ADVANCES IN ATMOSPHERIC SCIENCES, VOL. 32, AUGUST 2015, 1077–1091

Assessment of Indices of Temperature Extremes Simulated by Multiple CMIP5 Models over China

DONG Siyan, XU Ying*, ZHOU Botao, and SHI Ying

National Climate Center, China Meteorological Administration, Beijing 100081

(Received 11 August 2014; revised 6 January 2015; accepted 22 January 2015)

ABSTRACT

Given that climate extremes in China might have serious regional and global consequences, an increasing number of studies are examining temperature extremes in China using the Coupled Model Intercomparison Project Phase 5 (CMIP5) models. This paper investigates recent changes in temperature extremes in China using 25 state-of-the-art global climate models participating in CMIP5. Thirteen indices that represent extreme temperature events were chosen and derived by daily maximum and minimum temperatures, including those representing the intensity (absolute indices and threshold indices), duration (duration indices), and frequency (percentile indices) of extreme temperature. The overall performance of each model is summarized by a "portrait" diagram based on relative root-mean-square error, which is the RMSE relative to the median RMSE of all models, revealing the multi-model ensemble simulation to be better than individual model for most indices. Compared with observations, the models are able to capture the main features of the spatial distribution of extreme temperature during 1986–2005. Overall, the CMIP5 models are able to depict the observed indices well, and the spatial structure of the ensemble result is better for threshold indices than frequency indices. The spread amongst the CMIP5 models in different subregions for intensity indices is small and the median CMIP5 is close to observations; however, for the duration and frequency indices there can be wide disagreement regarding the change between models and observations in some regions. The model ensemble also performs well in reproducing the observational trend of temperature extremes. All absolute indices increase over China during 1961–2005.

Key words: temperature extremes, China, CMIP5, model evaluation

Citation: Dong, S. Y., Y. Xu, B. T. Zhou, and Y. Shi, 2015: Assessment of indices of temperature extremes simulated by multiple CMIP5 models over China. *Adv. Atmos Sci.*, **32**(8), 1077–1091, doi: 10.1007/s00376-015-4152-5.

1. Introduction

The implications of global warming on climatic extreme events are of particular concern internationally, particularly with respect to economic, ecological, and health impacts (Orlowsky and Seneviratne, 2012). The Intergovernmental Panel on Climate Change (IPCC) concluded in its Fifth Assessment Report (AR5) that climate change, whether driven by natural or human forcing, can lead to changes in the likelihood of the occurrence or strength of extreme weather and climate events, or both (Bindoff et al., 2013). Furthermore, some of the changes in weather and climate extreme events observed in the late 20th century are projected to continue into the future (Kharin et al., 2013).

In China, weather and climate events associated with warm and cold extremes, such as heavy rain, droughts, and typhoons since 1951, have also have shown increases in frequency, intensity, and regional extent. The combined effects of these changes could result in unprecedented extreme weather and climatic events, increasing the risk to humans of extreme climatic disasters (Xu, 2012). Extreme climate change has seriously affected China's agriculture, ecosystems, distribution of water resources, and human health; thus, the assessment of climate impact, particularly regarding climate extremes, has become an important area of research (Zheng et al., 2011).

The Coupled Model Intercomparison Project (CMIP), initiated by the World Climate Research Programme, is engaged in the simulation and prediction of the global climatic system. CMIP provides a community-based infrastructure in support of climate model diagnosis, validation, intercomparison, documentation and data access, based on state-of-the-art coupled global models for projecting and analyzing climate change. Today, the CMIP series has evolved to a fifth stage (CMIP5). CMIP5 is based on an improved CMIP3 model system, incorporating dynamic vegetation models, carbon cycle process models, and a more rational parameterization scheme, which through enhanced processing programs and coupling technology is expected to provide more detailed and more certain projections (Knutti and Sedláček, 2013). CMIP5 places greater emphasis on the consideration of land

* Corresponding author: XU Ying
Email: xuying@cma.gov.cn

use/cover forcing compared to CMIP3, and most models incorporate consistent land use/cover data (Taylor et al., 2012). Despite some limitations, CMIP5 simulation trends and observed trends can be determined, and the patterns of extreme temperature trends can be simulated reasonably well (Tebaldi et al., 2006). As Sillmann et al. (2013) revealed, the CMIP5 models are generally able to simulate climate extremes and their trends as representations of indices when compared with the HadEX2 gridded observational indices dataset (HadEX2 is a global land-based climate extremes dataset produced by the Met Office Hadley Centre, UK). The spread amongst the CMIP5 models for several temperature indices is reduced compared to CMIP3 models, despite the number of models participating in CMIP5 being larger (Sillmann et al., 2013). More recently, in an attempt to demonstrate its simulation capability, some researchers have used CMIP5 multi-model outputs to prove historical extremes of climate change over China, but using only parts of temperature extreme indices with a small number of models for the ensemble mean (e.g., Yao et al., 2012).

With a focus on the annual scale, this paper provides an assessment of the performance of CMIP5 multi-model ensemble techniques in simulating temperature extremes by comparing indices with gridded observations over China. The intention is for the indices calculated in this study to provide a useful reference for a greater understanding of the impacts of climate change on temperature extremes over China.

2. Datasets and methodology

2.1. *Global climate models*

The climate simulations of the 20th century by CMIP5, referred to as the historical experiment, are driven by observed changes in atmospheric composition (reflecting both natural and anthropogenic sources). Generally, the CMIP5 models possess higher horizontal and vertical resolutions in comparison to CMIP3. The purpose of the historical experiment is to evaluate a model's performance against present climate and observed climate change, provide initial conditions for experiments on future scenarios, and evaluate the anthropogenic impact on the 20th century climate (Taylor et al., 2012).

To assess climate model performances in simulating present temperature extremes, simulations of daily minimum and maximum near-surface temperatures in 25 state-of-the-art coupled general circulation climate models (GCMs) were collected from PCMDI (Program for Climate Model Diagnosis and Intercomparison) gateway websites. Table 1 provides an overview of the institutions and the related atmospheric

Table 1. Details of the 25 CMIP5 climate models.

Model name	Institution (or group)	Resolution (°, lon×lat)
ACCESS1-0	CSIRO (Commonwealth Scientific and Industrial Research Organisation, Australia), and BOM (Bureau of Meteorology, Australia)	192 × 145
ACCESS1-3	CSIRO (Commonwealth Scientific and Industrial Research Organisation, Australia), and BOM (Bureau of Meteorology, Australia)	192 × 145
BNU-ESM	College of Global Change and Earth System Science, Beijing Normal University	128 × 64
BCC-csm1	Beijing Climate Center, China Meteorological Administration	128 × 64
CanCM4	Canadian Centre for Climate Modelling and Analysis	128 × 64
CanESM2	Canadian Centre for Climate Modelling and Analysis	128 × 64
CCSM4	National Center for Atmospheric Research	288 × 192
CMCC-CESM	Centro Euro-Mediterraneo per i Cambiamenti Climatici	96 × 48
CMCC-CM	Centro Euro-Mediterraneo per i Cambiamenti Climatici	480 × 240
CMCC-CMS	Centro Euro-Mediterraneo per i Cambiamenti Climatici	192 × 96
CNRM-CM5	Centre National de Recherches Météorologiques/Centre Européen de Recherche et Formation Avancées en Calcul Scientifique	256 × 128
GFDL-ESM2G	Geophysical Fluid Dynamics Laboratory	144 × 90
GFDL-ESM2M	Geophysical Fluid Dynamics Laboratory	144 × 90
HadCM3	Geophysical Fluid Dynamics Laboratory	96 × 73
IPSL-CM5A-LR	L'Institut Pierre-Simon Laplace	96 × 96
IPSL-CM5A-MR	L'Institut Pierre-Simon Laplace	144 × 143
INMCM4	Institute for Numerical Mathematics	180 × 120
MIROC4h	Atmosphere and Ocean Research Institute (University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	640 × 320
MIROC5	Atmosphere and Ocean Research Institute (University of Tokyo), National Institute for Environmental Studies, and Japan Agency for Marine-Earth Science and Technology	256 × 128
MIROC-ESM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (University of Tokyo), and National Institute for Environmental Studies	128 × 64
MIROC-ESM-CHEM	Japan Agency for Marine-Earth Science and Technology, Atmosphere and Ocean Research Institute (University of Tokyo), and National Institute for Environmental Studies	128 × 64
MPI-ESM-LR	Max Planck Institute for Meteorology (MPI-M)	192 × 96
MRI-CGCM3	Meteorological Research Institute	320 × 160
NorESM1-M	Norwegian Climate Centre	144 × 96
GFDL-CM3	Geophysical Fluid Dynamics Laboratory	144 × 90

model components of those GCMs. Many GCM simulations can include multi-member ensembles, but for the purposes of model evaluation in this study, only the first ensemble member was used (1961–2005).

2.2. *Observations*

A 0.5° × 0.5° daily temperature dataset over China (Xu et al., 2009) (hereafter referred to as CN05.1) was used as observations. The gridded CN05.1 dataset has been widely used for comparing model-simulated indices. For the interpolation of the climatic field, the specific temperature information (Wu and Gao, 2013) was produced with reference to CRU (Climatic Research Unit) data (New et al., 1999, 2000) using thin-plate spline methods implemented in the ANUSPLIN software (Hutchinson, 1999). The number of observation stations used was 2416, which included national reference climatological stations, basic weather stations, and general national weather stations (Fig. 1). The dataset was subjected to quality control, including the deletion of any data that varied significantly in comparison with surrounding sites (Fig. 1, Table 2).

3. Methodology

3.1. *Indices of climatic extremes*

The WMO (World Meteorological Organization) Climate Committee recommended a set of indices of extreme weather events (http://cccma.seos.uvic.ca/ETCCDI/list27_indices.shtml). In addition, Frich et al. (2002) also defined a series of indices, and STARDEX (Statistical and Regional Dynamical Downscaling of Extremes for European Regions) defined extreme climate indices based on the statistics of extreme events in the European region and regional downscaling projects (Table 3). These indices were calculated from daily maximum and minimum temperatures and developed for assessing changes in the intensity (absolute indices and threshold indices), duration (duration indices), and frequency (percentile indices) of extreme climatic events.

STARDEX produced the FORTRAN software, which calculates indices of all extremes. The source code for the selected extreme climate indices was rewritten to fit the calculated grid data in this study. The extreme temperature

Table 2. Latitude and longitude range of the seven subregions over China.

Name	Abbreviation	Latitude range	Longitude range
Northeast China	NEC	39°–54°N	119°–134°E
North China	NC	36°–46°N	111°–119°E
East China	EC	27°–36°N	116°–122°E
Central China	CC	27°–36°N	106°–116°E
South China	SC	20°–27°N	106°–120°E
Southwest China	SWC	22°–36°N	98°–106°E
Northwest China	NWC	36°–46°N	75°–111°E

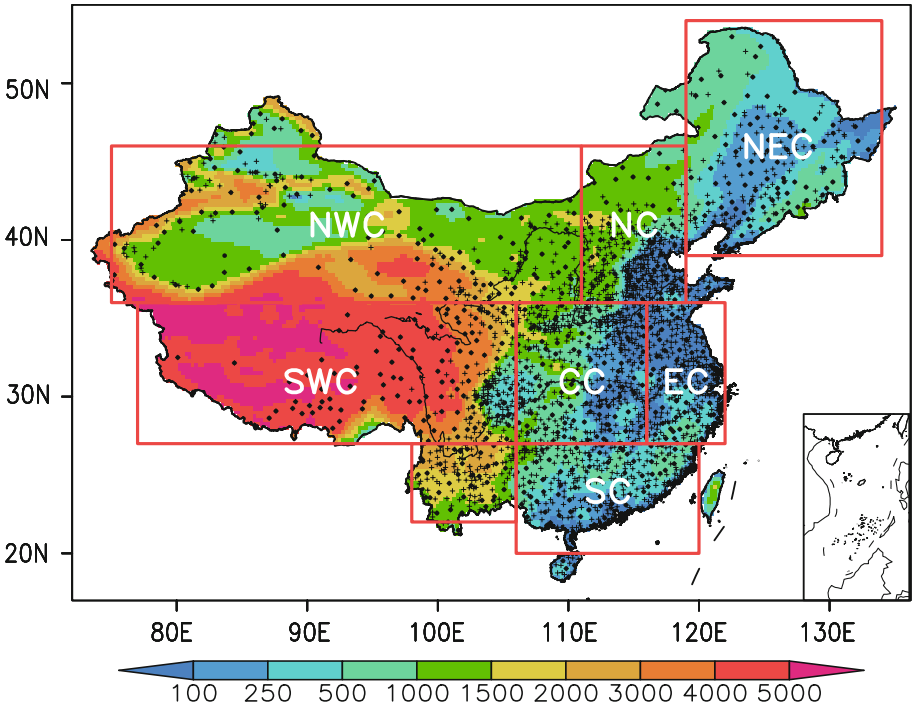

Fig. 1. The studied domains and their topography (units: m). Also shown are the seven subregions used for more detailed analysis: NEC [Northeast China (39°–54°N, 119°–134°E)]; NC [North China (36°–46°N, 111°–119°E)]; EC [East China (27°–36°N, 116°–122°E)]; CC [Central China (27°–36°N, 106°–116°E)]; SC [South China (20°–27°N, 106°–120°E)]; SWC [Southwest China (27°–36°N, 77°–106°E), (22°–27°N, 98°–106°E)]; NWC [Northwest China (36°–46°N, 75°–111°E)]. National reference climatological stations, basic weather stations (Dot mark), and general national weather stations (Cross mark).

Table 3. Definitions of the extreme temperature indices used in this study.

	Index	Indicator name	Definition	Units
Absolute indices	TXx	Max Tmax	The annual maxima of daily maximum	°C
	TNx	Max Tmin	The annual maxima of daily minimum	°C
	TNn	Min Tmin	The annual minima of daily minimum	°C
	TXn	Min Tmax	The annual minima of daily maximum	°C
Threshold indices	FD	Frost days	Number of days with minimum temperature less than 0°C	d
	SU	Summer days	Number of days with minimum temperature greater than 25°C	d
	TR	Tropical nights	Number of days with maximum temperature greater than 25°C	d
Duration indices	CWDI	Cold wave duration index	Longest number of days in interval of at least five consecutive days with daily minimum temperature 5°C less than the base period mean	d
	HWDI	Heat wave duration index	Longest number of days in interval of at least five consecutive days with daily maximum temperature 5°C greater than the base period mean	d
Percentile indices	TN10p	Cold nights	Percentage of days with daily minimum temperature less than the 10th percentile of the base period	%
	TN90p	Warm nights	Percentage of days with daily minimum temperature greater than the 90th percentile of the base period	%
	TX10p	Cold days	Percentage of days with daily maximum temperature less than the 10th percentile of the base period	%
	TX90p	Warm days	Percentage of days with daily maximum temperature greater than the 90th percentile of the base period	%

indices can be divided into four categories:

(1) Absolute indices, including annual maxima of daily maximum (TXx) and daily minimum (TNx) temperatures, and annual minima of daily maximum (TXn) and daily minimum (TNn) temperatures. Typically, TXx and TNx are referred to as warm extremes, while TNn and TXn are referred to as cold extremes (Wen et al., 2013).

(2) Threshold indices, including frost days (FD), summer days (SU), and tropical nights (TR) in this study, defined as the number of days on which a temperature value falls above or below a fixed threshold. These indices are often useful for studies of climate impact.

(3) Duration indices, including the cold wave duration index (CWDI) and heat wave duration index (HWDI), which are based on percentile thresholds calculated from the base period 1961–90. Frich et al. (2002) used a fixed threshold of 5°C above the climatology to calculate the duration indices.

(4) Percentile indices, including cold nights (TN10p), warm nights (TN90p), cold days (TX10p), and warm days (TX90p). The percentile indices are the percentage of days with daily minimum or maximum temperature more or less than the threshold percentile of the base period (1961–90).

3.2. *Processing*

CMIP5 models possess different resolutions, ranging from $96^\circ \times 48^\circ$ to $640^\circ \times 320^\circ$, but the observed data used were the 1961–2007 gridded daily temperatures on a $0.5^\circ \times 0.5^\circ$ latitude/longitude grid over China. For our analyses, we re-gridded all daily temperatures to a common resolution ($1^\circ \times 1^\circ$) using a bilinear method, and then chose the China region (1961–2005) to calculate the extreme temperature indices. For the multi-model ensemble, the arithmetic mean was used.

The studied domains included seven subregions, as shown in Fig. 1 and Table 2, based on China's climatological administrative boundaries and societal and geographical conditions, similar to those used in China's National Assessment Report on Climate Change (National Report Committee, 2007).

3.3. *Performance metrics*

3.3.1. *Model performance metrics*

For analyzing the performance metrics of different indices and models, a performance metric approach was used to assess model performance, based on the root-mean-square errors (RMSEs) of model climatologies (Flato et al., 2013; Sheffield et al., 2013; Sillmann et al., 2013), similar to that applied by Gleckler et al. (2008). This approach provides a general idea of relative model performance regarding various climatic parameters.

Each model and index was assessed with respect to CN05.1. In this analysis, RMSEs were calculated for the annual climatology of the indices of extremes over China and seven subregions of China. The RMSE was calculated as

$$\mathrm{RMSE} = \sqrt{(X-Y)^2}\,,$$

where X represents the model climatology of the extreme temperature index and Y is the corresponding extreme temperature in CN05.1.

All model RMSEs were then used to calculate the relative model error (RMSE') for each model, defined as

$$\mathrm{RMSE}' = (\mathrm{RMSE} - \mathrm{RMSE_M})/\mathrm{RMSE_M}\,,$$

where $\mathrm{RMSE_M}$ is the median RMSE of the 25 models.

To prevent abnormally large model errors (outliers) influencing the results, we chose the median rather than the mean value. For example, if the value of relative error was -0.2,

then the model's RMSE would be 20% smaller than $RMSE_M$. Conversely, if RMSE is equal to 0.2, then the RMSE would be 20% greater than $RMSE_M$ (Gleckler et al., 2008). Normalizing the calculations of RMSEs in this way produced a measure of the performance of a model compared with the typical model error.

3.3.2. *Regional performance metrics*

The Taylor diagram was originally devised to supply a concise statistical summary of how well spatial patterns match each other in terms of their correlation, root-mean-square difference, and ratio of their variances. Although the form of this diagram is generic, it is especially useful in evaluating complex models. These diagrams can also be applied in summarizing the changes in the performance of an individual model. In general, parameterization scheme revisions will affect all the fields simulated by a model, and amelioration in one aspect of a simulation might be offset by deterioration in another (Taylor, 2001). Thus, it can be useful to summarize on a diagram how well one model simulates a variety of subregions.

The Taylor diagram is used in this paper to compare the simulation performance for different subregions in two main aspects—spatial correlation coefficients and spatial standard deviation ratios—to compare the average simulated extreme temperature indices with the observed indices.

4. Results

The results of this work are presented in terms of (1) model performance, (2) spatial patterns of changes for 1986–2005, where box-and-whisker plots show regional changes and Taylor diagrams summarize variations between different regions, and (3) temporal evaluation of spatially averaged indices for the period 1961–2005.

4.1. *Model performance*

The performance of each model in simulating the climatology of indices in the period 1986–2005 is summarized in a "portrait" diagram, just as in Gleckler et al. (2008). First, the median of the model is obtained by calculating the multi-model median of each index and then obtaining its relative RMSE. The performance of each model is assessed with respect to CN05.1. Table 4 presents a summary of relative errors in the model by using "portrait" diagrams. In the table, different colors are used to characterize the magnitudes of the RMSEs; warmer colors indicate those models that perform worse, while colder colors indicated those that perform better (Gleckler et al., 2008). The portraits are arranged such that the columns are labeled by the name of the model and the rows by the extreme index name.

Consistent with the results of other multi-model studies (Flato et al., 2013; Gleckler et al., 2008; Sillmann et al., 2013; Sheffield et al., 2013), the ensemble mean generally outperforms individual models because part of the systematic errors of the individual models are offset in the multi-model mean. Most temperature indices are also captured reasonably well for most models, particularly in IPSL-CM5A-MR, MPI-ESM-LR, and CMCC-CMS. It is also noted that models with higher resolution often do not exhibit better performance than those with lower resolutions in simulations of indices of extreme temperatures (e.g., INMCM4, whose resolution is higher than NorESM1-M etc.).

For the construction of the percentile indices (TX90p, TX10p, TN90p, and TN10p), the performances of the models are generally good over China. The comparison between global results (Sillmann et al., 2013) and regions of China, based on $RMSE'$, further indicates that temperature-based percentile indices are generally better captured in China. The magnitude of the multi-model mean error over China, as measured by $RMSE'$, is generally larger for the threshold indices than the duration and absolute indices.

4.2. *Spatial pattern*

4.2.1. *Intensity extremes (threshold indices and absolute indices)*

The spatial patterns in simulating the threshold indices in the CMIP5 ensemble are shown in Fig. 2. These spatial pattern features of simulated threshold indices are similar to those of observations derived by daily minimum or maximum temperatures, e.g. FD or SU. Some fluctuations in extreme temperatures in the simulated indices, caused by regional topography, exhibit good performance, e.g., the low value area in the high temperature zone of the low elevation region in the northwest, which demonstrates the small-scale characteristics of the simulation. However, inadequacies are illustrated by the lower simulated value in the low elevation region in the southwest.

The spatial structure of absolute indices shows the distribution pattern is similar to the observations. However, there are still some divergences, particularly for high elevation terrain such as the Tibetan Plateau. In these regions, CN05.1 shows higher TXx and TXn values than the ensemble mean of CMIP5. The spatial coverage of the absolute indices of minimum temperature [TNn (0.95) and TNx (0.95)] is better than for maximum temperature [TXx (0.94) and TXn (0.94)] compared with CN05.1 (Fig. 3).

For a detailed evaluation of model performance, extreme temperature indices are analyzed using Taylor diagrams. The Taylor diagrams for extreme temperature indices during 1986–2005 over China are shown in Figs. 4 and 5. Each number corresponds to a region and the performance of its multi-model mean. Radial and angular coordinates indicate the magnitude of normalized standard deviation and correlation, respectively. The radial distance from the origin is proportional to the normalized standard deviation of a pattern. The seven subregions in the Taylor diagrams are shown in Fig. 1. Each field is normalized by the corresponding standard deviation of the reference data (hereafter referred to as NSD) (Gleckler et al., 2008), which allows the ensemble mean in the different subregions (distinguished by number) to be shown in each panel. In this figure, each numbered dot

Table 4. Performance of the relative spatial mean RMSEs –in the climatologies of temperature indices (column headings; see Table 3 for definitions) simulated by the 25 CMIP5 models (row headings; see Table 1 for further details) with respect to CN05.1 over China in the period 1986–2005.

	FD	TR	SU	CWDI_ANN	HWDI_ANN	TN10_ANN	TN90_ANN	TX10_ANN	TX90_ANN	TXx	TXn	TNn	TNx
ACCESS1-0	−0.20	0.14	0.09	0.19	0.36	0.01	0.03	0.04	0.08	0.10	−0.10	−0.11	−0.05
ACCESS1-3	−0.01	0.86	−0.02	−0.31	−0.07	0.13	−0.05	0.06	−0.07	0.19	−0.09	−0.15	0.06
BNU-ESM	0.16	0.52	0.95	0.29	0.20	−0.12	0.04	−0.10	−0.02	0.24	0.35	0.22	0.18
BCC-csm1	0.22	−0.08	0.61	0.08	0.03	−0.09	0.01	−0.09	−0.05	0.18	0.27	0.41	0.18
CanCM4	0.39	0.54	0.87	−0.04	0.10	−0.04	−0.10	0.02	0.01	0.41	−0.05	0.36	0.12
CanESM2	0.27	1.07	0.58	0.05	0.19	−0.18	−0.12	−0.04	−0.02	0.28	−0.11	0.27	0.21
CCSM4	0.03	0.71	−0.14	0.04	0.00	−0.09	−0.09	−0.03	−0.11	−0.20	0.17	0.33	−0.14
CMCC-CESM	0.15	0.31	0.96	0.01	0.36	0.15	0.10	0.15	0.01	0.19	0.03	0.07	0.29
CMCC-CM	−0.11	0.17	0.02	0.11	0.00	−0.03	0.03	−0.04	0.00	−0.15	−0.01	−0.08	0.00
CMCC-CMS	−0.17	−0.08	0.07	−0.12	0.09	−0.12	−0.01	−0.06	0.08	−0.11	−0.13	−0.11	−0.04
CNRM-CM5	0.71	−0.06	−0.01	0.05	0.00	0.08	0.03	0.09	0.11	0.19	−0.03	0.49	0.27
GFDL-ESM2G	−0.03	0.09	0.50	0.02	0.21	0.03	−0.05	−0.05	−0.01	0.17	0.03	−0.12	0.08
GFDL-ESM2M	−0.08	0.11	0.57	0.03	0.31	−0.01	−0.05	−0.05	0.00	0.23	0.01	−0.12	0.08
HadCM3	0.46	−0.16	0.16	0.67	0.09	0.15	0.01	0.11	0.06	0.01	0.05	0.29	0.09
IPSL-CM5A-LR	0.24	−0.13	0.18	−0.02	−0.03	−0.05	0.01	−0.03	0.12	0.26	0.09	−0.03	0.07
IPSL-CM5A-MR	0.07	0.06	−0.02	−0.02	0.01	−0.01	0.05	0.06	0.05	0.02	0.05	−0.09	−0.04
INMCM4	2.59	0.25	0.18	2.64	0.36	0.24	0.13	0.17	0.08	0.19	0.51	0.17	4.42
MIROC4h	−0.11	1.32	−0.15	−0.33	−0.10	−0.04	−0.03	−0.05	−0.03	0.36	−0.12	−0.06	0.13
MIROC5	−0.06	1.38	0.03	−0.11	−0.11	0.08	0.05	0.06	0.06	−0.03	−0.22	−0.08	0.04
MIROC-ESM	0.01	1.40	0.12	−0.15	0.00	0.17	0.04	0.09	0.02	0.44	0.08	0.07	0.16
MIROC-ESM-CHEM	0.02	1.69	0.13	0.05	−0.05	0.16	0.11	0.15	0.03	0.53	0.06	0.05	0.22
MPI-ESM-LR	−0.06	0.00	0.01	0.00	−0.08	0.03	0.01	−0.04	−0.01	−0.09	−0.22	−0.12	−0.11
MRI-CGCM3	0.04	0.18	−0.16	0.30	−0.11	0.20	0.10	0.18	0.02	−0.07	0.37	0.20	0.05
NorESM1-M	0.18	−0.08	0.20	0.00	0.02	−0.13	−0.02	−0.04	0.03	0.04	0.28	0.40	−0.02
GFDL-CM3	0.28	−0.06	0.29	0.37	0.10	0.14	0.03	0.06	0.06	0.27	0.11	−0.02	0.16
Ensemble	−0.18	−0.12	−0.20	−0.31	−0.27	−0.28	−0.26	−0.29	−0.31	−0.38	−0.14	−0.28	−0.26

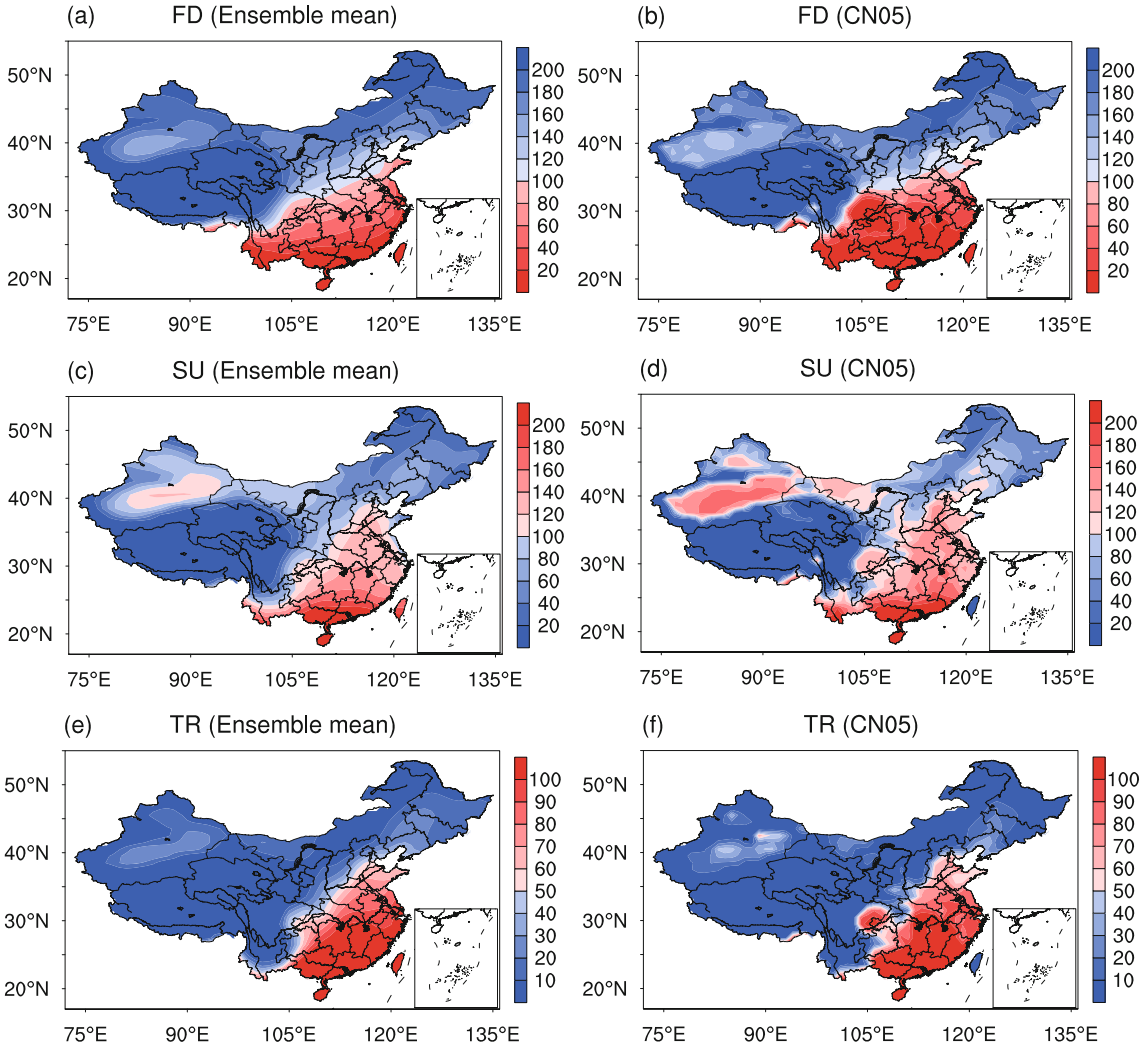

Fig. 2. 1986–2005 means of the (a, b) FD, (c, d) SU and (e, f) TR indices (see Table 3 for definitions) for the (a, c, e) CMIP5 multi-model ensemble mean and (b, d, f) CN05.1.

represents a subregion in the ensemble mean simulation, where each number represents a subregion about the ensemble mean. The nearest NSD to 1 in the extreme temperature can be found in the SU index. In contrast with other indices, the Taylor diagram for the SU indices indicate that in most regions they perform relatively well, as they are nearly close to the reference point (Fig. 4).

It is also clear that the accuracy of the model simulation depends on the extreme indices as well as the subregions. Generally, there is a much larger inter-index spread for the subregions (Fig. 4). In some simulated fields, FD shows correlations with the reference data of greater than 0.9 [e.g., the NEC (Northeast China), NC (North China) and SWC (Southwest China) subregions], whereas other subregions have much lower correlations [e.g., NWC (Northwest China) and EC (East China)]. The absolute indices over China are very good, and the absolute indices in the SWC subregion are better than in the other subregions (Fig. 5).

The multi-year mean of extreme temperature indices over each region of China during 1986–2005 were calculated for the different models and the observations. Temporal and spatial averages of extreme temperature indices are also summarized in box-and-whisker plots (Fig. 6). The colored solid mark within the box is the median of the multiple models (blue round solid mark within the box), and the blue dot is the observations. The interquartile model range is the range between the lower (25th) and upper (75th) percentiles of the total model ensemble, and the whiskers are the total inter-model range. It can be seen from Fig. 6 that the models compare well with CN05.1. In particular, the median of the CMIP5 models agrees well with CN05.1 in the representation of FD, SU, and TR, especially the latter, which is closer to CN05.1 compared with the other indices. There is also reasonable correspondence between the CMIP5 multi-model simulation of the median of the absolute indices and CN05.1, with differences typically within several degrees over most subregions of China. However, the results also show that the CMIP5 median of TXn is smaller than CN05.1 across all re-

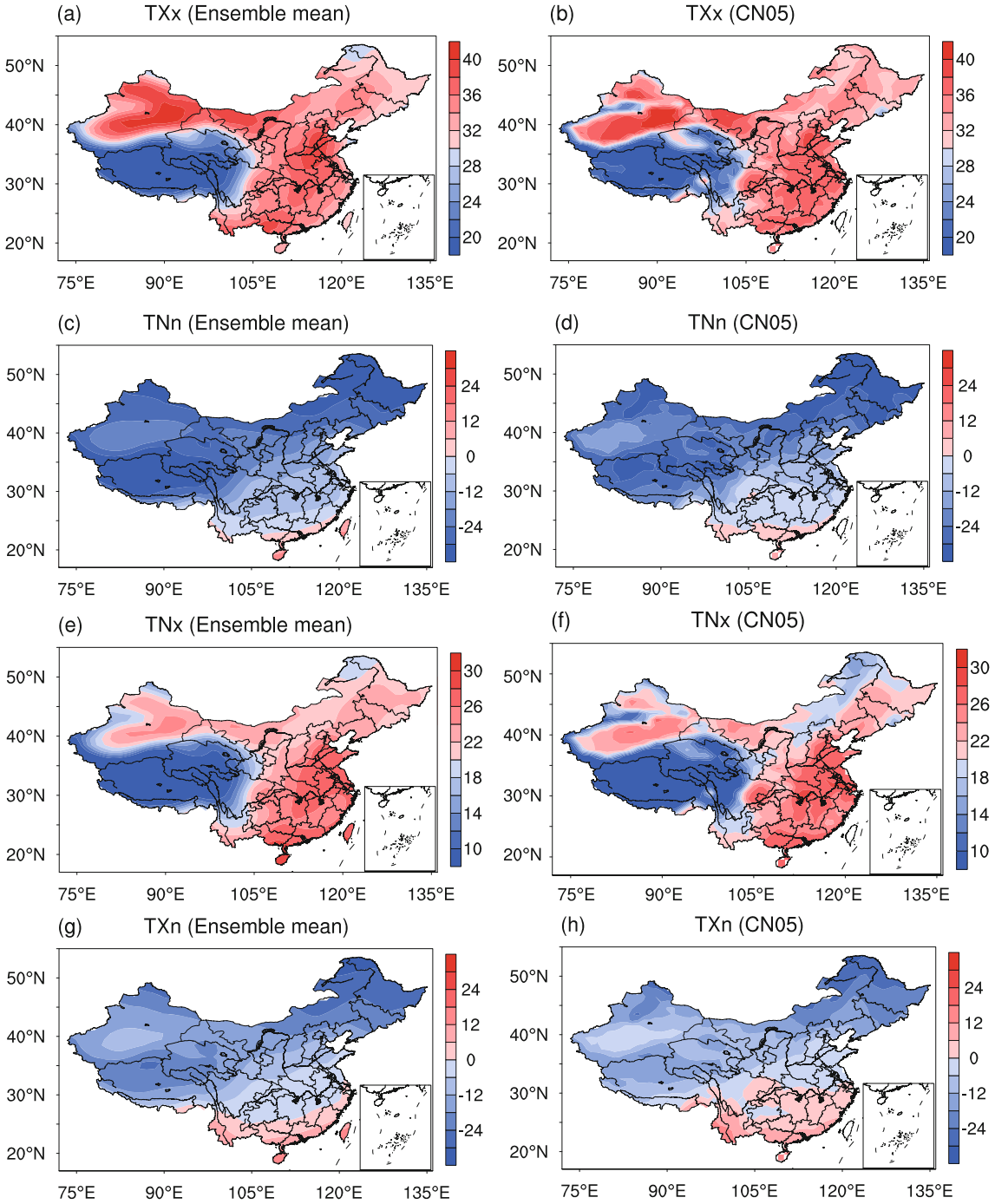

Fig. 3. 1986–2005 means of the (a, b) TXx, (c, d) TNn, (e, f) TNx and (g, h) TXn indices (see Table 3 for definitions) for the (a, c, e, g) CMIP5 multi-model ensemble mean and (b, d, f, h) CN05.1.

gions, especially in the SWC subregion.

4.2.2. *Duration extremes (duration indices)*

The CMIP5 ensemble mean HWDI is close to the observed distribution pattern, and the HWDI index of the ensemble simulation in northern China is lower compared with observations (Fig. 7). However, in northern, northeastern and northwestern China, as well as the Tibetan Plateau area, the simulated CWDI is higher than observed. The reason may be the difference between the modeled and actual topography, especially in the Tibetan Plateau. Also, the number of observation stations in the west is relatively small, such as in the northern part of the Tibetan Plateau to the northern foot of the Kunlun Mountains and Xinjiang's Taklimakan Desert hinterland, which may also have affected the results of the interpolation of the data.

Figure 8 clearly shows which regions exaggerate the amplitude of extreme temperature index (e.g., the NWC sub-

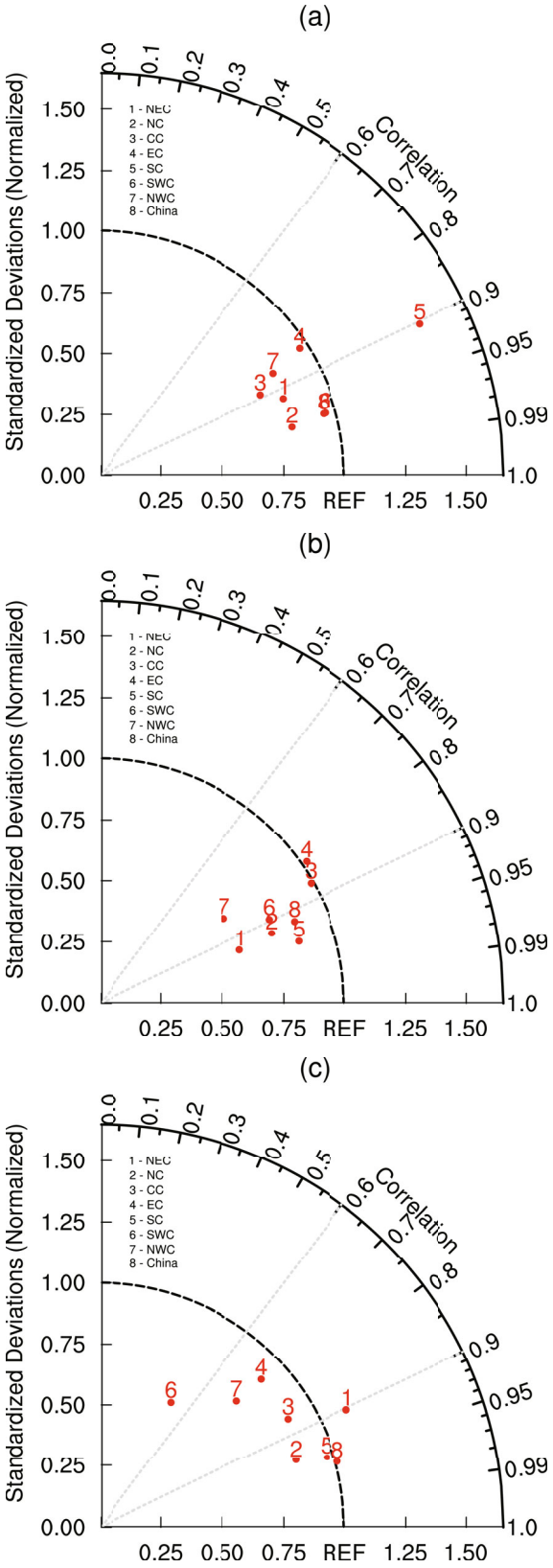

Fig. 4. Taylor diagrams of the CMIP5 ensemble mean in different regions (numbered 1–8; see Fig. 1 for definitions) (1986–2005): (a) FD index; (b) SU index; (c) TR index (see Table 3 for definitions). Each numbered dot represents an individual simulation made for a particular region.

region) and which models' ensembles grossly underestimate NSD in most regions. Basically, the largest NSD and the smallest correlation index in the temperature indices over China can be found in HWDI (Fig. 8).

These results show that, generally, the CMIP5 median of CWDI (Fig. 6) is larger than in CN05.1 across all regions; and thus, more cold waves are simulated in comparison with CN05.1. The CWDI index also produces the most significant outliers (biases) in the simulation of CWDI compared with CN05.1, and it also exhibits significant outliers in the SWC subregion in the CMIP5 ensemble for CWDI, and in the SC subregion for HWDI.

4.2.3. *Frequency extremes (percentile indices)*

The models also disagree in terms of the annual mean value over China, insofar as the percentile indices show much larger values compared to CN05.1 with respect to TN10p and TX10p, but much lower values compared to CN05.1 with respect to TN90p and TX90p (not shown). The Taylor diagrams for the percentile indices reveal that some fields show low correlation values and NSD with the reference data in most regions (not shown).

The results of the box-and-whisker plots (Fig. 6) show that the median of the CMIP5 models is generally underestimated with respect to TX90p in comparison with CN05.1 in all regions. The CMIP5 median tends to underestimate the TN10p and Tx90p indices over China compared with CN05.1 in all regions, and overestimates the TX10p index compared with CN05.1 in most regions. The models disagree with CN05.1 with respect to TX90p, showing much smaller values than the CMIP5 median. The discrepancy is most prominent for TN10p in the NEC subregion and TX90p in the CC (central China) subregion, for which the CN05.1 values are located far above the CMIP model range.

A comparison of the spatial structure between the CMIP5 models and observations shows that the main features of the spatial distribution of temperature extremes are captured well by the model ensemble percentile indices. As a whole, the CMIP5 models compare well with CN05.1, and the spatial structure of the ensemble result is better for indices of threshold extremes than for indices of intensity extremes (Alexander et al., 2006). In some regions, the number of observation stations in the west is relatively small, and in the northern part of the Tibetan Plateau to the northern foot of the Kunlun Mountains and Xinjiang's Taklimakan Desert hinterland, the basic distribution of observation sites, which also determines the interpolation of the data of these areas, has relatively large uncertainty. Warm extremes (SU, TR, TNx, HWDI), except TXx, are underestimated over high northern latitudes, particularly in the northwest, while cold extremes (FD, TXn, TNn, CWDI) are overestimated, which is in accordance with previous findings based on IPCC AR4 models (Wang et al., 2008).

4.3. *Temporal evolution of indices*

Sillmann et al. (2013) suggested that there is not a clear relationship between a model's spatial resolution and its representation of temperature indices, and thus the results from

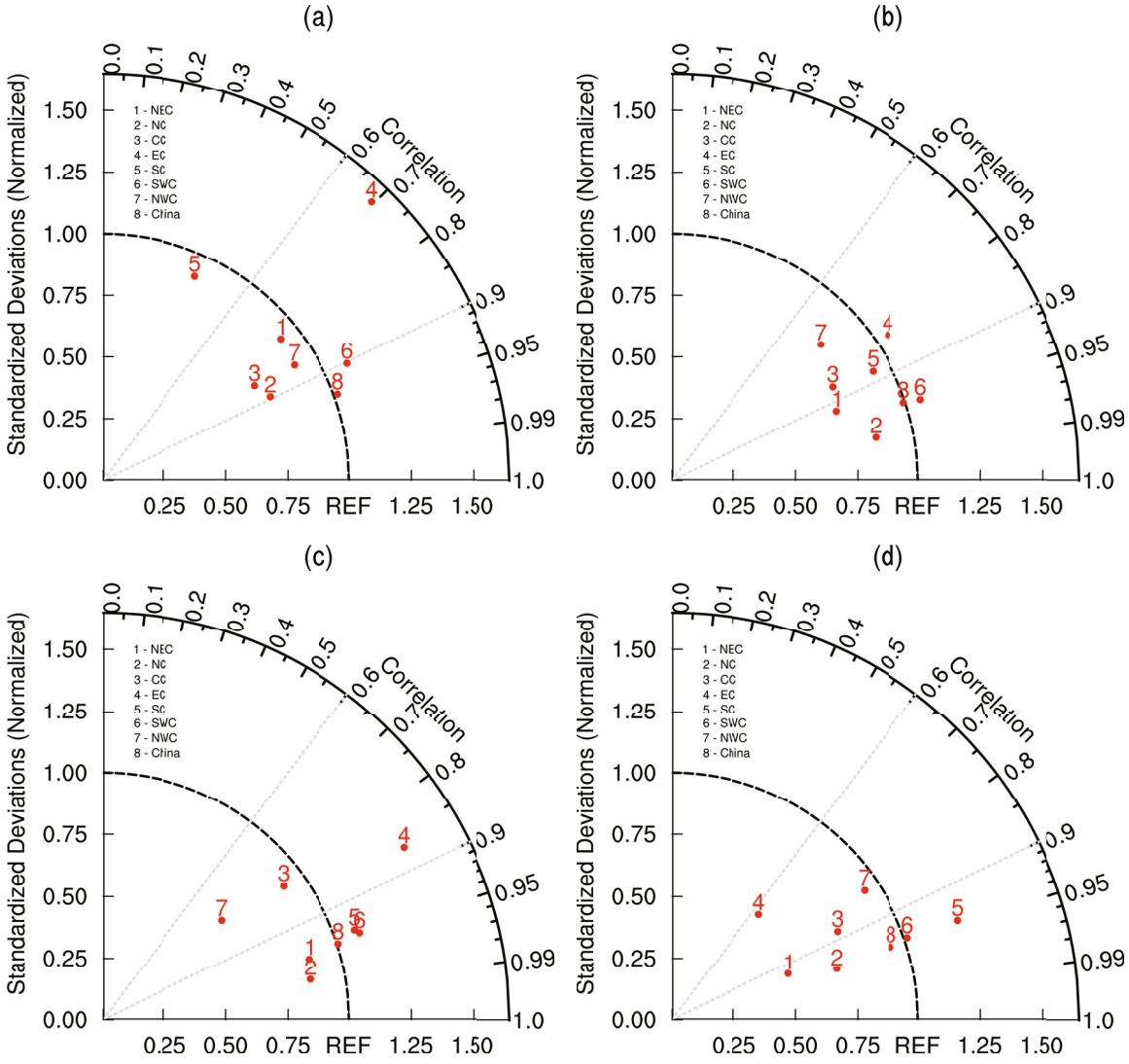

Fig. 5. Taylor diagrams of the CMIP5 ensemble mean in different regions (numbered 1–8; see Fig. 1 for definitions) (1986–2005): (a) TXx index; (b) TNx index; (c) TXn index; (d) TNn index (see Table 3 for definitions). Each numbered dot represents an individual simulation made for a particular region.

individual CMIP5 models were not shown. For historical trends, the CMIP5 models' ensemble generally captures well the observed trends in the indices of temperature extremes during 1961–2005. The long-term trends in simulating the historical temporal evolution of the indices for the anomalies are more distinct in Figs. 9 and 10.

4.3.1. *Intensity extremes (threshold indices and absolute indices)*

The modeled threshold index trends are consistent with CN05.1 (figures not shown). The threshold indices of the ensemble mean show corresponding trends, with increasing numbers of TR [1.31 d (10 yr)$^{-1}$] and SU [1.64 d (10 yr)$^{-1}$] and decreasing numbers of FD [−2.02 d (10 yr)$^{-1}$]. There are similar increasing trends in the models' ensemble mean compared with CN05.1 for SU [1.44 d (10 yr)$^{-1}$] and TR [1.00 d (10 yr)$^{-1}$] during 1961–2005.

The multi-model ensemble mean shows similar warming trends in the absolute indices starting in the 1960s, compared with CN05.1 (Fig. 9), with a general increase in TNn, TNx, TXx, and TXn. In China, the increases in both TNx [0.23°C (10 yr)$^{-1}$] and TNn [0.29°C (10 yr)$^{-1}$], which depend on the minimum temperature, are greater than the increases in both TXx [0.21°C (10 yr)$^{-1}$] and TXn [0.24°C (10 yr)$^{-1}$], which depend on maximum temperature.

4.3.2. *Duration extremes (duration indices)*

During 1961 to 2005, the mean CMIP5 models simulate a decrease in CWDI [−0.61 d (10 yr)$^{-1}$], while showing a strong increase in HWDI [1.38 d (10 yr)$^{-1}$] over China (not shown).

4.3.3. *Frequency extremes (percentile indices)*

The changes are much more pronounced in the percentile indices compared to the absolute indices, which are derived from annual extremes (Fig. 10). The construction of

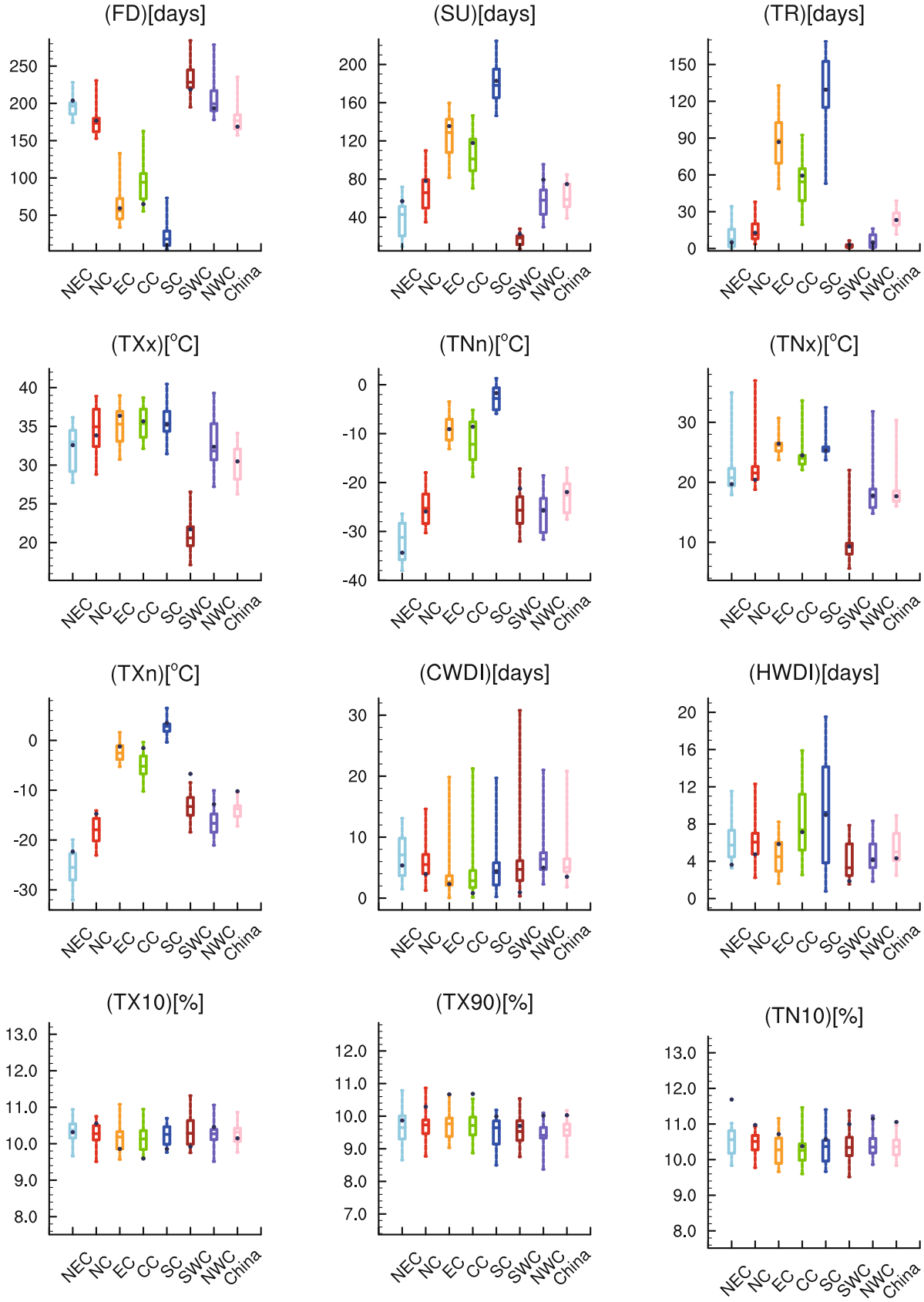

Fig. 6. Box-and-whisker plots of extreme temperature indices (panel titles; see Table 3 for definitions) calculated from the 25 CMIP5 models. The boxes indicate the interquartile model spread (range between the 25th and 75th quantiles); the solid color marks within the boxes show the multi-model median; and the whiskers indicate the total intermodal range (1986–2005). CN05.1 is indicated by the blue dot.

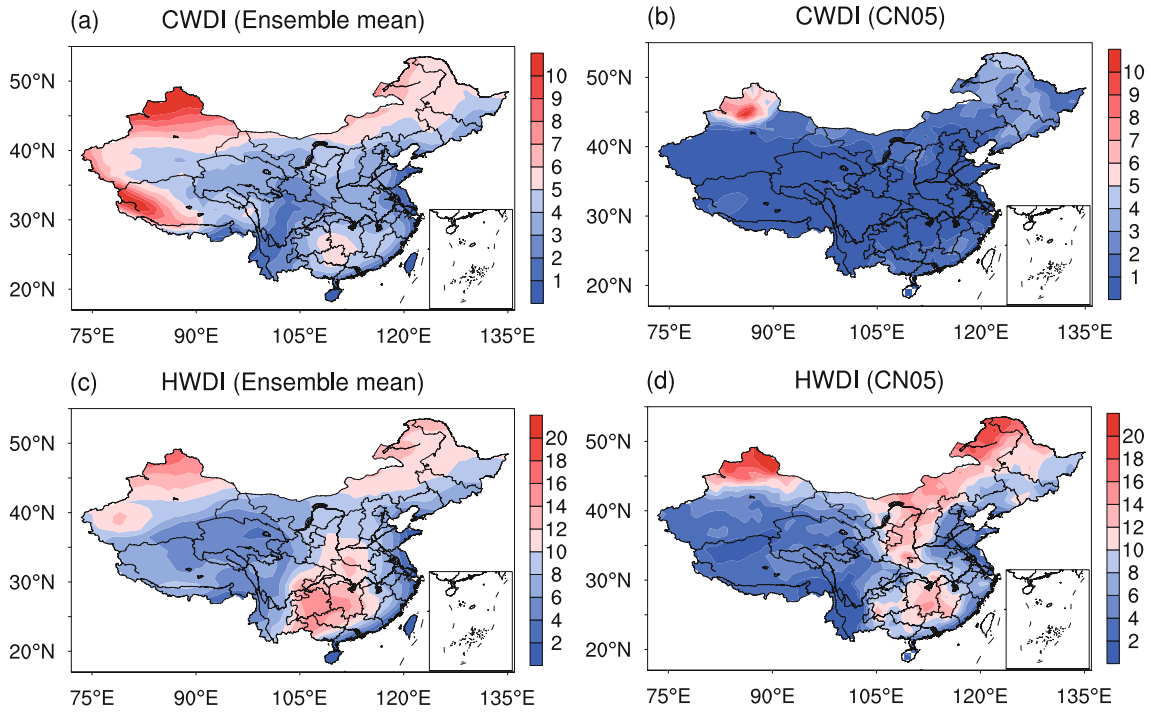

Fig. 7. 1986–2005 means of the (a, b) CWDI and (c, d) HWDI indices (see Table 3 for definitions) for the (a, c) CMIP5 multi-model ensemble mean and (b, d) CN05.1.

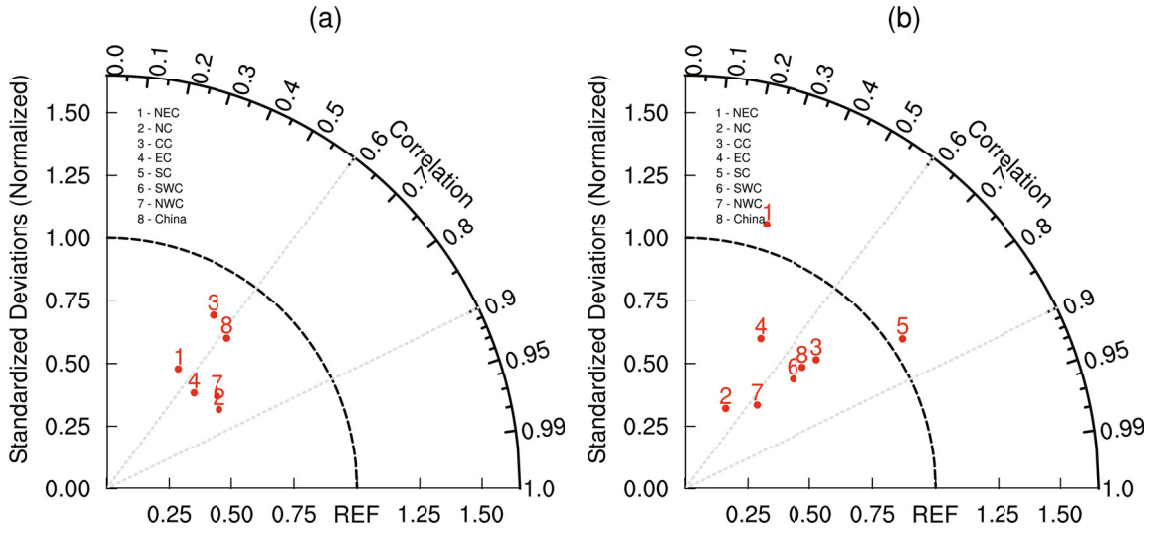

Fig. 8. Taylor diagrams of the CMIP5 ensemble mean in different regions (numbered 1–8; see Fig. 1 for definitions) (1986–2005): (a) CWDI index; (b) HWDI index (see Table 3 for definitions). Each numbered dot represents an individual simulation made for a particular region.

the percentile indices leads to adequate correlation (0.67–0.91) in the temporal evolution of the CMIP models and the observations. The average across China shows a decrease in terms of cold nights (TN10p) and cold days (TX10p), but an increase in warm nights (TN90p) and warm days (TX90p).

Differences in the percentile index lines are especially prominent between the most recent two decades and for those indices derived from minimum temperature. Increasing trends in warm days (TX90p) [1.21% (10 yr)$^{-1}$] and nights (TN90p) [1.70% (10 yr)$^{-1}$] and decreasing trends in cold days (TX10p) [-0.68% (10 yr)$^{-1}$] and nights (TN10p) [−0.96% (10 yr)$^{-1}$] can be seen in the CMIP5 models, which is consistent with globally observed changes (Alexander et

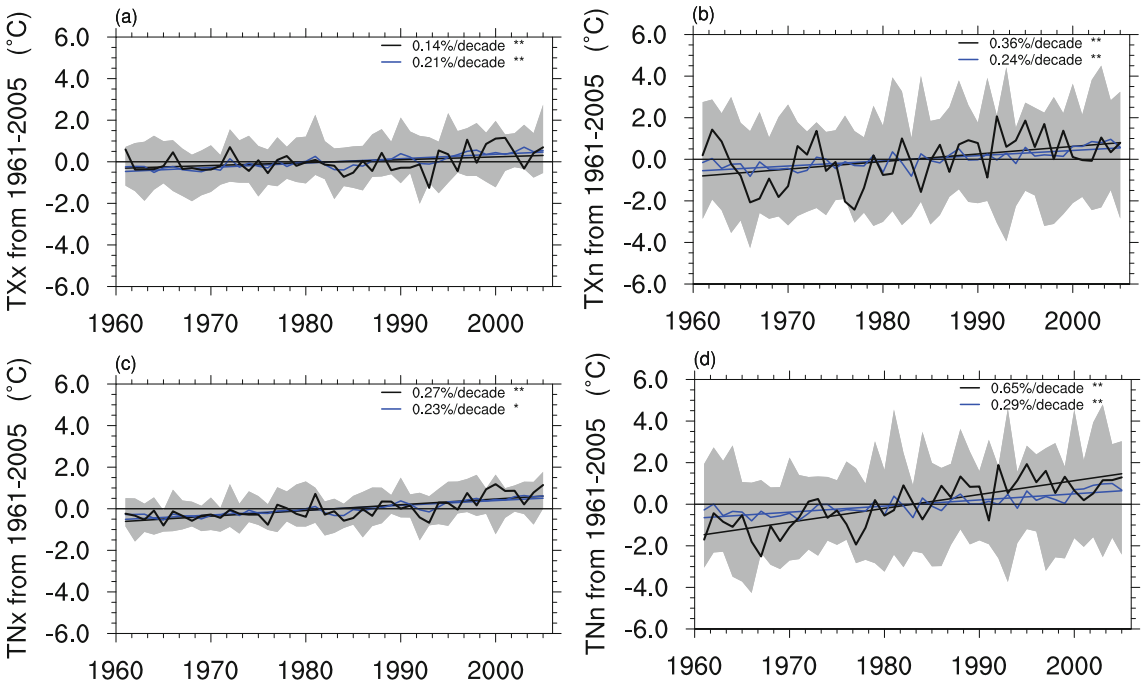

Fig. 9. Spatial means of anomalies of extreme temperature indices (TXx, TXn, TNx, and TNn; see Table 3 for definitions) over China from 1961–2005: ensemble mean of the 25 CMIP5 models (blue line) and CN05.1 (black line). Shading indicates the spread of the 25 CMIP5 models. * 95% confidence level; ** 99% confidence level (*t*-test).

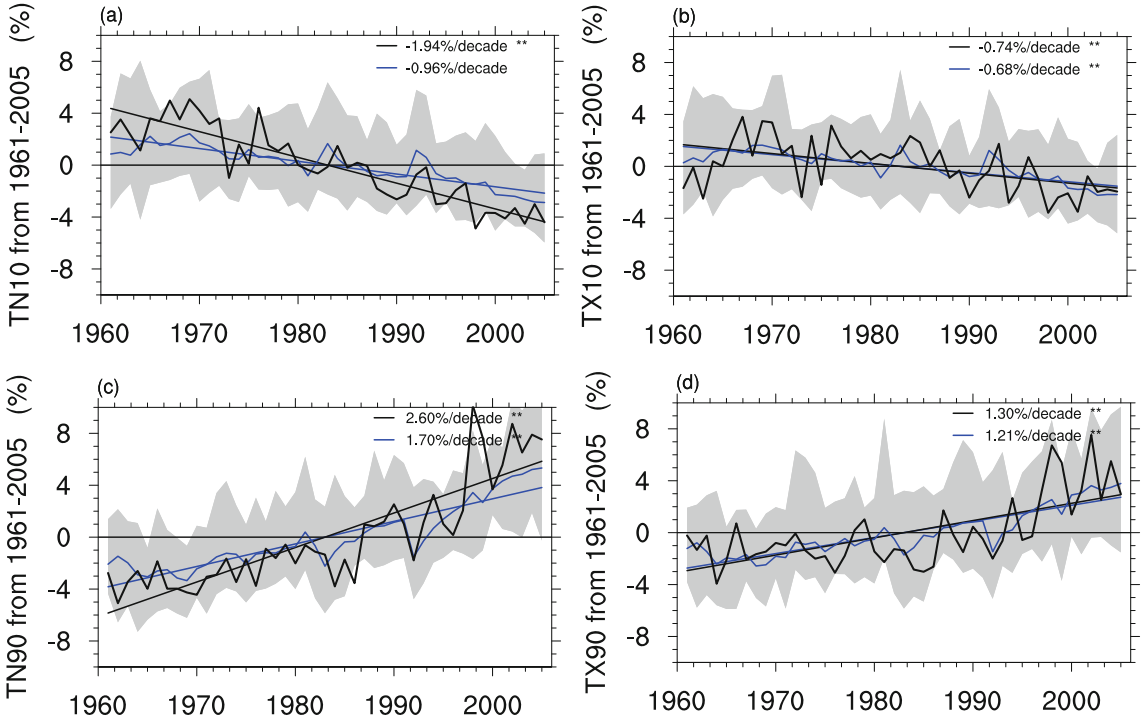

Fig. 10. Spatial means of anomalies of extreme temperature indices (TN10p, TX10p, TN90p, and TX90p; see Table 3 for definitions) over China from 1961–2005: ensemble mean of the 25 CMIP5 models (blue line) and CN05.1 (black line). Shading indicates the spread of the 25 CMIP5 models. *95% confidence level; **99% confidence level (*t*-test).

al., 2006; Donat et al., 2013) and global CMIP5 model results (Sillmann et al., 2013). This is also consistent with the general trend of extreme observations in China, which shows that the frequency of extreme cold temperatures and cold-temperature events is reducing (Ding et al., 2002; Ding et al., 2009).

5. Summary and concluding remarks

Based on simulations of the 20th century by the 25 climate models in the CMIP5 set of experiments, the performance in simulating the indices of present temperature extremes over China was assessed. The main findings can be summarized as follows:

(1) Extreme temperature simulation over China was derived from the 25 CMIP5 models and ensemble simulation and compared with observations. The difference between the models and observations is reduced by the multi-model ensemble method. The multi-model ensemble simulations are closer to observations in terms of relative RMSEs (RMSE relative to the median RMSE of all models) for most models and indices over China.

(2) Compared with observations, the models are able to capture the main features of the spatial distribution of temperature during 1986–2005. The Taylor diagrams show that divergences between the multi-model ensemble and observations in terms of the threshold temperature indices are smaller than other temperature indices. For the regional mean over China, the multi-model ensemble simulations of CMIP5 with respect to FD and CWDI show larger values than CN05.1, whereas the ensemble simulations with respect to SU, TXn and TX10p underestimate the actual extreme indices in CN05.1. Uncertainties in some extreme indices (such as TN10p, TN90p,TX90p, and TXn) are larger than in others, and there are also large spatial differences between threshold indices and absolute indices, which may be associated with the larger interquartile model spread due to the different climate sensitivities of the models and/or positive regional land–atmosphere feedback processes. For the duration and frequency indices, wide disagreement was found regarding the change between models and observations in some regions.

(3) In terms of the temporal changes of the indices of temperature extremes, the model ensemble performs well in reproducing the trend of change in the indices of temperature extremes. For example, for 1961–2005, CN05.1 shows an increase in warm days (TX90p) with a trend of 1.3% $(10\ \mathrm{yr})^{-1}$, while cold days (TX10p) shows a decrease of -0.74% $(10\ \mathrm{yr})^{-1}$. The CMIP5 ensemble results also show these trends, but with a smaller trend. These results show that the ensemble simulation tends to increase warm extremes (e.g., SU, TR, HWDI, TN90p, and TX90p) over China, and decease cold extremes (e.g., FD, CWDI, TN10p, and TX10p).

This study comprehensively compared model simulations and observations, but there was no in-depth analysis of the resolutions of the models, nor their physical or other processes. Avila et al. (2012) suggested that many extreme temperature indices can be affected by land use and land cover changes. Therefore, our model assessment should be extended by further studies of seasonal model performance and particular regions that are especially sensitive to climatic changes in extreme events over China.

Additionally, considering the uncertainties in the simulations, further research should focus on how to adopt a more rational and scientific approach to extracting useful information from the simulation results, and how to better use the simulation results for detecting and attributing climate change (Orlowsky and Seneviratne, 2012).

Acknowledgements. This study was supported by the National Basic Key Project (also called 973 Project, Grant Nos. 2010CB950501 and 2010CB950102), the R&D Special Fund for Public Welfare Industry (meteorology) (Grant No. GYHY 201306019), and the National Natural Science Foundation of China (Grant No. 41275078).

REFERENCES

Alexander, L. V., and Coauthors, 2006: Global observed changes in daily climate extremes of temperature and precipitation. *J. Geophys. Res.*, **111**, D05109, doi: 10.1029/2005JD006290.

Avila, F. B., A. J. Pitman, M. G. Donat, L. V. Alexander, and G. Abramowitz, 2012: Climate model simulated changes in temperature extremes due to land cover change. *J. Geophys. Res.*, **117**, D04108, doi: 10.1029/2011JD016382.

Bindoff, N. L., and Coauthors, 2013: Detection and attribution of climate change: From global to regional. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. F. Stocker et al., Eds., Cambridge University Press, 867–952,

Ding, T., W. H. Qian, and Z. W. Yan, 2009: Characteristics and changes of cold surge events over China during 1960–2007. *Atmos. Oceanic Sci. Lett.*, **2**(6), 339–344.

Ding, Y. H., J. Zhang, and Y. F. Song, 2002: Changes in weather and climate extreme events and their association with the global warming. *Meteorological Monthly*, **28**(3), 3–7. (in Chinese)

Donat, M. G., and Coauthors, 2013: Updated analyses of temperature and precipitation extreme indices since the beginning of the twentieth century: The HadEX2 dataset. *J. Geophys. Res.*, **118**, 2098–2118, doi: 10.1029/2012JD018606.

Flato, G., and Coauthors, 2013: Evaluation of climate models. *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. F. Stocker et al, Eds., Cambridge University Press, 741–866.

Frich, P., L. V. Alexander, P. Della-Marta, B. Gleason, M. Haylock, A. Tank, and T. Peterson, 2002: Observed coherent changes in climatic extremes during the second half of the twentieth century. *Climatic Research*, **19**, 193–212.

Gleckler, P. J., K. E. Taylor, and C. Doutriaux, 2008: Performance metrics for climate models. *J. Geophys. Res.*, **113**, D06104, doi: 10.1029/2007JD008972.

Hutchinson M. F., 1999: ANUSPLIN version 4. 0 user guide. Centre for Resources and Environmental Studies, Australian National Univeristy, Canberra ACT 0200, 51pp Australia.

Kharin, V. V., F. W. Zwiers, X. Zhang, and M. Wehner, 2013: Changes in temperature and precipitation extremes in the CMIP5 ensemble. *Climatic Change*, **119**(2), 345–357.

Knutti, R., and J. Sedláček, 2013: Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change*, **3**, 363–373, doi: 10.1038/nclimate1716.

National Report Committee, 2007: *China's National Assessment Report on Climate Change*. Science Press, 148 pp. (in Chi-

nese)

New, M., M. Hulme, and P. Jones, 1999: Representing twentieth-century space-time climate variability. Part I: Development of a 1961–90 mean monthly terrestrial climatology. *J. Climate*, **12**: 829–856.

New, M., M. Hulme, and P. Jones, 2000: Representing twentieth-century space-time climate variability. Part II: Development of 1901–96 monthly grids of terrestrial surface climate. *J. Climate*, **13**, 2217–2238.

Orlowsky, B., and S. I. Seneviratne, 2012: Global changes in extreme events: Regional and seasonal dimension. *Climatic Change*, **110**(3-4), 669–696.

Sheffield J., and Coauthors, 2013: North American climate in CMIP5 experiments. Part I: Evaluation of historical simulations of continental and regional climatology. *J. Climate*, **26**, 9209–9245. doi: http://dx.doi.org/10.1175/JCLI-D-12-00592.1.

Sillmann, J., V. V. Kharin, X. Zhang, F. W. Zwiers, and D. Bronaugh, 2013: Climate extremes indices in the CMIP5 multimodel ensemble: Part 1. Model evaluation in the present climate. *J. Geophys. Res.*, **118**(4), 1716–1733.

Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi: 10.1175/BAMS-D-11-00094.1.

Tebaldi, C., K. Hayhoe, J. M. Arblaster, and G. A. Meehl, 2006: Going to the extremes. *Climatic Change*, **79**(3-4), 185–211, doi: 10.1007/s10584-006-9051-4.

Wang, J., Z. H. Jiang, J. Song, and Y. G. Ding, 2008: Evaluating the simulation of the GCMS on the extreme temperature indices in China. *Acta Geographica Sinica*, **63**(3), 227–236. (in Chinese)

Wen, H. Q., X. B. Zhang, Y. Xu, and B. Wang, 2013: Detecting human influence on extreme temperatures in China. *Geophys. Res. Lett.*, **40**, 1171–1176, doi: 10.1002/grl.50285.

Wu, J., and X. J. Gao, 2013: A gridded daily observation dataset over China region and comparison with the other datasets. *Chinese Journal of Geophysics*, **56**(4), 1102–1111, doi: 10.6038/cjg20130406. (in Chinese)

Xu, Y., 2012: The projection and possible risks of future extreme weather events. *Climate Change Green Paper—Climate Financing and Low- Carbon Development*, 169–177.

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**(4), 763–772, doi: 10.1007/s00376-009-9029-z.

Yao, Y., Y. Luo, and J. B. Huang, 2012: Evaluation and projection of temperature extremes over China based on CMIP5 Model. *Advances in Climate Change Research*, **3**, 179–185.

Zheng, Y. H., and Coauthors, 2011: *The Second National Assessment Report on Climate Change*. Science Press, 710 pp. (in Chinese)