above six variables over the TP, which ensures the accuracy and reliability in GCMs assessment using ERA-Int as the reference over this region.

Spatial patterns for wind components, especially for V200, are not well simulated by all the GCMs. Distinct seasonal features in V200 are detected with overestimation in JJA over the TP, and underestimation in the South of the TP in DJF, which most likely has a strong impact on large-scale synoptic systems over the TP, especially the monsoon circulation. Remarkable cold biases are detected in T500 in most GCMs, especially in DJF, which led to the underestimation in Z500, underestimation in U200 in the North and overestimation in the South. Q600 is underestimated in the TP and pronouncedly underestimated in the southeast of the TP in JJA.

For distributional feature, closer agreements are found at 200 hPa than lower levels. Variables at the low levels possess more discrepancy compared with ERA-Int, especially for Q600. All the GCMs surfer from pronounced distributional biases for original datasets but these biases are significantly reduced when anomalies are applied in the KS test, suggesting the biases are mainly due to a climatic mean shift.

The spatial ranking scores show that CCSM4, CNRM-CM5, MPI-ESM-LR, NorESM1-M, MIROC4h, MPI_ESM_MR and CSIRO-MK are over other seven in terms of the spatial correlation efficient, ME, and SD averaged over the six variables in the study region. These models may be prioritized in further downscaling applications for this region.

It is worth noting that none of the GCMs performs absolutely best or worst at all aspects, although four GCMs stand out in terms of their scores for simulating the six atmospheric variables. As GCMs incorporate several components describing atmospheric, oceanic, land and cryospheric climate system components, it is not feasible to evaluate GCMs performance at all aspects concurrently. Therefore, it is hard to select a definitely best GCM as it may be good at certain aspects but have shortcomings in others, which have been pointed out by many previous studies (Sheffield *et al.*, 2013a, 2013b; Su *et al.*, 2013; Ou *et al.*, 2013). Even for a specific variable, performance of a GCM can be remarkably different depending on the region selected (e.g. Sheffield *et al.*, 2013a).

## Acknowledgements

We appreciate the free access of the ERA-Interim, NCEP-DOE reanalysis and GCM datasets, which are provided by the European Centre for Medium-Range Weather Forecasts (ECMWF, http://www.ecmwf.int/research/era/do/get/ERA-Interim), NCEP, the NOAA/OAR/ESRL PSD (http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis2.html), and ESGF web portals (http://pcmdi9.llnl.gov/esgf-web-fe/), respectively. This work is jointly supported by National Basic Research Programme of China (2013CB956004), National Natural Science Foundation of China (91537105 and 41322033). Deliang Chen is supported by Swedish VR, BECC, and MERGE.

## References

Alexandru A, De Elia R, Laprise R, Separovic L, Biner S. 2009. Sensitivity study of regional climate model simulations to large-scale nudging parameters. *Mon. Weather Rev.* **137**: 1666–1686, doi: 10.1175/2008mwr2620.1.

Bellenger H, Guilyardi E, Leloup J, Lengaigne M, Vialard J. 2014. ENSO representation in climate models: from CMIP3 to CMIP5. *Clim. Dyn.* **42**: 1999–2018, doi: 10.1007/s00382-013-1783-z.

Benestad RE, Hanssen-Bauer I, Chen D. 2008. *Empirical-Statistical Downscaling*. World Scientific: Singapore, 300.

Boos WR, Hurley JV. 2013. Thermodynamic bias in the multimodel mean boreal summer monsoon. *J. Clim.* **26**: 2279–2287, doi: 10.1175/JCLI-D-12-00493.1.

Brands S, Gutiérrez JM, Herrera S, Cofiño AS. 2012. On the use of reanalysis data for downscaling. *J. Clim.* **25**: 2517–2526, doi: 10.1175/jcli-d-11-00251.1.

Brands S, Herrera S, Fernandez J, Gutierrez JM. 2013. How well do CMIP5 Earth System Models simulate present climate conditions in Europe and Africa? *Clim. Dyn.* **41**: 803–817, doi: 10.1007/s00382-013-1742-8.

Bruyère C, Done J, Holland G. 2014. Bias corrections of global models for regional climate simulations of high-impact weather. *Clim. Dyn.* **43**(7–8): 1847–1856, doi: 10.1007/s00382-013-2011-6.

Cattiaux J, Douville H, Peings Y. 2013. European temperatures in CMIP5: origins of present-day biases and future uncertainties. *Clim. Dyn.* **41**: 2889–2907, doi: 10.1007/s00382-013-1731-y.

Chen L, Frauenfeld OW. 2014. A comprehensive evaluation of precipitation simulations over China based on CMIP5 multimodel ensemble projections. *J. Geophys. Res.-Atmos.* **119**(10): 5767–5786, doi: 10.1002/2013JD021190.

Chen L, Pryor S, Li D. 2012. Assessing the performance of Intergovernmental Panel on Climate Change AR5 climate models in simulating and projecting wind speeds over China. *J. Geophys. Res.-Atmos.* **117**: D24012, doi: 10.1029/2012JD017533.

Chen D, Xu B, Yao T, Guo Z, Cui P, Chen F, Zhang R, Zhang X, Zhang Y, Fan J, Hou Z, Zhang T. 2015. Assessment of past, present and future environmental changes on the Tibetan Plateau. *Chin. Sci. Bull.* **60**(32): 3025–3035, doi: 10.1360/N972014-01370. (in Chinese with English abstract).

Dee D, Uppala S. 2009. Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Q. J. R. Meteorol. Soc.* **135**: 1830–1841, doi: 10.1002/qj.493.

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars AC, van de Berg L, Bidlot J, Bormann N, Delsol, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette JJ, Park BK, Peubey C, de Rosnay P, Tavolato C, Thépaut JN, Vitart F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**: 553–597, doi: 10.1002/qj.828.

Dosio A, Panitz H-J, Schubert-Frisius M, Lüthi D. 2014. Dynamical downscaling of CMIP5 global circulation models over CORDEX-Africa with COSMO-CLM: evaluation over the present climate and analysis of the added value. *Clim. Dyn.* **44**(9–10): 2637–2661, doi: 10.1007/s00382-014-2262-x.

Douglass AR, Prather MJ, Hall TM, Strahan SE, Rasch PJ, Sparling LC, Coy L, Rodriguez JM. 1999. Choosing meteorological input for the global modeling initiative assessment of high-speed aircraft. *J. Geophys. Res.-Atmos.* **104**: 27545–27564, doi: 10.1029/1999jd900827.

Duan AM, Hu J, Xiao ZX. 2013. The Tibetan plateau summer monsoon in the CMIP5 simulations. *J. Clim.* **26**: 7747–7766, doi: 10.1175/jcli-d-12-00685.1.

Fan L, Chen D, Fu C, Yan Z. 2013. Statistical downscaling of summer temperature extremes in Northern China. *Adv. Atmos. Sci.* **30**: 1085–1095, doi: 10.1007/s00376-012-2057-0.

Feng J, Wei T, Dong W. 2014. CMIP5/AMIP GCM simulations of East Asian summer monsoon. *Adv. Atmos. Sci.* **31**: 836–850, doi: 10.1007/s00376-013-3131-y.

Feser F, Rockel B, von Storch H, Winterfeldt J, Zahn M. 2011. Regional climate models add value to global model data a review

 

and selected examples. *Bull. Am. Meteorol. Soc.* **92**: 1181–1192, doi: 10.1175/2011bams3061.1.

Gao L, Bernhardt M, Schulz K. 2012. Elevation correction of ERA-Interim temperature data in complex terrain. *Hydrol. Earth Syst. Sci.* **16**: 4661–4673, doi: 10.5194/hess-16-4661-2012.

Gao Y, Cuo L, Zhang Y. 2014. Changes in moisture flux over the Tibetan Plateau during 1979–2011 and possible mechanisms. *J. Clim.* **27**: 1876–1893, doi: 10.1175/JCLI-D-13-00321.1.

Gettelman A, Hegglin MI, Son SW, Kim J, Fujiwara M, Birner T, Kremser S, Rex M, Añe JA, Akiyoshi H, Austin J, Bekk S, Braesike P, Brühl C, Butchart N, Chipperfield M, Dameris M, Dhomse S, Garny H, Hardiman SC, Jöckel P, Kinnison DE, Lamarque JF, Mancini E, Marchand M, Michou M, Morgenstern O, Pawson S, Pitari G, Plummer D, Pyle JA, Rozanov E, Scinocca J, Shepherd TG, Shibata K, Smale D, Teyssèdre H, Tian W. 2010. Multimodel assessment of the upper troposphere and lower stratosphere: tropics and global trends. *J. Geophys. Res.-Atmos.* **115**: 898–907, doi: 10.1029/2009jd013638.

Giorgi F, Jones C, Asrar GR. 2009. Addressing climate information needs at the regional level: the CORDEX framework. *WMO Bull.* **58**: 175.

Gong H, Wang L, Chen W, Wu R, Wei K, Cui X. 2014. The climatology and interannual variability of the East Asian winter monsoon in CMIP5 models. *J. Clim.* **27**: 1659–1678, doi: 10.1175/JCLI-D-13-00039.1.

Gutiérrez JM, Sanmartín D, Brands S, Manzanas R, Herrera S. 2013. Reassessing statistical downscaling techniques for their robust application under climate change conditions. *J. Clim.* **26**: 171–188, doi: 10.1175/JCLI-D-11-00687.1.

Heikkila U, Sandvik A, Sorteberg A. 2011. Dynamical downscaling of ERA-40 in complex terrain using the WRF regional climate model. *Clim. Dyn.* **37**: 1551–1564, doi: 10.1007/s00382-010-0928-6.

Hong S, Kanamitsu M. 2014. Dynamical downscaling: fundamental issues from an NWP point of view and recommendations. *Asia-Pac. J. Atmos. Sci.* **50**(1): 83–104, doi: 10.1007/s13143-014-0029-2.

Immerzeel WW, Van Beek LP, Bierkens MF. 2010. Climate change will affect the Asian water towers. *Science* **328**: 1382–1385, doi: 10.1126/science.1183188.

IPCC. 2013. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, Stocker TF, Qin D, Plattner G-K, Tignor M, Allen SK, Boschung J, Nauels A, Xia Y, Bex V, Midgley PM (eds). Cambridge University Press: Cambridge, UK and New York, NY.

Jury M, Prein A, Truhetz H, Gobiet A. 2015. Evaluation of CMIP5 Models in the context of dynamical downscaling over Europe. *J. Clim.* **28**: 5575–5582, doi: 10.1175/JCLI-D-14-00430.1.

Kanamitsu M, Ebisuzaki W, Woollen J, Yang SK, Hnilo JJ, Fiorino M, Potter GL. 2002. NCEP-DOE AMIP-II reanalysis (R-2). *Bull. Am. Meteorol. Soc.* **83**: 1631–1643, doi: 10.1175/bams-83-11-1631.

Laprise R, Hernández-Díaz L, Tete K, Sushama L, Šeparović L, Martynov A, Winger K, Valin M. 2013. Climate projections over CORDEX Africa domain using the fifth-generation Canadian Regional Climate Model (CRCM5). *Clim. Dyn.* **15**(11–12): 3219–3246, doi: 10.1007/s00382-012-1651-2.

Liu XD, Chen BD. 2000. Climatic warming in the Tibetan Plateau during recent decades. *Int. J. Climatol.* **20**: 1729–1742, doi: 10.1002/1097-0088(20001130)20:14<1729::aid-joc556>3.0.co;2-y.

Ma Q, Wang K, Wild M. 2014. Evaluations of atmospheric downward longwave radiation from 44 coupled general circulation models of CMIP5. *J. Geophys. Res.-Atmos.* **119**: 4486–4497, doi: 10.1002/2013JD021427.

Meehl GA, Covey C, Taylor KE, Delworth T, Stouffer RJ, Latif M, McAvaney B, Mitchell JF. 2007. The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Am. Meteorol. Soc.* **88**: 1383–1394, doi: 10.1175/BAMS-88-9-1383.

Ou T, Chen D, Linderholm HV, Jeong JH. 2013. Evaluation of global climate models in simulating extreme precipitation in China. *Tellus A* **65**: 1393–1399, doi: 10.3402/tellusa.v65i0.19799.

Rangwala I, Sinsky E, Miller JR. 2013. Amplified warming projections for high altitude regions of the northern hemisphere mid-latitudes from CMIP5 models. *Environ. Res. Lett.* **8**: 279–288, doi: 10.1088/1748-9326/8/2/024040.

Sheffield J, Camargo SJ, Fu R, Hu Q, Jiang X, Karnauskas KB, Kim ST, Kinter J, Kumar S, Langenbrunner B, Maloney ED, Mariotti A, Meyerson JE, Johnson N, Neelin JD, Nigam S, Pan Z, Ruiz-Barradas A, Seager R, Serra YL, Sun D-Z, Wang C, Xie S-P, Yu J-Y, Zhang T, Zhao M. 2013a. North American climate in CMIP5 experiments. Part II: evaluation of historical simulations of intraseasonal to decadal variability. *J. Clim.* **26**: 9247–9290, doi: 10.1175/jcli-d-12-00592.1.

Sheffield J, Barrett AP, Colle B, Fernando DN, Fu R, Geil KL, Hu Q, Kinter J, Kumar S, Langenbrunner B, Lomardo K, Long LN, Maloney E, Mariotti A, Meyerson JE, Mo KC, Neelin JD, Nigam S, Pan Z, Ren T, Ruiz-barradas A, Serra YL, Seth A, Thibeault Jm, Stroeve JC, Yang Z, Yin L. 2013b. North American climate in CMIP5 experiments. part i: evaluation of historical simulations of continental and regional climatology. *J. Clim.* **26**: 9209–9245, doi: 10.1175/jcli-d-12-00593.1.

Soares PMM, Cardoso RM, Miranda PMA, Medeiros J, Belo-Pereira M, Espirito-Santo F. 2012. WRF high resolution dynamical downscaling of ERA-Interim for Portugal. *Clim. Dyn.* **39**: 2497–2522, doi: 10.1007/s00382-012-1315-2.

Sperber K, Annamalai H, Kang IS, Kitoh A, Moise A, Turner A, Wang B, Zhou T. 2013. The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Clim. Dyn.* **41**: 2711–2744, doi: 10.1007/s00382-012-1607-6.

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, doi: 10.1175/jcli-d-12-00321.1.

Su F, Zhang L, Ou T, Chen D, Yao T, Tong K, Qi Y. 2016. Hydrological response to future climate changes for the major upstream river basins in the Tibetan Plateau. *Glob. Planet. Chang.* **136**: 82–95, doi: 10.1016/j.gloplacha.2015.10.012.

Taylor KE, Stouffer RJ, Meehl GA. 2012. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**: 485–498, doi: 10.1175/bams-d-11-00094.1.

Wang AH, Zeng XB. 2012. Evaluation of multireanalysis products with in situ observations over the Tibetan Plateau. *J. Geophys. Res.-Atmos.* **117**: 214–221, doi: 10.1029/2011JD016553.

Wang S, Huijun J, Shuxun L, Lin Z. 2000. Permafrost degradation on the Qinghai-Tibet Plateau and its environmental impacts. *Permafr. Periglac. Process.* **11**: 43–53.

Waugh DW, Eyring V. 2008. Quantitative performance metrics for stratospheric-resolving chemistry-climate models. *Atmos. Chem. Phys.* **8**: 5699–5713, doi: 10.5194/acp-8-5699-2008.

Wei K, Xu T, Du Z, Gong H, Xie B. 2013. How well do the current state-of-the-art CMIP5 models characterise the climatology of the East Asian winter monsoon? *Clim. Dyn.* **43**: 1241–1255, doi: 10.1007/s00382-013-1929-z.

Wilks DS, Haman K. 2006. *Statistical Methods in the Atmospheric Sciences*, 2nd edn. Academic Press: Burlington, MA.

Xu Y, Xu C. 2012. Preliminary assessment of simulations of climate changes over China by CMIP5 multi-models. *Atmos. Oceanic Sci. Lett.* **5**: 489–494.

Xu Z, Yang Z. 2015. A new dynamical downscaling approach with GCM bias corrections and spectral nudging. *J. Geophys. Res.-Atmos.* **120**: 3063–3084, doi: 10.1002/2014JD022958.

Xue X, Guo J, Han B, Sun Q, Liu L. 2009. The effect of climate warming and permafrost thaw on desertification in the Qinghai-Tibetan Plateau. *Geomorphology* **108**: 182–190, doi: 10.1016/j.geomorph.2009.01.004.

Xue Y, Janjic Z, Dudhia J, Vasic R, De Sales F. 2014. A review on regional dynamical downscaling in intraseasonal to seasonal simulation/prediction and major factors that affect downscaling ability. *Atmos. Res.* **147**: 68–85.

Yao T, Pu J, Lu A, Wang Y, Yu W. 2007. Recent glacial retreat and its impact on hydrological processes on the Tibetan Plateau, China, and surrounding regions. *Arct. Antarct. Alp. Res.* **39**: 642–650, doi: 10.1657/1523-0430(07–510)[YAO]2.0.CO;2.

Yao T, Masson-Delmotte V, Gao J, Yu W, Yang X, Risi C, Sturm C, Werner M, Zhao H, He Y, Ren W, Tian L, Shi C, Hou S. 2013. A review of climate controls on delta O-18 in precipitation over the Tibetan Plateau: observations ansecd simulations. *Rev. Geophys.* **51**: 525–548, doi: 10.1002/rog.20023.

Yeh D, Gao Y. 1979. *Meteorology of the Tibetan Plateau*. Science Publication Agency: Beijing, 278 pp. (in Chinese).

 